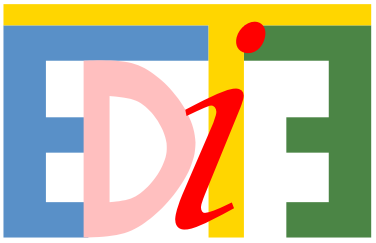


Thèse

présentée pour obtenir le grade de docteur

de l'Ecole Nationale Supérieure des Télécommunications

Spécialité : Signal et Images

Vincent DUVAL

Méthodes variationnelles et non locales en traitement d'images: une étude géométrique

Variational and non-local methods in image processing: a geometric study

Rapporteurs:	Antonin CHAMBOLLE	Ecole Polytechnique
	Daniel CREMERS	Technical University of Munich
	Jean-Michel MOREL	ENS Cachan
Examinateurs:	Jérôme DARBON	ENS Cachan
	Simon MASNOU	Université Lyon 1
	Gabriel PEYRÉ	Université Paris-Dauphine
	François ROUEFF	Télécom ParisTech
Directeurs:	Yann Gousseau	Télécom ParisTech
	Jean-François Aujol	Université Bordeaux 1

Remerciements

Tout d'abord je tiens à remercier chaleureusement mes directeurs de thèse, Yann Gousseau et Jean-François Aujol. Sur le plan scientifique, leur complémentarité ainsi que l'étendue de leurs connaissances m'ont permis de bénéficier de conseils éclairés à propos d'une très grande variété de problèmes, et j'ai beaucoup apprécié l'éthique et la rigueur qu'ils mettent dans leur travail. Sur le plan humain, j'ai pu mesurer chaque jour la chance que j'avais de travailler avec eux : il est plus facile de surmonter les difficultés de la thèse quand on a des directeurs avec autant d'humour et d'ouverture d'esprit!

Je remercie les membres du jury d'avoir accepté d'examiner cette thèse. C'est un très grand honneur que m'ont fait Antonin Chambolle, Jean-Michel Morel et Daniel Cremers en acceptant d'être rapporteurs de cette thèse. Aussi, je suis très heureux que Jérôme Darbon, Simon Masnou, Gabriel Peyré et François Roueff aient accepté d'en être examinateurs. Chacun des membres de ce jury a été un modèle pour l'apprenti chercheur que je suis. Je profite également de cette occasion pour exprimer toute ma gratitude envers Jean-Michel Morel, dont les cours de master ont considérablement influencé mon travail de recherche et mon approche du traitement d'images, ainsi qu'envers Gabriel Peyré : je n'aurais probablement jamais fait de thèse si je ne l'avais pas rencontré lors d'un enseignement d'approfondissement à l'Ecole polytechnique. Ses encouragements et ses conseils avisés m'ont été très précieux. Depuis mon stage à UCLA, j'ai eu la chance de croiser régulièrement Jérôme Darbon: merci à toi pour ta bienveillance, et pour avoir partagé avec moi (des pizzas au Palomino et) ta vision de la science, belle et sans concessions.

Cette thèse doit beaucoup à mes deux complices débruiteurs d'images, Charles et Joseph, que je remercie sincèrement. Charles, j'ai beaucoup appris à ton contact (notamment l'art et la manière de faire à peu près tout en utilisant uniquement un makefile!). Jo, ton érudition bibliographique et la rigueur avec laquelle tu traces les découvertes en traitement d'images m'ont beaucoup impressionné. Merci pour tout le travail de relecture que tu as accompli. Cela a été un plaisir de travailler avec vous deux, ne serait-ce que grâce à votre enthousiasme scientifique communicatif.

Même si nous n'avons jamais écrit d'articles ensemble, les conversations scientifiques avec Julien et Bruno resteront parmi mes plus beaux souvenirs de thèse. Quel plaisir de poser un problème au tableau et de l'attaquer ensemble! Merci à vous deux de m'avoir accueilli à bras ouverts dans la tribu (σ-algèbre?) des thésards de Yann. Bravo pour vos postes de MCF, et bonne chance pour la suite!

Parce qu'il n'y a pas que la recherche dans la vie, mais aussi l'enseignement, je tiens à remercier les membres de la "team" Hilbert-Fourier : Yann (à nouveau), Saïd (l'ami des chats et des maths), et Julie (l'amie des fraises tagada). Merci pour la confiance que vous m'avez accordée. La préparation de cet enseignement avec vous a été l'occasion de très bons moments. Merci aussi à Bertrand et Roland pour ma participation à OASIS et OPTIM.

Ces trois ans de thèse ont été très agréables grâce à la compagnie des autres thésards de Télécom. Un grand merci à Thierry et Baptiste : sous leur impulsion, la salle C07 restera probablement dans l'Histoire comme le bureau cool du LTCI. Les pains aux chocolats et croissants à l'improviste dégustés en leur compagnie ainsi que celle de Gui-Song, Alasdair et Cécilia me manqueront. Mais c'est toute la bande que je regretterai : Benoît et ses anecdotes culturelles, Charles et sa fameuse blague de 12h06, Flora (membre d'honneur de la C07), Julien C., Guillaume, Talita et Fang. Merci aussi aux plus anciens, qui sont déjà partis de Télécom: Ivan, Marie, Julien, Mihaï, Adrian, Aymen, Gabrielle, Guillaume ... Avec grand plaisir, j'ai aussi passé pas mal de temps à Cachan, pour discuter avec Julie, Eric, Bruno, Adina,

Zhongwei *et al.* . Merci à eux, et tout particulièrement à Adina qui m'a accordé de son temps quelques jours avant sa soutenance pour discuter de mes calculs sur le modèle TV+G.

Merci à Christine B. d'avoir toujours accepté mes missions avec le sourire, ainsi qu'à Patricia et à Corinne pour leur aide face à tous les petits soucis administratifs que j'ai pu rencontrer. Je suis aussi très reconnaissant à Jalal Fadili et à David Tschumperlé pour leur chaleureux accueil à Caen, ainsi qu'à Cécile Louchet pour son accueil tout aussi convivial à Orléans.

Last but not least, je remercie les membres de ma famille pour leurs encouragements, et tout particulièrement Diane pour son soutien inébranlable pendant ces années difficiles : j'espère pouvoir lui apporter une aide aussi précieuse pour la fin de sa propre thèse, et pour toutes les choses de la vie pendant de nombreuses années encore…

Contents

Résumé en français

D'une scène idéale, issue du monde physique, jusqu'à sa représentation sur nos écrans d'ordinateurs, de nombreuses étapes sont susceptibles de dégrader les images. Le processus d'acquisition, la transmission ou le stockage d'une image peuvent tous dégrader l'information originale, de sorte que les méthodes de restauration ou de simplification d'images sont devenues cruciales dans de nombreuses applications, que ce soit en diagnostic médical, imagerie satellitaire, en visioconférence ou dans les appareils photos de la vie quotidienne.

Pendant le processus d'acquisition, les capteurs CCD (*Charge-Coupled-Devices*) sont soumis au bruit de grenaille, qui est dû à la nature discrète de la quantité mesurée (c'est-à-dire le nombre de photons reçus), et qui suivent une loi de Poisson. L'effet du bruit de grenaille est d'autant plus important que le nombre de photons est faible, de sorte que les scènes insuffisamment éclairées apparaissent souvent bruitées. Le bruit thermique est aussi responsable de la dégration des images : en fonction de la température du capteur l'agitation thermique des électrons induit un courant qui s'ajoute au signal original. Mentionnons aussi les altérations dues à l'optique : certains objets sont parfois hors de focus, ou bien des aberrations chomatiques peuvent apparaître. Un objet en mouvement sera flou si le temps de pose est trop long. Pendant le processus de transmission, les images peuvent être dégradées par la perte de paquets, qui entraîne des valeurs des pixels incorrectes. Cela est particulièrement courant en imagerie satellitaire. Pour réduire le coût de stockage des images, des algorithmes de compression avec pertes sont employés, ce qui supprime des informations, induisant un "bruit de compression" d'autant plus important que le taux de compression est élevé.

Par conséquent, les images sont en général trop dégradées pour être montrées directement à un humain ou envoyées à un algorithme d'analyse d'images : il faut les filtrer. La communauté du traitement d'images a travaillé pendant plusieurs dizaines d'années sur des algorithmes permettant soit de restaurer les images de la façon la plus fidèle possible, soit de simplifier les images tout en conservant l'information la plus pertinente pour l'analyse automatique. La première méthode proposée consistait en un filtrage linéaire de l'image. Sous quelques hypothèses supplémentaires comme l'invariance par translation, cela revient à appliquer une convolution à l'image (voir par exemple [Guichard et al., 2004]). Le problème de la convolution est qu'il a tendance à rendre flous les contours qui existent dans les images. Pour lisser les images sans rendre les bords flous, les chercheurs et les ingénieurs ont proposé un large panel de méthodes qui correspondent toutes à un certain *a priori* sur les images.

Décrivons brièvement quelques grandes familles de modèles :

Equations aux Dérivées Partielles (EDP). Le filtrage linéaire peut être interprété asymptotiquement comme un pas de l'équation de la chaleur, qui diffuse l'information de manière anisotrope. L'idée proposée par Malik et Perona [Perona and Malik, 1990] est de diffuser l'information en fonction du contenu local de l'image, à travers des équations de la forme :

$$\frac{\partial u}{\partial t} = \text{div}\ \left(g(|\nabla u|)\nabla u\right), \tag{1}$$

où par exemple $g : \mathbb{R}^+ \to \mathbb{R}, s \mapsto \frac{1}{1+\lambda^2 s^2}$.

Dans les régions homogènes, la diffusion se comporte comme une équation de la chaleur à deux dimensions habituelle, mais près des contours (où $|\nabla u|$ est grand), la diffusion se comporte comme une

équation de la chaleur inverse dans la direction du gradient, et une équation de la chaleur dans la direction orthogonale au gradient. De nombreuses autres EDP ont été proposées, que ce soit pour le débruitage ou l'analyse de formes. Par exemple, le mouvement par courbure moyenne introduit en traitement d'images par [Kimia et al., 1995], est régi par l'équation :

$$\frac{\partial u}{\partial t} = |\nabla u| \text{div}\left(\frac{\nabla u}{|\nabla u|}\right). \tag{2}$$

Il s'agit d'une diffusion pure dans la direction orthogonale au gradient. Mentionnons aussi la diffusion proposée par Sochen, Kimmel et Maladi [Sochen et al., 1998] en lien avec les surfaces minimales :

$$\frac{\partial u}{\partial t} = \text{div}\left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right). \tag{3}$$

Il y a tant d'EDP paraboliques qui ont été proposées pour la restauration d'images qu'il est très facile de s'y perdre. En adoptant une approche axiomatique, Alvarez *et al.* [Alvarez et al., 1993] ont prouvé que les EDP invariantes par changement de contraste et isotropes (propriétés qui sont naturelles en traitement d'images) sont de la forme (voir aussi [Chen et al., 1991]) :

$$\frac{\partial u}{\partial t} = F(\text{div}\left(\frac{\nabla u}{|\nabla u|}\right), t)|\nabla u|. \tag{4}$$

Si l'on exige de plus l'invariance affine, cette famille se réduit à une unique équation appelée mouvement par courbure affine (see [Guichard et al., 2004]).

Les équations aux dérivées partielles en traitement d'images sont naturellement liées à la théorie de l'espace échelle (*scale-space*) [Witkin, 1983, Koenderink, 1984], qui associe une image à une famille d'images de plus en plus lisses. Alvarez *et al.* [Alvarez et al., 1993] ont montré que tout espace échelle causal correspond à une EDP. Le lecteur pourra se reporter aux références [Weickert, 1998, Guichard et al., 2004] pour plus de détails.

Morphologie Mathématique. La morphologie mathématique est une théorie d'analyse d'image développée par Matheron et Serra dans les années 1960, alors que Serra était chargé de faire un inventaire détaillé du gisement de fer en Lorraine (l'intéressante histoire de la naissance de la morphologie mathématique est racontée dans [Matheron and Serra, 2002]). Le point central de cette théorie est l'invariance par changement de contraste, et une conséquence notable de cette propriété est que les opérateurs morphologiques peuvent être considérés de manière équivalente comme agissant sur des fonctions (des images) ou sur des ensembles (leurs ensembles de niveau). Par exemple, deux des opérateurs morphologiques les plus connus sont respectivement l'érosion et la dilatation :

$$\forall X \subset \mathbb{R}^2, \quad \mathcal{E}_r X = \left\{x \in \mathbb{R}^2, B(x,r) \subset X\right\}, \quad \mathcal{D}_r X = \left\{y \in \mathbb{R}^2, \exists x \in E,\ y \in B(x,r)\right\}, \tag{5}$$

où $r > 0$ et $B(x,r)$ est la boule euclidienne fermée de rayon r. Ces opérateurs sur les ensembles sont associés aux opérateurs sur les fonctions :

$$\forall u \in \mathbb{R}^{\mathbb{R}^2},\ \forall x \in \mathbb{R}^2, \quad E_r u(x) = \inf_{y \in B(0,r)} u(x+r), \quad D_r u(x) = \sup_{y \in B(0,r)} u(x+r). \tag{6}$$

En fait, il est équivalent d'appliquer chaque opérateur fonctionnel à une fonction f ou d'appliquer l'opérateur correspondant sur les ensembles de niveau supérieurs de f, $F_t = \{x \in \mathbb{R}^2, f(x) \geq t\}$, pour obtenir les ensembles U_t avec $t \in \mathbb{R}$, puis de reconstruire u par la formule :

$$\forall x \in \mathbb{R}^2, \quad u(x) = \sup\{t \in \mathbb{R},\ x \in U_t\}. \tag{7}$$

A partir des dilatations et des érosions, deux autres opérations élémentaires sont construites : les ouvertures (définies comme $D_r \circ E_r$) et les fermetures (définies comme $E_r \circ D_r$). Les opérateurs d'ensembles correspondants peuvent être écrits comme :

$$\forall X \subset \mathbb{R}^2, \quad \mathcal{O}_r X = \bigcup \{B(x,r),\ B(x,r) \subset X\}, \quad \mathcal{F}_r X = \left(\bigcup \{B(x,r),\ B(x,r) \subset X^C\} \right)^C, \quad (8)$$

où $X^C = \mathbb{R}^2 \setminus X$.

Des références utiles sur la morphologie mathématiques sont par exemple le livre de Matheron [Matheron, 1975] et ceux de Serra [Serra, 1982, Serra, 1988]. A partir des axiomes établis par Serra et Matheron, et des opérations élémentaires comme les érosions, dilatations, ouvertures et fermetures, la morphologie mathématique s'est considérablement développée pour devenir une théorie bien établie, qui fournit des algorithmes de débruitage d'images [Cheng and Venetsanopoulos, 1992, Vincent, 1993], de segmentation [Beucher and Lantuéjoul, 1979, Bouraoui et al., 2008], d'analyse de texture [Vanrell and Vitria, 1993, Hanbury and Serra, 2002], etc.

Méthodes variationnelles. Une autre approche du traitement d'images consiste à chercher une image qui minimise une certaine énergie. Typiquement, si on observe une image bruitée f, l'image reconstruite est solution du problème :

$$\min_u G(u) + F(u, f), \quad (9)$$

Le terme $G(u)$ est le terme de régularisation, c'est-à-dire qu'il force les minimiseurs à être assez réguliers, tandis que le terme de fidélité $F(u, f)$ force le minimiseur à être cohérent avec l'observation f. Un exemple classique de problème variationnel en traitement d'images est celui de Rudin-Osher-Fatemi (ROF) [Rudin et al., 1992] :

$$\min_u \int_{\mathbb{R}^2} |\nabla u(x)| dx + \lambda \int_{\mathbb{R}^2} |f(x) - u(x)|^2 dx \quad (10)$$

Ici, le terme de régularisation est la variation totale, qui a l'avantage de lisser les images sans rendre les bords flous. Le terme de fidélité quadratique est typique lorsqu'on travaille avec du bruit additif gaussien. Ce modèle est une amélioration du modèle de Tikhonov [Tikhonov and Arsenin, 1977] plus ancien où $\int_{\mathbb{R}^2} |\nabla u(x)|^2 dx$ remplace $\int_{\mathbb{R}^2} |\nabla u(x)| dx$. L'énergie de Tikhonov a le défaut d'imposer aux minimiseurs d'être dans l'espace de Sobolev H^1. Contrairement aux fonctions à variations bornées (qui forment l'espace $\mathrm{BV}(\mathbb{R}^2)$ dans lequel sont les solutions du problème ROF), les fonctions de l'espace H^1 ne peuvent pas avoir de discontinuités le long d'hypersurfaces, ce qui empêche de travailler avec des images de type cartoon, composées de régions lisses délimités par des bords nets. En fait, beaucoup d'autres choix de termes de régularisation ont été proposés : fonction de Huber, variation totale régularisée (ce qui revient à mesurer la surface du graphe de u), ou diverses normes L^p du gradient. Ces fonctionnelles sont convexes, ce qui les rend plus faciles à étudier et à minimiser que les fonctionnelles non-convexes générales. Pourtant, l'intérêt pour les termes de régularisation non convexes est croissant. Dans [Bar et al., 2006a, Bar et al., 2006b, Cai et al., 2008], la fonctionnelle de Mumford-Shah [Mumford and Shah, 1989], bien que conçue à l'origine pour la segmentation, est utilisée à des fins de restauration et fournit des résultats meilleurs que la variation totale. Dans [Nikolova, 2005], Nikolova étudie le comportement de termes de régularisation non-convexes : elle montre que les petites discontinuités sont réduites, tandis que les grandes sont accrues. De nombreuses fonctions non-convexes ont été proposées, mais leur minimisation reste un problème ouvert.

Les modèles variationnels sont particulièrement intéressants puisqu'ils fournissent un cadre simple pour résoudre des problèmes inverses mal posés tels que la déconvolution ou la super-résolution. Par exemple, une adaptation naturelle du modèle de Rudin-Osher-Fatemi à la déconvolution est :

$$\min_u \int_{\mathbb{R}^2} |\nabla u(x)| dx + \lambda \int_{\mathbb{R}^2} |f(x) - h \star u(x)|^2 dx \quad (11)$$

où la convolution avec h modélise le flou.

Champs de Markov et estimation Maximum a Posteriori. Dans un cadre discret, une manière de modéliser les dépendances entre les valeurs des pixels est d'interpréter une image comme un champ de Markov, comme cela est proposé dans l'article fondateur [Geman and Geman, 1984]. Dans ce cadre, la probabilité d'observer une valeur à un pixel x conditionnellement aux valeurs de tous les autres pixels dépend uniquement des valeurs dans un voisinage de x. Le théorème d'Hammersley-Clifford indique que la probabilité jointe d'observer une image u peut s'écrire comme $P(u) = \frac{1}{Z}\exp(-\sum_{c\in\mathcal{C}} E_c(u))$, où E_c est appelée énergie de clique.

Pour le débruitage avec bruit additif blanc gaussien, on observe f et on cherche une version restaurée u. La formule de Bayes donne :

$$P(u|f) = \frac{P(u)P(f|u)}{P(f)} \tag{12}$$

où $P(u) = \frac{1}{Z}\exp(-\sum_{c\in\mathcal{C}} E_c(u))$ est l'*a priori* sur les images (par exemple un *a priori* de régularité $\frac{1}{Z}\exp(-\sum_{i,j}(|u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|))$, et $P(f|u)$ est la probabilité de f conditionnellement à u (par exemple $\frac{1}{Z'}\exp\left(-\frac{\|f-u\|^2}{2\sigma^2}\right)$).

En prenant le logarithme de cette expression, on voit que $\log P(u|f)$ est la somme d'une énergie de clique et d'un terme de fidélité. L'estimation Maximum A Posteriori (MAP) de u consiste à trouver un maximiseur de $\log P(u|f)$. De manière équivalente, on cherche un minimiseur de $-\log P(u|f)$, quantité qui est la somme d'un terme de fidélité et d'un terme de régularisation : les problèmes variationnels peuvent souvent être vus comme des estimations Maximum A Posteriori et vice-versa.

Un intérêt notable de la littérature sur les champs de Markov est sa multitude de connexions avec la physique statistique. En particulier, le modèle d'Ising, qui provient de l'étude du ferromagnétisme et de l'antiferromagnétisme, a été importé en traitement d'images dans sa version ferromagnétique :

$$E(u) = -\sum_{(s,t)/s\sim t} -\beta u_s u_t - \sum_s u_s \tag{13}$$

avec $\beta > 0$, et u une image binaire. Cette énergie implique que les images les plus probables ont peu de sauts. Ce modèle est lié au modèle TVL1 étudié dans cette thèse dans la mesure où ils coincident sur les images binaires d'ensembles convexes (voir le Chapitre 3 de cette thèse). On renvoie le lecteur à [Maruani et al., 1995] pour une étude de ses propriétés.

Mentionnons que l'estimation Maximum A Posteriori n'est pas nécessairement le moyen le plus raisonnable d'aborder un problème de débruitage. Dans [Louchet and Moisan, 2008, Louchet, 2008], Louchet et Moisan plaident en faveur de l'estimation aux moindres carrés (*Least-Square Error*, LSE), plutôt que du Maximum a Posteriori (qui revient au modèle ROF). L'idée est de trouver u qui minimise $\mathbb{E}_{u'|f}\|u - u'\|^2$. Cela mène à une somme d'images u' pondérées en fonction de leur probabilité $P(u'|f)$. L'intérêt de cet estimateur est que ses propriétés sont bien meilleures que celles de l'estimateur MAP (voir [Louchet, 2008]) : en particulier, il ne souffre pas de l'effet de *staircasing* (voir par exemple [Nikolova, 2000, Nikolova, 2004b] pour une étude de l'effet de *staircasing*).

Aussi, cette estimation LSE permet de dresser un parallèle entre le débruitage par variation totale avec LSE, et les Moyennes Non Locales [Buades et al., 2005]. Cela est lié à l'interprétation des NL-Means comme agrégation d'estimateurs par Salmon and Le Pennec [Salmon and Le Pennec, 2009a, Salmon and Le Pennec, 2009b]. Il est intéressant de noter que, dans un autre contexte, celui de la seconde partie de cette thèse (en particulier les chapitres 9 et 10) nous observons aussi qu'une somme pondérée d'estimateurs est parfois meilleure qu'un minimiseur particulier.

Filtres à voisinage et méthodes à base de patchs. Plutôt que de faire la moyenne des pixels en fonction de leur proximité spatiale, certains auteurs ont proposé d'utiliser la proximité entre leurs valeurs. C'est essentiellement le principe du filtre de Yaroslavky [Yaroslavsky, 1985] :

$$YNF_{h,\rho}u(x) = \frac{1}{C(x)}\sum_{y\in B(x,\rho)} e^{-\frac{|u(x)-u(y)|}{h^2}}u(y). \tag{14}$$

où $C(x)$ est une constante de normalisation , et $B(x, \rho)$ est une fenêtre de rayon ρ centrée autour de x. Cette idée est aussi au cœur des filtres sigma [Lee, 1983] et SUSAN [Smith and Brady, 1997]. Avec le filtre bilatéral [Tomasi and Manduchi, 1998], on utilise à la fois la valeur des pixels et leur position pour calculer les poids. Tandis qu'avec ces méthodes les pixels comparés appartiennent à une fenêtre fixe ou à une boule centrée autour du pixel à estimer, certains auteurs ont proposé de choisir de manière adaptative l'ensemble des pixels avec lesquels calculer cette moyenne [Polzehl and Spokoiny, 2000, Polzehl and Spokoiny, 2003] [Lepski, 1990].

Inspirés par les filtres à voisinage et par le travail d'Efros et Leung [Efros and Leung, 1999] en synthèse de texture, Buades, Coll et Morel [Buades et al., 2005] ont proposé de comparer de petits voisinages autour de chaque pixel appelés patchs, plutôt que les valeurs des pixels. Le filtre "Non-Local Means" a constitué une percée en matière de débruitage d'images, et depuis cinq ou six ans la majorité des algorithmes de débruitage proposés s'appuie sur la notion de patch. L'algorithme BM3D [Dabov et al., 2007], qui se classe aujourd'hui parmi les meilleures méthodes de débruitage en est un exemple.

Ondelettes et parcimonie. Un autre *a priori* sur les images est de supposer qu'elles ont une représentation parcimonieuse dans une certaine base ou trame. Donoho et Johnstone [Donoho and Johnstone, 1994] ont montré qu'un seuillage dur est asymptotiquement optimal au sens minimax dans la classe des estimateurs diagonaux, et que le risque associé est d'autant plus faible que la représentation de l'image est parcimonieuse. De manière informelle, la parcimonie implique que le signal est concentré sur quelques grands coefficients, tandis que le bruit est uniformément réparti sur les différentes composantes. Ainsi, un simple seuillage permet de supprimer la plupart du bruit sans beaucoup dégrader le signal. Les méthodes à base d'ondelettes sont typiques de cette approche : puisque les coefficients d'ondelettes encodent les variations de l'image plutôt que chaque valeur absolue, les images régulières par morceaux ont une représentation creuse dans des bases d'ondelettes. Pourtant, les bases d'ondelettes n'exploitent pas la géométrie des images, et notamment le fait que les grands coefficients sont répartis le long de courbes régulières. Des trames plus complexes ont été proposées pour tirer parti de cette propriété : les curvelettes [Starck et al., 2002], bandelettes [Le Pennec and Mallat, 2005, Mallat and Peyré, 2007] or contourlettes [Do and Vetterli, 2005] entre autres.

Plutôt que de considérer un dictionnaire fixé à l'avance, certaines méthodes consistent à apprendre un dictionnaire adapté à une base d'images. Etant donné un dictionnaire surcomplet, la *matching pursuit* [Mallat and Zhang, 1993, Bergeaud and Mallat, 1995] permet de trouver une sélection d'atomes qui mènent à une représentation aussi parcimonieuse que possible d'une famille de signaux. L'analyse en composantes principales (PCA) permet de trouver une base qui concentre sur les premiers vecteurs la plupart de l'énergie de la plupart des signaux d'une certaine famille. Ces outils, et la parcimonie en général ont produit de nombreux algorithmes puissants de débruitage. En particulier, l'adaptation de ces idées aux patchs dans les images dans [Aharon et al., 2006] et sa variante utilisant la parcimonie structurée [Mairal et al., 2009] atteint l'état de l'art en matière de débruitage.

La frontière entre toutes ces familles de modèles n'est pas aussi nette qu'il n'y paraît, et de nombreux algorithmes ont un équivalent dans une autre catégorie. Par exemple :

- L'équivalence asymptotique entre opérateurs morphologiques et EDP est établie dans [Lax, 1965] pour les érosions et les dilatations, et plus généralement dans [Alvarez et al., 1993] (voir [Guichard et al., 2004] pour un exposé complet de la théorie).
- Les modèles variationnels mènent souvent à des EDP puisque leur minimisation est liée au flot de leur gradient : par exemple, Andreu *et al.* [Andreu et al., 2001b] ont étudié le flot de la variation totale en lien avec le problème ROF.
- La limite continue du modèle de segmentation bayésien [Geman and Geman, 1984] mène à la fonctionnelle de Mumford-Shah[Mumford and Shah, 1989] (voir [Chambolle, 2000]).
- Steidl *et al.* ont montré qu'en dimension 1, le débruitage par variation totale et le seuillage par

ondelettes sont équivalents [Steidl et al., 2004].
– Le filtrage bilatéral est asymptotiquement équivalent à une étape de l'équation de Malik et Perona (voir [Durand and Dorsey, 2002, Buades et al., 2005]).

Aspects Morphologiques du modèle TVL1

La première partie de cette thèse explore une passerelle similaire entre un problème variationnel et la morphologie mathématique. Elle est dédiée à l'étude du modèle TVL1 qui consiste, étant donnée $f \in L^1(\mathbb{R}^2)$, à résoudre :

$$\min_{u \in \mathbf{BV}(\mathbb{R}^2)} \int |Du| + \lambda \int_{\mathbb{R}^2} |f(x) - u(x)| dx. \tag{15}$$

Ce modèle a été introduit par Alliney pour les signaux à 1 dimension [Alliney, 1992, Alliney, 1996, Alliney, 1997] : l'auteur montre que dans ce cas, un filtre médian récursif permet de calculer son minimiseur. L'intérêt de ce modèle pour le traitement des images a été mis en évidence par Nikolova dans [Nikolova, 2002] : elle montre que le terme de fidélité non régulier permet une meilleure détection et une meilleure suppression des valeurs aberrantes que des termes de fidélité réguliers. En conséquence, ce modèle est particulièrement bien adapté à la suppression du bruit impulsionnel, et l'article [Nikolova, 2004a] en donne des illustrations numériques frappantes.

Dans [Chan and Esedoglu, 2005] le modèle est étudié dans un cadre continu. Chan et Esedoglu pointent le fait que malgré des similarités frappantes avec le modèle ROF, le modèle TVL1 a des propriétés bien spécifiques. Par exemple, suivant [Strong et al., 1996, Meyer, 2001] ils calculent la solution du problème quand la donnée est une fonction f caractéristique d'un disque : $f = \mathbb{1}_{B(0,R)}$. La solution est :

$$u = \begin{cases} \mathbb{1}_{B(0,R)} & \text{si } \lambda \geq \frac{2}{R}, \\ c\mathbb{1}_{B(0,R)}, & \forall c \in [0,1] \text{ et } \lambda = \frac{2}{R}, \\ 0 & \text{autrement.} \end{cases} \tag{16}$$

En d'autres termes, il n'y a pas de perte de contraste contrairement à ROF (voir [Meyer, 2001]), et certains objets disparaissent soudainement en fonction de leur échelle. A la suite de [Chambolle, 2004a, Alter et al., 2005a], où le modèle ROF est relié à une famille de problèmes géométriques, Chan et Esedoglu remarquent que l'énergie du modèle TVL1 peut être écrite comme la somme d'énergies sur les ensembles de niveau :

$$\mathbb{E}(u) = \int_{-\infty}^{+\infty} \text{Per}\,\{x, u(x) > \mu\} + \lambda |\{x, u(x) > \mu\} \Delta \{x, f(x) > \mu\}| d\mu, \tag{17}$$

où Per X désigne le périmère de $X \subset \mathbb{R}^2$, $|X|$ sa mesure de Lebesgue deux-dimensionnelle, et $X \Delta Y = (X \setminus Y) \cup (Y \setminus X)$. Cette formule est d'une importance particulière, puisqu'elle confère au modèle la propriété d'invariance par changement de contraste. La même observation a été faite simultanément par Darbon dans [Darbon, 2005], où le lien avec la morphologie mathématique est mis en évidence pour la première fois. Cela permet aussi à Darbon et Sigelle de proposer un algorithme de graph-cut qui a fourni pendant de nombreuses années le moyen le plus rapide pour résoudre le problème TVL1 [Darbon and Sigelle, 2006] (jusqu'au travail de Pock *et al.* [Pock et al., 2008] qui ont adapté l'algorithme [Aujol et al., 2006] aux GPUs). Mentionnons aussi le travail de Yin *et al.* [Yin et al., 2007b] qui ont étudié la monotonie du problème sur les ensembles de niveau ainsi que l'invariance par changement de contraste. Dans la littérature, le modèle TVL1 ou ses variantes a été appliqué à des problèmes très divers : suppression du bruit impulsionnel [Nikolova, 2004a, Koko and Jehan-Besson, 2010, Liu et al., 2010], segmentation [Chan et al., 2006, Bresson et al., 2007], détection de motifs en imagerie biomédicale [Cunha et al., 2007], registration 3D MRI [Chen et al., 2005], décomposition cartoon-texture [Aujol et al., 2006, Yin et al., 2007a, Haddad, 2007], calcul du flux optique dans les vidéos [Wedel et al., 2008a, Wedel et al., 2008b], analyse de formes [Vixie et al., 2010]...

Notre contribution est d'étudier le problème géométrique dans les détails, de manière à donner plus d'informations sur les solutions de TVL1. Après avoir précisé l'équivalence avec le problème géométrique, nous caractérisons les ensembles qui disparaissent soudainement de la même façon que le disque (16). En appliquant des résultats de la théorie géométrique de la mesure [Ambrosio, 1997, Ambrosio and Paolini, 1998], nous montrons que la frontière de la solution est régulière et que sa courbure est égale à $-\lambda$ à l'intérieur de la donnée F et $+\lambda$ à l'extérieur (quand cette frontière coïncide avec celle de F, la courbure est bornée en valeur absolue par λ). Puis nous observons que lorsque la donnée est un ensemble C convexe, le problème se simplifie en :

$$\min_{X \subset C} \text{Per } X - \lambda |X| \tag{18}$$

Nous en déduisons que les solutions de TVL1 sont données par une ouverture de rayon $\frac{1}{\lambda}$, suivie d'un seuillage sur le rapport périmètre / aire. Plus précisement, si l'on écrit $C_{1/\lambda} = \mathcal{O}_{1/\lambda} C$ (où $\mathcal{O}_r$ est définie dans (8)), on montre que :

- si $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} < \lambda$, la solution est $C_{1/\lambda}$,
- si $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} > \lambda$, la solution est $\emptyset$,
- si $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} = \lambda$, $C_{1/\lambda}$ et $\emptyset$ sont solutions.

Insistons sur le fait que ce résultat s'appuie pour une large mesure sur les travaux d'Alter, Caselles, Chambolle [Alter et al., 2005a, Alter et al., 2005b] et plus généralement sur ceux de Bellettini, Novaga, Paolini [Bellettini et al., 2001, Bellettini et al., 2002, Bellettini et al., 2005a]. En quelques mots, l'étude des ensembles calibrables, i.e. des ensembles qui évoluent à vitesse constante par le flux de la variation totale, montre que les ensembles calibrables sont solutions du problème variationnel suivant :

$$\min_{X \subset C} \text{Per } X - \lambda_C |X| \quad \text{ with } \lambda_C = \frac{\text{Per } C}{|C|}. \tag{19}$$

En observant que les ensembles de niveau F_t de la solution du problème ROF :

$$\min_{u \in \mathbf{BV}(\mathbb{R}^2)} \int |Du| + \frac{\lambda}{2} \int_{\mathbb{R}^2} (u(x) - \mathbb{1}_C(x))^2 dx \tag{20}$$

sont solutions de la famille de problèmes géométriques :

$$\min_{X \subset C} \text{Per } X - \lambda(1-t)|X|, \tag{21}$$

Alter *et al.* étudient la dépendance en λ de cette famille de problèmes, par des arguments d'approximation, ils caractérisent la courbure des ensembles C qui sont invariants par (18). Nous avons utilisé cette caractérisation en lien avec la caractérisation des ouvertures dans le plan par la courbure pour obtenir le résultat, mais nous aurions pu utiliser le champ de vecteurs construit dans [Alter et al., 2005b] pour le flot de la variation totale comme "calibration" (voir le chapitre 2). L'étude du flot de la variation totale fournit aussi les solutions exactes du problème ROF pour des données de la forme $\mathbb{1}_C$ (voir [Alter et al., 2005b]), et des ouvertures sont aussi impliquées (d'une manière plus compliquée).

Dans le cas de TVL1, la classe de fonctions pour lesquelles on obtient des solutions exactes est légèrement plus grande (fonctions quasi-convexes au lieu d'indicatrices d'ensembles convexes) que pour ROF, mais surtout, le lien avec les ouvertures est clair. Cela vient compléter l'observation de Darbon [Darbon, 2005] que TVL1 est un opérateur morphologique. Mentionnons aussi que le résultat sur TVL1 et les ouvertures peut être vu comme un cas particulier de l'étude générale, mais difficile, des problèmes $TV + L^p$ mené par Allard ([Allard, 2008], voir aussi [Allard, 2007] and [Allard, 2009]), bien que nous l'ayons obtenu de manière indépendante. Pour l'illustrer numériquement, nous avons construit un algorithme s'appuyant sur des ouvertures et la Transformée Rapide en Ensembles de Niveaux (FLST, *Fast*

Level-Set Transform, [Monasse and Guichard, 2000]), qui donne la solution exacte dans les cas simples, et une solution approchée autrement, en un temps de calcul très compétitif (voir le Chapitre 4).

Une autre contribution de cette thèse est d'interpréter la comportement de TVL1 dans le problème de décomposition cartoon-texture à la lumière de la morphologie mathématique. On commence par comparer les décompositions fournies par TVL1 avec celles du modèle TV+G, le premier modèle de décomposition cartoon-texture proposé [Meyer, 2001]. On sait depuis [Haddad, 2007, Yin et al., 2007a] que TVL1 se comporte bien pour la décomposition cartoon-texture. Nous observons qu'avec le modèle TV+G, les traces des contours apparaissent dans la partie texture, ainsi qu'une sorte de halo, ce qui nous conduit à étudier ce modèle plus en détail.

Une idée couramment répandue à propos des normes favorisant les oscillations est que si une image est composée d'une partie cartoon, disons l'indicatrice d'un ensemble régulier, et d'une partie oscillante, disons une sinusoïde, le modèle sera capable de la décomposer puisque la variation totale des indicatrices d'ensembles réguliers (resp. sinusoïdes) est faible (resp. forte) et que la norme G d'une sinusoïde (resp. indicatrices d'ensembles réguliers) est faible (resp. forte). Pourtant, il n'est pas trivial que la décomposition soit parfaite, et à notre connaisance, aucun exemple explicite de décomposition parfaite n'a été donné jusqu'à présent. En utilisant le cadre à symétrie radiale introduit dans [Haddad, 2007] et la caractérisation du sous-différentiel de la norme G établie au Chapitre 1, nous prouvons que l'indicatrice d'un disque plus une sinusoïde (légèrement perturbée) est parfaitement décomposée par le modèle TV+G. La bonne nouvelle est donc que le modèle fait ce qu'il est supposé faire, au moins sur un exemple. Toutefois, un examen approfondi de cet exemple montre qu'il est nécessaire que la texture ne s'annule pas au voisinage de la frontière du disque.

Nous expliquons pourquoi les traces des bords et des halos peuvent parfois apparaître dans des images plus générales. Dans le cas de l'indicatrice d'un convexe, et avec quelques hypothèses techniques supplémentaires sur la partie texture, nous montrons qu'une décomposition où la partie texture s'annule au voisinage des bords ne peut pas être solution du modèle TV+G. Par conséquent, il nous semble que l'apparition des bords dans la partie texture est une propriété intrinsèque des décompositions TV+G.

Après avoir étudié les propriétés des décompositions TV+G, nous retournons au modèle TVL1. Nous exhibons un exemple similaire de décomposition parfaite d'une indicatrice et d'une sinusoïde. Les limites mises en évidence pour le modèle TV+G n'apparaissent pas avec le modèle TVL1 : les parties cartoon et textures peuvent être totalement indépendantes. Enfin, pour expliquer le bon comportement du modèle dans les décompositions cartoon-texture nous les relions aux granulométries utilisées en morphologie mathématique : tandis que le modèle TV+G s'appuie sur la notion de fréquence, le modèle TVL1 travaille avec la notion d'échelle.

Puisque l'étude du "problème stationnaire" a donné un peu d'intuition sur l'effet du paramètre de fidélité λ, nous étudions un modèle où λ varie spatialement, comme dans [Gilboa et al., 2006] pour le modèl ROF. Régler ce paramètre localement permet de préserver les régions texturées comme dans [Gilboa et al., 2006] ou de supprimer le bruit impulsionnel comme dans [Koko and Jehan-Besson, 2010] sans modifier les pixels qui n'ont pas été affectés par le bruit. Nous examinons d'abord le modèle d'un point de vue théorique : nous montrons que le modèle se comporte comme on s'y attend de manière intuitive. Le problème fonctionnel est équivalent à une famille de problèmes géométriques, la valeur de λ contrôle la courbure de la solution, et en choisissant λ assez grand dans certaines régions, il est possible de préserver des formes très irrégulières, même avec des coins.

Ensuite, nous illustrons l'utilisation du modèle TVL1 adapatif dans le contexte de la suppression du bruit impulsionnel. La plupart des méthodes de suppression du bruit impulsionnel sont divisées en deux étapes [Chen and Wu, 2001a, Pok et al., 2003, Crnojevic et al., 2004, Chan et al., 2004, Dong et al., 2007] : une phase de détection et une phase de débruitage. On rappelle le principe de quelques détecteurs de bruit, notamment le *Adaptive Center-Weighted median Filter* (ACWMF) [Chen and Wu, 2001a], le *Rank Ordered Absolute differences* (ROAD) [Garnett et al., 2005] et le *Rank Ordered Logarithmic Differences* (ROLD) [Dong et al., 2007] avant de montrer le résultat du débruitage utilisant TVL1 adaptatif.

La performance est comparable aux méthodes de l'état de l'art telles que [Chan et al., 2004] mais pas

véritablement meilleure. Pourtant, un avantage de ce cadre est sa simplicité. En fait la principale difficulté de la suppression du bruit impulsionnel est la détection du bruit. Une fois cette étape proprement réalisée, la plupart des méthodes de débruitage donnent des résultats similaires.

La fin de cette première partie est dédiée à la justification des observations empiriques du Chaptitre 4, c'est-à-dire que les solutions de TVL1 quand la variation totale est calculée en utilisant la norme ℓ^1 (plutôt que ℓ^2) du gradient (par exemple dans les méthodes de type graph-cut) est donnée par une ouverture où les boules sont remplacées par des carrés. En fait, on montre que pour toute anisotropie cristaline, la solution est donnée par une ouverture avec la forme de Wulff suivie d'un test sur le rapport périmètre / aire. Ce résultat est une conséquence directe des travaux de Belletini *et al.* [Bellettini et al., 2001].

Méthodes Non-Locales pour la restauration d'Images

La seconde partie de cette thèse est consacrée au filtre *Non-Local Means* [Buades et al., 2005] à partir duquel ont émergé quelques unes des méthodes de débruitage les plus performantes. On suppose que l'on veut retrouver l'image $\mathbf{f}$ qui a été contaminée par du bruit additif blanc Gaussien :

$$\mathbf{u}(x) = \mathbf{f}(x) + \boldsymbol{\varepsilon}(x) \text{ où } \varepsilon(x) \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.} \tag{22}$$

Une méthode simple pour retrouver $\mathbf{f}(x)$ est de faire la moyenne de N observations indépendantes et identiquement distribuées (i.i.d.) du pixel bruité $\{\mathbf{u}_i(x),\ i = 1 \dots N\}$. Ainsi on obtient un estimateur sans biais de $\mathbf{f}(x)$ de variance :

$$\mathrm{Var}\left(\mathbf{f}(x) - \frac{1}{N}\sum_{i=1}^{N} \mathbf{u}_i(x)\right) = \frac{\sigma^2}{N}. \tag{23}$$

En pratique, une seule réalisation de $\mathbf{u}(x)$ est observée. Une solution courante est de considérer que l'image ne varie pas beaucoup dans un voisinage de x (au moins dans une direction) de manière à faire la moyenne des intensités dans ce voisinage. Cela revient *grosso modo* à ce que font la plupart des méthodes présentées ci-dessus. La limite de cette approche est claire : on considère en général de petits voisinages de sorte que les valeurs $\mathbf{f}(y)$ ne varient pas beaucoup, et la réduction de variance qui en résulte est donc limitée.

La belle idée de Buades *et al.* est de tirer parti des auto-similarités à travers toute l'image : en calculant la moyenne de motifs qui se répètent, on peut réduire de manière drastique la variance du bruit sans rendre flous les détails. Ces similarités peuvent apparaître à grande distance, d'où le terme "non-local" pour qualifier le problème. En fait, un ingrédient crucial de cette méthode est la comparaison d'un voisinage de x avec celui de y plutôt que de comparer simplement leurs intensités.

Ces voisinages, habituellement carrés, sont appelés des patchs, et ont rapidement été adoptés par la communauté du traitement d'images depuis le travail d'Efros et Leung en synthèse de textures [Efros and Leung, 1999]. Depuis la percée des Non-Local Means (ou du filtre UINTA [Awate and Whitaker, 2006] proposé simultanément), de nombreuses variantes ont été proposées : certains auteurs donnent une interprétation variationnelle au filtre [Brox and Cremers, 2007, Kindermann et al., 2005, Buades et al., 2006, Gilboa and Osher, 2007, Gilboa and Osher, 2008, Azzabou et al., 2007c], ce qui les amène à itérer l'algorithme. D'autres auteurs se sont intéressés à la notion de patch, et ont interprété l'algorithme comme une diffusion dans l'espace des patchs [Szlam, 2006, Tschumperlé and Brun, 2009, Peyré, 2008, Peyré, 2009, Singer et al., 2009] : le débruitage d'image revient à l'équation de la chaleur ! Puisque l'algorithme Non-Local Means original est très coûteux en temps de calcul, en obtenir une version rapide est un problème difficile. Une approximation des Non-Local Means peut être rapidement calculée en utilisant des techniques de préselection [Mahmoudi and Sapiro, 2005, Bilcu and Vehvilainen, 2008], ou un arbre d'agrégats pour calculer les patchs similaires [Brox et al., 2008], ou un sous-échantillonnage des distances de patch [Pang et al., 2009]. Des calculs rapides et exacts ont été proposés à partir de la Transformée de Fourier Rapide [Wang et al., 2006], ou des tables de sommes cumulées [Darbon et al., 2008].

Précisons les notations. La sortie du filtre est :

$$\mathrm{NL}\,\mathbf{u}(x) = \frac{\sum_{y\in\Omega} e^{-\frac{\|U(x)-U(y)\|^2}{2h^2}}\mathbf{u}(y)}{\sum_{y'\in\Omega} e^{-\frac{\|U(x)-U(y')\|^2}{2h^2}}}. \tag{24}$$

où Ω est l'image entière (souvent remplacée par une fenêtre de recherche autour de x), $U(x)$ désigne les valeurs des pixels dans un patch carré de taille $s \times s$ autour de x, et la distance est la distance euclidienne normalisée :

$$\|U(x) - U(y)\|^2 = \frac{1}{s^2} \sum_{|j|\leq\frac{s-1}{2}} (\mathbf{u}(x+j) - \mathbf{u}(y+j))^2.$$

Le paramètre $h > 0$ est un paramètre de largeur de bande qui détermine l'importance relative des pixels dans la moyenne, en fonction de leur similarité.

Au fur et à mesure que diverses améliorations des Non-Local Means ont été proposées, les limitations du filtre ont été progressivement comprises par la communauté : la sélection des patchs n'est pas optimale, et Mahmoudi et Sapiro [Mahmoudi and Sapiro, 2005] ont proposé d'ajouter une information sur le gradient dans la sélection des pixels. Une autre façon de rendre la sélection des patchs plus pertinente est d'utiliser une analyse en composantes principales comme le proposent Azzabou *et al.* [Azzabou et al., 2007a] ou Tasdizen [Tasdizen, 2009] (en particulier si l'intensité du bruit est très forte). L'influence de la fenêtre de recherche sur la qualité du résultat est mise en évidence par Kervrann et Boulanger [Kervrann and Boulanger, 2006] qui choisissent localement sa taille en fonction d'un compromis biais-variance. Le choix de la fonction exponentielle pour les poids est remis en question par Goossens *et al.* dans [Goossens et al., 2008], et l'importance trop grande donnée au pixel x dans sa propre estimation est discutée dans [Salmon, 2010b]. De plus, les Non-Local Means échouent quand il n'y a pas de patch similaire et certains auteurs ont proposé des astuces pour résoudre ce problème [Gilboa and Osher, 2007, Brox and Cremers, 2007, Brox et al., 2008, Louchet and Moisan, 2010, Salmon and Strozecki, 2010a].

Toutes ces bribes d'informations sont plutôt dispersées dans la littérature, et nous proposons dans le Chapitre 8 une "visite guidée" des Non-Local Means qui met en évidence certaines de ces propriétés. Le principe de cette visite est d'étudier des modèles très simples pour donner de l'intuition au lecteur qui ne serait pas familier avec les Non-Local Means. En particulier, nous illustrons le fait que :

- une grande taille de patch tend à rendre flous les objets,
- les grands patchs sont plus robustes au bruit,
- il y a une perte de contraste qui dépend de la répétition de chaque motif,
- moins les détails sont contrastés, plus ils sont dégradés, et cette relation est fortement non linéaire,
- même les images périodiques sont dégradées,
- la fenêtre de recherche a un impact sur la qualité visuelle du résultat,
- un poids à support compact plutôt qu'exponentiel permet de réduire le biais.

Si l'on décompose l'erreur du filtre en la somme d'un terme de biais et de variance, le biais est relié à une forme de régularité que nous essayons de caractériser dans l'espace des patchs.

Toutes ces considérations plaident en faveur d'un choix local des paramètres, puisque la régularité mise en évidence est fortement dépendante de la position de l'image.

Nous construisons d'abord un oracle qui est un Non Local Means avec des poids uniformes (similaire à la version des Non-Local Means de [Brox and Cremers, 2007], à la différence que le nombre de pixels est choisi à partir de l'image sans bruit) : il montre le comportement idéal d'un filtre à choix local de h. Etonnament, la distinction entre textures et régions lisses n'est pas le critère prépondérant de la décision de cet oracle.

En effet, le critère le plus important est si un pixel appartient ou non à un patch qui intersecte un bord. Les patchs qui sont les plus proches des bords, ou les détails très contrastés, ont habituellement peu de copies similaires , de sorte que les moyennes non locales laissent un halo de bruit. L'oracle force un paramètre h élevé le long des bords pour réduire ce halo, et un algorithme réel devrait en faire de même. Nous construisons donc un alorithme pratique en nous appuyant sur l'estimateur sans biais du

risque (SURE)[Stein, 1981], proposé pour la première fois dans le contexte des Non Local Means par [Van De Ville and Kocher, 2009]. Puisque cet estimateur du risque n'est pas robuste lorsqu'on l'utilise ponctuellement, nous proposons de régulariser le risque par une moyenne locale, et ensuite de garder la valeur du paramètre h qui minimise cet estimation régularisée. Un algorithme pratique est donné, qui nécessite un temps de calcul modéré, et de nombreuses expériences sont montrées. Cette procédure permet de réduire le halo près des bords et de mieux préserver les détails faiblement contrastés qu'avec les Non-Local Means. Pourtant, la décision est un peu brutale, et quand l'estimation du risque est mauvaise, de petites taches de bruit peuvent rester, comme avec les Non Local Means classiques. Inspiré par le travail de [Salmon and Le Pennec, 2009b], nous utilisons une agrégation à poids exponentiels (EWA) pour créer une combinaison convexe locale des filtres avec différentes valeurs de h en fonction de leur risque estimé. Cette procédure préserve bien mieux les textures que la méthode précédente, ne souffre pas du problème de halo et ne laisse pas de tache de bruit. Les images produites semblent légèrement plus bruitées, mais sont plus agréables. Le PSNR est légèrement inférieur à celui de la méthode précédente, mais le résultat visuel est convaincant.

Une autre variante des Non-Local Means est proposée dans le dernier chapitre, qui est le fruit d'une collaboration avec Joseph Salmon et Charles Deledalle. Le but de cet algorithme est de supprimer l'effet de patch rare, c'est-à-dire le halo de bruit qui entoure les bords contrastés ou les détails. L'idée provient du travail de Salmon et Strozecki [Salmon and Strozecki, 2010b], où il est proposé de translater les patchs de manière à obtenir des copies similaires près des bords. Ici, nous poussons l'idée plus loin en remplaçant les patchs décentrés par des formes arbitraires. L'idée est d'utiliser des formes qui peuvent s'introduire dans les régions les plus étroites de manière à trouver un maximum de pixels similaires. Nous proposons un algorithme s'appuyant sur la Transformée de Fourier Rapide pour calculer le résultat des Non-Local Means avec une forme arbitraire, indépendamment de la taille de la forme.

En utilisant un dictionnaire de formes fixé à l'avance, nous obtenons autant d'estimateurs que le nombre de formes. Comme dans le chapitre précédent, le fait de ne garder que l'estimateur qui minimise l'estimée du risque est trop brutal, c'est pourquoi nous agrégeons les résultats en utilisant la pondération basée sur la variance [Salmon and Strozecki, 2010b] ou l'agrégation à poids exponentiels [Leung and Barron, 2006]. L'algorithme résultant est rapide, et produit de très belles images.

Description des chapitres

Partie I : Morphological Aspects of the TVL1 Model

Chapter 1 : Notations and Preliminaries. Le but de ce chapitre est de rappeler quelques propriétés élémentaires des fonctions à variations bornées, ainsi qu'une formule de Green étendue, due à [Anzellotti, 1983] qui joue un rôle crucial dans cette thèse. Le sous-différentiel d'une fonction convexe s.c.i. positivement homogène est caractérisé, ce qui servira dans l'étude des modèles TVL1 et TV+G. Enfin, une brève introduction au problème de Cheeger et aux ensembles calibrables est donnée.

Chapter 2 : How to study the TVL1 problem? Ce chapitre résume différentes approches du modèle TVL1 proposées dans la littérature : caractérisation utilisant la norme G dans [Kindermann et al., 2006], le problème dual dans [Haddad, 2007], la *flat norm* dans [Morgan and Vixie, 2007], et une méthode inspirée des calibrations de la théorie géométrique de la mesure exposée dans ce chapitre. Nous montrons que toutes ces méthodes peuvent être interprétées comme les équation d'Euler-Lagrance pour la fonctionnelle TVL1. La "méthode des calibrations" exposée ici est liée aux travaux d'Alter, Bellettini, Caselles, Chambolle, Novaga et Paolini. Elle consiste à utiliser le champ de vecteur construit dans [Alter et al., 2005b] pour le flot de la variation totale pour montrer qu'une fonction est optimale pour TVL1. Une fois que le champ de vecteur est construit, la preuve est élémentaire.

Chapter 3 : The TVL1 model: a geometric point of view. Le problème TVL1 peut être étudié en utilisant le principe de Cavalieri, en décomposant une image en ses ensembles de niveau. D'un problème fonctionnel, nous sommes amenés à étudier une famille de problèmes géométriques. Nous proposons une preuve dans le cadre continu de cette équivalence remarquée pour TVL1 dans [Chan and Esedoglu, 2005, Darbon, 2005]. Cette équivalence rend le modèle TVL1 invariant par changement de contraste. En fait, comme cela est remarqué dans [Darbon, 2005], TVL1 est un opérateur morphologique.

Plutôt que d'étudier le problème fonctionnel, nous nous intéressons au problème géométrique. A l'aide d'arguments très simples, nous montrons que les ensembles qui disparaissent soudainement, comme le disque [Chan and Esedoglu, 2005], sont nécessairement Cheeger en eux-mêmes, et si l'ensemble est convexe, cette condition est suffisante. De plus, les ensembles calibrables généraux disparaissent soudainement.

L'approche géométrique est d'autant plus intéressante qu'elle permet d'utiliser des théorèmes puissants de la théorie géométrique de la mesure. Les lignes de niveau sont lisses et leur courbure est bornée par λ. Dans le cas d'un ensemble donné convexe, l'énergie se reformule en remarquant que la solution sera nécessairement à l'intérieur de l'ensemble. En nous appuyant sur les résultats de [Andreu-Vaillo et al., 2002, Alter et al., 2005a] pour le flot de la variation totale et la caractérisation des ensembles convexes calibrables, nous montrons que la solution du problème géométrique est donnée par une ouverture de rayon $1/\lambda$, et un test sur le rapport périmètre / aire.

Cela nous conduit à discuter le lien entre la morphologie mathématique et le problème TVL1. Dans le cas non-convexe, seule une borne sur les solutions est obtenue, mais des "candidats raisonnables" pour la solution de TVL1 sont proposés à partir de filtres alternés séquentiels. La validité d'un tel choix est examinée empiriquement dans le Chapitre 4.

Chapter 4 : Algorithms and numerical experiments. Le but de ce chapitre est d'appuyer les conclusions théoriques du Chapitre 3 à l'aide d'expériences numériques. Nous rappelons d'abord quelques algorithmes pour résoudre le problème TVL1. Puis, nous en proposons deux nouveaux. Le premier est naïf et lent : il consiste à appliquer un filtre alterné séquentiel puis de seuiller les composantes connexes de chaque ensemble de niveau. Le second consiste à appliquer la même tâche en utilisant l'arbre des formes fourni par la transformée en ensembles de niveau rapide (*Fast Level Set Transform*, FLST) [Monasse and Guichard, 2000]. Cela fournit un algorithme rapide, comparable à [Darbon and Sigelle, 2006]. Les expériences numériques sur des images binaires et naturelles montrent

que les algorithmes proposés produisent des résultats qui sont très proches des véritables solutions de TVL1 (fournies par descente de gradient par exemple).

Pour comparer notre méthode avec une minimisation exacte de TVL1 qui satisfait la formule de la coaire, nous utilisons l'algorithme de Darbon-Sigelle [Darbon and Sigelle, 2006]. Cet algorithme est basé sur un schéma anisotrope pour la variation totale. Une propriété remarquable est que si l'on remplace dans notre algorithme les ouvertures avec des boules par des ouvertures avec des carrés (et le calcul du périmètre associé), nous obtenons aussi des résultats qui sont très proches du modèle TVL1 avec variation totale anisotrope ! Cela suggère que la propriété est encore vraie pour les périmètres anisotropes, ce que nous étudions dans le Chapitre 7.

Chapter 5 : Cartoon, Textures and Granulometries. Le lien avec la morphologie mathématique mis en évidence dans les Chapitres 3 et 4 pour le modèle TVL1 nous permet d'interpréter son comportement dans le problème de décomposition cartoon-texture. D'abord, nous rappelons le principe de telles décompositions, et nous montrons des expériences numériques comparant les résultats fournis par TVL1 avec ceux fournis par TV+G, le premier modèle de décomposition cartoon-texture proposé [Meyer, 2001]. Avec le modèle TV+G, nous observons que des traces des contours apparaissent dans la partie texture, ainsi qu'une sorte de halo, ce qui nous amène à étudier ce problème en détail.

En utilisant le cadre radialement symétrique de [Haddad, 2007] et la caractérisation du sous-différentiel de la norme G établie dans le Chapitre 1, nous montrons qu'une fonction indicatrice d'un disque plus une sinusoïde (légèrement perturbée) est parfaitement décomposée par le modèle TV+G. Au passage, nous remarquons dans cet exemple que la texture ne peut pas s'annuler au voisinage de la frontière du disque. Nous expliquons pourquoi des traces des contours et du halo peuvent apparaître dans des images plus générales. Nous montrons que dans le cas de l'indicatrice d'un convexe et avec des hypothèses techniques supplémentaires sur la partie texture, une décomposition où la partie texture s'annule au voisinage des contours ne peut pas être solution du modèle TV+G.

Après avoir étudié les propriétés du modèle TV+G, nous revenons au modèle TVL1. A l'aide de l'équation d'Euler-Lagrange, nous montrons un exemple similaire de décomposition parfaite d'une indicatrice et d'un sinus. Nous constatons que les limitations du modèle TV+G ne s'appliquent pas à TVL1. Enfin, pour expliquer le bon comportement de ce modèle pour la décomposition cartoon-texture, nous le relions aux granulométries utilisées en morphologie mathématique : tandis que le modèle TV+G s'appuie sur la notion de fréquence, le modèle TVL1 travaille avec la notion d'échelle.

Chapter 6 : Spatially adaptive TVL1. A partir de l'étude du problème "stationnaire" des chapitres précédents, nous avons compris l'effet du paramètre de fidélité λ. Dans ce chapitre, nous utilisons cette intuition pour régler localement la valeur de λ en fonction de l'image, comme dans [Gilboa et al., 2006], ou pour supprimer le bruit impulsionnel comme dans [Koko and Jehan-Besson, 2010]. La première partie de ce chapitre est dédiée à l'étude théorique du modèle TVL1 où λ varie localement. Elle montre essentiellement que le modèle se comporte comme on s'y attend intuitivement : il est équivalent à une famille de problèmes géométriques, la valeur de λ contrôle la courbure de la solution et en choisissant λ assez grand dans certaines régions, il est possible de préserver des formes très irrégulières, présentant par exemple des coins.

La seconde partie illustre l'utilisation du modèle TVL1 adaptatif dans le contexte de la suppression du bruit impulsionnel. La plupart des méthodes pour supprimer le bruit impulsionnel procèdent en deux étapes : une phase de détection et une phase de suppression. Nous rappelons le principe de quelques détecteurs, avant de montrer le résultat du débruitage en utilisant TVL1 adaptatif. Les résultats sont comparables aux méthodes les plus efficaces telles que [Chan et al., 2004] mais pas significativement meilleurs.

Chapter 7 : Extension to the anisotropic case. Le but de ce court chapitre est d'expliquer l'observation empirique du Chapitre 4 concernant la variation totale anisotrope, par exemple quand $\int |Du|_2$ est remplacée par $\int |Du|_1$. En d'autres termes, nous montrons que la solution du problème géométrique

pour un ensemble convexe est donnée par une ouverture avec un carré (au lieu d'un disque), suivie d'un seuillage sur le rapport périmètre (anisotrope)/aire. Cela provient essentiellement d'un résultat sur les facettes calibrables de [Bellettini et al., 2001].

Le chapitre commence avec la définition de la variation totale anisotrope. Quelques propriétés standard sont rappelées. Puis, nous montrons que tous les arguments du Chapitre 3 s'appliquent encore, et que la solution de TVL1 avec périmètre anisotrope est donnée par une ouverture par la forme de Wulff, suivie d'un seuillage. Le résultat vaut pour les variations totales cristallines.

Partie II : Non-local Methods for Image Restoration

Chapter 8 : Introduction to the Non Local Means. Ce chapitre est une introduction au filtre Non-Local Means [Buades et al., 2005]. Après une revue rapide de l'histoire du filtre et des filtres similaires, nous proposons au lecteur une "visite guidée" des moyennes non locales. Pour donner plus d'intuition sur le filtre, nous décrivons les effets de chaque paramètre avec l'aide de trois exemples jouets dans lesquels nous donnons une expression analytique de la sortie du filtre. Nous illustrons l'apparition du biais, même sur des signaux périodiques, en fonction de la taille de patch et de la fenêtre de recherche, ou du noyau utilisé dans les poids. Ces exemples donnent des illustrations concrètes des diverses observations empiriques faites dans la littérature, et les expressions analytiques les rendent indépendants du cadre expérimental (paramètres de l'algorithme, réalisations du bruit...). En particulier nous montrons que l'utilisation de poids à support compact réduit l'influence de la fenêtre de recherche.

Pour étudier l'apparition du biais dans les Non-Local Means et la notion de régularité associée, nous faisons une brève excursion dans l'espace des patchs, inspirée par [Peyré, 2009, Tschumperlé and Brun, 2009, Singer et al., 2009]. Nous ne supposons pas que les images reposent sur une variété dans l'espace des patchs, mais nous insistons sur l'importance d'un moment de la distribution des patchs. Ce point de vue élémentaire est utilisé pour expliquer le comportement du filtre quand la fenêtre de recherche ou la taille de patch varient.

Pour résumer, ce chapitre expose les défauts et les limites des Non-Local Means. Il montre aussi que dans la plupart des cas, un choix judicieux des paramètres comme la largeur de bande h ou la taille de patch permet de limiter l'artefact.

Chapter 9 : Spatially adaptive choice of the bandwidth h. On s'attache à concevoir un filtre Non-Local Means doté d'un choix local de largeur de bande, pour réduire les artefacts mis en évidence dans le Chapitre 8. D'abord, nous construisons un oracle basé sur un compromis biais-variance qui nous indique le comportement idéal d'un tel filtre. En fait, la distinction entre textures et régions lisses n'est pas primordiale dans la décision de l'oracle. En effet, les variations les plus importantes dépendent surtout de la présence ou non d'un contour dans le patch. Les patchs qui intersectent des contours ou qui comportent des détails très contrastés ont en général peu de copies similaires, de sorte que les Non-Local Means y laissent un halo de bruit. L'oracle force une grande largeur de bande h dans ces zones, et un algorithme pratique devrait en faire de même.

Puis nous construisons un tel algorithme pratique, en nous appuyant sur l'estimation sans biais du risque (SURE)[Stein, 1981], proposée pour la première fois dans le contexte des Non-Local Means dans [Van De Ville and Kocher, 2009]. Puisque cet estimateur du risque n'est pas robuste quand il est utilisé de manière ponctuelle, nous proposons de régulariser le risque par une moyenne locale, et ensuite de garder la valeur de h qui minimise cette estimation régularisée. Un algorithme pratique est donné pour réaliser ces opérations en un temps raisonnable, et de nombreuses expériences sont montrées. La procédure permet de réduire le halo près des bords, et elle préserve les détails peu contrastés bien mieux que les Non-Local Means classiques. Pourtant, la décision est un peu brutale, et quand l'estimation du risque échoue, de petites taches de bruit sont laissées par l'algorithme. Inspiré par [Salmon and Le Pennec, 2009b], nous utilisons une agrégation à poids exponentiels (EWA) pour calculer une combinaison convexe des filtres associés à différentes largeurs de bande en fonction de leur risque estimé. Cette procédure préserve bien mieux les textures que la méthode précédente, ne souffre pas du halo de bruit, et ne laisse pas de tache de bruit. Les images produites sont un peu plus bruitées, mais leur aspect est plus agréable, malgré un PSNR un peu en dessous de la méthode précédente.

Chapter 10 : From patches to shapes: NLM-SAP. Ce chapitre relate un travail réalisé en commun avec Charles Deledalle et Joseph Salmon. Il s'agit d'une tentative de réduire le halo de bruit dans les Non-Local Means. L'idée prend sa source dans le travail de Salmon et Strozecki [Salmon and Strozecki, 2010b], où il est proposé de translater les patchs pour trouver des copies similaires près des contours. Dans ce

chapitre, nous poussons ce concept plus loin en remplaçant les patchs translatés par des formes arbitraires. L'intérêt est d'utiliser des formes qui sont adaptées à la géométrie locale et qui peuvent s'insérer dans les régions étroites. Nous proposons un algorithme basé sur la Transformée de Fourier Rapide pour calculer le résultat des Non-Local Means avec une forme arbitraire, indépendamment de la taille de la forme.

La question est ensuite de choisir la forme à chaque pixel. Nous utilisons un dictionnaire de formes fixés à l'avance, et à chaque pixel nous devons donc décider comment combiner nos estimateurs. Comme dans le chapitre précédent, ne garder que l'estimateur ayant le risque estimé le plus faible est trop brutal, donc nous agrégeons les résultats en utilisant la pondération basée sur la variance [Salmon and Strozecki, 2010b] ou l'agrégation à poids exponentiels [Leung and Barron, 2006]. L'algorithme qui en résulte est rapide, et fournit une très bonne qualité visuelle, même si en terme de PSNR, il se classe un peu derrière [Dabov et al., 2007, Goossens et al., 2008].

Publications

Le contenu des Chapitres 3, 4 et une partie du Chapitre 6 a été publié dans le *SIAM Journal on Multiscale Modelling and Simulation (MMS)* [Duval et al., 2009]. Il a été présenté à la *SIAM Conference on Imaging Sciences 2010*. Une petite portion (une page) du Chapitre 5 est extraite d'un article publié dans le *Journal of Mathematical Imaging and Vision (JMIV)* [Duval et al., 2010]. Un article incluant approximativement les Chapitres 8 et 9 a été accepté par le *SIAM Journal of Imaging Science (SIIMS)* [Duval et al., 2011]. Le contenu du Chapitre 10 est l'objet d'un article écrit en commun avec Charles Deledalle et Joseph Salmon qui a été accepté au *Journal of Mathematical Imaging and Vision (JMIV)*[Deledalle et al., 2011]. Il a été présenté à la *Conference on Scale-Space and Variational Methods (SSVM) 2011*.

Introduction

From an ideal perfect scene occurring in the physical world to its display on our computer screens or other imaging devices, many steps induce alterations of images. The acquisition process, the transmission or the storage of an image may all degrade the original information, so that efficient methods to restore or simplify images have become crucial in many applications, whether medical diagnosis, satellite imaging, videoconference or even every-day life digital cameras.

During the acquisition process, the *Charge-Coupled-Devices* (CCD) sensors are subject to the shot noise, which is due to the discrete nature of the measured quantity (the number of incoming photons), and which follows a Poisson distribution. The effect of the shot noise is all the stronger as the number of photons is small, so that scenes with insufficient lighting often appear noisy. The thermal noise is also responsible for degraded images: depending on the temperature of the captor, the thermal agitation of the electrons induces a current that adds to the original signal. Let us also mention degradations induced by the optics: some objects may be out-of-focus or chromatic aberrations may appear. A fast moving object may be blurred if the shutter speed is too slow. During the transmission process, images may be degraded by the loss of several packets, leading to wrong pixel values. This is especially common in satellite imaging. To reduce the storage cost of images, lossy compression algorithms are employed, which removes some of the information, inducing a "compression noise" which is all the more visible as the compression rate is high.

As a consequence, images are usually too degraded to be directly shown to a human or to be sent to an image analysis algorithm: they have to be filtered. The image processing community has been working for decades on designing algorithms either to restore images as faithfully as possible, or to simplify images while keeping the most relevant information for automatic analysis. The first proposed method consists in linearly filtering the image. With additional assumptions such as translation invariance this amounts to applying a convolution to the image (see for instance [Guichard et al., 2004]). The problem with convolution is that it tends to blur the sharp edges that exist in most natural images. To smooth images without blurring edges, researchers and engineers have proposed a wide variety of methods which all correspond to a certain a priori on images. Let us briefly describe a few families of models:

Partial Differential Equations (PDE). The linear filtering can be interpreted asymptotically as one step of the heat equation, which diffuses information in an isotropic way. The idea proposed by Perona and Malik [Perona and Malik, 1990] is to diffuse information depending on the local content of the image, using equations of the form:

$$\frac{\partial u}{\partial t} = \text{div}\ \left(g(|\nabla u|)\nabla u\right), \tag{25}$$

where *e.g.* $g : \mathbb{R}^+ \to \mathbb{R}, s \mapsto \frac{1}{1+\lambda^2 s^2}$. In homogeneous regions, the diffusion behaves like a regular two-dimensional heat equation, but near edges (where $|\nabla u|$ is large), the diffusion is a reverse heat equation in the direction of the gradient, and a heat equation in the orthogonal direction. As a consequence the edges are enhanced and the other regions are smoothed. Many other PDEs have been proposed, whether for denoising or analyzing shapes. For instance, the mean curvature motion, introduced in image processing

in [Kimia et al., 1995], is governed by the equation:

$$\frac{\partial u}{\partial t} = |\nabla u| \text{div}\left(\frac{\nabla u}{|\nabla u|}\right). \tag{26}$$

It is a pure diffusion in the direction orthogonal to the gradient. Let us also mention the diffusion proposed by Sochen, Kimmel and Maladi [Sochen et al., 1998] in connection with minimal surfaces:

$$\frac{\partial u}{\partial t} = \text{div}\left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right). \tag{27}$$

So many parabolic PDEs have been proposed for image restoration that one might easily get lost. Using an axiomatic approach, Alvarez *et al.* [Alvarez et al., 1993] have proved that the PDEs which have the contrast invariance property and isotropic invariance (which are quite relevant properties for image processing) are of the form (see also [Chen et al., 1991]):

$$\frac{\partial u}{\partial t} = F(\text{div}\left(\frac{\nabla u}{|\nabla u|}\right), t)|\nabla u|. \tag{28}$$

Adding the requirement of affine invariance allows to single out a specific equation called affine curvature motion (see [Guichard et al., 2004]).

Partial Differential Equations in image processing are naturally linked to the theory of scale-space [Witkin, 1983, Koenderink, 1984], which associates an image with a family of gradually smoothed images. Alvarez *et al.* [Alvarez et al., 1993] have shown that every causal scale-space corresponds to a PDE. Depending on authors (see for instance [Babaud et al., 1986, Olver et al., 1994, Weickert, 1996]), various sets of axioms for scale-spaces are imposed. We refer the reader to the monographs [Weickert, 1998, Guichard et al., 2004] for more details.

Mathematical morphology. Mathematical morphology is a theory of image analysis developed by Matheron and Serra in the 1960s, while Serra was in charge of making a detailed inventory of the iron deposit in Lorraine (the interesting story of the birth of mathematical morphology is related in [Matheron and Serra, 2002]). Its main focus is the contrast-invariance in image processing, and the consequence is that morphological operators can be equivalently seen as operators on functions (images) or on sets (their level-sets). For instance, some of the most famous morphological operators are respectively the erosion and dilation:

$$\forall X \subset \mathbb{R}^2, \quad \mathcal{E}_r X = \left\{x \in \mathbb{R}^2, B(x,r) \subset X\right\}, \quad \mathcal{D}_r X = \left\{y \in \mathbb{R}^2, \exists x \in E,\ y \in B(x,r)\right\}, \tag{29}$$

where $r > 0$ and $B(x,r)$ is the Euclidean closed ball with radius r. These operators on sets are associated with functional operators:

$$\forall u \in \mathbb{R}^{\mathbb{R}^2},\ \forall x \in \mathbb{R}^2, \quad E_r u(x) = \inf_{y \in B(0,r)} u(x+r), \quad D_r u(x) = \sup_{y \in B(0,r)} u(x+r). \tag{30}$$

As a matter of fact, it is equivalent to apply each functional operator to a function f or to apply the corresponding set operator on the upper level-sets of f, $F_t = \{x \in \mathbb{R}^2, f(x) \geq t\}$, to obtain sets U_t for $t \in \mathbb{R}$ and then reconstruct u by the formula:

$$\forall x \in \mathbb{R}^2, \quad u(x) = \sup\{t \in \mathbb{R},\ x \in U_t\}. \tag{31}$$

From dilations and erosions, two other basic operators are constructed: openings (defined as $D_r \circ E_r$) and closings (defined as $E_r \circ D_r$). The corresponding set operators may be written as:

$$\forall X \subset \mathbb{R}^2, \quad \mathcal{O}_r X = \bigcup\{B(x,r),\ B(x,r) \subset X\}, \quad \mathcal{F}_r X = \left(\bigcup\{B(x,r),\ B(x,r) \subset X^C\}\right)^C, \tag{32}$$

where $X^C = \mathbb{R}^2 \setminus X$.

Useful references about mathematical morphology are the book by Matheron [Matheron, 1975] and those by Serra [Serra, 1982, Serra, 1988]. From the fundamental axioms established by Matheron and Serra, and the basic operations such as erosions, dilations, openings and closings, mathematical morphology has grown into a well-established theory from which derive algorithms for image denoising [Cheng and Venetsanopoulos, 1992, Vincent, 1993], segmentation [Beucher and Lantuéjoul, 1979, Bouraoui et al., 2008], texture analysis [Vanrell and Vitria, 1993, Hanbury and Serra, 2002], etc.

Variational methods. Another approach to image processing is to find an image which minimizes a certain energy. Typically, if one observes a noisy image f, the restored image is the solution of the problem:

$$\min_u G(u) + F(u, f), \tag{33}$$

The term $G(u)$ is the regularization term, i.e. it forces the minimizers to be smooth enough, whereas the fidelity term, $F(u, f)$, forces the minimizers to be coherent with the observation f. A canonical example of variational problem in image processing is the Rudin-Osher-Fatemi (ROF) problem [Rudin et al., 1992]:

$$\min_u \int_{\mathbb{R}^2} |\nabla u(x)| dx + \lambda \int_{\mathbb{R}^2} |f(x) - u(x)|^2 dx \tag{34}$$

Here, the regularization term is the total variation, which has the advantage of smoothing images without blurring edges. The quadratic fidelity term is typical when working with additive Gaussian noise. This model is an improvement of the older Tikhonov regularization problem [Tikhonov and Arsenin, 1977] where $\int_{\mathbb{R}^2} |\nabla u(x)|^2 dx$ replaces $\int_{\mathbb{R}^2} |\nabla u(x)| dx$. The Tikhonov energy has the drawback that its minimizers belong to the Sobolev space H^1. Contrary to the functions with bounded variations (which form the space $\mathrm{BV}(\mathbb{R}^2)$ in which the solutions of the ROF problem lie), functions in the space H^1 cannot have discontinuities along hypersurfaces, which prevents from working with cartoon-like images, composed of smooth regions delimited with sharp edges. In fact, many other choices of regularization terms have been proposed : the Huber function, the regularized total variation, which amounts to measuring the surface of the graph of u, or various L^p norms of the gradient. Those functionals are convex, so that they are easier to study and to minimize than general non-convex functionals. Yet, there is a growing interest in non-convex regularization terms. In [Bar et al., 2006a, Bar et al., 2006b, Cai et al., 2008], the Mumford-Shah functional [Mumford and Shah, 1989], although it was originally designed to perform segmentation, is used in restoration problems with better results than with total variation. In [Nikolova, 2005], Nikolova investigates the behavior of non-convex regularization terms: she shows that small discontinuities are shrunk, and large discontinuities are amplified. A large panel of non-convex functions has been proposed, but the fast minimization of corresponding functionals is still an open problem.

Variational problems are especially interesting since they provide a simple framework to solve ill-posed inverse problems like deconvolution or super-resolution. For instance, the natural adaptation of the Rudin-Osher-Fatemi model to deconvolution is:

$$\min_u \int_{\mathbb{R}^2} |\nabla u(x)| dx + \lambda \int_{\mathbb{R}^2} |f(x) - h \star u(x)|^2 dx \tag{35}$$

where the convolution with h models the blur operator.

Markov Random fields and Maximum a Posteriori Estimation. In a discrete setting, a way to model the dependencies between the pixel values is to interpret an image as a Markov random field, as proposed in the seminal paper [Geman and Geman, 1984]. The probability of observing one value at pixel x conditionally to the values of all other pixels actually only depends on the values in a neighborhood of x. Hammersley-Clifford's theorem states that the joint probability of observing an image u can be written as $P(u) = \frac{1}{Z} \exp(-\sum_{c \in \mathcal{C}} E_c(u))$, where E_c is called a clique energy.

For the denoising problem with additive Gaussian white noise, we observe f and we look for the restored version u. The Bayes formula gives:

$$P(u|f) = \frac{P(u)P(f|u)}{P(f)} \tag{36}$$

where $P(u) = \frac{1}{Z}\exp(-\sum_{c\in\mathcal{C}} E_c(u))$ is the prior on images (for instance the smoothness prior $\frac{1}{Z}\exp(-\sum_{i,j}(|u_{i+1,j}-u_{i,j}|+|u_{i,j+1}-u_{i,j}|))$ and $P(f|u)$ is the conditional probability of f given u (for instance $\frac{1}{Z'}\exp\left(-\frac{\|f-u\|^2}{2\sigma^2}\right)$). Taking the logarithm of this expression we see that $\log P(u|f)$ is the sum of a clique energy and a fidelity term. The Maximum A Posteriori (MAP) estimation of u consists in finding a maximizer of $\log P(u|f)$. Equivalently one looks for a minimizer of $-\log P(u|f)$ which is the sum of a "fidelity term" and a "smoothing term": variational problems can be seen as Maximum A Posteriori estimation problems and vice-versa.

The literature on Markov Random Fields in image processing has the notable interest of its connections with statistical physics. In particular, Ising's model, which originated from the study of ferromagnetism and antiferromagnetism, was imported to image processing in its ferromagnetic version:

$$E(u) = -\sum_{(s,t)/s\sim t} -\beta u_s u_t - \sum_s u_s \tag{37}$$

with $\beta > 0$, and u is a binary image. This energy implies that the images with most probability have few jumps. This model is related to the TVL1 model studied in this thesis inasmuch as they coincide on binary images of convex sets (see Chapter 3 of this thesis). We refer the reader to [Maruani et al., 1995] for a study of its properties.

Let us mention that the Maximum A Posteriori estimation is not necessarily the most reasonable way to tackle a denoising problem. In [Louchet and Moisan, 2008, Louchet, 2008], Louchet and Moisan advocate for the total variation with Least-Square Error (LSE) estimation instead of the Maximum a Posteriori (which amounts to the ROF model). The idea is to find u which minimizes $\mathbb{E}_{u'|f}\|u-u'\|^2$. This leads to a weighted sum of images u' depending on their probability $P(u'|f)$. The point with this estimator is that its properties are much better than those of the MAP estimator (see [Louchet, 2008]): in particular it does not suffer from the staircasing effect (see for instance [Nikolova, 2000, Nikolova, 2004b] for a study of the staircasing effect). Also, this LSE estimation allows to draw a parallel between the total variation with LSE denoising and the Non-Local Means [Buades et al., 2005]. This is related to the interpretation of NL-Means as the aggregation of estimators by Salmon and Le Pennec [Salmon and Le Pennec, 2009a, Salmon and Le Pennec, 2009b]. Interestingly enough, in a different context, in the second part of this thesis (especially Chapters 9 and 10) we also observe that a weighted sum of estimators is sometimes better than one particular minimizer.

Neighborhood filters and patch-based methods. Instead of averaging pixels depending on their spatial proximity, several authors have proposed to use the proximity between their values. This is in essence the principle of the Yaroslavsky filter [Yaroslavsky, 1985]:

$$YNF_{h,\rho}u(x) = \frac{1}{C(x)}\sum_{y\in B(x,\rho)} e^{-\frac{|u(x)-u(y)|}{h^2}}u(y). \tag{38}$$

where $C(x)$ is a normalization constant, and $B(x,\rho)$ is a window with radius ρ centered around x. This idea is also at the core of the sigma filter [Lee, 1983] and the SUSAN filter [Smith and Brady, 1997]. With the bilateral filter [Tomasi and Manduchi, 1998], both the pixel values and the pixel positions are used to design the weights. Whereas in these methods the compared pixels always belong to a fixed window or ball around the pixel to denoise, some authors have proposed to choose adaptively the set of pixels with which to compute the average [Polzehl and Spokoiny, 2000, Polzehl and Spokoiny, 2003].

The idea is to rely on a bias/variance tradeoff and to use the monotony of the variance as the set of pixels increases according to Lepski's method [Lepski, 1990].

Inspired by neighborhood filters and by the work of Efros and Leung [Efros and Leung, 1999] in texture synthesis, Buades, Coll and Morel [Buades et al., 2005] have proposed to compare small neighborhoods around pixels called patches, rather than the pixel values only. This "Non-Local Means" filter led to a breakthrough in image processing, and for half a decade the majority of the proposed denoising algorithms have relied on the notion of patch. The BM3D algorithm [Dabov et al., 2007], which currently ranks among the best denoising methods, is one of them.

Wavelet and sparse methods. Another a priori on images is that they have a sparse representation in a well-chosen basis or frame. Donoho and Johnstone [Donoho and Johnstone, 1994] have shown that hard thresholding is asymptotically optimal in the minimax sense in the class of diagonal estimators, and that the associated risk is all the smaller as the image representation is sparse. Informally, sparsity implies that the signal is concentrated on a few large coefficients, whereas the white noise is uniformly spread on the different components. Thus, a simple thresholding allows one to remove most of the noise without degrading much the signal. Wavelet-based methods are typical of this approach: since wavelet coefficients encode the variations of the image instead of each absolute value, piecewise smooth images have a sparse representation in wavelet bases. Yet, wavelet bases do not take advantage of the geometry of images: the large coefficients are spread along edges but this structure is not exploited by wavelet decompositions. More sophisticated frames have been proposed to take advantage of this property: curvelets [Starck et al., 2002], bandlets [Le Pennec and Mallat, 2005, Mallat and Peyré, 2007] or contourlets [Do and Vetterli, 2005] among others.

Rather than considering a fixed dictionary, several methods consist in learning an adapted dictionary on a database. Given an overcomplete dictionary, the Matching pursuit [Mallat and Zhang, 1993, Bergeaud and Mallat, 1995] algorithm allows to find a selection of atoms which yields a representation as sparse as possible of a family of signals. The Principal Component Analysis (PCA) allows to find a basis which concentrates most of the energy of most signals in a database on the first vectors of this basis. These tools and sparsity in general have helped produce many powerful denoising algorithms. In particular, the adaptation of these ideas to the use of patches in images in [Aharon et al., 2006] and its variant using structured sparsity [Mairal et al., 2009] lead to state-of-the art performance in image denoising.

The boundary between these families of models is not as sharp as it may seem, and many algorithms have an equivalent in another category. For instance:

- The asymptotic equivalence between morphological operators and PDEs is established in [Lax, 1965] for erosions and dilations, and more generally in [Alvarez et al., 1993] (see [Guichard et al., 2004] for a thorough exposition of the theory).

- Variational models often lead to PDEs, as their minimization is related to the flow of their gradient: for instance, Andreu *et al.* [Andreu et al., 2001b] have studied the total variation flow in connection with the ROF problem.

- The continuous limit of the Bayesian segmentation model [Geman and Geman, 1984] leads to the Mumford-Shah functional [Mumford and Shah, 1989] (see [Chambolle, 2000]).

- Steidl *et al.* have proved that in dimension 1, the total variation denoising and wavelet shrinkage are equivalent [Steidl et al., 2004].

- The bilateral filter is asymptotically equivalent to one step of the Perona Malik equation (see [Durand and Dorsey, 2002, Buades et al., 2005]).

Morphological Aspects of the TVL1 model

The first part of this thesis explores a similar connection between a variational problem and mathematical morphology. It is devoted to the study of the TVL1 model, which consists, given $f \in L^1(\mathbb{R}^2)$, in solving:

$$\min_{u \in \mathrm{BV}(\mathbb{R}^2)} \int |Du| + \lambda \int_{\mathbb{R}^2} |f(x) - u(x)| dx. \tag{39}$$

This model was introduced by Alliney for one-dimensional signals [Alliney, 1992, Alliney, 1996, Alliney, 1997]: the author proves that in this case, a recursive median filter allows to compute its minimizers. The interest of the model for image processing is pointed out by Nikolova in [Nikolova, 2002]: she shows that the non-smooth data fidelity term provides a better detection and removal of outliers than smooth fidelity terms. As a result, the model is particularly well suited to remove impulse noise, and the paper [Nikolova, 2004a] gives striking numerical illustrations of this fact.

In [Chan and Esedoglu, 2005] the model is studied in a continuous framework. Chan and Esedoglu point out that although the similarity between ROF and TVL1 is striking, the model has its own interesting properties. For instance, in the spirit of [Strong et al., 1996, Meyer, 2001] they compute the solution of the problem when the input function f is the characteristic function of a disc: $f = \mathbb{1}_{B(0,R)}$. The solution is:

$$u = \begin{cases} \mathbb{1}_{B(0,R)} & \text{if } \lambda \geq \frac{2}{R}, \\ c\mathbb{1}_{B(0,R)}, & \forall c \in [0,1] \text{ and } \lambda = \frac{2}{R}, \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

In other words, there is no loss of contrast contrary to ROF (see [Meyer, 2001]), and some objects vanish suddenly depending on their scale. Following [Chambolle, 2004a, Alter et al., 2005a], where the ROF model is connected to a family of geometric problems, Chan and Esedoglu notice that the energy of the TVL1 model can be written as the sum of energies on level sets:

$$\mathbb{E}(u) = \int_{-\infty}^{+\infty} \text{Per}\,\{x, u(x) > \mu\} + \lambda |\{x, u(x) > \mu\} \Delta \{x, f(x) > \mu\}| d\mu, \tag{41}$$

where Per X refers to the perimeter of $X \subset \mathbb{R}^2$, $|X|$ refers to its Lebesgue two-dimensional measure, and $X \Delta Y = (X \setminus Y) \cup (Y \setminus X)$. This formula is of particular importance, since it confers on the model the contrast invariance property. The same observation is made simultaneously by Darbon in [Darbon, 2005], where a connection to mathematical morphology is drawn for the first time. It also allows Darbon and Sigelle to propose a graph-cut based algorithm which has provided for many years the highest speed to solve the TVL1 problem [Darbon and Sigelle, 2006] (until the work of Pock *et al.* [Pock et al., 2008] who adapted the algorithm [Aujol et al., 2006] to GPUs). Let us also mention the work of Yin *et al.* [Yin et al., 2007b] who have studied the monotonicity of the problem on level-sets as well as the contrast invariance. In the literature, the TVL1 model or its variants have been applied to various problems: impulse noise removal [Nikolova, 2004a, Koko and Jehan-Besson, 2010, Liu et al., 2010], segmentation [Chan et al., 2006, Bresson et al., 2007] feature detection in biomedical imaging[Cunha et al., 2007], 3D MRI registration [Chen et al., 2005], cDNA microarray data filtering [Yin et al., 2005], cartoon-texture decomposition [Aujol et al., 2006, Yin et al., 2007a, Haddad, 2007], computation of the optical flow in videos [Wedel et al., 2008a, Wedel et al., 2008b], shape analysis [Vixie et al., 2010]...

Our contribution is to study the geometric problem in depth so as to give more information on the solutions of TVL1. After making the equivalence with the geometric problem precise, we characterize the sets that suddenly vanish in the same way as the disc (40). Applying results from geometric measure theory, [Ambrosio, 1997, Ambrosio and Paolini, 1998], we show that the boundary of the solution is smooth and that its curvature is equal to $-\lambda$ in the interior of the data F and $+\lambda$ in its exterior (when it coincides with the boundary of F, the curvature should be bounded by λ). Then we observe that when the input set C is convex, the problem can be simplified as:

$$\min_{X \subset C} \text{Per}\, X - \lambda |X| \tag{42}$$

We deduce that the solutions of TVL1 are given by an opening of radius $\frac{1}{\lambda}$, followed by a thresholding on the ratio perimeter/area. More precisely, if we write $C_{1/\lambda} = \mathcal{O}_{1/\lambda}C$ (where $\mathcal{O}_r$ is defined in (32)), we prove that:

- if $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} < \lambda$, the solution is $C_{1/\lambda}$,
- if $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} > \lambda$, the solution is $\emptyset$,
- if $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} = \lambda$, both $C_{1/\lambda}$ and $\emptyset$ are solutions.

Let us emphasize that this result relies to a large extent on the work of Alter, Caselles, Chambolle [Alter et al., 2005a, Alter et al., 2005b] and more generally Bellettini, Novaga, Paolini [Bellettini et al., 2001, Bellettini et al., 2002, Bellettini et al., 2005a]. In a few words, the study of calibrable sets, i.e. sets that evolve at constant speed by the total variation flow, shows that convex calibrable sets are solutions of the following variational problem:

$$\min_{X \subset C} \text{Per } X - \lambda_C |X| \quad \text{with } \lambda_C = \frac{\text{Per } C}{|C|}. \tag{43}$$

Observing that the level sets F_t of the solution of the ROF problem:

$$\min_{u \in \mathbf{BV}(\mathbb{R}^2)} \int |Du| + \frac{\lambda}{2} \int_{\mathbb{R}^2} (u(x) - \mathbb{1}_C(x))^2 dx \tag{44}$$

are solutions of the family of geometric problems:

$$\min_{X \subset C} \text{Per } X - \lambda(1-t)|X|, \tag{45}$$

Alter *et al.* study the dependence on λ of this family of problems, and by approximation arguments, they characterize by their curvature the sets C that are invariant by problem (42). We have used this characterization in connection with a characterization by their curvature of openings in the plane to get the result, but we could also have used the vector-field constructed in [Alter et al., 2005b] for the total variation flow as a "calibration" (see chapter 2). As a by-product of the study of the total variation flow, exact solutions of the ROF problem for data of the form $\mathbb{1}_C$ are given in[Alter et al., 2005b], and openings are also involved (in a more complicated way).

In the case of TVL1, the class of functions for which the exact solutions are obtained is slightly larger (quasi-convex functions instead of indicator functions of convex sets) than with ROF, but more importantly, the link with openings is straightforward. This gives some insight on the observation of Darbon [Darbon, 2005] that TVL1 is a morphological operator. Let us also mention that the result on TVL1 and openings can be seen as a particular case of the general but difficult study of the $TV + L^p$ problems by Allard ([Allard, 2008], see also [Allard, 2007] and [Allard, 2009]), although we obtained it independently. To illustrate it numerically, we have built an algorithm based on openings and the Fast Level-Set Transform [Monasse and Guichard, 2000], which gives the exact solution in simple cases and an approximate solution of TVL1 otherwise in a very competitive computing time (see Chapter 4).

Another contribution of this thesis is to interpret the behavior of the TVL1 model for the cartoon-texture decomposition problem in light of mathematical morphology. We begin by comparing the decompositions provided by TVL1 with those provided by the TV+G model, the first proposed cartoon-texture decomposition model [Meyer, 2001]. The fact that the TVL1 model performs well at this task has been known since [Haddad, 2007, Yin et al., 2007a]. With the TV+G model, we notice that traces of edges appear in the texture part, plus some sort of halo, which brings us to study this problem more in depth.

A common idea when dealing with norms that favor oscillations is that if an image is composed of a cartoon part, say the indicator function of a smooth set, and an oscillating part, say a sine function, the model will be able to decompose it since the total variation of smooth sets (resp. sine) is low (resp.

high) and the G-norm of a sine (resp. smooth set) is low (resp. high). Yet, it is not trivial that the decomposition will be perfect, and to our knowledge, no explicit example of a perfect decomposition has ever been given. Using the radially symmetric framework of [Haddad, 2007] and the characterization of the subdifferential of the G-norm established in Chapter 1, we prove that the indicator function of a disc plus a (slightly perturbed) sine is perfectly decomposed by the TV+G model. The good news is, the model does indeed what it is supposed to do, at least on one example. However, a careful examination of the example shows the necessity that the texture does not vanish in the neighborhood of the boundary of the disc. We explain why traces of edges and a halo may appear in more general images. In the case of the indicator function of a convex set and with additional technical hypotheses on the texture part, we show that a decomposition where the texture part vanishes in a neighborhood of the edges cannot be a solution of the TV+G model. As a consequence, we believe that the apparition of edges in the texture part is an intrinsic property of TV+G decompositions.

Having studied the properties of TV+G decompositions, we then turn again to the TVL1 model. We show a similar example of a perfect decomposition of an indicator function and a sine. The limitations exhibited for the TV+G model do not hold with the TVL1 model: the cartoon part and the texture part may be totally independent. Eventually, to explain the good behavior of the model in cartoon-texture decomposition, we relate it to the granulometries used in mathematical morphology: whereas the TV-G model relies on the notion of frequency, the TVL1 model works with the notion of scale.

As the study of the "stationary problem" has given some intuition on the effect on the fidelity parameter λ, we study a model where λ varies spatially, as in [Gilboa et al., 2006] for the ROF model. Tuning locally this parameter allows to preserve textured areas as in [Gilboa et al., 2006] or to remove impulse noise as in [Koko and Jehan-Besson, 2010] without modifying the pixels that are not affected by the noise. We first examine the model from a theoretical point of view: we essentially show that the model behaves as one would intuitively expect. The variational problem is equivalent to a family of geometric problems, the value of λ controls the curvature of the solution, and by choosing λ large enough in some regions, it is possible to preserve very irregular shapes, even those with corners. Then, we illustrate the use of the adaptive TVL1 model in the context of impulse noise removal. Most methods to remove impulse noise are divided in two steps [Chen and Wu, 2001a, Pok et al., 2003, Crnojevic et al., 2004, Chan et al., 2004, Dong et al., 2007]: a detection phase and a denoising phase. We recall the principle of a few noise detectors, namely the Adaptive Center-Weighted median Filter (ACWMF) [Chen and Wu, 2001a], Rank Ordered Absolute differences (ROAD) [Garnett et al., 2005] and Rank Ordered Logarithmic Differences (ROLD) [Dong et al., 2007] before showing the result of the denoising using adaptive TVL1. The performance is comparable to state-of-the-art methods like [Chan et al., 2004] but not clearly better. Yet, one advantage of the framework is its simplicity. In fact, most of the difficulty of the impulse noise removal comes from the noise detection. Once this step is done properly most denoising methods provide similar results.

The end of this first part is devoted to the justification of the empirical observations of Chapter 4, namely that the solution of TVL1 when the total variation is computed using the ℓ^1 norm (instead of ℓ^2) of the gradient (e.g. in graph-cut based methods) is given by an opening where balls are replaced with squares. In fact, we show that for a given crystalline anisotropy, the solution is given by an opening with the Wulff shape followed by a test on the ratio perimeter/area. This result is a direct consequence of the work of Belletini *et al.* [Bellettini et al., 2001].

Non Local Methods for Image Restoration

The second part of this thesis is devoted to the Non-Local Means filter [Buades et al., 2005] from which originate some of the best denoising methods. We assume that we want to recover the image $\mathbf{f}$ which has been contaminated by additive Gaussian white noise:

$$\mathbf{u}(x) = \mathbf{f}(x) + \boldsymbol{\varepsilon}(x) \text{ where } \varepsilon(x) \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.} \tag{46}$$

A simple method to recover $\mathbf{f}(x)$ is to average N independent identically distributed (i.i.d.) observations of the noisy pixel $\{\mathbf{u}_i(x),\ i = 1 \dots N\}$. Thus one gets an unbiased estimate of $\mathbf{f}(x)$ with variance:

$$\mathrm{Var}\left(\mathbf{f}(x) - \frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_i(x)\right) = \frac{\sigma^2}{N}. \tag{47}$$

In practice, only one realization of $\mathbf{u}(x)$ is observed. A common solution is to assume that the image does not vary much in a neighborhood of x (at least in one direction) so as to average the intensity values in this neighborhood. This is roughly what most of the methods exposed above do. The limitation of this method is clear: one usually considers only small neighborhoods so that the values $\mathbf{f}(y)$ do not vary much, and the resulting reduction of variance is therefore limited.

The beautiful idea of Buades *et al.* is to take advantage of self-similarities all across the image: when averaging patterns that are repeated in the image, one may reduce drastically the variance of the noise without blurring the details. These similarities may appear at large distances, hence the term "non-local" to qualify the algorithm. In fact, a crucial ingredient for this method is the comparison of a neighborhood of x with the one of y instead of simply comparing their intensity value. These neighborhoods, usually square, are called patches, and they have quickly been adopted by the image processing community since the work of Efros and Leung in texture synthesis [Efros and Leung, 1999]. From the breakthrough of the Non-Local Means (or the simultaneous UINTA filter [Awate and Whitaker, 2006]), many variants or studies have been proposed: some authors give a variational interpretation of the filter [Brox and Cremers, 2007, Kindermann et al., 2005, Buades et al., 2006, Gilboa and Osher, 2007, Gilboa and Osher, 2008, Azzabou et al., 2007c], which leads them to iterate the algorithm. Other authors have focussed on the notion of patch, and they interpret the algorithm as diffusion in the patch space [Szlam, 2006, Tschumperlé and Brun, 2009, Peyré, 2008, Peyré, 2009, Singer et al., 2009]: image denoising comes back to the heat equation! As the original Non-Local Means filter is computationally expensive, its fast computation is a challenging problem. An approximation of the Non-Local Means can be quickly computed using preselection techniques [Mahmoudi and Sapiro, 2005, Bilcu and Vehvilainen, 2008], or a cluster tree to compute similar patches [Brox et al., 2008], or a spatial subsampling of the patch distances [Pang et al., 2009]. Fast and exact computations are proposed using the Fast Fourier Transform [Wang et al., 2006], or summed area tables [Darbon et al., 2008].

Let us fix the notations. The output of the filter is:

$$\mathrm{NL}\,\mathbf{u}(x) = \frac{\sum_{y\in\Omega} e^{-\frac{\|U(x)-U(y)\|^2}{2h^2}}\mathbf{u}(y)}{\sum_{y'\in\Omega} e^{-\frac{\|U(x)-U(y')\|^2}{2h^2}}}. \tag{48}$$

where Ω is the whole image (often replaced by a search window around x), $U(x)$ denotes the pixel values in a square patch of size $s \times s$ around x, and the distance is the normalized Euclidean distance:

$$\|U(x) - U(y)\|^2 = \frac{1}{s^2}\sum_{|j|\leq\frac{s-1}{2}} (\mathbf{u}(x+j) - \mathbf{u}(y+j))^2.$$

The parameter $h > 0$ is a bandwidth parameter which determines the relative importance of pixels in the mean, depending on their similarity.

As various improvements of the Non-Local Means have been proposed, the limitations of the filter have progressively been understood by the community: the selection of patches is not optimal, and Mahmoudi and Sapiro [Mahmoudi and Sapiro, 2005] have proposed to add gradient information in the selection of the pixels. Another way to make the selection of patches more relevant is to use a principal component analysis as proposed by Azzabou *et al.* [Azzabou et al., 2007a] or Tasdizen [Tasdizen, 2009] (especially if the noise intensity is strong). The influence of the search window on the quality of the result is highlighted by Kervrann and Boulanger in [Kervrann and Boulanger, 2006]

where they choose locally its size depending on a bias-variance trade-off. The choice of the exponential function for the weights is questioned by Goossens *et al.* in [Goossens et al., 2008], and the too large importance given to the pixel x in its own estimation is discussed in [Salmon, 2010b]. Moreover, the Non-Local Means fail when there are no similar patches and several authors have proposed tricks to solve this problem [Gilboa and Osher, 2007, Brox and Cremers, 2007, Brox et al., 2008, Louchet and Moisan, 2010, Salmon and Strozecki, 2010a].

All these pieces of information are rather scattered in the literature, and we propose in Chapter 8 a "guided tour" of the Non-Local Means which highlights some of these properties. The principle of this tour is to study very simple models to give some intuition to the reader who is not familiar with the Non-Local Means. In particular we illustrate that:

- A large patch size tends to blur objects,
- Large patches are more robust to noise,
- There is a loss of contrast depending on the occurrence of each pattern,
- The less contrasted the details, the more they are degraded, and this relation is highly non linear,
- Even periodic images are altered,
- The search window has an impact on the visual quality of the result,
- A weight with compact support instead of exponential allows to reduce the bias.

If one decomposes the error of the filter as the sum of a bias and variance term, the bias term is connected to some regularity of images that we try to characterize in the patch space.

All those considerations advocate for a local choice of the parameters, since the exhibited regularity is highly dependent on the position in the image. We first build an oracle which is a Non-Local Means with uniform weights (similar to the one in [Brox and Cremers, 2007] except that the number of pixels is chosen using the noise-free image): it shows the ideal behavior of a filter with local choice of the bandwidth h. Surprisingly, the distinction between textures and smooth areas is not the prominent criterion in the behavior of the oracle. Indeed, the most important variation depends on whether the pixel belongs to a patch that intersects an edge or not. Patches that are close to edges or details with high contrast usually have few similar copies, so that the Non-Local Means leave them with a noisy halo. The oracle forces a large bandwidth to reduce this halo, and so should a practical algorithm with local choice of the bandwidth. Then we build such a practical algorithm, relying on Stein's Unbiased Risk Estimate (SURE)[Stein, 1981], first proposed in the context of the Non-Local Means in [Van De Ville and Kocher, 2009]. Since this estimator of the risk is not robust when used pixel-wise, we propose to regularize the risk by a local average, and then to keep the value of the bandwidth that minimizes the regularized estimate. A practical algorithm is given to perform this in reasonable time, and many experiments are shown. The procedure allows to reduce the halo near edges, and it preserves details with little contrast much better than the regular Non-Local Means. Yet, the decision is a bit brutal, and when the estimation of the risk is flawed, little spots of noise may be left, like with the global Non-Local Means. Inspired by the work [Salmon and Le Pennec, 2009b], we use an Exponentially Weighted Aggregation (EWA) to make a local convex combination of the filters with various bandwidth in function of their estimated risk. This procedure preserves textures much better than the previous method, does not suffer from the noisy halo and does not leave any spot of noise. The produced images look slightly noisier, but more pleasant. The PSNR is a little below the one of the previous method, but the visual result is worth it.

Another variant of the Non-Local Means is proposed in the last chapter, which is the outcome of a collaboration with Joseph Salmon and Charles Deledalle. The purpose of this algorithm is to remove the rare patch artifact, namely a noise halo which appears along contrasted edges or details. The idea stems from the work of Salmon and Strozecki [Salmon and Strozecki, 2010b], where it is proposed to

shift patches in order to find similar copies near edges. Here, we push this idea further by replacing shifted patches with arbitrary shapes. The idea is to use shapes that can fit in narrow regions so as to find a maximal number of similar pixels. We propose an algorithm based on the Fast-Fourier Transform to compute the result of the Non-Local Means with an arbitrary patch shape, regardless of the size of the shape.

Using a fixed dictionary of shapes, we obtain as many different estimators as the number of shapes. As in the previous chapter, keeping only the estimator with the smallest estimate of the risk is too brutal, so we aggregate the results using the Weighted Average based on Variance [Salmon and Strozecki, 2010b] or the Exponentially Weighted Aggregation [Leung and Barron, 2006]. The resulting algorithm is fast, and it provides very neat images.

Chapter description

Part I: Morphological Aspects of the TVL1 Model

Chapter 1: Notations and Preliminaries. The aim of this chapter is to recall the basic properties of functions of bounded variations, together with an extension of Green's formula due to [Anzellotti, 1983] which plays a crucial role in this thesis. The subdifferential of convex l.s.c. positive homogeneous functions is characterized, which will be useful in the study of the TVL1 and the TV+G models. Eventually a short introduction to the Cheeger problem and calibrable sets is given.

Chapter 2: How to study the TVL1 problem? This chapter summarizes the different approaches used in the literature: the characterization using the G-norm in [Kindermann et al., 2006], the dual problem in [Haddad, 2007], the flat norm in [Morgan and Vixie, 2007], and a method inspired from calibrations of geometric measure theory exposed in this chapter. We show that all these methods can be interpreted as the Euler-Lagrange equation for the TVL1 functional. The "method of calibrations" explained here is in fact related to the work of Alter, Bellettini, Caselles, Chambolle, Novaga and Paolini. It consists in using the vectorfield constructed in [Alter et al., 2005b] for the Total Variation Flow to show that a function is optimal for TVL1. Once the vectorfield is built, the proof is straightforward.

Chapter 3: The TVL1 model: a geometric point of view. The TVL1 problem may be studied using Cavalieri's principle, by decomposing an image into its level sets. From a functional problem, we are led to study a family of geometric problems. We propose a proof in the continuous framework of this equivalence noticed for TVL1 in [Chan and Esedoglu, 2005, Darbon, 2005]. This equivalence makes the TVL1 model a contrast invariant filter. In fact, as noted in [Darbon, 2005], TVL1 is a morphological operator.

Instead of studying the functional problem, we focus on the geometric problem. Using very simple arguments, we show that the sets that suddenly vanish, like the disc [Chan and Esedoglu, 2005], are necessarily Cheeger in themselves, and if the set is convex, this condition is sufficient. Moreover, general calibrable sets suddenly vanish.

The geometric approach is all the more fruitful as it allows to use powerful theorems from geometric measure theory. The level lines are smooth and their curvature is bounded by λ. In case the input set is convex, the energy can be reformulated by noticing that the solution will necessarily be inside the input set. Relying on the results in [Andreu-Vaillo et al., 2002, Alter et al., 2005a] for the total variation flow and the characterization of convex calibrable sets, we prove that the solution of the geometric problem is given by an opening of radius $1/\lambda$, and a test on the ratio perimeter/area.

This leads us to discuss the link between mathematical morphology and the TVL1 problem. In the non-convex case, only a bound on the solutions is obtained, but "reasonable candidates" for the solution of TVL1 are proposed using Alternate Sequential Filters. The opportunity of such a choice is validated empirically in Chapter 4.

Chapter 4: Algorithms and numerical experiments. The aim of this chapter is to support the theoretical conclusions of Chapter 3 with numerical experiments. We first recall several classical methods to solve the TVL1 problem. Then, we propose two algorithms to solve the TVL1 problem. The first one is naive and slow: it consists in decomposing an image in performing an alternate sequential filter and then threshold the connected components of each level set. The second one consists in performing the same tasks using the tree of shapes and the Fast Level Set Transform [Monasse and Guichard, 2000]. This leads to a fast algorithm, comparable to [Darbon and Sigelle, 2006]. The numerical experiments on binary and natural images show that the proposed algorithms provide results that are very close to the real solutions of TVL1 (provided by gradient descent for instance).

To compare our method with an exact minimization of TVL1 which satisfies the coarea formula, we use the Darbon-Sigelle algorithm [Darbon and Sigelle, 2006]. This algorithm is based on an anisotropic

scheme for the total variation. A remarkable property is that if we replace in our algorithm openings with balls with openings with squares (and the associated computation of the perimeter), we also obtain results that are very close to the TVL1 model with anisotropic total variation! This suggests that the property is still true for anisotropic perimeters, which will be investigated in Chapter 7.

Chapter 5: Cartoon, Textures and Granulometries. The connection with mathematical morphology drawn in Chapters 3 and 4 for the TVL1 model allows us to interpret its behavior in the cartoon-texture decomposition problem. First, we recall the principle of such decompositions, and we show numerical experiments comparing the results using TVL1 with those provided by the TV+G model, the first proposed cartoon-texture decomposition model [Meyer, 2001]. With the TV+G model, we notice that traces of edges appear in the texture part, plus some sort of halo, which brings us to study this problem in detail.

Using the radially symmetric framework of [Haddad, 2007] and the characterization of the subdifferential of the G-norm established in Chapter 1, we prove that the indicator function of a disc plus a (slightly perturbed) sine is perfectly decomposed by the TV+G model. Incidentally, we notice that in this example the texture cannot vanish in the neighborhood of the boundary of the disc. We explain why traces of edges and a halo may appear in more general images. In the case of the indicator function of a convex set and with additional technical hypotheses on the texture part, we show that a decomposition where the texture part vanishes in a neighborhood of the edges cannot be a solution of the TV+G model.

Having studied the properties of the TV+G model, we then turn again to the TVL1 model. Using the Euler-Lagrange equation, we show a similar example of a perfect decomposition of an indicator function and a sine. We observe that the limitations exhibited for the TV+G model do not hold with the TVL1 model. Eventually, to explain the good behavior of the model for cartoon-texture decomposition, we relate it to the granulometries used in mathematical morphology: whereas the TV-G model relies on the notion of frequency, the TVL1 model works with the notion of scale.

Chapter 6: Spatially adaptive TVL1. From the study of the "stationary" problem in the previous chapters, we have understood the main effects of the fidelity parameter λ. In this chapter we use this insight to locally tune the value of λ in function of the image as in [Gilboa et al., 2006], or to remove impulse noise as in [Koko and Jehan-Besson, 2010]. The first part of the chapter is devoted to the theoretical study of the TVL1 model where λ varies locally. It essentially shows that the model behaves as one would expect intuitively: it is equivalent to a family of geometric problems, the value of λ controls the curvature of the solution, and by choosing λ large enough in some regions, it is possible to preserve very irregular shapes, even those with corners.

The second part illustrates the use of the adaptive TVL1 model in the context of impulse noise removal. Most methods to remove impulse noise are divided in two steps: a detection phase and a denoising phase. We recall the principle of a few noise detectors before showing the result of the denoising using the adaptive TVL1. The results are comparable to the most efficient methods like [Chan et al., 2004] but not clearly better.

Chapter 7: Extension to the anisotropic case. The aim of this short chapter is to explain the empirical observations of Chapter 4 regarding the anisotropic total variation, *e.g.* when $\int |Du|_2$ is replaced with $\int |Du|_1$. In other words, we show that the solution of the geometric problem for a convex set is given by an opening with a square (instead of a disc), followed by a thresholding on the ratio (anisotropic) perimeter/area. This follows essentially from a result on calibrable facets in [Bellettini et al., 2001].

The chapter begins with the definition of the anisotropic total variation. Some standard properties are recalled. Then, we show that all the arguments of Chapter 3 still apply, and that the solution of TVL1 with an anisotropic total variation is given by an opening with the Wulff shape, followed by a thresholding. The result holds for any crystalline total variation.

Part II: Non-local Methods for Image Restoration

Chapter 8: Introduction to the Non Local Means. This chapter is an introduction to the Non-Local Means filter [Buades et al., 2005]. After a brief review of its history and of related filters, we propose the reader a "guided tour" of the Non-Local Means. To give more insight on the filter, we describe the effect of each parameter with the help of three toy examples for which we give a closed-form expression for the output of the filter. We illustrate the apparition of bias, even on periodic signals, depending on the patch size and on the search window, or the kernel of the weights. These examples give concrete illustrations of various empirical observations made in the literature, and the closed-form expressions make them independent of the experimental setting (parameters of the algorithm, realization of the noise...). In particular, we show that the use of weights with compact support reduces the influence of the search window.

To study the apparition of bias with the Non-Local Means and the associated notion of regularity, we make a brief excursion in the patch space, inspired by [Peyré, 2009, Tschumperlé and Brun, 2009, Singer et al., 2009]. We do not assume that images lie on a manifold in the patch space, but we emphasize the importance of a weighted moment of the distribution of patches. This basic point of view is used to explain the behavior of the filter when the search window or the patch size vary.

To sum up, this chapter exposes the drawbacks and limitations of the Non-Local Means. It also shows that in most cases, an adequate choice of a parameter like the bandwidth or the patch size allows to limit the impact of the artifact.

Chapter 9: Spatially adaptive choice of the bandwidth h. We focus on designing a Non-Local Means filter with a local choice of the bandwidth, in order to circumvent the artifacts exhibited in Chapter 8. First, we build an oracle based on the bias variance trade-off which tells us the ideal behavior of such a filter. In fact, the distinction between textures and smooth areas is not the prominent criterion in the behavior of the oracle. Indeed, the most important variation depends on whether the pixel belongs to a patch that intersects an edge or not. Patches that are close to edges or details with high contrast usually have few similar copies, so that the Non-Local Means leave them with a noisy halo. The oracle forces a large bandwidth to reduce this halo, and so should a practical algorithm with local choice of the bandwidth.

Then we build such a practical algorithm, relying on Stein's Unbiased Risk Estimate (SURE)[Stein, 1981], first proposed in the context of the Non-Local Means in [Van De Ville and Kocher, 2009]. Since this estimator of the risk is not robust when used pixel-wise, we propose to regularize the risk by a local average, and then to keep the value of the bandwidth that minimizes the regularized estimate. A practical algorithm is given to perform this in reasonable time, and many experiments are shown. The procedure allows to reduce the halo near edges, and it preserves details with little contrast much better than the regular Non-Local Means. Yet, the decision is a bit brutal, and when the estimation of the risk is sometimes flawed, little spots of noise may be left, like with the global Non-Local Means. Inspired from the work [Salmon and Le Pennec, 2009b], we use an Exponentially Weighted Aggregation (EWA) to make a local convex combination of the filters with various bandwidth in function of their estimated risk. This procedure preserves the textures much better than the previous method, does not suffer from the noisy halo, and does not leave any spot of noise. The produced images look slightly noisier, but more pleasant, despite the fact that the PSNR is a little below the one of the previous method.

Chapter 10: From patches to shapes: NLM-SAP. This chapter relates a joint work with Charles Deledalle and Joseph Salmon. It is an attempt to reduce the noisy halo in the Non-Local Means. The idea stems from the work of Salmon and Strozecki [Salmon and Strozecki, 2010b], where it is proposed to shift patches in order to find similar copies near edges. In this chapter, we push this idea further by replacing shifted patches with arbitrary shapes. The point is to use shapes that are adapted to the local geometry and that can fit in narrow regions. We provide an algorithm based on the Fast-Fourier

Transform to compute the result of the Non-Local Means with an arbitrary patch shape, regardless of the size of the shape.

The question is then how to choose the shape at each pixel. We use a fixed dictionary of shapes, and at each pixel we must therefore decide what to do with the different estimators. As in the previous chapter, keeping only the estimator with the smallest SURE is too brutal, so we aggregate the results using the Weighted Average based on Variance [Salmon and Strozecki, 2010b] or the Exponentially Weighted Aggregation [Leung and Barron, 2006]. The resulting algorithm is fast, and it provides a good visual quality, even though in terms of PSNR it falls behind [Dabov et al., 2007, Goossens et al., 2008].

Publications

The material covered by Chapters 3, 4 and a part of Chapter 6 has been published in the SIAM Journal on Multiscale Modelling and Simulation (MMS) [Duval et al., 2009]. It was presented at the SIAM Conference on Imaging Sciences 2010. A tiny part (one page) of Chapter 5 is extracted from a paper published in the Journal of Mathematical Imaging and Vision (JMIV) [Duval et al., 2010]. A paper including approximately Chapters 8 and 9 has been accepted by the SIAM Journal of Imaging Science (SIIMS) [Duval et al., 2011]. The material contained in Chapter 10 is the subject of a paper jointly written with Charles Deledalle and Joseph Salmon which has been accepted by the Journal of Mathematical Imaging and Vision (JMIV) [Deledalle et al., 2011]. It will be presented at the Conference on Scale-Space and Variational Methods (SSVM) 2011.

Part I

Morphological Aspects of the TVL1 Model

Chapter 1

Notations and Preliminaries

1.1 Functions of bounded variation and the Gauss-Green formula

1.1.1 Functions of bounded variation

In this section, we recall some definitions and properties about functions of bounded variation. All considered functions will be defined on the plane and we shall denote by L^1 the space of integrable functions on $\mathbb{R}^2$. For more details on functions with bounded variation, we refer the reader to [Ambrosio et al., 2000] or [Ambrosio, 1997].

Definition 1.1.1. A function $u \in L^1$ is said to be of bounded variation if its distributional gradient is a vector valued Radon measure with finite total variation. The total variation of Du on an open set $\Omega \subset \mathbb{R}^2$ is equal to:

$$|Du|(\Omega) = \sup \left\{ \int_{\mathbb{R}^2} u \operatorname{div} \varphi / \ \varphi \in C_c^1(\Omega, \mathbb{R}^2), \forall x \in \Omega, \ |\varphi(x)| \leq 1 \right\} \tag{1.1}$$

(where for a vector $|v| = (v_1, v_2) \in \mathbb{R}^2$, we set $|v|^2 = v_1^2 + v_2^2$). When $\Omega = \mathbb{R}^2$, this quantity is called the total variation of u, and will be denoted by $\int |Du|$ (or sometimes $J(u)$).

The total variation of a Borel set $B \subset \mathbb{R}^2$ is defined as $\inf\{|Du|(A), \ A \text{ open }, B \subset A\}$. The total variation is lower semi-continuous with respect to L^1 convergence.

If $u = \mathbb{1}_E$ is the characteristic function of a measurable set $E \subset \mathbb{R}^2$ and has bounded variation, we say that E is a set of finite perimeter, and we write Per $E = |D\mathbb{1}_E|(\mathbb{R}^2)$, or Per $(E, \Omega) = |D\mathbb{1}_E|(\Omega)$. Observe that Per $(E, \Omega) =$ Per $(\mathbb{R}^2 \setminus E, \Omega)$.

The coarea formula below provides a link between the total variation of a function and the perimeter of its level sets.

Theorem 1.1.2 (Coarea Formula)**.** *Let $u \in BV$. Then for any open set $\Omega \subset \mathbb{R}^2$:*

$$|Du|(\Omega) = \int_{-\infty}^{+\infty} \text{Per } (\{x \in \Omega, u(x) > t\}, \Omega) \, dt. \tag{1.2}$$

For $u \in L^1(\Omega)$, if the right member of the equation is finite, then $u \in BV(\Omega)$ and the last equality holds.

If E is a set of finite perimeter, its reduced boundary $\partial^* E$ is defined as the set of all $x \in \text{Spt } D\mathbb{1}_E$ such that the limit $\nu(x) := \lim_{r \to 0+} -\frac{D\mathbb{1}_E(B(x,r))}{|D\mathbb{1}_E|(B(x,r)}$ exists with $|\nu(x)| = 1$. The vector $\nu(x)$ is called a generalized exterior normal of E at $x \in \partial^* E$. The reduced boundary is $\mathcal{H}^1$ countably rectifiable with finite one-dimensional Hausdorff measure $\mathcal{H}^1$, and $|D\mathbb{1}_E|$ coincides with the restriction of $\mathcal{H}^1$ to $\partial^* E$.

The following property is crucial when studying monotonicity properties of minimization problems with the total variation:

Proposition 1.1.3 (Submodularity of the perimeter)**.** *Let $E, F \subset \mathbb{R}^2$ be measurable sets, and $\Omega \subset \mathbb{R}^2$ be an open set. Then:*

$$\text{Per } (E \cap F, \Omega) + \text{Per } (E \cup F, \Omega) \leq \text{Per } (E, \Omega) + \text{Per } (F, \Omega)\,. \tag{1.3}$$

In this thesis, we only consider the case where the dimension of the ambient space is $N = 2$. We thus have: $BV(\Omega) \subset L^2(\Omega)$ (with continuous embedding). This is ensured by the isoperimetric inequality, which is also fundamental for the Cheeger problem.

Proposition 1.1.4 (Isoperimetric inequality)**.** *Let $f \in BV(\mathbb{R}^2)$. Then $f \in L^2$ and:*

$$\|f\|_2 \leq \frac{1}{2\sqrt{\pi}} \int |Df|. \tag{1.4}$$

If E is a set of finite perimeter, then either E or $\mathbb{R}^2 \setminus E$ has finite Lebesgue measure and:

$$\min\left(\sqrt{|E|}, \sqrt{|\mathbb{R}^2 \setminus E|}\right) \leq \frac{1}{2\sqrt{\pi}} \text{Per } E. \tag{1.5}$$

There is equality if and only if E or $\mathbb{R}^2 \setminus E$ is a disc.

1.1.2 Precise representative of a Lebesgue-measurable set

If $E \subset \mathbb{R}^2$ is a Lebesgue measurable set, and $x \in \mathbb{R}^2$, the upper and lower densities of E at x are respectively defined by:

$$\overline{D}(x, E) := \limsup_{r \to 0} \frac{|B(x, r) \cap E|}{|B(x, r)|}, \quad \underline{D}(x, E) := \liminf_{r \to 0} \frac{|B(x, r) \cap E|}{|B(x, r)|}.$$

If both quantities are equal, the common value $D(x, E)$ will be called the density of E at x. Lebesgue's density theorem states that the density of any measurable set is 1 at almost every point of this set. As a consequence, the set of points x for which $D(x, E) = 1$ (which does not depend of the chosen representative E) is itself a representative of E. In this thesis, the precise representative we choose for E is the set of points $x \in \mathbb{R}^2$ where E has density 1.

It is clear that if E' is in the Lebesgue class of equivalence of E, then $\partial E \subset \partial E'$. Therefore the topological boundary of E is the smallest (in the inclusion sense) topological boundary of the representatives of E. This definition is coherent with the notion of boundary used in [Ambrosio, 1994].

Notice that we always have $\partial^* E \subset \partial E$, but in general the inclusion is strict (see [Ambrosio, 1994] for a counterexample).

1.1.3 An extension of Green's formula

In this thesis, the integration by parts formulae play a crucial role, so that we need a version of the Green theorem with as few assumptions as possible on the smoothness of the considered functions and vectorfields. Following the framework of [Andreu-Vaillo et al., 2002], we rely on the results of [Anzellotti, 1983].

Let $\Omega \subset \mathbb{R}^2$ be an open subset, and for $p \in [1, +\infty]$,

$$X_p(\Omega) := \{z \in L^\infty(\Omega, \mathbb{R}^2),\ \text{div } z \in L^p(\Omega)\}. \tag{1.6}$$

If $z \in X_p(\Omega)$ and $w \in L^q(\Omega) \cap \text{BV}(\Omega)$ with $p^{-1} + q^{-1} = 1$, we define the functional $(z, Dw) : C_c^\infty(\Omega) \to \mathbb{R}$ by the formula:

$$\langle (z, Dw), \varphi \rangle := -\int_\Omega w\varphi \text{div } z dx - \int_\Omega wz \cdot \nabla\varphi dx. \tag{1.7}$$

In fact, (z, Dw) is a Radon measure in Ω, and for all Borel set $B \subset \Omega$:

$$\left|\int_B (z, Dw)\right| \leq \int_B |(z, Dw)| \leq \|z\|_\infty \int_B |Dw|. \tag{1.8}$$

Moreover, if $w \in L^q(\Omega) \cap W^{1,1}$, $\int_\Omega (z, Dw) = \int_\Omega z \cdot \nabla w dx$.

The following theorem is proved in [Anzellotti, 1983]:

Theorem 1.1.5. *Let $\Omega \subset \mathbb{R}^2$ be a bounded open set with Lipschitz boundary. Let $u \in BV(\Omega) \cap L^q(\Omega)$ and $z \in X_p(\Omega)$. Then there exists a function $[z \cdot \nu^\Omega] \in L^\infty(\partial\Omega)$ such that $\|[z \cdot \nu^\Omega]\|_{L^\infty(\partial\Omega)} \leq \|z\|_{L^\infty(\Omega,\mathbb{R}^2)}$, and*

$$\int_\Omega u \text{div } z dx + \int_\Omega (z, Du) = \int_{\partial\Omega} [z \cdot \nu^\Omega] u d\mathcal{H}^1.$$

When $\Omega = \mathbb{R}^2$, we have the following formula [Anzellotti, 1983], for $z \in X_p(\mathbb{R}^2)$ and $w \in L^q(\mathbb{R}^2) \cap \text{BV}(\mathbb{R}^2)$:

$$\int_{\mathbb{R}^2} w \text{div } z dx + \int_{\mathbb{R}^2} (z, Dw) = 0. \tag{1.9}$$

1.2 Subdifferential of a convex function

In this section, V is a normed space and V^* is its topological dual space. We recall here some basic facts of convex analysis. We refer the reader to [Ekeland and Temam, 1999] for more details. The goal of this section is to characterize the subdifferential of the total variation as in [Andreu-Vaillo et al., 2002].

1.2.1 Convex functions and the Legendre-Fenchel transform

Definition 1.2.1. Let $F : V \to \overline{R}$, we say that:

- F is convex if for all $t \in [0, 1]$, $\quad F(tu + (1-t)v) \leq tF(u) + (1-t)F(v)$, whenever the right hand-side is defined.
- F is proper convex if it is convex and F nowhere takes the value $-\infty$ and is not identically equal to $+\infty$.
- $F \in \Gamma(V)$ if F is the pointwise supremum of continuous affine functions.

In fact, the functions in $\Gamma(V)$ are almost the convex lower semi-continuous functions:

Proposition 1.2.2. *The following properties are equivalent:*

(i) $F \in \Gamma(V)$.

(ii) *F is convex lower semi-continuous (l.s.c.) and if F takes the value $-\infty$ then F is identically equal to $-\infty$.*

In the study of subdifferentials, the Legendre-Fenchel transform will be helpful.

Definition 1.2.3 (Legendre-Fenchel transform)**.** Let $F : V \to \overline{\mathbb{R}}$. The conjugate function of F is $F^* : V^* \to \overline{\mathbb{R}}$ defined by:

$$\forall p \in V^*, \quad F^*(p) = \sup_{u \in V} \{\langle u, p\rangle - F(u)\}. \tag{1.10}$$

A straightforward consequence of that definition is that $F^* \in \Gamma(V^*)$ (it is therefore convex l.s.c.) and the following inequality, called Fenchel's inequality, holds:

$$\langle u, p\rangle \leq F(u) + F^*(p), \tag{1.11}$$

for all $u \in V, \forall p \in V^*$ such that the right hand-side is defined.

An interesting property is that functions of $\Gamma(V)$ may be written as their biconjugate:

Proposition 1.2.4. *If $F \in \Gamma(V)$, then $F = F^{**}$.*

1.2.2 Subdifferential of a convex l.s.c. function

Definition 1.2.5 (Subdifferential of F)**.** Let $u \in V$ and $p \in V^*$. We say that $p \in \partial F(u)$ if $F(u)$ is finite and:

$$\forall v \in V,\ F(v) \geq F(u) + \langle p, v - u \rangle. \tag{1.12}$$

Notice that the subdifferential of a convex function is a (possibly empty) closed convex set.

The interest of the subdifferential in variational problems comes from Fermat's rule (or Euler-Lagrange equation):

$$F(u) = \min_{v \in V} F(v) \text{ if and only if } 0 \in \partial F(u). \tag{1.13}$$

As F will often be the sum of two functionals, it is useful to know the subdifferential of a sum:

Proposition 1.2.6. *Let $G, H : V \mapsto \mathbb{R} \cup +\infty$ be two convex, l.s.c., proper functions. If dom $G \cap$ int dom $H \neq \emptyset$, then:*

$$\partial(G + H) = \partial G + \partial H. \tag{1.14}$$

Moreover, the link between the subdifferential and the Legendre-Fenchel transform is described in the next proposition:

Proposition 1.2.7. *Let $F : V \to \overline{\mathbb{R}}$, and $u \in V$. Then $p \in \partial F(u)$ if and only if $\langle u, p \rangle = F(u) + F^*(p)$.*

The above condition being symmetric in u and p, we see that, for $F \in \Gamma(V)$, $p \in \partial F(u)$ if and only if $u \in \partial F^*(p)$.

1.2.3 Convex positively homogeneous functions

Let us consider a set $A \subset V^*$. Its support function, $S_A : V \to \mathbb{R} \cup \{+\infty\}, u \mapsto \sup_{p \in A} \langle u, p \rangle$ is convex, l.s.c and positively homogeneous, i.e.:

$$\forall u \in V, \forall t > 0, \quad S_A(tu) = tS_A(u). \tag{1.15}$$

Notice that A and its closed convex hull $\overline{\text{co}}A$ have the same support function.

Conversely, any function $F \in \Gamma(V)$ which is positively homogeneous can be written as the support function of some closed convex set $A \subset V^*$. Indeed, let us define $A = \{p \in V^*,\ \forall u \in V,\ \langle p, u \rangle \leq F(u)\}$ and observe that:

$$F^*(p) = \sup_{u \in V} \sup_{t > 0} \left(\langle p, tu \rangle - F(tu) \right) \tag{1.16}$$

$$= \sup_{u \in V} \sup_{t > 0} t \left(\langle p, u \rangle - F(u) \right) \text{ by homogeneity,} \tag{1.17}$$

$$= \chi_A(p), \quad \text{with } \chi_A(p) = \begin{cases} 0 & \text{if } p \in A, \\ +\infty & \text{otherwise.} \end{cases} \tag{1.18}$$

χ_A is called the indicator function of A. Since $F = F^{**}$, we see that F is the support function of A.

The duality between indicator functions and support functions allows to compute the subdifferential of a convex l.s.c. positively homogeneous function. Indeed, by Proposition 1.2.7, $p \in \partial F(u)$ if and only if $\langle u, p \rangle = F(u) + F^*(p)$. Let $A \subset V^*$ be a closed convex set such that F is the support function of A, then $F^* = \chi_A$. As a result,

$$p \in \partial F \text{ if and only if } p \in A \text{ and } \langle u, p \rangle = F(u). \tag{1.19}$$

1.2.4 Applications

1.2.4.1 Subdifferential of the L^1-norm

Let $V = L^1$, so that $V^* = L^\infty$. Since the L^1 norm can be written as:

$$\forall u \in L^1, \quad \|u\|_1 = \sup\{\int_{\mathbb{R}^2} u(x)p(x)dx,\ p \in L^\infty, \|p\|_\infty \leq 1\}, \tag{1.20}$$

we see that it is the support function of the set $A := \{p \in L^\infty, \|p\|_\infty \leq 1\} \subset V^*$ which is obviously closed and convex. As a consequence, if $F = \|\cdot\|_1$,

$$p \in \partial F(u) \text{ if and only if } p \in L^\infty,\ \|p\|_\infty \leq 1 \text{ and } \int_{\mathbb{R}^2} p(x)u(x)dx = \int_{\mathbb{R}^2} |u(x)|dx. \tag{1.21}$$

Roughly speaking, the subdifferential of the L^1-norm at u is the sign of u.

1.2.4.2 Subdifferential of the total variation

Let $V = L^1$, so that $V^* = L^\infty$. The total variation may be written as:

$$\forall u \in L^1, \quad \int |Du| = \sup\{\int_{\mathbb{R}^2} u(x)\text{div } z(x),\ z \in X_\infty, \|z\|_\infty \leq 1\}, \tag{1.22}$$

where $X_\infty = \{z \in L^\infty(\mathbb{R}^2, \mathbb{R}^2),\ \text{div } z \in L^\infty\}$. As a consequence, the total variation is the support function of $A := \{\text{div } z,\ z \in X_\infty,\ \|z\|_\infty \leq 1\} \subset L^\infty$.

The fact that A is convex is clear. To see that A is closed, consider a sequence of elements $p_n = \text{div } z_n \in A$ that converges to $p \in L^\infty$. Since $\{z \in L^\infty, \|z\|_\infty \leq 1\}$ is compact for the weak-* topology, we can extract a subsequence $z_{n'}$ which weakly-* converges to some $z \in L^\infty$ with $\|z\|_\infty \leq 1$. Then in the distribution sense $\text{div } z = p$, so that $p \in A$.

As a result, if $F : u \mapsto \int |Du|$:

$$p \in \partial F(u) \text{ if and only if } p \in L^\infty,\ \exists z \in L^\infty,\ \|z\|_\infty \leq 1,\ p = \text{div } z \text{ and } \int_{\mathbb{R}^2} \text{div } z(x)u(x)dx = \int |Du|. \tag{1.23}$$

The last equality may be rewritten: $\int_{\mathbb{R}^2}(z, Du) = \int |Du|$. This means that z is orthogonal to the level lines of u.

1.2.4.3 Subdifferential of the G-norm

Let $V = L^2$, so that $V^* = L^2$. The G-norm can be defined as :

$$\|u\|_G = \sup\{\int_{\mathbb{R}^2} v(x)u(x)dx,\ v \in \text{BV}(\mathbb{R}^2), \int |Dv| \leq 1\}. \tag{1.24}$$

It is therefore the support function of $A := \{v \in \text{BV}(\mathbb{R}^2),\ \int |Dv| \leq 1\} \subset V^*$, and A is closed and convex. As a consequence, if $F = \|\cdot\|_G$,

$$p \in \partial F(u) \text{ if and only if } p \in \text{BV}(\mathbb{R}^2),\ \int |Dp| \leq 1 \text{ and } \int_{\mathbb{R}^2} p(x)u(x)dx = \|u\|_G. \tag{1.25}$$

1.3 Calibrable sets and the Cheeger problem

1.3.1 Calibrable sets and the total variation flow

Calibrable sets appear naturally in the study of the total variation flow:

$$\frac{\partial u}{\partial t} = \text{div}\left(\frac{Du}{|Du|}\right), \quad \text{in } (0,+\infty)\times\Omega \tag{1.26}$$

$$u(t,x) = 0 \text{ on } (0,+\infty)\times\partial\Omega \tag{1.27}$$

$$u(0,x) = u_0(x) \text{ for } x\in\Omega \tag{1.28}$$

where Ω is a bounded open set with Lipschitz boundary, $u_0 \in L^1(\Omega)$. In [Andreu et al., 2001a], Andreu *et al.* show existence and uniqueness of solutions in the weak sense for this equation. In [Andreu et al., 2002], it is proved that there exists a finite extinction time $T(u_0)$, and that the rescaled functions $\frac{u(t)}{T(u_0)-t}$ converge along subsequences to a solution of the eigenvalue problem:

$$\begin{cases} -\text{div}\left(\frac{Du}{|Du|}\right) = u \\ u = 0 \text{ on } \partial\Omega \end{cases}. \tag{1.29}$$

This eigenvalue problem was studied by Belletini *et al.* [Bellettini et al., 2005b]. To avoid boundary conditions they studied this problem in the plane:

$$-\text{div}\left(\frac{Du}{|Du|}\right) = u, \quad u\in L^1_{loc}(\mathbb{R}^2). \tag{1.30}$$

In the case where u is the characteristic function of a set $E\subset\mathbb{R}^2$, E is said to be *calibrable*. Calibrable sets are the sets that evolve at constant speed by the total variation flow. More precisely, such sets satisfy the condition $\lambda_E \mathbb{1}_E \in \partial J(\mathbb{1}_E)$ for some $\lambda_E\in\mathbb{R}$, where ∂J is the subdifferential of the total variation.

Definition 1.3.1. Let $E\subset\mathbb{R}^2$ be a bounded set with finite perimeter. We say that E is calibrable if there exists a vector field $\xi\in X_\infty$ with $\|\xi\|_\infty\le 1$ such that $-\text{div}\,\xi = \lambda_E\mathbb{1}_E$ in $\mathcal{D}'(\mathbb{R}^2)$ for some λ_E, and $(\xi, D\mathbb{1}_E) = |D\mathbb{1}_E|$ as measures in $\mathbb{R}^2$.

In that case, λ_E is uniquely determined by $\lambda_E = \frac{\text{Per}\,E}{|E|}$ where $|E|$ denotes the two-dimensional (Lebesgue) area. In the rest of this thesis, we shall use this notation to refer to $\frac{\text{Per}\,E}{|E|}$ whether $E\subset\mathbb{R}^2$ is calibrable or not.

Calibrable sets were characterized in [Bellettini et al., 2002] using the following two propositions:

Proposition 1.3.2 ([Bellettini et al., 2002])**.** *Let $F\subset\mathbb{R}^2$ be a calibrable set. Then:*

(i) $\frac{\text{Per}\,F}{|F|} \le \frac{\text{Per}\,U}{|U\cap F|}$ *for all* $U\subset\mathbb{R}^2$.

(ii) Each connected component of F is convex.

Proposition 1.3.3. *Let $C\subset\mathbb{R}^2$ be a bounded convex set with finite perimeter. The following assertions are equivalent:*

(i) C is calibrable,

(ii) C is a solution of the problem:

$$\min_{X\subset C} \text{Per}\,X - \lambda_C|X| \quad \left(\textit{where } \lambda_C = \frac{\text{Per}\,C}{|C|}\right). \tag{1.31}$$

A very interesting property of convex calibrable sets is the following characterization, proved in [Giusti, 1978, Theorem A.1], [Alter et al., 2005a], [Bellettini et al., 2002], [Kawohl and Lachand-Robert, 2006]. Let us recall that if E is of class $C^{1,1}$, the curvature of ∂E is defined $\mathcal{H}^1$ almost everywhere. It will be denoted by κ.

Proposition 1.3.4. *Let C be a bounded convex set. Then C is calibrable if and only if the following two conditions hold:*

- *∂C is of class $C^{1,1}$*
- $\operatorname{ess\,sup}_{x\in\partial C} |\kappa(x)| \leq \frac{\operatorname{Per} C}{|C|}$.

Example: Using the above characterization, Kawohl and Lachand-Robert [Kawohl and Lachand-Robert, 2006] proved that an ellipse with eccentricity lower than $\bar{e} \approx 0.7192$ is calibrable.

We refer the reader to the monograph [Andreu-Vaillo et al., 2002] for a self-contained study of the total variation flow and calibrable sets in the plane.

1.3.2 The Cheeger problem

A related topic is the Cheeger problem. Given a nonempty open bounded set $\Omega \subset \mathbb{R}^2$, find:

$$h_\Omega := \inf_{X\subset\Omega} \frac{\operatorname{Per} X}{|X|}, \tag{1.32}$$

where we only consider sets X with positive Lebesgue measure. The existence of a minimizer follows from the direct method of the calculus of variations. Any minimizer X of this problem is called a Cheeger set of Ω.

The Cheeger constant was introduced in [Cheeger, 1970], where it was shown that for a general bounded open set $\Omega \subset \mathbb{R}^2$ with Lipschitz boundary, the Cheeger constant gives a lower bound on the first eigenvalue of the Laplacian:

$$h_\Omega \geq \left(\frac{\mu_2(\Omega)}{2}\right)^2, \tag{1.33}$$

where $\mu_2(\Omega) = \min_{0\neq v\in W_0^{1,2}(\Omega)} \frac{\int_\Omega |\nabla v|^2 dx}{\int_\Omega |v|^2 dx}$. A similar inequality was proved in [Lefton and Wei, 1997] where the exponent 2 is replaced with $p \in (1, +\infty)$, and $\mu_p(\Omega) = \min_{0\neq v\in W_0^{1,p}(\Omega)} \frac{\int_\Omega |\nabla v|^p dx}{\int_\Omega |v|^p dx}$.

If Ω itself is a minimizer of (1.32), we say that Ω is Cheeger in itself. For instance, a disc $\Omega = B(0,R)$ is Cheeger in itself. This is a consequence of the isoperimetric inequality (1.5):

$$\forall X \subset \Omega, \quad \frac{\operatorname{Per} X}{|X|} \geq \frac{2\sqrt{\pi}\sqrt{|X|}}{|X|} \geq \frac{2\sqrt{\pi}}{\sqrt{|\Omega|}} = \frac{2}{R} = \frac{\operatorname{Per}\Omega}{|\Omega|}.$$

Notice that Ω is Cheeger in itself if and only if it is solution of the problem: $\min_{X\subset\Omega} \operatorname{Per} X - h_\Omega |X|$. As a consequence, by Proposition 1.3.2, a calibrable set is Cheeger in itself, and **the converse is true when the set is convex** (by Proposition 1.3.3).

Whereas the existence of a Cheeger set is straightforward, its uniqueness is a difficult problem. In general the Cheeger set of Ω is not unique (see [Kawohl and Lachand-Robert, 2006]), but it is unique if Ω is a convex body, i.e. a bounded convex set which contains the origin in its interior (the result was proved in dimension N [Caselles et al., 2007b] for a C^2 uniformly convex set, and in the general case in [Alter and Caselles, 2009]):

Theorem 1.3.5 ([Caselles et al., 2007b], [Alter and Caselles, 2009])**.** *Let C be a non-trivial convex body (i.e. a nonempty compact convex subset of $\mathbb{R}^2$). Then there is a unique Cheeger set inside C. This set is convex and of class $C^{1,1}$.*

This Cheeger set is given explicitly by the following theorem (see [Alter et al., 2005b] and [Kawohl and Lachand-Robert, 2006]):

Theorem 1.3.6. *Let C be a non-trivial convex body, then there exists a unique value $t = t* > 0$ such that the Cheeger set of C is given by an opening with radius t^*. This value t^* is the unique value of t such that the area of the eroded set of C with radius t is equal to πt^2.*

Whereas the proof in [Kawohl and Lachand-Robert, 2006] relies on calculations on polygons and a passage to the limit, the one in [Alter et al., 2005b] relies on argument linked with image processing: if $f = \mathbb{1}_C$, the level sets $(U_t)_{t\in[0,1]}$ of the solution of the ROF problem:

$$\min_{u\in \mathbf{BV}(\mathbb{R}^2)} \int |Du| + \frac{\mu}{2}\int_{\mathbb{R}^2}(u-f)^2, \tag{1.34}$$

minimize the following geometric problems: Per $U - \mu(1-t)|U|$ for $U \subset C$. As a consequence the set $\{x, u(x) \geq \|u\|_\infty\}$ is the Cheeger set of C.

Let us mention that the Cheeger problem has also been studied with weights on the total variation or on the area: in that case the uniqueness of the Cheeger set may not hold, but Buttazzo *et al.* [Buttazzo et al., 2007] have shown the existence of a unique maximal Cheeger set, (see also [Carlier and Comte, 2007], and [Carlier et al., 2009] where numerical experiments are carried out). Moreover, the results of this section have been extended to the anisotropic case: Moll has studied the well-posedness of the anisotropic total variation flow [Moll, 2005], and the characterization of calibrable sets with respect to anisotropic norms is given in [Bellettini et al., 2001] in dimension 2 and in [Caselles et al., 2008] for dimension $N \geq 2$. In [Kawohl and Novaga, 2008], Kawohl and Novaga show the existence and uniqueness of the Cheeger set of a convex set. In dimension 2 the Cheeger set inside a convex set is given by an opening ([Bellettini et al., 2001, Kawohl and Novaga, 2008]). Applications of the anisotropic Cheeger problem to segmentation (with the geodesic active contours) and colorization can be found in [Caselles et al., 2009]. Eventually, a variant of the Cheeger problem in infinite dimension with Gauss measure is studied in [Caselles et al., 2010].

Chapter 2

How to study the TVL1 problem?

Since the introduction of the TVL1 model in signal processing by Alliney [Alliney, 1992, Alliney, 1997] and in image processing by Nikolova [Nikolova, 2002, Nikolova, 2004a], many different tools have been used to study its solutions: introduction of the dual problem in [Haddad, 2007], optimality conditions regarding the G-norm of the sign of $u - f$ [Haddad, 2007, Kindermann et al., 2006], maximizing forms in [Morgan and Vixie, 2007] ...

The aim of this chapter is to give an overview of some of these methods and to show that they essentially amount to the Euler-Lagrange equation for the TVL1 functional. Although this thesis (and Chapter 3 in particular) emphasizes the geometrical aspects of the TVL1 problem, it is actually also connected to the Euler Lagrange equation via the results of [Bellettini et al., 2002, Bellettini et al., 2005b, Alter et al., 2005a] which are related to the theory of calibrations.

2.1 Introduction

The TVL1 problem consists, given $f \in L^1$, in finding a minimizer of the energy:

$$\int |Du| + \lambda \int_{\mathbb{R}^2} |f(x) - u(x)| dx. \tag{2.1}$$

This model was introduced by Alliney for one-dimensional signals [Alliney, 1992, Alliney, 1996, Alliney, 1997]. It was Nikolova who introduced it in image processing in [Nikolova, 2002]: studying the detection and removal of outliers in variational methods, she shows that non-smooth data fidelity terms behave better than smooth ones. Consequently, she notices that the TVL1 model is well adapted to the removal of impulse noise, which she illustrates in [Nikolova, 2004a] with striking numerical results. Since then, many authors have studied the TVL1 model, each focussing on a particular aspect. Chan and Esedoglu [Chan and Esedoglu, 2005] study the evolution of the energy as the fidelity parameter increases. Yin *et al.* [Yin et al., 2007b] study the monotonicity of the geometric problem, in view of the geometric equivalence, whereas Morgan and Vixie [Morgan and Vixie, 2007] insist on the link with geometric measure theory by drawing a connection with the flat norm (see [Federer, H., 1969]). In [Vixie and Esedoglu, 2007], Vixie and Esedoglu give geometric bounds on the solutions of TVL1. Allard also employs tools from geometric measure theory to study all the $TV + L^p$ problems $p \geq 1$: in [Allard, 2007] he studies the smoothness of the boundary of the level sets of the solution, in [Allard, 2008] he studies the problem where the input is the characteristic function of convex set. Notice that he states in a more general setting similar results as those of Chapter 3, but that we obtained these results independently, relying essentially on the work of Alter, Caselles, Chambolle [Alter et al., 2005a, Alter et al., 2005b]. In [Allard, 2009], he gives particular examples of solutions.

Haddad also investigated the problem: he characterizes radial solutions, and he also studies the dual problem of (2.1) in [Haddad, 2007]. Kindermann *et al.* [Kindermann et al., 2006] have characterized the optimality of a function using the G-norm introduced in [Meyer, 2001].

Let us also mention the many applications of the model: impulse noise removal [Nikolova, 2004a, Koko and Jehan-Besson, 2010, Liu et al., 2010], segmentation [Chan et al., 2006, Bresson et al., 2007] feature detection in biomedical imaging[Cunha et al., 2007], 3D MRI registration [Chen et al., 2005], cDNA microarray data filtering [Yin et al., 2005], or cartoon-texture decomposition [Aujol et al., 2006, Yin et al., 2007a, Haddad, 2007], computation of the optical flow in videos [Wedel et al., 2008a, Wedel et al., 2008b], or shape analysis [Vixie et al., 2010].

All these authors have used various techniques. Yet, we illustrate in this chapter that such methods essentially amount to the Euler-Lagrange equation.

2.2 Euler Lagrange equation

A classical way to characterize the minimizers u of a convex energy $\mathbb{E}$ is to apply Fermat's rule, i.e. write $0 \in \partial\mathbb{E}(u)$. In the calculus of variations, this leads to the Euler-Lagrange equation which often gives precious information about the solution of a variational problem. In the case of TVL1, since both terms of the functional are non-smooth, the formal equation:

$$-\operatorname{div}\left(\frac{Du}{|Du|}\right) + \lambda \operatorname{sign}\ (u-f) = 0 \tag{2.2}$$

should be properly written using the subdifferentials of the total variation and the L^1 norm. Applying the framework of [Andreu-Vaillo et al., 2002] (see Section 1.2), we see that a function $u \in L^1$ is solution if and only if there exists a vectorfield $z \in L^\infty(\mathbb{R}^2, \mathbb{R}^2)$ such that div $z \in L^\infty(\mathbb{R}^2)$, a function $p \in L^\infty(\mathbb{R}^2)$ such that:

$$|z| \leq 1 \text{ a.e.}, \qquad \int |Du| = \int_{\mathbb{R}^2} u \operatorname{div} z, \tag{2.3}$$

$$|p| \leq 1 \text{ a.e.}, \qquad \int_{\mathbb{R}^2} |u-f| = \int_{\mathbb{R}^2} (u-f)p, \tag{2.4}$$

$$\text{and} \qquad \operatorname{div} z + \lambda p = 0. \tag{2.5}$$

The condition on the right hand-side of Equation (2.3) may be rewritten $\int |Du| = -\int (z, Du)$, or simply $-(z, Du) = |Du|$, $|Du|$-a.e., which essentially means that z is orthogonal to the level lines of u.

Thus, in the case where u is the characteristic function of a set $U \subset \mathbb{R}^2$, z must coincide with the outer unit normal of U, $\mathcal{H}^1$ a.e. in $\partial^* U$.

2.3 The dual problem

In [Haddad, 2007], A. Haddad derived the dual problem of TVL1. From the primal problem:

$$\inf_{u \in BV(\mathbb{R}^2)} \int |Du| + \lambda \int_{\mathbb{R}^2} |f-u| = \inf_{u \in BV(\mathbb{R}^2)} \sup_{\left\{ \begin{smallmatrix} \varphi \in X_\infty, & p \in L^\infty, \\ |\varphi| \leq 1, & |p| \leq 1 \end{smallmatrix} \right\}} \int_{\mathbb{R}^2} u \operatorname{div} \varphi + \lambda \int_{\mathbb{R}^2} (f-u)p, \tag{2.6}$$

one defines the dual problem as:

$$\left\{ \begin{array}{l} \sup \\ \varphi\in X_\infty, p\in L^\infty, \\ |\varphi|\leq 1, |p|\leq 1 \end{array} \right. \inf_{u\in BV(\mathbb{R}^2)} \int_{\mathbb{R}^2} u\text{div }\varphi + \lambda \int_{\mathbb{R}^2} (f-u)p = \left\{ \begin{array}{l} \sup \\ \varphi\in X_\infty, p\in L^\infty, \\ |\varphi|\leq 1, |p|\leq 1 \end{array} \right. \inf_{u\in BV(\mathbb{R}^2)} \int_{\mathbb{R}^2} u(\text{div }\varphi - \lambda p) + \int_{\mathbb{R}^2} fp$$

$$= \left\{ \begin{array}{l} \sup \\ \varphi\in X_\infty, p\in L^\infty, \\ |\varphi|\leq 1, |p|\leq 1 \\ \text{div }\varphi - \lambda p = 0 \end{array} \right. \inf_{u\in BV(\mathbb{R}^2)} \lambda \int_{\mathbb{R}^2} fp$$

$$= \left\{ \begin{array}{l} \sup \\ \varphi\in X_\infty, \\ |\varphi|\leq 1, |\text{div }\varphi|\leq \lambda \end{array} \right. \int_{\mathbb{R}^2} f\text{div }\varphi \tag{2.7}$$

Using a minimax theorem, Haddad shows that the two problems have the same values. Moreover, the equality in the minimax inequality implies that their solutions are related by the following condition:

$$\int_{\mathbb{R}^2} \overline{u}\text{div }\overline{\varphi} = \int |D\overline{\varphi}| \quad \text{and } \lambda \int_{\mathbb{R}^2} |f-\overline{u}| = \int (f-\overline{u})\text{div }\overline{\varphi} \tag{2.8}$$

In other words, at least formally:

$$-\text{ div }\left(\frac{D\overline{u}}{|D\overline{u}|}\right) = \text{div }\overline{\varphi} = \lambda \text{ sign }(f-\overline{u}) \tag{2.9}$$

(which is nothing but the Euler-Lagrange equation (2.2)).

2.4 The flat norm

Although the origin of the TVL1 model in signal processing is usually attributed to the work of Alliney [Alliney, 1992, Alliney, 1997], it appears in earlier works under different forms, notably in Geometric Measure Theory. Not to mention the link with the Prescribed Mean Curvature problem [Ambrosio, 1997] (which will be discussed in Chapter 3), a correspondence between the geometric problem (3.26) and the Flat Norm (see [Federer, H., 1969, 4.1.12]) was established by Morgan and Vixie in [Morgan and Vixie, 2007]. We will need some basic concepts of the theory of currents (see [Federer, H., 1969] for more details).

Let $\mathbb{D}^m$ be the space of C^∞ differential forms of degree m with compact support, endowed with the topology defined in [Federer, H., 1969, 4.1.1]. Its dual space, $\mathbb{D}_m$, is the space of m-dimensional currents. The mass norm is defined on $\mathbb{D}_m$ by:

$$\forall T\in\mathbb{D}_m, \quad M(T) = \sup\{T(\phi),\ \phi\in\mathbb{D}^m \text{ and } \|\phi\|\leq 1\}. \tag{2.10}$$

Now, given a current $T\in\mathbb{D}_m$, one defines its flat norm as:

$$\mathbb{F}(T) := \inf_{S\in\mathbb{D}_{m+1}} M(T-\partial S) + M(S). \tag{2.11}$$

The flat norm induces a topology on the space $\mathbb{D}_m$ of m-dimensional currents that is stronger than the weak topology but weaker than the mass norm topology. It is generally acknowledged to give a good indication of when surfaces are geometrically close to one another, contrary to the mass norm (see [Morgan, 2009]). Recently, Glaunï¿½s [Glaunès, 2005, Vaillant and Glaunes, 2005, Glaunès and Joshi, 2006] proposed to use the flat norm in image processing to compute distances between shapes.

The remark of Morgan and Vixie in [Morgan and Vixie, 2007] is the following: given a set $F \subset \mathbb{R}^2$, the infimum of the geometric energy $\mathcal{E}_G(U)$ over all $U \subset \mathbb{R}^2$ can be written as the flat norm of its boundary:

$$\inf_{U \subset \mathbb{R}^2} \text{Per}\, U + |U \Delta F| = \mathbb{F}(\partial F) \tag{2.12}$$

This link with Geometric Measure Theory is interesting for two reasons. First, it provides a theoretical framework for the denoising of objects that are "thinner" than with TVL1. Indeed, the theory of functions with bounded variations is adapted to deal with codimension 1 surfaces. The theory of currents has no such restriction on the dimensionality: one may consider the denoising of curves in the plane, or surfaces in the Euclidean space. Second, one may apply the dual formula derived in [Federer, H., 1969, 4.1.12]:

$$\mathbb{F}(T) = \sup \{T(\phi),\ \phi \in \mathbb{D}^m, \|\phi\| \leq 1, \|d\phi\| \leq 1\}, \tag{2.13}$$

to obtain directly the dual problem of TVL1. To be more precise, Morgan and Vixie propose a scaled version of the Flat Norm corresponding to the geometric problem with parameter λ, and they show that:

$$\mathbb{F}_\lambda(T) := \inf_{S \in \mathbb{D}_{m+1}} M(T - \partial S) + \lambda M(S) = \sup \{T(\phi),\ \phi \in \mathbb{D}^m, \|\phi\| \leq 1, \|d\phi\| \leq \lambda\}. \tag{2.14}$$

Using the correspondence between vectorfields and differential forms of degree $N-1$ in $\mathbb{R}^N$, let us observe that this may be rewritten as:

$$\inf_{U \subset \mathbb{R}^2} \text{Per}\, U + \lambda |U \Delta F| = \mathbb{F}_\lambda(\partial F) = \sup \left\{ \int_F \text{div}\, \varphi,\ \varphi \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2), |\varphi| \leq 1,\ |\text{div}\, \varphi| \leq \lambda \right\}, \tag{2.15}$$

which is precisely the dual problem (2.7).

The formulation (2.14) leads Morgan and Vixie to study several examples with the help of maximizing forms on [Morgan and Vixie, 2007].

2.5 Calibrations

Calibrations are a tool used in geometry to prove that a surface minimizes area in its homological class. Whereas the more common (indirect) way to prove that a surface is area minimizing usually consists in:

- establishing the existence of a minimizer,
- stating necessary conditions on the minimizer (curvature, symmetry, optimal junctions angles...),
- comparing the few candidates of the remaining "short list",

the calibration method is more direct. Given a hypersurface[1] with boundary $S \subset \mathbb{R}^N$, the idea is to build a vectorfield z such that:

- $|z(x)| \leq 1$ for all $x \in \mathbb{R}^N$,
- $z \cdot \nu = 1$ for all $x \in S$, where ν is the positive unit normal,
- $\text{div}\, z = 0$ in $\mathbb{R}^N$.

[1] The method is not restricted to hypersurfaces, but we only mention this case for simplicity.

Then, for any other hypersurface $S' \subset \mathbb{R}^N$ such that $\partial S' = \partial S$, by Stokes' theorem:

$$\mathcal{H}^{N-1}(S') \geq \int_{S'} z \cdot \nu d\mathcal{H}^{N-1} = \int_S z \cdot \nu d\mathcal{H}^{N-1} = \mathcal{H}^{N-1}(S). \tag{2.16}$$

Therefore, given a calibration, it is straightforward to show that a hypersurface minimizes area. The most difficult part is generally to build the calibration. In the example below, we use the minimal surface equation to build the calibration, and such a construction is always feasible when the surface is the graph of a function defined on a convex domain D.

The history of calibrations is summarized in [Morgan, 2009]. The first uses of calibrations can be found implicitly in the work of Wirtinger concerning complex analytic varieties (1936), and explicitly in the work of de Rham concerning complex analytic subvarieties (1957), then Federer (1967) who proves that the holomorphic submanifolds of a Kï¿½hler manifold are area minimizing. The term calibration comes from the founding paper of Harvey and Lawson [Harvey and Lawson, 1982]. Since then, the use of calibrations has spread and many extensions have been proposed. Let us mention in particular the paired calibrations by Lawlor and Morgan [Morgan and Lawlor, 1994] and Brakke [Brakke, 1991], which allows to deal with the interface of more than two immiscible fluids, and which have strong connections with the segmentation problem [Chambolle et al., 2008]. Also, Lawlor recently proposed a generalization of calibrations called metacalibrations which allows to deal with volume constraints, boundary constraints or length constraints [Lawlor, 2010]. An important work in connection with image processing is the contribution of Alberti *et al.* [Alberti et al., 2003] to the study of the Mumford-Shah functional and other non-convex functionals: by considering the subgraph of the solution, they adapt the calibration method to free discontinuity problems so that short and easy minimality results can be derived. The idea is at the core of a state-of-the-art algorithm by Pock *et al.* [Pock et al., 2010] that minimizes energies that are the sum of a convex regularization term and a non-convex fidelity term.

Below are given two examples. The first one illustrates calibrations as a traditional tool to prove that a surface is area minimizing. The second one comes back to the topic of the TVL1: it gives the solution of the TVL1 problem for the characteristic function of a disc, using an argument "in the spirit of calibrations".

Example: Catenoids are locally area minimizing surfaces In $\mathbb{R}^3$, the catenoid is defined by the equation:

$$\cosh z = \sqrt{x^2+y^2}, \text{ or } z = f_\pm(x,y) \text{ with:} \tag{2.17}$$

$$f_+(x,y) = \log(\sqrt{x^2+y^2} + \sqrt{x^2+y^2-1}) \text{ for } z \geq 0,$$
$$f_-(x,y) = \log(\sqrt{x^2+y^2} - \sqrt{x^2+y^2-1}) \text{ for } z > 0.$$

As an illustration of calibrations, let us show that the part of catenoid

$$S := \{(x,y,f_+(x,y)), (x,y) \in D\} \text{ where } D = \left\{(x,y) \in \mathbb{R}^2,\ 1 \leq \sqrt{x^2+y^2} \leq R\right\},$$

is area minimizing among the graphs of smooth functions $g : D \to \mathbb{R}$, with boundary condition $g(x,y) = 0$ for $\sqrt{x^2+y^2} = 1$, $g(x,y) = \log(R + \sqrt{R^2-1})$ for $\sqrt{x^2+y^2} = R$.

Straightforward calculations show that:

$$\frac{\partial f_+}{\partial x} = \frac{x}{\sqrt{x^2+y^2}\sqrt{x^2+y^2-1}}, \quad \frac{\partial f_+}{\partial y} = \frac{y}{\sqrt{x^2+y^2}\sqrt{x^2+y^2-1}},$$
$$\text{and } \sqrt{1+|\nabla f_+|^2} = \frac{\sqrt{x^2+y^2}}{\sqrt{x^2+y^2-1}},$$
$$\text{so that} \quad 0 = \text{div}\left(\frac{1}{\sqrt{1+|\nabla f_+|^2}}\nabla f_+\right).$$

The last equation is the minimal surface equation: it means that the surface has a vanishing mean curvature. It may also be interpreted as the fact that the vectorfield

$$p : \begin{array}{ccc} D \times \mathbb{R} & \longrightarrow & \mathbb{R}^3 \\ (x,y,z) & \longmapsto & \frac{1}{\sqrt{1+|\nabla f|^2}} \begin{pmatrix} -\nabla f_+ \\ 1 \end{pmatrix} \end{array} \tag{2.18}$$

is a calibration for the catenoid. For any admissible surface S' which is the graph of a function g, if we define $\underline{S'} = \{(x,y,z) \in D \times \mathbb{R},\ 0 \leq z \leq g(x,y)\}$:

$$\int_{\partial \underline{S'}} p \cdot \nu d\mathcal{H}^2 = \int_{\underline{S'}} \operatorname{div} p = 0 = \int_{\underline{S}} \operatorname{div} p = \int_{\partial \underline{S}} p \cdot \nu d\mathcal{H}^2. \tag{2.19}$$

Decomposing the integration domain $\partial \underline{S'} = S' \cup (D \times \{0\}) \cup \left(S^1(0,R) \times [0, \log(R + \sqrt{R^2-1})]\right)$, and noticing that $\mathcal{H}^2(S') \geq \int_{S'} p \cdot \nu d\mathcal{H}^2$ with equality for $S' = S$, we get:

$$\mathcal{H}^2(S') \geq \mathcal{H}^2(S). \tag{2.20}$$

S is therefore area minimizing.

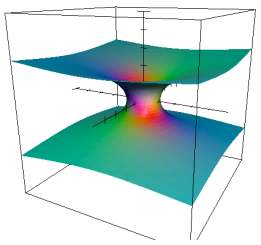


Figure 2.1 – *The catenoids are the only smooth minimal surfaces of revolution in* $\mathbb{R}^3$.

Example: The characteristic function of a disc. In [Chan and Esedoglu, 2005], Chan and Esedoglu show that if the input function is $f = \mathbb{1}_{B(0,R)}$ the output of TVL1 is given by:

$$u(x) = \begin{cases} \mathbb{1}_{B(0,R)}(x) & \text{if } \lambda > \frac{2}{R}, \\ c\mathbb{1}_{B(0,R)}(x) & \text{with } c \in [0,1], \text{ if } \lambda = \frac{2}{R}, \\ 0 \quad \text{if } 0 < \lambda < \frac{2}{R}. \end{cases} \tag{2.21}$$

To obtain this result, one may use symmetry arguments to see that the solution is radial and reduce the problem to the study of a function of one variable. The general solutions of TVL1 for radial input functions are given in [Haddad, 2007].

Here we give a proof that is in the spirit of calibrations and which is directly inspired from [Bellettini et al., 2001, Theorem 5.2]. It is also close to the argument given by Meyer in the study of the ROF model when the input is a characteristic function of a disk [Meyer, 2001], and it is a constructive example of the (general) argument given in [Chan and Esedoglu, 2005, Theorem 5.6] to show that the characteristic function of a C^2 domain is preserved by TVL1 for λ large enough. In fact it is a quite standard technique in the study of geometric problems [Bellettini et al., 2001, Alter et al., 2005a].

Consider the vectorfield $z(x) = \alpha(r)e_r$, where:

$$\alpha(r) = \begin{cases} \frac{R}{r} & \text{for } r > R \\ \frac{r}{R} & \text{for } 0 \leq r \leq R. \end{cases} \tag{2.22}$$

We see that $|z(x)| \leq 1$ for all $x \in \mathbb{R}^2$ and $z \cdot \nu = 1$ on $\partial B(0,R)$, where ν is the outer unit normal of the disk. Moreover,

$$\text{div } z(x) = \frac{1}{r}\frac{\partial\left(r\alpha(r)\right)}{\partial r} = \begin{cases} 0 & \text{for } r > R \\ \frac{2}{R} & \text{for } 0 \leq r < R \end{cases} . \tag{2.23}$$

The conclusion follows from a simple integration by parts:

- For $\lambda \geq \frac{2}{R}$:

$$\begin{aligned} E(u) - E(\mathbb{1}_{B(0,R)}) &= \int |Du| + \lambda \int_{\mathbb{R}^2} |u - \mathbb{1}_{B(0,r)}| - \int |D\mathbb{1}_{B(0,R)}| \\ &\geq \int_{\mathbb{R}^2} u\text{div } z + \lambda \int_{B(0,R)} |u-1| + \lambda \int_{\mathbb{R}^2\setminus B(0,R)} |u| - 2\pi R \\ &\geq \int_{B(0,R)} \underbrace{\left(\frac{2}{R}(u-1) + \lambda|u-1|\right)}_{\geq 0} + \lambda \int_{\mathbb{R}^2\setminus B(0,R)} |u| \geq 0. \end{aligned}$$

- For $\lambda \leq \frac{2}{R}$:

$$\begin{aligned} E(u) - E(0) &= \int |Du| + \lambda \int_{\mathbb{R}^2} |u - \mathbb{1}_{B(0,r)}| - \lambda\pi R^2 \\ &\geq \int_{\mathbb{R}^2} u\text{div } z + \lambda \int_{B(0,R)} (|u-1| - 1) + \lambda \int_{\mathbb{R}^2\setminus B(0,R)} |u| \\ &\geq \int_{B(0,R)} \left[\frac{2}{R}u + \lambda(|u-1| - 1)\right] + \lambda \int_{\mathbb{R}^2\setminus B(0,R)} |u| \\ &\geq \int_{B(0,R)} \underbrace{\left(\frac{2}{R} - \lambda\right)}_{\geq 0} u + \lambda \underbrace{((u-1) + |u-1|)}_{\geq 0} + \lambda \int_{\mathbb{R}^2\setminus B(0,R)} |u| \geq 0. \end{aligned}$$

  (here we have assumed that $u \geq 0$, which does not reduce generality since $E(\max(u,0)) \leq E(u)$ for all $u \in \text{BV}(\mathbb{R}^2)$). The equality case implies $u = \mathbb{1}_{B(0,R)}$ for $\lambda > \frac{2}{R}$, $u = 0$ for $\lambda < \frac{2}{R}$.

- For $\lambda = \frac{2}{R}$: we already have that 0 and $\mathbb{1}_{B(0,R)}$ are solutions, and by convexity so are $c\mathbb{1}_{B(0,R)}$ for $c \in (0,1)$. Now if u is a solution, the equality in the above inequality implies that $u_{|\mathbb{R}^2\setminus B(0,R)} = 0$ and $0 \leq u \leq 1$ in $B(0,R)$. Moreover we must have $\int |Du| = \int_{\overline{B(0,R)}} (Du, z)$. Since $|z(x)| < 1$ for $|x| < R$, $|Du|(B(0,R)) = 0$ so that u is constant on $B(0,R)$.

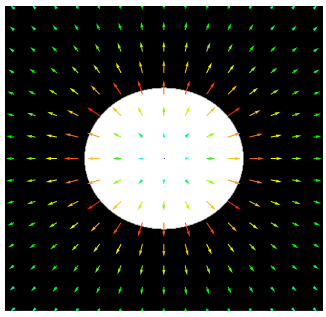


Figure 2.2 – *The "calibration" z constructed for the disc (Eq. (2.22)).*

This proof might look a bit tedious, because we have not use any thresholding argument to ensure that the solution is a characteristic function (although we could have done so). One should remember

that in [Chan and Esedoglu, 2005] and [Haddad, 2007], simplification arguments like radial symmetry or the search for a solution of the form $c\mathbb{1}_{B(0,R)}$) are invoked, which make their proof more elegant.

Yet, the main advantage of this proof is that it can be immediately generalized to calibrable sets (except that more solutions may exist for $\lambda = \frac{\text{Per } F}{|F|}$), and a slight adaptation allows to give a direct proof of Theorem 3.3.9 which states that the result of TVL1 is given by an opening: it consists in considering the vector field built in [Alter et al., 2005b] for the total variation flow, which satisfies:

$$|z(x)| \leq 1, \quad z \cdot \nu = 1 \text{ on } \partial C_r \tag{2.24}$$

$$\text{div } z(x) = \begin{cases} 0 & \text{for } x \in \mathbb{R}^2 \setminus C, \\ \frac{1}{r(x)} & \text{for } x \in C \setminus C_{1/\lambda^*}, \\ \lambda^* & \text{for } x \in C_{1/\lambda^*}. \end{cases} \tag{2.25}$$

where λ^* is defined in Proposition 3.3.7, C_r is the opening of C with radius r, ν is its outer unit normal, and $r(x) = \sup\{r > 0, x \in C_r\}$.

In our case, this method directly inspired from [Bellettini et al., 2001, Theorem 5.2] is also another way of writing the Euler Lagrange equation.

Chapter 3

The TVL1 model: a geometric point of view

The aim of this chapter is to study the TVL1 model in a continuous framework. Given $f \in L^1$, we are interested in finding u in $\mathrm{BV}(\mathbb{R}^2)$ which solves the following minimization problem:

$$\inf_{u \in \mathrm{BV}(\mathbb{R}^2)} \left\{ \int_{\mathbb{R}^2} |Du| + \lambda \int_{\mathbb{R}^2} |f-u| \right\} \tag{3.1}$$

The philosophy we have adopted in this chapter is to rely as much as possible on basic geometric arguments, to make the exposition as clear as possible (in fact, despite their simplicity, some of the results we use rely on the profound works [Bellettini et al., 2002, Bellettini et al., 2005b, Alter et al., 2005a]).

The organization of the chapter is the following. First, following [Chan and Esedoglu, 2005, Darbon, 2005], we show the connection between the TVL1 problem and a family of geometric problems which make the model contrast invariant. Then we study this geometric problem in detail: we relate the sudden disparition of sets with the Cheeger problem, and we show that the solutions of the geometric problem are smooth. In the case of convex sets, we give a simplified version of the energy, and we prove that the exact solution is given by an opening. The end of the chapter is devoted to the link between the TVL1 model and mathematical morphology.

3.1 The geometric equivalence theorem and its consequences

3.1.1 The geometric equivalence

Because of the coarea formula (1.2), many problems involving the total variation can be reformulated as a geometric problem on the level sets of the function. In [Chambolle, 2004a], Chambolle proposes an algorithm for the mean curvature flow using the Rudin-Osher-Fatemi problem associated with the signed distance to a set. He shows that this algorithm is equivalent at each step to the flat curvature flow algorithm by Almgren *et al.* [Almgren et al., 1993], by observing that the zero-level set of the solution of ROF is a minimizer of a geometric problem. In [Alter et al., 2005a], a similar observation is used to relate the solutions of the Rudin-Osher-Fatemi problem with the indicator function of a convex set C with the family of geometric problems:

$$\min_{X \subset C} \mathrm{Per}\, X - \lambda |X|. \tag{3.2}$$

Relying on this equivalence, the authors study the dependence of the solutions on the parameter λ to eventually characterize the calibrable sets of $\mathbb{R}^N$. We will use some of their results in Section 3.3.1. Let us also mention that the same principle is used in [Caselles et al., 2007a] to study the discontinuity set of the solutions of the Rudin-Osher-Fatemi problem, and that this method can be extended to variational

problems of the form:

$$\int |Du| + \int_{\mathbb{R}^2} G(x, u(x))dx, \tag{3.3}$$

where G is measurable in the first variable and C^1 convex in the second variable, as explained in [Chambolle et al., 2009].

In this section, after presenting a few basic facts about the TVL1 functional, we state a similar geometric equivalence for the TVL1 problem.

For $f \in L^1$, we want to solve:

$$\inf_{u \in \mathrm{BV}(\mathbb{R}^2)} \mathbb{E}(u) := \int |Du| + \lambda \int_{\mathbb{R}^2} |f(x) - u(x)|dx. \tag{3.4}$$

The following result is well-known: it is a standard application of the direct method of calculus of variations. We refer the reader to ([Yin et al., 2007b], [Chan and Esedoglu, 2005]) or to Chapter 6 for a proof in the more general case of adaptive TVL1.

Proposition 3.1.1. *There exists at least one solution to the Problem (3.4). Let us define:*

$$T : L^1 \to \mathcal{P}(L^1) \tag{3.5}$$

the operator which maps f to the set of solutions $Tf = \{u \in L^1, u \text{ is solution of } (3.4)\}$. The set Tf is a convex closed set in L^1.

Let us notice that, since the functional is not strictly convex, many solutions may exist (see Section 2.5 for the example of the disc). This is the reason why we need to define T as a set-valued map.

Inspired by the geometric equivalences used in [Chambolle, 2004a, Alter et al., 2005a], Chan and Esedoglu noticed in [Chan and Esedoglu, 2005] that, using the coarea formula, the energy $\mathbb{E}(u)$ could be reformulated as a sum of energies over the level sets of u:

$$\mathbb{E}(u) = \int_{-\infty}^{+\infty} \text{Per}\,\{x, u(x) > \mu\} + \lambda |\{x, u(x) > \mu\} \Delta \{x, f(x) > \mu\}| d\mu. \tag{3.6}$$

In view of this formula, one may consider the family of geometric problems, given a function $f \in L^1$:

$$\inf_{U \subset \mathbb{R}^2} \text{Per}\, U + \lambda |U \Delta F_\mu| \qquad (P_\mu) \tag{3.7}$$

where $\mu \in \mathbb{R}$ and $F_\mu = \{x, f(x) > \mu\}$.

Existence of solutions for each of these problems is standard, since it can be proved by the direct method of the calculus of variations. Here again, we can define a multivalued map $\mathcal{T}$ that maps F_μ to the set of solution sets $\mathcal{T}F_\mu$ for Problem (P_μ). Actually, in [Chan and Esedoglu, 2005], the authors show that if the data is the characteristic function of a set ($f = \mathbb{1}_F$), then every level set of a solution u of (3.4) is a solution of the geometric problem associated to F. In [Yin et al., 2007b], it was pointed out that stacking solutions of the geometric problem led to a solution of (3.4). The following theorem states the equivalence between the family of geometric problems on level sets and the functional TVL1 problem. Let us observe that this result is implicitly assumed in order to establish several properties in [Darbon, 2005, Yin et al., 2007b], such as the contrast invariance of the TVL1 model. A proof of this result in the discrete case is given in [Darbon and Sigelle, 2006]. We propose the following formulation and proof for the continuous case. It follows roughly the same scheme as the one in [Chambolle, 2004a], namely:

- Observe that the geometric problem is monotone, so that its solutions are nested.
- Build a function with the solutions of the geometric problem, and compare it with the minimizer of the functional problem.

Theorem 3.1.2 (Geometric equivalence)**.** *Let $f \in L^1$.The following assertions are equivalent:*

(i) u is solution of (3.4).

(ii) Almost every level set U_μ of u is a solution of (3.7).

We shall need the following monotonicity result. We state it in a similar way as Yin *et al.* [Yin et al., 2007b], but let us mention that it was also established in terms of Gibbs energy by Darbon and Sigelle [Darbon and Sigelle, 2006] in order to derive their graph-cut algorithm. For a proof of this result, we refer the reader to ([Yin et al., 2007b, Theorem 3.1], [Darbon and Sigelle, 2006, Lemma 1]). In Chapter 6, we will generalize this monotonicity result to the adaptive TVL1.

Theorem 3.1.3 (Pseudo-monotonicity of the geometric problem)**.** *Let $S_1 \subset S_2 \subset \mathbb{R}^2$, and $U_i, i = 1, 2$ be a solution the geometric problem associated to S_i. Then:*

- *$U_\wedge = U_1 \cap U_2$ is a solution associated to S_1,*
- *$U_\vee = U_1 \cup U_2$ is a solution associated to S_2.*

Remark 3.1.4. Notice that it is equivalent to consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ or to consider a family of sets $(U_\mu)_{\mu\in\mathbb{R}}$ that have the nesting property:

$$U_\mu \subset \bigcap_{\nu<\mu} U_\nu \tag{3.8}$$

This fact is at the core of the following proof.

Proof of Theorem 3.1.2. Since the implication $(ii) \Rightarrow (i)$ is clear by the coarea formula, we only focus on proving $(i) \Rightarrow (ii)$.

Let us endow the collection of Borel subsets of $\mathbb{R}^2$ with the topology induced by the pseudo-metric $d(A, B) = \min(1, |A\Delta B|)$. Notice that the mapping $t \mapsto F_t$ is non-increasing for the inclusion, therefore it is continuous everywhere but on a countable set of points (which is contained in the union of $\{0\}$ and the discontinuity sets of the real-valued monotone functions $t \mapsto |F_t|$ for $t > 0$ and $t \mapsto |F_t^C|$ for $t < 0$). We shall denote this set $\mathcal{J}$. Now, let us consider a countable dense set $D \subset \mathbb{R}$ (for instance $D = \mathbb{Q}$).

First step: We are going to construct by induction a family $(U_q)_{q\in D}$. To this end, let $(q_n)_{n\in\mathbb{N}}$ be an enumeration of D. We choose an optimal set U_{q_0} for problem (P_{q_0}). Let us assume that we have chosen U_{q_j} $(j = 0 \dots n)$ satisfying the nesting property, and let us consider q_{n+1}. Three cases are possible. If there exists $m, p \leq n$ such that $q_m < q_{n+1} < q_p$ (we can assume that q_m (resp. q_p) is the largest (resp. smallest) such element), then let us consider $U \subset \mathbb{R}^2$ solution of $(P_{q_{n+1}})$, and set:

$$U_{q_{n+1}} = (U \cap U_{q_m}) \cup U_{q_p} \tag{3.9}$$

By induction hypothesis $U_{q_p} \subset U_{q_{n+1}} \subset U_{q_m}$, and therefore the family $(U_{q_j})_{j=0\dots n+1})$ satisfies the nesting property. Theorem 3.1.3 ensures that $U_{q_{n+1}}$ is solution of $(P_{q_{n+1}})$.

The other two cases ($\forall j \leq n,\ q_j < q_{n+1}$ or $\forall j \leq n,\ q_j > q_{n+1}$) can be dealt with similarly (take $U \cap U_{q_m}$ or $U \cup U_{q_p}$).

Therefore we have built a countable family of optimal sets satisfying the nesting property.

Second step: Now let us define the family $(U_t^*)_{t\in\mathbb{R}}$ by:

$$U_t^* = \bigcap_{q<t, q\in D} U_q. \tag{3.10}$$

This family has the nesting property; let us show that it is a solution of problem (P_t) for almost every $t \in \mathbb{R}$.

Let $t \in \mathbb{R} \setminus \mathcal{J}$, and (t_n) an increasing sequence in D converging to t.

For all $V \subset \mathbb{R}^2$, Per $V + \lambda|V\Delta F_{t_n}| \geq$ Per $U_{t_n} + \lambda|U_{t_n}\Delta F_{t_n}|$. We take the lower-limit as $n \to +\infty$ and get:

$$\text{Per } V + \lambda|V\Delta F_t| \geq \text{Per } U_t^* + \lambda|U_t^*\Delta F_t| \tag{3.11}$$

where we used the continuity of $s \mapsto F_s$ at t, and the lower semicontinuity of the perimeter. Eventually, the elements of the family $(U_t^*)_{t\in\mathbb{R}}$ are solutions of (P_t), except maybe for $t \in \mathcal{J}$.

Conclusion: We can then build a function u^* by the formula:

$$u(x) = \sup\{t \in \mathbb{R}, x \in U_t^*\}. \tag{3.12}$$

Let then v be a solution of problem (3.4), and let us call V_t its level sets. By the coarea formula, it is clear that u is a solution too, thus:

$$\int_{-\infty}^{+\infty} \text{Per } V_t + \lambda|V_t\Delta F_t|dt = \int_{-\infty}^{+\infty} \text{Per } U_t^* + \lambda|U_t^*\Delta F_t|dt. \tag{3.13}$$

with the inequality for almost every t:

$$\text{Per } V_t + \lambda|V_t\Delta F_t| \geq \text{Per } U_t^* + \lambda|U_t^*\Delta F_t|. \tag{3.14}$$

We deduce that there is actually equality for almost every t. □

In the rest of this section, we will state a few properties of the TVL1 model that are desirable when processing images. Some of them (Propositions 3.1.7, 3.1.9, 3.1.10,3.1.11) are well-known ([Darbon, 2005], [Yin et al., 2007b]).

3.1.2 Maximum principle and monotonicity

Proposition 3.1.5 (Maximum principle)**.** *Let $M, m \in \mathbb{R}$ and assume $m \leq f \leq M$ almost everywhere. Then:*

$$\forall u \in Tf,\ m \leq u \leq M \text{ almost everywhere.} \tag{3.15}$$

Proof. Let us assume by contradiction that $u \in Tf$ with $u(x) > M$ on a set with non-zero Lebesgue measure. Then the truncated function $u_M = \min(u, M)$ has energy strictly lower than u. Indeed:

$$\int |Du_M| + \lambda \int_{\mathbb{R}^2} |f - u_M| = \int_{-\infty}^{M} \text{Per } \{x, u(x) > \mu\} + \lambda|\{x, u(x) > \mu\}\Delta\{x, f(x) > \mu\}|d\mu \tag{3.16}$$

whereas:

$$\begin{aligned}\int |Du| + \lambda \int_{\mathbb{R}^2} |f - u| = &\int_{-\infty}^{M} \text{Per } \{x, u(x) > \mu\} + \lambda|\{x, u(x) > \mu\}\Delta\{x, f(x) > \mu\}|d\mu \\ &+ \underbrace{\int_{M}^{+\infty} \text{Per } \{x, u(x) > \mu\}d\mu}_{\geq 0} + \underbrace{\int_{M}^{+\infty} \lambda|\{x, u(x) > \mu\}|d\mu}_{=\lambda\int_{\mathbb{R}^2}(u(x)-M)^+dx>0}\end{aligned} \tag{3.17}$$

which contradicts the fact that u is a minimizer. □

Because of the non-uniqueness of the solution, we have to precise the notion of monotonicity we will be using.

Definition 3.1.6. We say that an operator $\mathcal{A} : L^1 \mapsto \mathcal{P}(L^1)$ is pseudo-monotone if, given $f, g \in L^1$ with $f \leq g$ a.e., we have:

$$\forall u_1 \in \mathcal{A}f,\ \forall u_2 \in \mathcal{A}g,\ \max(u_1, u_2) \in \mathcal{A}g \text{ and } \min(u_1, u_2) \in \mathcal{A}f \tag{3.18}$$

A straightforward consequence of Theorems 3.1.2 and 3.1.3 is the following result:

Proposition 3.1.7 (Pseudo-monotonicity)**.** *T is pseudo-monotone.*

Yet, it can be handful to define the notion of "largest" and "smallest " solutions in order to take advantage of the classical notion of monotonicity.

Definition 3.1.8. Let us define the operators T^+ and $T^- : L^1 \to L^1$ by:

$$T^- f = \arg\min \left\{ \int_{\mathbb{R}^2} u,\ u \in Tf \right\} \tag{3.19}$$

$$T^+ f = \arg\max \left\{ \int_{\mathbb{R}^2} u,\ u \in Tf \right\} \tag{3.20}$$

Each operator defines a unique function, and is monotone. Moreover, this function is a solution of (3.4).

Proof that $T^{\pm} f$ are well defined. Let us denote by K the minimal value of the energy (3.4). We shall prove the result for T^- (the result for T^+ will follow since $T^+ f = -T^-(-f)$).

The infimum of $\{\int_{R^2} u,\ u \in Tf\}$ is finite since:

$$\forall u \in Tf,\ \int_{\mathbb{R}^2} (u - f) \geq -\left(\int |Du| + \lambda \int_{\mathbb{R}^2} |f(x) - u(x)| dx \right) = -K. \tag{3.21}$$

Now let u_n be a minimizing sequence. By Proposition 3.1.7, we can always assume,by replacing u_{n+1} by $\min(u_{n+1}, u_n)$, that the sequence $u_n(x)$ is non-increasing for almost every $x \in \mathbb{R}^2$.

Let $u^*(x)$ be the pointwise limit of $u_n(x)$. By Beppo-Levi's theorem (monotone convergence) we have that u_n converges to u^* in L^1, and we have:

$$\int_{\mathbb{R}^2} u^* = \inf \left\{ \int_{\mathbb{R}^2} u,\ u \in Tf \right\}. \tag{3.22}$$

Since the set of solutions of (3.4) is closed in L^1, u^* is a solution.

Now, we need to prove that u^* does not depend on the choice of the sequence u_n. Let v_n be another minimizing sequence, then the associated limit v^* is a solution of (3.4). By Proposition 3.1.7, $\min(u^*, v^*)$ is a solution too, and its integral is lower than those of u^* and v^*. This is only possible if $u(x) = v(x)$ almost everywhere.

Eventually, monotonicity is a consequence of Proposition 3.1.7. □

Let us notice that $T^- f$ actually defines the lowest solution, and $T^+ f$ the greatest. Problem (3.4) has a unique solution if and only if $T^- f = T^+ f$.

In a similar way, we can define the notion of largest and smallest solution to the geometric problem, and denote them by $\mathcal{T}^+, \mathcal{T}^-$.

3.1.3 Commutation with constants and affine invariance

In this subsection and only in this subsection, we shall consider data functions $f \in L^1 + C$, that is, sums of integrable functions and constants. Indeed, it is straightforward to extend the problem to such functions since the total variation is not affected by the addition of constants. The proofs of the next two propositions are elementary and omitted for brevity.

Proposition 3.1.9. *The operator T commutes with the addition of constants:*

$$\forall f \in L^1, \forall C \in \mathbb{R},\ T(f + C) = T(f) + C, \tag{3.23}$$

and is self-dual:

$$\forall f \in L^1,\ T(-f) = -T(f). \tag{3.24}$$

Proposition 3.1.10. *The operator T commutes with translations and rotations:*

$$\forall f \in L^1, \forall R \in SO_2,\ T(f \circ R) = T(f) \circ R \tag{3.25}$$

3.1.4 Contrast Invariance

As observed in [Darbon, 2005, Yin et al., 2007b], the following result is a direct consequence of the geometric equivalence Theorem 3.1.2. This is probably the most remarkable property of the TVL1 model.

Proposition 3.1.11 (Contrast invariance)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ a Lipschitz homomorphism. Then: $T(g \circ f) = g(Tf)$.*

Proof. It is a straightforward consequence of Theorem 3.1.2. g being a homomorphism, the level sets of $g \circ f$ (resp. $g \circ u$ where $u \in Tf$) are the same as the level sets of f (resp. u). Since g is Lipschitz, $g \circ u \in \mathrm{BV}(\mathbb{R}^2)$, and g has the Lusin (N) property, i.e. g maps every negligible set to a negligible set. Therefore since almost every level set of $u \in Tf$ is a solution of problem (3.7), almost every level set of $g \circ u$ is a solution of the geometric problem associated to $g \circ f$. □

Notice that one usually uses contrast invariance of a functional operator to derive the existence of a level set formulation [Guichard et al., 2004]. Interestingly enough, in the case of TVL1 we had to take the converse way.

3.2 The geometric problem

As it was shown in the last section, the study of the TVL1 problem amounts to solving the following geometric problem:

$$\inf_{U \subset \mathbb{R}^2} \mathbb{E}_G(U) := \mathrm{Per}\, U + \lambda |U \Delta F| \tag{3.26}$$

where $F \subset \mathbb{R}^2$ is a measurable set. In fact, F and U stand for a given level set of respectively the data and the solution. As it is equivalent to consider F and U on the one hand, or $\mathbb{R}^2 \setminus F$ and $\mathbb{R}^2 \setminus U$ on the other hand, we may always assume that F has finite Lebesgue measure. Indeed, we may take $F = F_t = \{x, f(x) \geq t\}$ for $t > 0$, and $F = \mathbb{R}^2 \setminus \{x, f(x) \geq t\}$ for $t < 0$.

This section is devoted to the study of problem (3.26). We begin with a series of simple calculations that show the link between the sudden disappearance of shapes and the Cheeger problem before giving more elaborate results about the regularity of the solutions.

3.2.1 Suddenly vanishing sets

An interesting property of the TVL1 model is the fact that some sets suddenly vanish. For instance, in Section 2.5, we have seen that the model preserves characteristic functions of discs with radius R if $\lambda > 2/R$; below this value, the solution is the null function. In this subsection we focus on sets which have the same property. Those sets are necessarily Cheeger in themselves.

Proposition 3.2.1. *Let $F \subset \mathbb{R}^2$ be a non empty set (not necessarily convex). Let us assume that F has the thresholding property, i.e. there is some critical value $\tilde{\lambda} > 0$ such that:*

- *F is a solution of the geometric problem (3.26) for $\lambda > \tilde{\lambda}$.*
- *$\emptyset$ is a solution of the geometric problem (3.26) for $0 < \lambda < \tilde{\lambda}$.*

Then F is Cheeger in itself, and $\tilde{\lambda} = \lambda_F := \frac{\mathrm{Per}\, F}{|F|}$.

Proof. Let us first notice that $\emptyset$ and F are solutions of the geometric problem for $\lambda = \tilde{\lambda}$ (consider a monotone sequence $(\lambda_n)_{n \in \mathbb{N}}$ converging to $\tilde{\lambda}$ and pass to the limit in the inequality). Therefore, both sets have the same energy: $\mathrm{Per}\, \emptyset + \tilde{\lambda}|F| = \mathrm{Per}\, F - 0$, and we deduce that $\tilde{\lambda} = \frac{\mathrm{Per}\, F}{|F|} = \lambda_F$.

Then, considering only sets U that are included in F, we notice that F is a non empty solution of the problem: $\inf_{U \subset F} \mathrm{Per}\, U - \lambda_F |U|$, which means that F is Cheeger in itself. □

The converse is true in the convex case (in which case Cheeger in itself is equivalent to calibrable), but in general, not every Cheeger set in itself has the thresholding property. Here is a sufficient condition though:

Proposition 3.2.2. *Let $F \subset \mathbb{R}^2$ be a calibrable set. Then F has the thresholding property. More precisely:*

- *F is the unique solution of the geometric problem (3.26) for $\lambda > \lambda_F$*
- *Both $\emptyset$ and F are solutions for $\lambda = \lambda_F$*
- *$\emptyset$ is the unique solution for $\lambda < \lambda_F$*

where $\lambda_F := \frac{\text{Per } F}{|F|}$.

Notice that, when $\lambda = \tilde{\lambda}$, other solutions than F and $\emptyset$ may exist. For instance, consider the union of two discs ($F = D_1 \cup D_2$) with same radius R. Provided they are far enough from each other, they behave independently (see Section 3.3.4) and for $\lambda = 2/R$ the solutions are $\emptyset$, D_1, D_2 and $D_1 \cup D_2$.

Proof. Let us recall that, for all V with finite perimeter, $\frac{\text{Per } F}{|F|} \leq \frac{\text{Per } V}{|V \cap F|}$.

- For $\lambda < \lambda_F$, we have:

$$\begin{aligned}\mathbb{E}_G(V) &= \text{Per } V - \lambda|V \cap F| + \lambda|V \cup F| \\ &= \underbrace{\text{Per } V - \lambda_F|V \cap F|}_{\geq 0} + \underbrace{(\lambda_F - \lambda)}_{>0}|V \cap F| + \lambda|V \cup F| \\ &\geq \lambda|F| = E(\emptyset)\end{aligned}$$

  and this inequality is strict if $|V \cap F| \neq 0$. But, if $|V \cap F| = 0$, then this quantity is strictly minimized when $|V| = 0$.

- For $\lambda = \lambda_F$, let us write:

$$\begin{aligned}\mathbb{E}_G(V) &= \underbrace{\text{Per } V - \frac{\text{Per } F}{|F|}|V \cap F|}_{\geq 0} + \frac{\text{Per } F}{|F|}|V \cup F| \\ &\geq \text{Per } F \frac{|F \cup V|}{|F|} \\ &\geq \text{Per } F = \mathbb{E}_G(F).\end{aligned}$$

  Therefore F is a solution.

  Moreover $\emptyset$ is a solution (passing to the limit with a sequence $\lambda_n \nearrow \lambda_F$).

- For $\lambda > \lambda_F = \frac{\text{Per } F}{|F|}$:

$$\begin{aligned}\mathbb{E}_G(V) &= \underbrace{\text{Per } V + \lambda_F|V \Delta F|}_{\geq \text{Per } F} + (\lambda - \lambda_F)|V \Delta F| \\ &\geq \text{Per } F + \underbrace{(\lambda - \lambda_F)}_{>0}|V \Delta F| \\ &\geq \mathbb{E}_G(F)\end{aligned}$$

  and this inequality is strict if $V \neq F$.

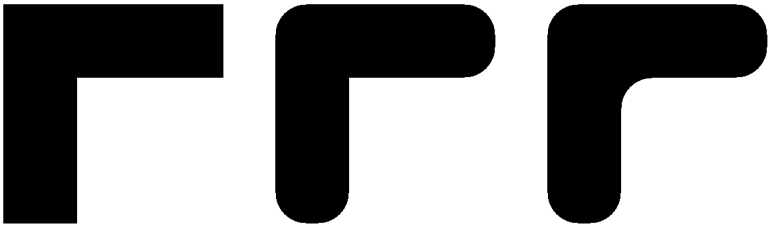


Figure 3.1 – *The property Cheeger in itself imposes regularity "from the inside" (an upper bound on the curvature) whereas the property of being invariant by TVL1 requires regularity from "both sides" (the curvature is lower and upper-bounded). Therefore some Cheeger sets do not have the suddenly vanishing property. From left to right: a set with corners, its Cheeger set, the solution of TVL1 with same curvature.*

□

As a conclusion, every calibrable set suddenly vanishes by TVL1, and every suddenly vanishing set is Cheeger in itself. **The converse implications are false.** In [Yin et al., 2007b], it is shown that an annulus with large enough inner radius suddenly vanishes. It is connected and not convex, therefore it is not calibrable. Moreover, since suddenly vanishing sets are solutions of TVL1, they are generally smoother than Cheeger sets in themselves (see Figure 3.1).

Example 3.2.3 (Stack of calibrable sets)**.** *The above sudden disappearance properties allow to exhibit simple examples of solutions of TVL1. For instance one may rederive the computations of [Haddad, 2007] when the input is a radial function $f(x) = \tilde{f}(r)$* ***with the additional assumption that f is non-increasing****. In that case the level sets are discs, which are known to be calibrable. The use of convex calibrable sets allows slight variations of this case as we illustrate here using ellipses.*

In [Kawohl and Lachand-Robert, 2006], Kawohl and Lachand-Robert prove that an ellipse with eccentricity lower than $\bar{e} \approx 0.7192$ is calibrable. Equivalently, the ratio $\alpha = \frac{a}{b}$ must satisfy $\alpha \leq \overline{\alpha} \approx 1.635$ (where a is the half large axis, b the half small one). Using Ramanujan's approximation for the perimeter $\text{Per}\, C \approx \pi \left(3(a+b) - \sqrt{(3a+b)(a+3b)}\right)$*, the condition of Proposition 3.2.2 amounts to:*

$$a \gtrsim \frac{1}{\lambda}\left(3(\alpha+1) - \sqrt{(3\alpha+1)(\alpha+3)}\right). \tag{3.27}$$

The ratio $\alpha = \frac{a}{b}$ being fixed in $[1, \overline{\alpha}]$, this means that the scale of the ellipse has to be large enough to be preserved by TVL1.

Now, considering the input function:

$$f : (x, y) \mapsto e^{-(\frac{x^2}{a^2} + \frac{y^2}{b^2})} \tag{3.28}$$

where a and b are chosen such that $\alpha = \frac{a}{b} \in [1, \overline{\alpha}]$, these geometric considerations allow to give an exact solution of TVL1. The level sets of f for $t \in (0, 1)$ are the following ellipses:

$$\{(x, y) \in \mathbb{R}^2, f(x, y) \geq t\} = \left\{(x, y) \in \mathbb{R}^2,\ \frac{x^2}{(-\log t)a^2} + \frac{y^2}{(-\log t)b^2} \leq 1\right\} \tag{3.29}$$

Level sets such that:

$$t \leq \exp\left(-\frac{1}{(\lambda a)^2}\left(3(\alpha+1) - \sqrt{(3\alpha+1)(\alpha+3)}\right)^2\right) \tag{3.30}$$

are preserved, the other vanish. The function f was thresholded. This example is illustrated in Figure 3.2.

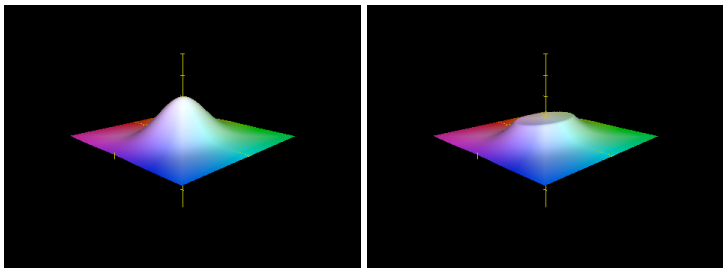


Figure 3.2 – *Graph of the analytic example 3.2.3 (left: original input, right: predicted solution). When the input is a stack of calibrable sets, TVL1 simply thresholds the function depending on the ratio perimeter/area of the level sets.*

3.2.2 Smoothness of the boundary

The study of the regularity of the solutions of problem (3.7) is made easier by the following remark:

Remark 3.2.4. As noticed by Alliney [Alliney, 1996], if u is a solution of the TVL1 problem associated to f, then u is a solution to the TVL1 problem associated to itself, since:

$$\int |Dv| + \lambda \int |u - v| \geq \int |Dv| + \lambda \int (|f - v| - |f - u|) \geq \int |Du|. \tag{3.31}$$

In other words, TVL1 is idempotent. As a consequence, it is sufficient to study the sets that are invariant by TVL1 in order to describe all the sets that are solutions of the geometric problem for some data.

Proposition 3.2.5. *Let $F \subset \mathbb{R}^2$, and $U \subset \mathbb{R}^2$ be a solution of the TVL1 problem associated to F. Then ∂U coincides with the reduced boundary $\partial^* U$ of U and it is a $C^{1,1}$ hypersurface.*

Proof. We rely on results from [Ambrosio, 1997, Ambrosio and Paolini, 1998] (see also [Andreu-Vaillo et al., 2002, Theorem 2.25]). Let us notice that U is a strong λ-minimizer ([Ambrosio, 1997], Definition 4.7.3) in $\mathbb{R}^2$, that is:

$$\text{Per}\,(U, \Omega) \leq \text{Per}\,(U', \Omega) + \lambda |U \Delta U'| \tag{3.32}$$

for every open set Ω and every U' with locally finite perimeter such that $U \Delta U' \subset\subset \Omega$.

Indeed, let Ω and U' be such sets: then $|D\mathbb{1}_U|(\mathbb{R}^2 \setminus \Omega) = |D\mathbb{1}_{U'}|(\mathbb{R}^2 \setminus \Omega)$. By the above remark, U is a solution of TVL1 associated to U. Hence (3.32).

The result stems from Theorems 4.7.1 and 4.7.4 in [Ambrosio, 1997]. □

The next proposition shows that at points where the boundary of the solution associated to F does not coincide with the boundary of F, the curvature is equal to $\pm\lambda$.

Proposition 3.2.6. *Let $F \subset \mathbb{R}^2$, $U \subset \mathbb{R}^2$ a solution of the geometric problem (3.26) associated to F, $x \in \partial U$, $\kappa(x)$ the curvature of ∂U at x, and V a neighborhood of x.*

- *If $V \cap F = \emptyset$ then $\kappa(x) = \lambda$*
- *If $V \cap F = V$ then $\kappa(x) = -\lambda$.*

Notice that, if no such V can be found, then ∂F and ∂U coincide at x. The idea is that, restricted to V, the geometric problem is equivalent to the famous mean curvature problem [Ambrosio, 1997].

Proof. Without loss of generality, and up to a rotation of the axes, we can assume that $V = D \times I$ where $D \subset \mathbb{R}$ and $I \subset \mathbb{R}$ are bounded open intervals and $V \cap U$ is the hypograph of a $C^{1,1}$ function $\phi : D \to I$.

The geometric energy classically reformulates in case (i):

$$\mathbb{E}_G(U) = C + \int_D \sqrt{1+|\nabla\phi(x)|^2}dx + \lambda\int_D \phi(x)dx \tag{3.33}$$

and in case (ii):

$$\mathbb{E}_G(U) = C + \int_D \sqrt{1+|\nabla\phi(x)|^2}dx + \lambda\int_D (\sup I - \phi(x))dx \tag{3.34}$$

where C is the energy outside V.

This is the prescribed mean curvature problem and the result follows by ([Ambrosio, 1997], Theorem 1.1.3). For the convenience of the reader, we recall the proof, working with case (i). We consider a small perturbation of ϕ: $(\phi + \epsilon\varphi)$ where $\varphi \in C_c^1(D)$. Then the energy becomes:

$$\int_D \sqrt{1+|\nabla\phi(x)+\epsilon\nabla\varphi|^2}dx + \lambda\int_D (\phi+\epsilon\varphi(x))dx \tag{3.35}$$

and since it is minimal in $\epsilon = 0$, its derivative vanishes:

$$\forall\varphi \in C_c^1(D), \quad \int_D \frac{\langle\nabla\phi,\nabla\varphi\rangle}{\sqrt{1+|\nabla\phi|^2}} + \lambda = 0, \tag{3.36}$$

which means:

$$\text{div}\left(\frac{\nabla\phi}{\sqrt{1+|\nabla\phi|^2}}\right) = \lambda. \tag{3.37}$$

The left term is exactly $\kappa(x)$. □

Remark Using similar techniques one can prove that in fact:

$$\operatorname*{ess\,sup}_{x\in\partial U} |\kappa(x)| \leq \lambda. \tag{3.38}$$

3.3 The geometric problem for a convex set

This section deals with the simplifications that appear in the geometric problem when the input is a convex set.

3.3.1 Reformulation of the energy for a convex set

To begin with, let us point out that if F is convex, the energy can be simplified.

Proposition 3.3.1. *Let us assume that $F \subset \mathbb{R}^2$ is convex. Then, the geometric problem (3.26) associated to F reformulates:*

$$\inf_{U\subset F} \text{Per}\, U - \lambda|U|. \tag{3.39}$$

Proof. The idea is to remark that every solution U of the geometric problem (3.26) is contained in F (modulo a Lebesgue negligible set).

Indeed, let $U \subset \mathbb{R}^2$. By contradiction, if $|U \setminus F| > 0$, let us prove that the geometric energy $\mathbb{E}_G(U \cap F)$ is strictly lower than $\mathbb{E}_G(U)$. Since F is convex[1]: Per $(U \cap F) \leq$ Per U (see [Ambrosio, 1994]). We add $\lambda|F \setminus U|$ on both sides, and we get:

[1] A similar argument ensures that even in the non convex case, solutions of TVL1 are contained in the closed convex hull of F.

$$\text{Per}\,(U\cap F)+\lambda|F\setminus U|<\text{Per}\,U+\lambda(|F\setminus U|+|U\setminus F|)\text{ since }|U\setminus F|>0. \tag{3.40}$$

Which precisely means $\mathbb{E}_G(U\cap F)<\mathbb{E}_G(U)$, and contradicts the fact that U is a solution. Thus, the geometric energy reformulates:

$$\begin{aligned}\text{Per}\,U+\lambda|U\Delta F| &= \text{Per}\,U+\lambda(|U|+|F|-2|U\cap F|\\ &= \text{Per}\,U-\lambda|U|+\lambda|F|.\end{aligned}$$

The last term is constant. It has no influence on the choice of the solution. □

Let us also recall a monotonicity result regarding the fidelity parameter λ. The following proposition can be found in ([Alter et al., 2005a], Lemma 4) or [Barozzi et al., 1987]:

Proposition 3.3.2 (Monotonicity with λ)**.** *Let $0<\lambda<\mu$ and U^λ, U^μ be solutions of Problem (3.39) with fidelity parameter λ (resp. μ) . Then $U^\lambda\subset U^\mu$.*

In the convex case, calibrable is equivalent to Cheeger in itself. Therefore we can state the following proposition (uniqueness is provided by Theorem 1.3.5).

Proposition 3.3.3 (Suddenly vanishing convex sets)**.** *Let $F\subset\mathbb{R}^2$ be a convex set. Then F has the thresholding property if and only if it is calibrable. Moreover, when $\lambda=\lambda_F$, $\{\emptyset,F\}$ is exactly the set of solutions.*

3.3.2 Convex sets that are invariant by TVL1

In view of Propositions 3.2.6 and 3.3.1, it is very easy to see that, given any convex set that is invariant by TVL1, both its curvature and its ratio perimeter/area are less than λ. A very interesting property is that the converse is true, as proved in [Alter et al., 2005a]. Indeed, Problem (3.39) was studied in [Alter et al., 2005a] in dimension N, in order to characterize convex calibrable sets. We shall rely on some of their results (notably the next theorem), but our working in dimension 2 makes some simplifications that allow to draw a parallel between the TVL1 model and mathematical morphology.

Theorem 3.3.4 ([Alter et al., 2005a], Theorem 9)**.** *Let C be a convex $C^{1,1}$ bounded set of $\mathbb{R}^2$ and, as before, $\lambda_C=\text{Per}\,C/|C|$. Let $\Lambda=\operatorname{ess\,sup}_{x\in\partial C}|\kappa(x)|$. Then C is a solution of the geometric problem (3.39) associated to C (with fidelity parameter λ) if and only if $\max(\lambda_C,\Lambda)\leq\lambda$.*

In view of Remark 3.2.4, let us notice that this theorem describes all convex solutions of the TVL1 problem: **a convex set C is solution of TVL1 for some F if and only if** $\max(\lambda_C,\Lambda)\leq\lambda$. Notice that this criterion combines both a local and a global term.

3.3.3 Characterization of solutions for a convex data

In this subsection, we explain how one can obtain the exact solution of the geometric problem using openings and a simple thresholding procedure. By the geometric equivalence Theorem 3.1.2, this gives the exact solution of the TVL1 problem in case the data f is convex or quasi-convex (i.e. its level sets are convex).

Definition 3.3.5. Let $C\subset\mathbb{R}^2$. Let us define the erosion and the opening with radius $r>0$ of C by:

$$C^r:=C\ominus B(0,r):=\{x\in C,\ B(x,r)\subset C\}, \tag{3.41}$$

$$C_r:=C^r\oplus B(0,r)=\bigcup_{B(x,r)\subset C}B(x,r). \tag{3.42}$$

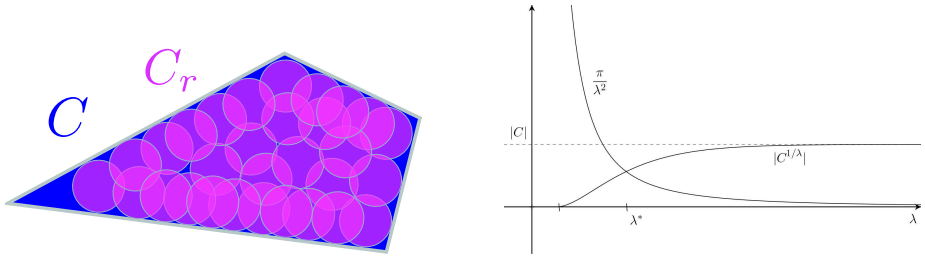


Figure 3.3 – *Left: the opening C_r of a set C with radius r is the union of all the balls of radius r that are contained in C. Right: General profile of the function $\lambda \mapsto |C^{1/\lambda}|$ for the erosion of a convex set C. The value λ^*, at the intersection with the graph of the function π/λ^2, is such that the opening C_{1/λ^*} of C is the unique Cheeger set of C.*

The following result can be found in [Andreu-Vaillo et al., 2002], and is a consequence of Lemma 9.2 in [Bellettini et al., 2001].

Lemma 3.3.6 ([Andreu-Vaillo et al., 2002] Lemma 4.27)**.** *Let $C \subset \mathbb{R}^2$ be an open convex bounded set. The following assertions are equivalent:*

- *There exists some $\rho > 0$ such that $C = C_\rho$.*
- *∂C is of class $C^{1,1}$, and $\sup_{p\in\partial C} \operatorname{ess}|\kappa_{\partial C}(p)| \leq \frac{1}{\rho}$.*

Next proposition describes the evolution of the ratio perimeter/area when performing openings.

Proposition 3.3.7 ([Alter et al., 2005b], [Kawohl and Lachand-Robert, 2006])**.** *The mapping $\lambda \mapsto |C^{\frac{1}{\lambda}}|$ is continuous, increasing on $\mathbb{R}_+^*$, vanishes in a neighborhood of 0 and tends to $|C|$ in $+\infty$. Moreover: $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} < \lambda$ if and only if $|C^{1/\lambda}| < \frac{\pi}{\lambda^2}$. Eventually, there exists a unique λ^* such that $\frac{\text{Per } C_{1/\lambda^*}}{|C_{1/\lambda^*}|} = \lambda^*$ (which is equivalent to $|C^{1/\lambda^*}| = \frac{\pi}{\lambda^{*2}}$).*

Notice that the value λ^* is precisely the one for which the opening C_{1/λ^*} is the Cheeger set of a convex body C (see Section 1.3).

Proof. For the continuity of the mapping, we refer the reader to [Alter et al., 2005b]. We only show here why the term $\frac{\pi}{\lambda^{*2}}$ appears. By Steiner's formulas:

$$\forall r > 0, \text{Per } (C^r \oplus B(0,r)) = \text{Per } C^r + 2\pi r \tag{3.43}$$

$$|C^r \oplus B(0,r)| = |C^r| + r\text{Per } C^r + \pi r^2. \tag{3.44}$$

Therefore:

$$\frac{\text{Per } (C^r \oplus B(0,r))}{|C^r \oplus B(0,r)|} > \frac{1}{r} \Leftrightarrow r\text{Per } C^r + 2\pi r^2 > |C^r| + r\text{Per } C^r + \pi r^2. \tag{3.45}$$

□

Lemma 3.3.8. *Let $C \subset \mathbb{R}^2$ be a convex set, and $\lambda > 0$. The largest convex $C^{1,1}$ open set included in C whose curvature is lower than λ is $C_{1/\lambda}$.*

Proof. • Let $\tilde{C}$ be an open $C^{1,1}$ convex set contained in C with curvature lower than λ. Then, by Lemma 3.3.6, $\tilde{C}_{1/\lambda} = \tilde{C}$.

But, by monotonicity of the opening: $\tilde{C}_{1/\lambda} \subset C_{1/\lambda}$. Thus: $\tilde{C} \subset C_{1/\lambda}$.

- It is clear that $C_{1/\lambda}$ is convex (since if C contains two balls it contains their convex hull) and it is open. By Lemma 3.3.6, $C_{1/\lambda}$ is $C^{1,1}$ and its curvature is lower than λ.

□

We are now in a position to state the main result of this section, giving the solutions of the binary TVL1 problem in the convex case by means of operators from mathematical morphology.

Theorem 3.3.9 (Solutions of the convex problem)**.** *Let C be a bounded, convex set, and λ^* defined by Proposition (3.3.7) (i.e. such that $|C^{1/\lambda^*}| = \frac{\pi}{\lambda^{*2}}$)*

- *For $\lambda > \lambda^*$, the set $C_{1/\lambda}$ is the unique solution of the geometric problem (3.39).*
- *For $\lambda = \lambda^*$, the set of solutions is exactly $\{C_{1/\lambda^*}, \emptyset\}$, and C_{1/λ^*} is the Cheeger set of C.*
- *For $\lambda < \lambda^*$ the unique solution of Problem (3.39) is $\emptyset$.*

Proof. **For** $\lambda > \lambda^*$, let us define $C^* = \mathcal{T}^+ C$ (see Definition 3.1.8). Then C^* is $C^{1,1}$ (Proposition 3.2.5) and its curvature is lower than λ (Proposition 3.3.4).

$C_{1/\lambda}$ is the largest convex $C^{1,1}$ open set contained in C with curvature lower than λ (see Lemma 3.3.8), so that $\text{int}\, C^* \subset C_{1/\lambda} \subset C$, and as a result:

$$\mathcal{T}^+(\text{ int } C^*) \subset \mathcal{T}^+ C_{1/\lambda} \subset \mathcal{T}^+ C^*. \tag{3.46}$$

But the extreme terms are equal to C^* (because we can identify $\text{int}\, C^*$ and C^*, since the $C^{1,1}$ boundary has null measure), therefore modulo a Lebesgue negligible set , $\mathcal{T}^+ C_{1/\lambda} = C^*$.

Eventually, $C_{1/\lambda}$ is invariant by TVL1 (since $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} \leq \lambda$ and $|\kappa| \leq \lambda$), hence:

$$C_{1/\lambda} = \mathcal{T}^+ C_{1/\lambda} = C^* \tag{3.47}$$

and $C_{1/\lambda}$ is solution of the geometric problem.

We still have to prove uniqueness. Let $U \subset C$ be a solution associated with the value λ. Let λ_n (resp. μ_n) be an increasing (resp. decreasing) sequence with limit λ. The sets C_{1/λ_n} are solutions of (3.39) associated to λ_n, thus, by monotonicity in λ (Proposition (3.3.2)):

$$\bigcup_n C_{1/\lambda_n} \subset U \subset \bigcap_n C_{1/\mu_n}. \tag{3.48}$$

Both extreme terms are equal to $C_{1/\lambda}$. Hence uniqueness.
For $\lambda = \lambda^*$, we compare the energy to the Cheeger problem. If $|U| > 0$:

$$\text{Per } U - \lambda^* |U| = |U| \left(\frac{\text{Per } U}{|U|} - \lambda^* \right) > 0 \tag{3.49}$$

except if $U = C_{1/\lambda^*}$, in which case the energy is equal to 0, as well as the energy of $\emptyset$.
For $\lambda < \lambda^*$:

$$\begin{aligned} \text{Per } U - \lambda |U| &= \text{Per } U - \lambda^* |U| + (\lambda^* - \lambda)|U| \\ &> 0 \text{ except if } U = \emptyset. \end{aligned}$$

□

From Theorem 3.3.9, we can deduce the solution of TVL1 on convex sets by a simple opening followed by a thresholding (see Corollary 3.4.3). At this point we can look back at the thresholding property of convex calibrable sets (Proposition 3.3.3). These sets suddenly disappear because when

$\lambda > \lambda^*$ they are invariant to an opening with radius $1/\lambda$ (since, by characterization of convex calibrable sets, $|\kappa(x)| \leq \frac{\text{Per } C}{|C|} = \lambda^*$).

In the general convex case, the last shape we see when decreasing λ is the Cheeger set of C. One should relate this with the Rudin-Osher-Fatemi model and the following result of [Alter et al., 2005a], [Caselles et al., 2008] which is the key to the study of Cheeger sets in dimension N. Given a data $f = \mathbb{1}_C$, denote by u the unique solution of this problem, then the set $\{x, u(x) \geq \|u\|_\infty\}$ is the Cheeger set of C. The similarity between the two problems stems from the fact that in the convex case, the ROF model behaves like a "stack of geometric problems": the level sets $(U_t)_{t\in[0,1]}$ minimize the quantity $\text{Per } U - \lambda(1-t)|U|$.

This explains why Theorem 3.3.4, which was proved in the study of Cheeger sets, turned out to be so helpful to the study of the TVL1 problem.

3.3.4 Separated convex components

We are now in a position to deal with the case of several convex connected components, provided they are far enough from one another.

Proposition 3.3.10. *Let $C_{(1)}, \dots, C_{(m)} \subset \mathbb{R}^2$, be non empty connected bounded sets such that $\bar{C}_{(i)} \cap \bar{C}_{(j)} = \emptyset$ for $i \neq j$, and $F = \bigcup_{i=1}^m C_{(i)}$. Assume the sets $C_{(i)}$ have the following property:*

$$(\mathcal{C}) \quad \forall U \supset F, \text{Per}\,(U) \geq \sum_{i=1}^m \text{Per}\,(C_{(i)}). \tag{3.50}$$

Let us define a partition of every set $U \subset \mathbb{R}^2$ by $U = \bigcup_{i=0}^m U_i$ with $U_0 \subset F^C$ and $U_i \subset C_{(i)}$. Then, the geometric problem (3.26) can be split into independent problems:

$$\forall i = 1, \dots, m, \ \min_{U_i} \text{Per}\, U_i + \lambda |C_{(i)} \setminus U_i| \tag{3.51}$$

Condition $(\mathcal{C})$ intuitively means that the components $C_{(i)}$ are far enough one from another. It also implies that the $C_{(i)}$'s are convex. We can therefore apply results from the last subsection to characterize the solution in this case.

Proof. Notice that by submodularity:

$$\text{Per}\,(U \cup \bigcup_{i=1}^m C_{(i)}) + \text{Per}\,(U \cap \bigcup_{i=1}^m C_{(i)}) \leq \text{Per}\, U + \underbrace{\text{Per}\,(\bigcup_{i=1}^m C_{(i)})}_{=\sum_{i=1}^m \text{Per}\, C_{(i)}} \tag{3.52}$$

(since the distance between the C_i is strictly positive). But $\text{Per}\,(U \cup \bigcup_{i=1}^m C_{(i)}) \geq \sum_{i=1}^m \text{Per}\, C_{(i)}$ by condition (C). Thus, the inequality can hold only if: $\text{Per}\,(U \cap \bigcup_{i=1}^m C_{(i)}) \leq \text{Per}\, U$. Moreover:

$$|(U \cap \bigcup_{i=1}^m C_{(i)}) \Delta F| = \sum_{i=1}^m |C_{(i)} \setminus U_i| \leq |U_0| + \sum_{i=1}^m |C_{(i)} \setminus U_i| \leq |U \Delta F|. \tag{3.53}$$

We can infer that $\mathbb{E}_G(U \cap \bigcup_{i=1}^m C_{(i)}) \leq \mathbb{E}_G(U)$. The geometric problem (3.26) amounts then to: $\min \sum_{i=1}^m (\text{Per}\, U_i + \lambda |C_{(i)} \setminus U_i|)$. These problems are independent because $U_i \subset C_{(i)}$ and the $C_{(i)}$'s are distant from one another. □

Notice that Condition (C), which is here only a sufficient condition, is very similar to conditions that appear naturally when studying calibrable sets (see [Bellettini et al., 2002], Theorem 6).

Remark 3.3.11. Although Condition (C) may not be optimal, notice that a condition on the distance between connected components appears necessary. This is easy to check in the particular case of two discs of radius r. Assume the distance between their centers is $d \geq 2r$, then their convex hull has lower energy than the two discs when $2r \leq d \leq \frac{\pi r}{2}\left(1+\frac{1}{1+\lambda r}\right)$. One can choose such d as soon as $\lambda < \frac{1}{r}\left(\frac{2\pi-4}{4-\pi}\right) \approx \frac{2.6}{r}$. For $\lambda \in \left(\frac{2}{r}, \frac{1}{r}\left(\frac{2\pi-4}{4-\pi}\right)\right)$, the independent evolution predicts that the solution is two discs, which is contradicted by the fact that the convex hull has a lower energy.

3.4 Mathematical morphology almost everywhere...

As noticed in [Darbon, 2005] and discussed in Section 3.1, the minimization of the TVL1 model yields a morphological operator. In this Section we analyze further the link between the model and the openings introduced in Proposition 3.3.9.

In the convex case, we have seen that TVL1 is equivalent to an opening followed by a thresholding. When the connected components are non convex or too close to one another, things get more complicated and one may wonder if a simple morphological operator can describe or approximate the TVL1 model. K. R. Vixie and S. Esedoglu prove in [Vixie and Esedoglu, 2007] the following result by geometric arguments:

Proposition 3.4.1 ([Vixie and Esedoglu, 2007])**.** *Let $F \subset \mathbb{R}^2$ be a bounded measurable set, and U be a solution of the corresponding geometric problem (3.26). Assume that a ball $B(x, \frac{2}{\lambda})$ lies completely in F. Then, $U \cup B(x, \frac{2}{\lambda})$ is a solution of the geometric problem (3.26). Moreover, if $B(x', \frac{2}{\lambda}) \subset F^c$, then $(B(x', \frac{2}{\lambda}) \cup U^c)^c$ is a solution too.*

This result is all the more interesting as it gives the following geometrical bounds on some solution of TVL1:

$$\bigcup_{x:B(x,\frac{2}{\lambda})\subset F} B(x, \frac{2}{\lambda}) \subset U \subset \left(\bigcup_{x:B(x,\frac{2}{\lambda})\subset F^C} B(x, \frac{2}{\lambda}) \right)^C . \tag{3.54}$$

Writing $\mathcal{T}$ for the TVL1 operator, $\mathcal{O}_{2/\lambda}$ and $\mathcal{F}_{2/\lambda}$ respectively for the opening and closing operators with radius $\frac{2}{\lambda}$, one can formally write:

$$\mathcal{O}_{2/\lambda} \subset \mathcal{T} \subset \mathcal{F}_{2/\lambda}. \tag{3.55}$$

This result is somehow optimal since a ball with radius $r < \frac{2}{\lambda}$ vanishes by TVL1. Therefore, it is hopeless to try to get a finer inequality by opening with balls with a smaller radius. Nevertheless, the result is coarse in the sense that it implies a curvature of $\pm\frac{\lambda}{2}$ in regions where the boundary of U is distinct from the boundary of F, and not $\pm\lambda$ as predicted by Proposition 3.2.6.

In fact, we can generalize Proposition 3.4.1 and give a simple proof of it by noticing that it is a consequence of the monotonicity Theorem 3.1.3 (since balls of radius $\frac{2}{\lambda}$ are invariant by TVL1). In the framework of the present paper, one sees that **Proposition 3.4.1 is true when replacing $B(x, \frac{2}{\lambda})$ by any set that is invariant by TVL1.** As a consequence, one can replace balls in Equation 3.55 by any set that is invariant by TVL1. In order to have better approximations of the TVL1 solutions, we can therefore consider a larger family of structuring elements than just discs.

For instance one may consider ellipses, or any convex calibrable set satisfying the hypothesis of Theorem 3.2.2, and try to refine the previous inequality. In theory, we are guaranteed that, taking a large enough family, one could describe exactly operators $\mathcal{T}^+$ or $\mathcal{T}^-$: this is indeed a consequence of a famous theorem by Matheron [Matheron, 1975] that we now recall.

Theorem 3.4.2 (Matheron). *Let $\mathcal{T} : \mathcal{P}(\mathbb{R}^2) \to \mathcal{P}(\mathbb{R}^2)$ be a translation invariant standard monotone operator. Let us consider $\mathcal{B} = \{B \subset \mathbb{R}^2 / 0 \in \mathcal{T}B\}$. Then:*

$$\mathcal{T}X = \{x \in \mathbb{R}^2, x + B \subset X \text{ for some } B \in \mathcal{B}\}. \tag{3.56}$$

Unfortunately, the family $\mathcal{B}$ of structuring elements given by this theorem is far too large for a practical use. Indeed any solution of TVL1, up to a translation, is a structuring element! One may thus wonder if a reasonable family could lead to useful approximations of TVL1, yielding curvature values close to λ. However, the choice of a simple approximating family of structuring elements appears to depend on the considered set.

On the other hand, in the convex case, an immediate corollary to Theorem 3.3.9 yields an alternative expression of TVL1 using operators from mathematical morphology. Indeed,

Corollary 3.4.3. *Let C be a bounded, convex set, and $\lambda > 0$. Let $C_{1/\lambda}$ be the opening of C with radius $\frac{1}{\lambda}$. Then:*

- *If $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} < \lambda$, the solution is $C_{1/\lambda}$.*
- *If $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} > \lambda$, the solution is $\emptyset$.*
- *If $\frac{\text{Per } C_{1/\lambda}}{|C_{1/\lambda}|} = \lambda$, both $C_{1/\lambda}$ and $\emptyset$ are solutions.*

In short, TVL1 amounts to a thresholding following an opening. The condition over the ratio perimeter/area appearing in this result essentially means that the energy of the set after the opening is less than the energy of the empty set. Let us mention that a similar thresholding operation has already been proposed in a completely different context, in the framework of morphological connected filters, see [Salembier et al., 1998].

In the non-convex case, Corollary 3.4.3 is non longer true, but one may wonder how to extend the corresponding morphological operator. In Chapter 4, we push the discussion further relying on empirical considerations: a numerical scheme is proposed to approximate TVL1, and we illustrate it by various experiments on both synthetic and natural images.

Before proceeding to the next chapter, let us give a brief motivation for this algorithm. Of course, the approach that would apply openings to non-convex sets is not satisfactory, because it would obviously induce errors in locally concave parts. In order to overcome this difficulty, we consider the operators given by an opening followed by a closing or vice versa ($\mathcal{F}_{1/\lambda}\mathcal{O}_{1/\lambda}$ or $\mathcal{O}_{1/\lambda}\mathcal{F}_{1/\lambda}$) with the idea that in regions where the shape is locally convex or concave, Theorem 3.3.9 applies and these operators would behave correspondingly. Notice also that both operators provide curvature $\pm\lambda$ and also satisfy the same inequality as TVL1, Formula (3.55), namely:

$$\mathcal{O}_{2/\lambda} \subset \mathcal{O}_{1/\lambda}\mathcal{F}_{1/\lambda} \subset \mathcal{F}_{2/\lambda} \text{ and } \mathcal{O}_{2/\lambda} \subset \mathcal{F}_{1/\lambda}\mathcal{O}_{1/\lambda} \subset \mathcal{F}_{2/\lambda}. \tag{3.57}$$

After applying operator $\mathcal{F}_{1/\lambda}\mathcal{O}_{1/\lambda}$ (or $\mathcal{O}_{1/\lambda}\mathcal{F}_{1/\lambda}$), we consider every connected component U_i of the result, and we test if its energy is lower than the energy of the empty set i.e.:

$$\text{Per } U_i + \lambda \left(|U_i| - 2|U_i \cap F|\right) \leq 0. \tag{3.58}$$

We remove the connected components accordingly.

One may object that the above procedure is not self-dual, contrary to TVL1. It is true, but the two points should be noted:

- When the solution of TVL1 is not unique, a choice must be made which is not necessarily self-dual. For instance, one may consider $\mathcal{T}^+$ or $\mathcal{T}^-$ which are not self-dual.

- In mathematical morphology, the lack of self-duality of the opening followed by a closing is often resolved by using an alternating sequential filter [Soille, 2003]. This filter consists, given a small value r_0, in performing alternatively opening and closing of radius nr_0 until $nr_0 \geq 1/\lambda$. When r_0 is small enough, the operator is much more symmetric than a single opening-closing step, and starting with either an opening or a closing makes little difference. Empirically we have found that we obtain results that are closer to the exact solution. Therefore we use this procedure in the following.

Chapter 4

Algorithms and numerical experiments

In this chapter, we support the previous theoretical study with numerical experiments. First, we recall several classical ways to solve the TVL1 problem. Then, we propose two algorithms based on morphological openings to approximate these solutions. Both are isotropic, and the second one is especially fast. Let us emphasize that we do not claim to perform an exact minimization of the TVL1 functional in the discrete case, as is done in [Darbon and Sigelle, 2006]. Our purpose here is to illustrate the fact that the TVL1 minimization behaves (exactly in many cases, and at least not very differently in general) like simple morphological operators such as openings or closings.

4.1 Classical ways to numerically solve the TVL1 problem

4.1.1 Gradient descent

In [Chan and Esedoglu, 2005], the method employed to numerically solve the TVL1 model is the most straightforward. The authors employ the following gradient descent scheme:

$$\frac{u_{i,j}^{n+1}-u_{i,j}^{n}}{\delta t} = D_x^- \left(\frac{D_x^+ u_{i,j}^n}{\sqrt{\epsilon + (D_x^+ u_{i,j}^n)^2 + (D_y^+ u_{i,j}^n)^2}} \right) + D_x^- \left(\frac{D_y^+ u_{i,j}^n}{\sqrt{\epsilon + (D_x^+ u_{i,j}^n)^2 + (D_y^+ u_{i,j}^n)^2}} \right) + \lambda \frac{u_{i,j}^n - f_{i,j}}{\sqrt{\delta + (u_{i,j}^n - f_{i,j})^2}},$$

where $\epsilon > 0$ and $\delta > 0$ are small constants. The operators D^+ and D^- refer to the forward and backward differences.

Despite its simplicity, this scheme has many drawbacks. First, since the total variation and the L^1 norm are not smooth, one has to regularize them, hence the constants ϵ and δ. Second, the convergence of this scheme is very slow, especially with binary images. Indeed, in flat regions the gradient is zero. As a consequence, for pixels in a flat region to evolve, one has to wait for the propagation of the evolution of the edges. Last, the underlying energy does not satisfy the coarea formula, so that new gray levels are created, even at convergence. As a result, the solutions obtained when the input images are binary input look blurred.

4.1.2 Alternate minimization schemes

4.1.2.1 Aujol *et al.*

In the case of the Rudin-Osher-Fatemi problem, Chambolle has proposed a duality based algorithm [Chambolle, 2004b] that has widely and quickly been adopted by the image processing community.

Its main advantages are that it is simple to implement, its convergence is theoretically proved, and it converges faster than the gradient descent.

In [Aujol et al., 2006], Aujol *et al.* propose an algorithm to solve the TVL1 problem that relies on alternating Chambolle's algorithm and a shrinkage. It consists in solving alternatively the two following problems:

- v being fixed,
$$\inf_u \left(J(u) + \frac{1}{2\alpha}\|f-u-v\|_2^2 \right), \tag{4.1}$$
- u being fixed:
$$\inf_v \frac{1}{2\alpha}\|f-u-v\|_2^2 + \lambda\|v\|_1. \tag{4.2}$$

Here, J and $\|\cdot\|_1$ refer respectively to the discrete total variation and discrete L^1 norm:

$$J(u) = \sum_{1\leq i,j\leq N} \sqrt{(\nabla_x u)_{i,j}^2 + (\nabla_y u)_{i,j}^2}, \qquad \|u\|_1 = \sum_{1\leq i,j\leq N} |u_{i,j}|$$

The discrete gradient is given by:

$$(\nabla_x u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } 1 \leq i \leq N-1 \\ 0 \text{ if } i = N \end{cases}, \quad (\nabla_y u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } 1 \leq j \leq N-1 \\ 0 \text{ if } j = N \end{cases}. \tag{4.3}$$

The solution of the first problem is obtained using Chambolle's algorithm, whereas the solution of the second one is given by a shrinkage of the pixel values. The authors show that the solutions (u_n, v_n) converge to the solutions of the problem:

$$\inf_{u,v} \left(J(u) + \frac{1}{2\alpha}\|f-u-v\|_2^2 + \lambda\|v\|_1 \right). \tag{4.4}$$

Now, for $\alpha \to 0+$, they state that $u^{(\alpha)}$ converges to a solution of the TVL1 problem:

$$\inf_u \left(J(u) + \lambda\|u\|_1 \right). \tag{4.5}$$

This scheme is a bit difficult to use since in practice one has to choose a value α, and the obtained solution is actually a solution of (4.4). Nevertheless, Pock *et al.* [Pock et al., 2008] have proposed a GPU-based implementation of this algorithm which is faster than graph-cut techniques (which, as far as we know, are difficult to adapt to GPU programming).

4.1.2.2 Chambolle-Pock

Very recently, Chambolle and Pock [Chambolle and Pock, 2011] proposed an algorithm which, as far as we know, achieves the current best convergence speed for problem (4.5).

To solve a problem of the form:

$$\min_{u\in X} F(Au) + G(u), \tag{4.6}$$

where X, Y are a Euclidean spaces, $F : Y \to [0, +\infty)$ and $G : X \to [0, +\infty)$ are proper, convex, l.s.c. functions, and $A : X \to Y$ is a linear map, the authors write it as a primal-dual problem:

$$\min_{u\in X} \max_{y\in Y} \langle y, Au\rangle - F^*(y) + G(u). \tag{4.7}$$

$$\tag{4.8}$$

They propose to solve it by alternate minimizations:

$$\left\{\begin{array}{lcl} y^{n+1} & = & (I+\sigma\partial F^*)^{-1}(y^n+\sigma A\overline{u}^n) \\ u^{n+1} & = & (I+\tau\partial G)^{-1}(u^n-\tau A^* y^{n+1}) \\ \overline{u}^{n+1} & = & 2u^{n+1}-u^n. \end{array}\right. \tag{4.9}$$

This is an alternate minimization scheme since $(I+\tau\partial F)^{-1}(y)=\arg\min_x\left(\frac{1}{2}\|x-y\|^2+F(x)\right)$.

In the particular case of the TVL1 algorithm, we have $X=\mathbb{R}^{N\times N}$, $Y=X\times X$, $A=\nabla$, $F=\|\cdot\|_1$, and $G=\|\cdot-f\|_1$. The scheme amounts to:

$$\left\{\begin{array}{lcl} y^{n+1} & = & \frac{y^n+\sigma\nabla\overline{u}^n}{\max(1,|y^n+\sigma\nabla\overline{u}^n|)} \\ u^{n+1} & = & f+\mathrm{ST}_{\tau\lambda}(u^n+\tau\times(\mathrm{div}\,(y^{n+1}))-f) \\ \overline{u}^{n+1} & = & 2u^{n+1}-u^n, \end{array}\right. \tag{4.10}$$

where $\mathrm{ST}_{\lambda\tau}$ is the soft-thresholding operator: $\mathrm{ST}_{\lambda\tau}(x)=\left\{\begin{array}{ll} x-\lambda\tau & \text{if } x>\lambda\tau \\ x+\lambda\tau & \text{if } x<\lambda\tau \\ 0 & \text{otherwise} \end{array}\right.$.

The advantages of this scheme are that there is a proof of convergence, and that it is very fast (see [Chambolle and Pock, 2011] for a comparison).

4.1.3 Second-Order Cone programming

In [Yin et al., 2007a], Yin *et al.* use a second order cone programming method to minimize the TVL1 energy. They write the problem as:

$$\begin{aligned} \min_{s,t,u,\nabla_x u,\nabla_y u} & \sum_{1\le i,j\le N} t_{i,j}+\lambda s \\ \text{s.t. } & (\nabla_x u)_{i,j}=u_{i+1,j}-u_{i,j} \quad \forall 1\le i,j\le N, \\ & (\nabla_y u)_{i,j}=u_{i,j+1}-u_{i,j} \quad \forall 1\le i,j\le N, \\ & \sum_{1\le i,j\le N}(f_{i,j}-u_{i,j})\le s, \\ & \sum_{1\le i,j\le N}(u_{i,j}-f_{i,j})\le s, \\ & (t_{i,j},(\nabla_x u)_{i,j},(\nabla_y u)_{i,j})\in\mathcal{K}, \quad \forall 1\le i,j\le N, \end{aligned}$$

where $\mathcal{K}=\{(a,b,c)\in\mathbb{R}^3,\ a\ge\sqrt{b^2+c^2}\}$.

This formulation allows the authors to use a commercial solver to obtain the solutions of TVL1. We will not focus on this approach in this thesis.

4.1.4 The Darbon-Sigelle algorithm

The computer vision community has been using graph-cuts to solve minimization problems for decades [Greig et al., 1989, Hochbaum, 2001, Boykov and Kolmogorov, 2004, Chambolle, 2005]. In [Darbon and Sigelle, 2006], Darbon and Sigelle have proposed an very efficient way to solve the TVL1 problem[1]. The energy they minimize is of the form:

$$\mathbb{E}(u):=\left(\sum_{k=0}^{N-1}\sum_{l=0}^{N}|u_{k+1,l}-u_{k,l}|+\sum_{k=0}^{N}\sum_{l=0}^{N-1}|u_{k,l+1}-u_{k,l}|\right)+\sum_{1\le k,l\le N}|u_{k,l}-f_{k,l}|. \tag{4.11}$$

[1] Let us also mention the similar algorithm proposed simultaneously in [Chambolle, 2005] for ROF.

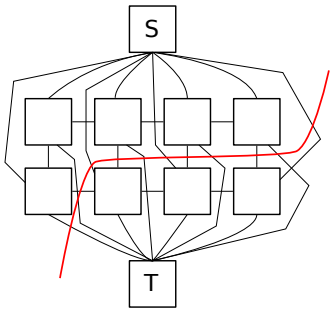


Figure 4.1 – *The graph associated to the image. The nodes S and T are special. Then encode the fidelity term.*

We mention here the 4-connected total variation scheme but other submodular regularization terms are also considered, including a more anisotropic 8-connected scheme. The choice of the authors is justified by the fact that such discrete energies satisfy a coarea formula: let $U_j = \{(k,l) \in [\![0,N]\!]^2,\ u_{k,l} \geq j\}$ and $F_j = \{(k,l) \in [\![0,N]\!]^2,\ f_{k,l} \geq j\}$ be the upper level sets of u,

$$\mathbb{E}(u) = \sum_{j=0}^{254} \left(\text{Per}_1\ U_j + \lambda \sum_{(k,l)\in[\![0,N]\!]^2} \left(\mathbb{1}_{U_j}[k,l] + \mathbb{1}_{F_j}[k,l] - 2 \cdot \mathbb{1}_{U_j}[k,l]\mathbb{1}_{F_j}[k,l] \right) \right), \tag{4.12}$$

where Per_1 denotes the anisotropic perimeter:

$$\begin{aligned} \text{Per}_1\ U_j = &\sum_{k=0}^{N-1}\sum_{l=0}^{N} \left(\mathbb{1}_{U_j}[k,l] + \mathbb{1}_{U_j}[k+1,l] - 2 \cdot \mathbb{1}_{U_j}[k,l]\mathbb{1}_{U_j}[k+1,l] \right) \\ &+ \sum_{k=0}^{N}\sum_{l=0}^{N-1} \left(\mathbb{1}_{U_j}[k,l] + \mathbb{1}_{U_j}[k,l+1] - 2 \cdot \mathbb{1}_{U_j}[k,l]\mathbb{1}_{U_j}[k,l+1] \right). \end{aligned}$$

Then each geometric problem can be seen as a cut problem in a graph, and therefore can be solved using a graph-cut approach [Ford and Fulkerson, 1956]. Let us consider a directed graph where the nodes are the pixels of the image plus two additional nodes called the source (s) and the sink (t). The pixels are related by the 4-connectivity of the image, and each pixel is connected to the source and the sink. If a, b are two nodes of the graph corresponding to pixels, we set $C(a,b) = 1$. For any pixel a of the image, we set $C_{s\to a} = 0$ and $C_{a\to t} = \lambda$ if $a \in F_j$, and $C_{s\to a} = \lambda$ and $C_{a\to t} = 0$ if $a \notin F_j$

A cut is a partition $(\mathcal{V}_s, \mathcal{V}_t)$ of the set of nodes such that $s \in \mathcal{V}_s$, $t \in \mathcal{V}_t$. Its cost is defined by:

$$E(\mathcal{V}_s, \mathcal{V}_t) = \sum_{a\in\mathcal{V}_s, b\in\mathcal{V}_t} C_{a\to b}. \tag{4.13}$$

Then one can see that $E(\mathcal{V}_s, \mathcal{V}_t)$ is exactly the geometric energy of $\mathcal{V}_t \setminus \{t\}$ in Eq.(4.12). As a consequence any minimizer of the minimum cut problem is a minimizer of the geometric problem (4.12) and conversely.

The theorem by Ford and Fulkerson [Ford and Fulkerson, 1956] ensures that a minimum cut problem can be solved by computing the maximal flow σ on a graph associated to the image, which can be done very efficiently. We refer the reader to [Boykov and Kolmogorov, 2004] for more details about such algorithms, but it is interesting to look at the properties of the constructed flow. In [Strang, 1983, Strang, 2009], Strang studies the continuous analog of discrete minimum cut problems. Depending on the source terms on the boundary or inside the domain, he finds geometric problems involving the perimeter, including the Cheeger problem. Since then, many authors have focussed on the continuous maximum flow problem [Chambolle and Darbon, 2009, Yuan et al., 2010]. It is interesting to use this

framework to draw a parallel between the constructed maximal flow and the vectorfield z of the Euler-Lagrange equation (2.2) for the geometric problem, i.e..

$$\exists z \in X_\infty, \quad |z| \leq 1 \text{ a.e.}, \quad (D\mathbb{1}_U, z) = |D\mathbb{1}_U|, \quad |D\mathbb{1}_U| \text{ a.e.} \tag{4.14}$$

$$\exists p \in L^\infty, \quad |p| \leq 1 \text{ a.e.}, \quad |U \Delta F| = \int_{\mathbb{R}^2} (\mathbb{1}_U - \mathbb{1}_F) p, \tag{4.15}$$

$$\text{and} \quad \operatorname{div} z + \lambda p = 0. \tag{4.16}$$

The first equation means that $z(x) \cdot \nu = 1$, $\mathcal{H}^1$ a.e. on ∂U (where ν is the outer unit normal to ∂U), and the second one that $p = \operatorname{sign}(\mathbb{1}_U - \mathbb{1}_F)$. The following correspondence between the maximum flow problem and the optimality conditions holds:

	Maximal flow	Euler-Lagrange
Capacity constraint	$0 \leq \sigma_{a\to b} \leq 1 = C_{a\to b}$	$\lvert z(x)\rvert \leq 1$
Saturation of the capacity of an edge of the cut	$\sigma_{a\to b} = 1$ if $a \in \mathcal{V}_s$ and $b \in \mathcal{V}_t$	$z(x) \cdot \nu = 1$ on ∂U
Equilibrium (Kirchhoff)	$\sum_{b\in\mathcal{N}(B)}(\sigma_{a\to b} - \sigma_{b\to a}) - \sigma_{s\to a} + \sigma_{a\to t} = 0$ $0 \leq \sigma_{s\to i} \leq C_{s\to a} = \lambda \mathbb{1}_{F_j^C}$ $0 \leq \sigma_{i\to t} \leq C_{a\to t} = \lambda \mathbb{1}_{F_j}$	$\operatorname{div} z + \lambda \operatorname{sign}(\mathbb{1}_U - \mathbb{1}_C) = 0$

4.2 Opening-based algorithms

We now propose two algorithms based on openings to compute approximate solutions of TVL1. The first one is naive and works on each level set after applying an alternate sequential filter. The second one takes advantage of the tree of shapes [Monasse and Guichard, 2000] to threshold the level sets in a more elegant manner.

4.2.1 Naive algorithm

Following the discussion of Section 3.4, the simplest algorithm we propose consists in performing an opening then a closing($\mathcal{FO}$), or vice versa ($\mathcal{OF}$), or an alternate sequential filter (ASF). We then threshold the connected components of the level sets. Since in the non convex case the thresholding driven by Equation (3.58) might not be monotone with respect to the inclusion, in order to avoid visible artifacts when applying such a scheme to images, we impose the monotonicity, by removing all connected components of level sets that are included in a removed component. This algorithm is described in Figure 4.2.

Observe that the thresholding procedure is self-dual and that the only possible cause of non self-duality stems from the morphological filter ($\mathcal{OF}$, $\mathcal{FO}$ or ASF). This procedure could lead to contrast inversion in case of connected components with holes. However, we never observed such a behavior in numerical experiments. Actually, one may show that this phenomenon never happens in the case of concentric circles (annulus) because of the preliminary morphological filter.

4.2.2 FLST-based algorithm

Connected components of upper level sets have a tree structure, as well as connected components of lower level sets. Although the previous algorithm can be implemented by imposing monotonicity, it does not fully benefit from this tree structure, and a more elegant approach can be proposed using the Fast Level Set Transform (FLST).

Before proceeding, we recall briefly what this transform consists in. In order to merge the trees of lower and upper level sets, Monasse *et al.* , [Monasse and Guichard, 2000], introduced the concept of

Algorithm Level-set TVL1

Inputs: Image f
Parameters: Fidelity parameter λ
Output: Approximate solution u

Perform Alternate Sequential Filter: $a := ASF(f)$
for all $j \in \{0, \dots, 255\}$ **do**
  Compute the upper level set of a: $A_j := \{(x, y),\ a(x, y) \geq j\}$
  $U_j := A_j$;
  for all connected component C of A^j **do**
    if the pixels of C were thresholded in U_{j_0} ($j_0 < j$) or $\Delta\mathbb{E}(C) < 0$ **then**
      (Threshold) $U_j := U_j \setminus C$.
    end if
  end for
end for
Reconstruct u from $\{U_j\}_{0 \leq j \leq 255}$.
for all $j \in \{255, \dots, 0\}$ **do**
  Compute the lower levelset of a: $A^j := \{(x, y),\ a(x, y) \leq j\}$
  Compute the lower levelset of u: $U^j := \{(x, y),\ a(x, y) \leq j\}$
  for all connected component C of A^j **do**
    if the pixels of C were thresholded in U^{j_0} ($j_0 > j$) or $\Delta\mathbb{E}(C) < 0$ **then**
      (Threshold) $U^j := U^j \setminus C$.
    end if
  end for
end for
Reconstruct u form its level sets $\{U^j\}_{0 \leq j \leq 255}$.

Figure 4.2 – *Naive algorithm for TVL1 using openings and level sets.*

shape, and of tree of shapes (see Figure 4.3). A shape is defined as the union of a connected component of an upper or lower set together with its holes. The holes of a set A are defined as the connected components of the complementary set of A which do not intersect the boundary of the image. It can be shown that these shapes have a tree structure, where the parent-children relationship is given by the inclusion of shapes. This structure results from the fact that two shapes are either disjoint or one is included in the other. To each shape is associated its type indicating whether it stems from a connected component of an upper level set ($\geq \lambda$) or of a lower level set ($\leq \lambda$), as well as the value of the associated gray level. An important property of this representation is its self-duality, that is, its invariance with respect to the operation $u \mapsto -u$. This implies in practice that light and dark objects are treated in the same way. A fast algorithm to compute this tree is proposed in [Monasse and Guichard, 2000].

In order to take advantage of the tree of shapes we first notice that the energy of the TVL1 model can be written in a FLST-friendly way. In the discrete case, the energy reads:

$$\mathbb{E}(u) = \sum_{l=0}^{255} (\text{Per } U_l + \lambda |U_l \Delta F_l|) \tag{4.17}$$

where U_l and F_l are the upper level sets of respectively u and f. Incidentally, the length of the boundary of U_l that coincides with the boundary of the image is not counted in the perimeter Per U_l (which amounts to assuming the Neumann conditions $\frac{\partial u}{\partial n} = 0$). Now, for a given shape s, define $\tilde{s}$ as the set of its *proper pixels*, that is, *pixels which belong to s and not to its children*. We obviously have either $\tilde{s} \subset U_l$ or

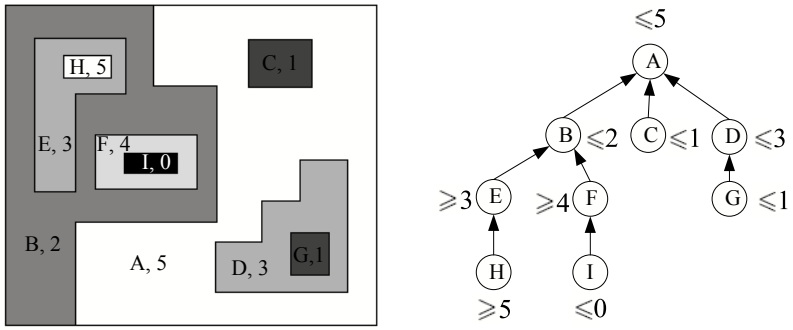


Figure 4.3 – *The Fast Level Set Transform on an example. Left: an image composed of different shapes. Right: the corresponding tree. This illustration was drawn by Gui-Song Xia based on the example given in [Monasse and Guichard, 2000]*

$\tilde{s} \subset U_l^C$. One even has:

$$U_l = \bigcup_{\tilde{s} \subset U_l} \tilde{s} \text{ and } U_l^C = \bigcup_{\tilde{s} \subset U_l^C} \tilde{s} \tag{4.18}$$

Given a gray level l, the perimeter of a shape s contributes to the perimeter Per U_l if and only if l lies in the interval between the value V_s of s and the value $V_{p(s)}$ of its parent. Let us denote this interval by co $(V_s, V_{p(s)})$.

Now, define $F_{sl}^C = \begin{cases} F_l^C \text{ if } \tilde{s} \subset U_l \\ F_l \text{ if } \tilde{s} \subset U_l^C \end{cases}$ and observe that :

$$\tilde{s} \cap F_{sl}^C = \begin{cases} \{x \in \tilde{s} / u(x) \geq l \text{ and } f(x) < l\} \text{ if } \tilde{s} \subset U_l \\ \{x \in \tilde{s} / u(x) < l \text{ and } f(x) \geq l\} \text{ if } \tilde{s} \subset U_l^C \end{cases}.$$

Then, the energy can be written as:

$$\mathbb{E}(u) = \sum_{l=0}^{255} \sum_{s \text{ shape}} \left(\text{Per } s \cdot \mathbb{1}_{l \in \text{co } (V_s, V_{p(s)})} + \lambda |\tilde{s} \cap F_{sl}^C| \right) \tag{4.19}$$

$$= \sum_{s \text{ shape}} \left(\sum_{l=0}^{255} (\text{Per } s \cdot \mathbb{1}_{l \in \text{co } (V_s, V_{p(s)})} + \lambda |\tilde{s} \cap F_{sl}^C|) \right) \tag{4.20}$$

The energy $\mathbb{E}(u)$ is therefore a sum of energies over all the shapes. These energies are not quite independent since the value of a parent has an effect on the energy of its children. As a consequence, one cannot minimize these energies independently. However, one can decrease the global energy by minimizing the energy of the shapes while visiting the tree from the leaves to the root.

More precisely, the thresholding performed at each gray level in the previous algorithm corresponds to modifying the gray value associated with a shape. Given each shape s, and assuming its children are fixed (they have already been visited by the algorithm), we can move the value V_s of the shape towards the value of its parent by one level, if and only if it decreases the energy ($\Delta\mathbb{E} < 0$). The variation of energy is:

$$\Delta\mathbb{E} = -\text{Per } s + \lambda |\tilde{s} \cap F_{s,V_s \pm 1}^C| + \sum_{\substack{c \text{ child of s} \\ \text{with same type}}} \text{Per } c - \sum_{\substack{c \text{ child of s} \\ \text{with opposite type}}} \text{Per } c \tag{4.21}$$

Algorithm FLST-based TVL1

Inputs: Image f
Parameters: Fidelity parameter λ
Output: Approximate solution u

Perform Alternate Sequential Filter: $a := ASF(f)$
Compute the Level Set Transform: $T := FLST(a)$
for all shape s of the tree T **do**
  Remove the children t of s such that $V_t = V_s$.
  if $(\Delta\mathbb{E} < 0)$ (Equation 4.21) **then**
    Remove s.
  end if
end for
Reconstruct $u := FLST^{-1}(T)$.

Figure 4.4 – *An algorithm to compute an approximate solution of TVL1 using openings and the Fast Level Set Transform. Notice that the tree is visited from the leaves to the root.*

where $V_s \pm 1$ is the new gray value to be given to s. Recall that if a shape s is of inferior type, then the value of its parent is above the value of s, and if s is of superior type, the value of its parent is below. Therefore the gray value of s is getting closer to the values of the children with opposite type, but is getting further from the values of the children with same type.

In practice, a quantization step of one is imposed when computing the FLST, so that the difference between the value of every shape and its parent is ± 1. Therefore, instead of moving the value V_s towards the value of the parent, we simply remove the shape s (which amounts to merging s with its parent). Every time a child c is removed, its proper area (that is, $|\tilde{c}|$) is added to the proper area of its parent s, and its children become the children of s. Thus, in Equation (4.21), Per c is replaced by the perimeters of the children of c, with appropriate sign.

Ordering the shapes from the leaves to the root so that each node is visited after all its children have been handled, we perform the algorithm described in Figure 4.4.

Observe that at the beginning of each iteration (in the for loop) we remove the children that have the same value as s but have not been removed yet (this may happen when the grand-children of s have the same value as s and their parent is removed).

Notice that this algorithm is very similar to the one proposed by Dibos and Koepfler in [Dibos and Koepfler, 2001] in order to simulate the TV flow by using the FLST, and later by Darbon in [Darbon, 2005] for TVL1. However the difference is that we start with the output of an ASF filter and compare it to the original image. A first consequence is that the criterion to decide whether or not the gray level associated with a shape should be modified is different. A second difference is that our approach is legitimated by our theoretical study. In particular, whereas Dibos and Darbon do not impose any condition on the curvature of the output, we know that the shapes provided by the ASF filter are somehow "good candidates" to be solutions of TVL1. However, it seems difficult to quantify or bound the error between the exact solution and the result of this algorithm. Assuming that the ASF filter really gives the good candidate for the solution, the algorithm we propose, by testing if the energy decreases when thresholding, follows a greedy approach: it solves successive geometric problems *with, at each step, the constraint that the children are left identical*. Although the solution of TVL1 is a global solution of these constrained problems, this algorithm might lead to a local minimizer.

Both the naive and the FLST-based algorithms give visually very similar results. The second one is much faster and even competitive with Darbon-Sigelle's algorithm [Darbon and Sigelle, 2006]: using square as structuring element it took $1.27s$ to filter a 512×512 image with $\lambda = 1/3$, using an Intel

Core2 Duo E8400, 4Gb of RAM. Notice that in this experiment, half the computation time is due to the openings and closings (ASF filter). As a comparison, Darbon-Sigelle's algorithm takes $0.96s$ using 4-connectivity. Using isotropic schemes (i.e. balls as structuring elements versus 8-connectivity), the trend is inverted since Darbon-Sigelle's algorithm takes longer using a more complicated connectivity ($1.71s$) whereas the FLST -based scheme is not affected at all ($1.27s$). However, let us recall again that we do not perform exact minimization.

4.2.3 Commentary

In this section we discuss the results of the FLST-based algorithm. It should be noted that in order to compare this algorithm with existing TVL1 algorithms, we should focus on schemes that satisfy the coarea formula. For instance, when using a gradient descent formulation, numerical instabilities compels us to regularize the total variation by replacing $\sqrt{u_x^2 + u_y^2}$ with $\sqrt{\epsilon^2 + u_x^2 + u_y^2}$ in the divergence term. Although ϵ is assumed to be very small, this implies that the scheme does not satisfy the coarea formula. As a matter of fact, in the discrete case, even when $\epsilon = 0$, the coarea formula is not satisfied. As a consequence, new grey levels appear and characteristic functions of sets are blurred. Therefore, it is not possible to experimentally check e.g. curvature values because of uncertainty on the boundary of the smoothed sets (see Figure 4.5).

We therefore compare our results with Darbon-Sigelle's algorithm [Darbon and Sigelle, 2006] since, to our knowledge, it is the only algorithm that satisfies the coarea formula. Observe however that this algorithm is anisotropic. Its authors propose two different schemes: one relying on 4-connectivity and the other one relying on 8-connectivity, the latter being more isotropic. In case of 4-connectivity, the scheme is actually consistent with an anisotropic total variation formally defined as: $\int(|u_x| + |u_y|)$ instead of $\int \sqrt{u_x^2 + u_y^2}$. We refer the reader to Chapter 7 for more details about anisotropic TVL1. For now, let us mention that this kind of anisotropic, crystalline, total variation is the framework of intensive research (among which [Bellettini et al., 2001],[Caselles et al., 2008], [Bellettini et al., 2009]). A didactic introduction to this subject and a study of the anisotropic ROF model are provided in [Esedoglu and Osher, 2004]. Let us mention that one can define an anisotropic perimeter as well as an anisotropic curvature and that similar results as those of Lemma 3.3.6 can be stated [Bellettini et al., 2001]. In our case the associated Wulff shape is a square. It is not the purpose of this chapter to explore the links between anisotropic total variation and the results of Section 3.1, but for now let us observe that our results seem to hold with anisotropic total variation when replacing opening with balls by opening with squares. This property will be proved in Chapter 7. Therefore, the reader should keep in mind, while looking at the result with Darbon-Sigelle's algorithm connectivity 4, that these correspond to openings relying on squares instead of balls.

Figure 4.5 shows the evolution of simple convex and non-convex shapes. First, notice that results of rows 2 and 3 (respectively from Darbon-Sigelle's algorithm and our approach when using square structuring elements) are remarkably similar (although a few pixels may differ). Observe also that, as it is well known, the finite difference implementation of the fourth row exhibits blur. Last, we can notice that results displayed on the last row, using the approach introduced in this paper with balls as structuring elements, are the only ones being both isotropic and with sharp edges.

Figure 4.6 and 4.7 illustrate the ability of the algorithm to provide isotropic results when using balls as structuring elements and provides comparisons with Darbon-Sigelle's algorithm using 8-connectivity on natural images. Some details may differ but on the whole the results look very similar. Incidentally, let us notice that using 8-connectivity with Darbon-Sigelle's algorithm yields, as it is well-known, slightly anisotropic results, but this anisotropy becomes visible only at very low values of λ (which are unrealistic when denoising images). The isotropy of the FLST-based algorithm, is a direct consequence of the isotropy of the structuring element used to perform the openings. In this thesis, we use a straightforward scheme relying on the euclidean distance. However, implementations relying on more sophisticated distances could be considered.

Figure 4.8 displays another experiment on a natural image. In practice, the alternating sequential

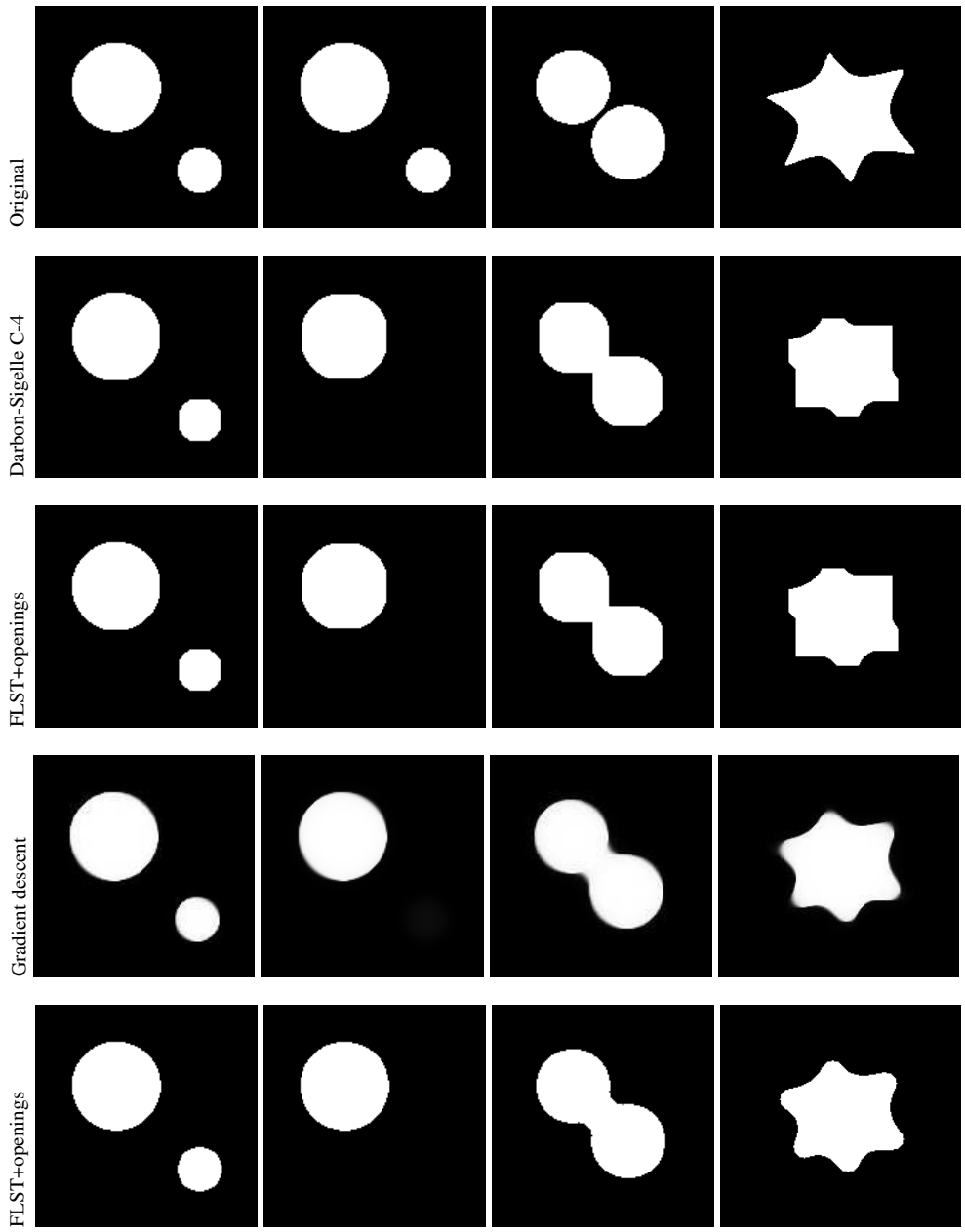


Figure 4.5 – *From top to bottom.* ***First row:*** *original image.* ***Second row:*** *result obtained with Darbon-Sigelle's algorithm (connectivity 4).* ***Third row:*** *result of the FLST based algorithm, i.e. an alternating sequential filter (alternated opening/closing with increasing radius, the structuring element being a square of side $2/\lambda$) completed with a thresholding on the shapes.* ***Fourth row:*** *result of the gradient descent scheme (300 000 iterations, $dt = 0.1$).* ***Fifth row:*** *result of the FLST-based algorithm with euclidean balls as structuring elements. All pictures were computed with $\lambda = 1/9$, except the first column ($\lambda = 1/6$, the small component is kept, contrary to the second column). Notice the blurry effect introduced by the the gradient descent.*

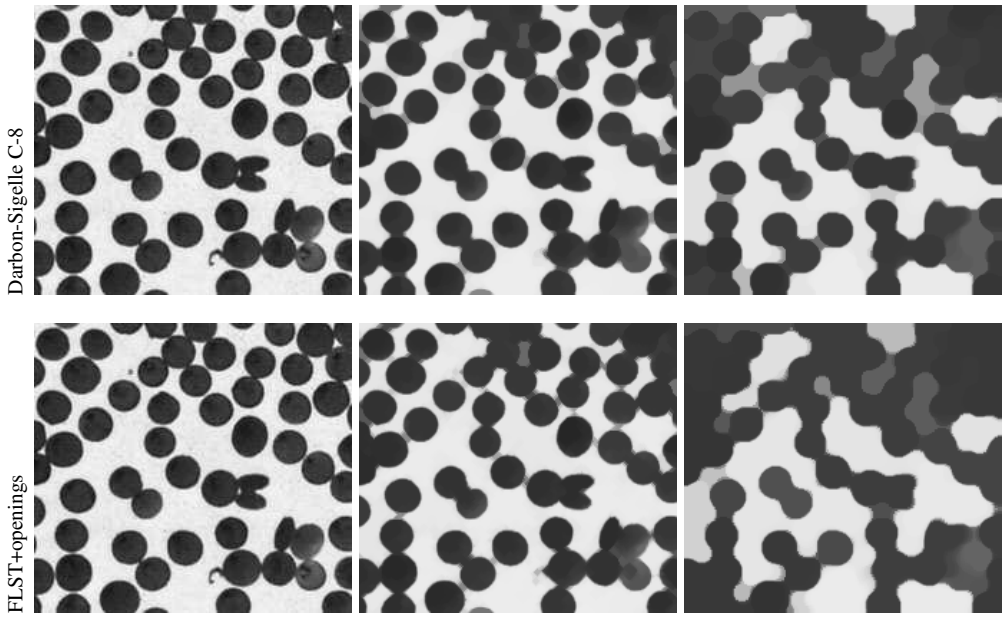


Figure 4.6 – *Isotropy. The upper row illustrates the behavior the Darbon-Sigelle algorithm using 8-connectivity on a detail of a Brodatz texture (from left to right: original image, result of the algorithm with increasing values of λ) Lower row: behavior of the FLST-based filter using euclidean balls followed by a thresholding. Notice that the anisotropy of Darbon-Sigelle's scheme is not very visible, even at low values of λ.*

filter is not exactly self-dual since a finite step $r_0 = 1$ has to be chosen. As a consequence, the result in some specific areas may depend on the order of the ASF.

Figure 4.9 shows the "granulometry property" of the algorithm: details vanish in function of their scale. Eventually, we compare on Figures 4.10 and 4.11 the scale-spaces of the proposed algorithm and the TVL1 algorithm. Although slight differences may appear (the Wulff shape is a disc in the first case, an octagon in the second), the result is remarkably similar.

Figure 4.7 – *Isotropy again. Left: original image. Middle: result of Darbon Sigelle's algorithm using 8-connectivity ($\beta = 4.6$). Using 8-connectivity provides more isotropic results than using 4-connectivity. However, especially for very low values of λ, it is not as isotropic as the scheme proposed in this paper when using balls as structuring element, as shown on the right image ($\lambda = 1/11$).*

Figure 4.8 – *Chessboard dilemma. From left to right: original image, results of the FLST based algorithm ($\lambda = 1/3$ in the first case the alternating sequential filter starts with an opening, and in the second one it starts with a closing). The ASF is not exactly self-dual and in some specific cases the result visibly depends on the chosen order. However, this does not necessarily contradict the TVL1 model since in case of the chessboard dilemma, the solution of the model is not unique.*

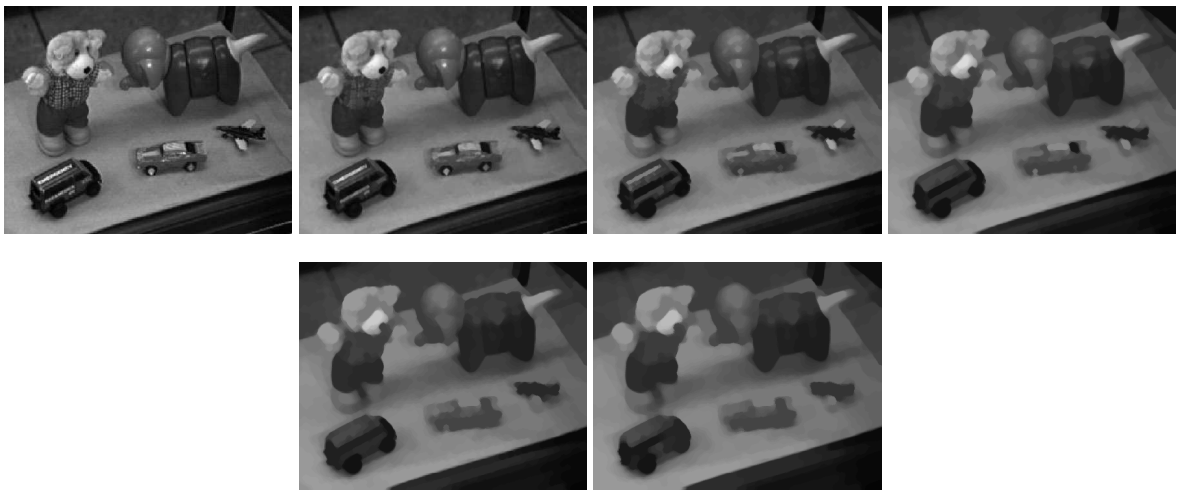


Figure 4.9 – *Scale space using the FLST-based algorithm. From left to right, top to bottom: original image, result of the FLST-based algorithm ($\lambda = 1$, $1/2$, $1/3$,$1/4$, $1/5$). Observe how details vanish one after the other, in function of their size: the shirt, the eyes of the teddy-bear, one wing of the plane... This illustrates the "improved granulometry" property of the TVL1 model: objects evolve according to an opening/closing and then vanish depending on a geometric criterion.*

(a) ASF+FLST (b) TVL1 [Darbon and Sigelle, 2006]

Figure 4.10 – *Comparison of algorithm 4.4 with discs and [Darbon and Sigelle, 2006] with 8-connectivity. The parameters are respectively* $\lambda \in \{1, 1/2, 1/3, 1/4, 1/5\}$*, and* $\beta \in \{0.28, 0.65, 1.08, 1.45, 1.7\}$.

(a) ASF+FLST (b) TVL1 [Darbon and Sigelle, 2006]

Figure 4.11 – *Comparison of algorithm 4.4 with discs and [Darbon and Sigelle, 2006] with 8-connectivity. The parameters are respectively* $\lambda \in \{1, 1/2, 1/3, 1/4, 1/5\}$*, and* $\beta \in \{0.28, 0.65, 1.08, 1.45, 1.7\}$.

Chapter 5

Cartoon, Textures and Granulometries

In this chapter, we compare two cartoon-texture decomposition models introduced in the literature: the TV-G model and the TVL1 model. We begin with a reminder on the principle of such decompositions and, for once, we show the visual results before getting into the theoretical details. In particular we highlight an artifact that appears with the TV-G model, namely the fact that edges appear in the texture part.

In the second section, we study the TV-G model. A typical motivation of the TV-G model is that indicator functions of smooth sets have a small total variation whereas oscillatory signals like sine functions have a small G-norm. Yet, to our knowledge, no example was ever given that the TV-G model is able to separate such signals. We provide here an example where the model gives the expected decomposition when the input is the indicator function of a disc plus a (slightly perturbed) sine function. The study of this example leads us to discuss the presence of edges in the texture part. They do not necessarily appear but it seems that the texture part cannot be empty in a neighborhood of the edges of the cartoon part.

The third Section is devoted to the TVL1 model: we show that it is also able to separate an indicator function from a sine function, and that it does not suffer from the apparition of edges in the texture part. Eventually, we explain the good performance of the TVL1 model for cartoon-texture decomposition by drawing a connection with the granulometries of mathematical morphology.

5.1 Cartoon-Texture decomposition: introduction and visual results

5.1.1 Introduction

The problem of cartoon-texture decomposition originates from the seminal work of Y. Meyer [Meyer, 2001]. In this book, the author proposes to separate a signal into two components by solving a variational problem of the form:

$$\min_{(u,v)\in X_1\times X_2} \{F_1(u) + \lambda F_2(u) \;/\; u+v=f\}\,, \tag{5.1}$$

where $\lambda > 0$ is a tuning parameter, $F_1, F_2 \geq 0$ are functionals and X_1, X_2 are function spaces such that $X_1 = \{u, :\, F_1(u) < +\infty\}$ and $X_2 = \{v :\ F_2(v) < +\infty\}$. Typically F_1 and F_2 are norms or semi-norms. They are chosen so that if u is a cartoon and v is an oscillating function (which is the model we adopt for textures), $F_2(u) \gg F_1(u)$ and $F_1(v) \gg F_2(v)$. If this property is satisfied, one expects any minimizer (u, v) to be a clear separation of f into a cartoon and an oscillating function.

The choice of the total variation for F_1 is quite natural, since it penalizes oscillations and does not penalize much piecewise smooth functions with sharp edges. As a consequence, the image processing community has mainly focussed on the choice of F_2, and consequently on the choice of a function space to model textures. In [Meyer, 2001], three spaces were originally proposed: E, F and G which respectively stand for $\dot{B}^{-1}_{\infty,\infty}$ (the dual of the Besov space $\dot{B}^1_{1,1}$), the divergence of vector fields in BMO, and the divergence of functions in L^∞. Many more have been proposed by the community, and yet, as noted in [Buades et al., 2010], no specific function space has really prevailed in the modelling of textures.

To our knowledge, the first attempt to perform a numerical cartoon-texture decomposition was made by Vese and Osher [Vese and Osher, 2003], using an approximation of the L^∞ norm with the L^p norm ($p \geq 1$) in the definition of the G norm. In [Osher et al., 2003], Osher, Solï¿½ and Vese have proposed a variant of this model for the case $p = 2$, which corresponds to the space $W^{-1,2}$. The Osher-Solï¿½-Vese model is a particular case of $TV + \mathcal{H}$ decompositions, where $\mathcal{H}$ is a Hilbert space. Such decompositions were investigated in [Aujol and Chambolle, 2005, Aujol and Gilboa, 2006]. In the latter work, the texture term being actually the square of a Hilbert norm, the authors propose a fast algorithm inspired from the Chambolle algorithm [Chambolle, 2004b]. Negative Sobolev spaces have drawn a lot of attention: see [Aujol and Chambolle, 2005, Lieu and Vese, 2009, Kim and Vese, 2009]. The space E and general homogeneous Besov spaces were investigated in [Aujol and Chambolle, 2005, Garnett et al., 2007], and the space F in [Le and Vese, 2005].

Several works have focussed on the G-space: in [Aujol et al., 2005, Aubert and Aujol, 2005], Aujol *et al.* propose an adaptation of the G-space for functions defined on a bounded open set Ω, and they give a numerical algorithm that provides a solution of the TV-G problem. Since the G-norm has no integral expression, even its computation is an issue: different algorithms were proposed in [Aujol and Chambolle, 2005, Kindermann et al., 2006]. Haddad and Meyer [Haddad, 2007] have investigated the model further and they have shown that for a certain class of functions, called extremal functions, the solution of the TV-G problem is not necessarily unique. This class of functions turns out to be the one independently studied by Belletini *et al.* in [Bellettini et al., 2002] as the solutions of the eigenvalue problem:

$$-\operatorname{div}\left(\frac{Du}{|Du|}\right) = u. \tag{5.2}$$

When the image has been contaminated by noise, it is necessary to split the image into three parts: cartoon, texture, and noise. In [Aujol and Chambolle, 2005], it is proposed to use the space G to model textures and the space E for the noise, whereas in [Gilles, 2007], both textures and noise are modelled by the G space, but the noise is assumed to have the smaller G norm (since it is more oscillating). To increase the performance, a system of spatially varying weights is added. Recently, Gilles and Meyer have studied a $TV - L^2 - G$ model in [Gilles and Meyer, 2010] in order to perform road detection in satellite images.

The above bibliography is far from being exhaustive. Let us also mention wavelet or frame-based decomposition methods like the Morphological Component Analysis [Starck et al., 2005] or the variational approach of Daubechies and Teschke [Daubechies and Teschke, 2005]. Also, in recent works [Kindermann et al., 2005, Gilboa and Osher, 2008], the classical functional analysis spaces are replaced with their non-local analogous inspired from [Buades et al., 2005].

In fact, the problem of the cartoon-texture decomposition should not necessarily be addressed in the classical setting of functional analysis: the recent work [Buades et al., 2010] shows that a carefully designed filter may perform as well (if not better) as a variational model. In that paper the authors build a spatially varying filter depending on the relative reduction rate of a localized total variation (which measures the local oscillations of the image). It might be a bit disappointing to abandon the idea of finding a function space that models textures, but to some extent, this is also what we will be doing here by using the TVL1 model to perform decompositions. Following [Haddad, 2007, Yin et al., 2007a] we compare the TVL1 and TV-G models. Our contribution is to highlight some visual artifacts and to explain them on simple models.

5.1.2 Visual results

To compare the decomposition results provided by the TV-G and the TVL1 models, we have used the algorithms proposed in [Aujol et al., 2005, Aujol et al., 2006]. Let us mention that faster approaches could be adopted (for instance [Chambolle and Pock, 2011]).

To make this manuscript more cheerful, we have also added experiments performed on color images using a straightforward extension of the models (see [Duval et al., 2010]). The total variation is defined

so that:

$$\int |Du| = \int_{\mathbb{R}^2} \sqrt{|\nabla u_R(x)|^2 + |\nabla u_G(x)|^2 + |\nabla u_B(x)|^2} dx \tag{5.3}$$

for smooth functions, where u_R, u_G, u_B are the red, green and blue color channels. This choice is better than a simple channel by channel decomposition (see [Bresson and Chan, 2008]), but more sophisticated models have been proposed [Sapiro and Ringach, 1996, Sochen et al., 1998, Blomgren and Chan, 1998]: in particular let us mention the use of a chromaticity-brightness color space for total variation filtering (see [Chan et al., 2000, Aujol and Kang, 2006]) and the extension of the total variation using Jacobians (see [Federer, H., 1969]) proposed by Goldluecke and Cremers [Goldluecke and Cremers, 2010].

In all the experiments, we choose the parameters that provide satisfactory visual results, although it is sometimes difficult to define what the expected cartoon-texture decomposition should be. Figure 5.1 shows four test images for the decomposition (the bottom left one is from the Kodak true color image test set, the bottom right one was downloaded from `www.ipol.im`). We begin with Figure 5.2 which deals with the classical decomposition of the Barbara image. The reader may observe the difference between the two models, in particular the fact that with the A2BC model, the texture part looks "inflated": all the relief indicated by the lighting of the scene is taken to the v part so that the woman seems to have plump cheeks and the floor looks irregular. Moreover, most edges are visible in the texture part. We wish to emphasize that this is not because of a wrong choice of parameters. In the next Section, we will give theoretical arguments that justify the apparition of this artifact. For now, let us mention that the TV-G decompositions of Barbara shown in the literature [Aujol and Gilboa, 2006, Gilles, 2009] share the same aspect. By comparison, the TVL1 decomposition looks flatter, and the edges are less prominent. Yet, the details and small patterns have been sent to the v part as expected.

On Figure 5.3 and 5.4, the decompositions provided by the two algorithms are similar: the small details are sent to the v part whereas the structure is sent to the u part. On the contrary, the decompositions displayed on Figure 5.5 look rather different. In the v part of the A2BC algorithm, the caps look inflated, and shadows of their edges are clearly visible. The v part of the TVL1 model looks flatter, and the edges are far less visible. A detail of this decomposition is given in Figure 5.6.

5.2 The TV-G model

5.2.1 Definition

For consistency with our study of the TVL1 model, we adopt a similar framework to the one in [Meyer, 2001, Kindermann et al., 2006], i.e. images are defined on the plane. We refer the reader to [Aubert and Aujol, 2005] for a definition of the G space with functions defined on a bounded open set $\Omega \subset \mathbb{R}^2$. As in [Kindermann et al., 2006], we regard G as a subspace of L^2:

Definition 5.2.1. A function $f \in L^2(\mathbb{R}^2)$ belongs to G if and only if there exists a vectorfield $g \in L^\infty(\mathbb{R}^2, \mathbb{R}^2)$ such that div $g = f$. The G-norm is defined as:

$$\|f\|_G := \inf \left\{ \|g\|_\infty,\ f = \text{div } g, g \in L^\infty(\mathbb{R}^2, \mathbb{R}^2) \right\}. \tag{5.4}$$

Notice that this infimum is in fact a minimum, i.e. there exists $g \in L^\infty(\mathbb{R}^2, \mathbb{R}^2)$ such that div $g = f$ and $\|g\|_\infty = \|f\|_G$. In that case we say that g is adapted to v.

The G-norm can be written in a "dual" form :

Proposition 5.2.2 ([Haddad and Meyer, 2007, Kindermann et al., 2006])**.** *Let $f \in L^2(\mathbb{R}^2)$. Then:*

$$\|f\|_G = \sup_{u \in BV(\mathbb{R}^2) \setminus \{0\}} \frac{\int_{\mathbb{R}^2} fu}{\int |Du|}. \tag{5.5}$$

Figure 5.1 – *Test images for the cartoon-texture decomposition.*

This expression of the G norm and the isoperimetric inequality lead to the following relation:

Proposition 5.2.3 ([Haddad and Meyer, 2007, Kindermann et al., 2006])**.** *Let $f \in L^2$, then $f \in G$ and:*

$$\|f\|_G \leq \frac{1}{2\sqrt{\pi}}\|f\|_2 \leq \frac{1}{4\pi}\int |Df| \tag{5.6}$$

(with $\int |Df| = +\infty$ if $f \notin BV(\mathbb{R}^2)$).

The motivation for the G-norm in decomposition models is that it is small for oscillating functions. For instance, Y. Meyer [Meyer, 2001] shows that if $f \in L^2$ is defined by $f(x_1, x_2) = \mathbb{1}_{[0,1]^2}(x_1, x_2)\cos(N\pi x_1)$, then $\|f\|_G = \frac{1}{N}$. More generally, the following result holds:

Proposition 5.2.4 ([Haddad and Meyer, 2007], Corollary 3.4)**.** *Let $Q = [0,1]^2$, $f \in L^2$ and $\mu \in L^\infty$ such that $\int_{Q+k} \mu(x)dx = 0$, for all $k \in \mathbb{Z}^2$. Then, $\|f(\cdot)\mu(N\cdot)\|_G \to 0$ when $N \to +\infty$.*

Having observed this behavior on oscillating functions, Y. Meyer proposed to decompose images using the following decomposition TV-G model in [Meyer, 2001]:

$$\inf_{u \in \mathrm{BV}(\mathbb{R}^2)} \int |Du| + \lambda\|f - u\|_G, \tag{5.7}$$

given $f \in L^2$. The existence of a solution follows from Propositions (5.2.2) and (5.2.3) and the direct method of the calculus of variations. An interesting point is the fact that the solution is not necessarily unique, which was proved by Haddad and Meyer [Haddad and Meyer, 2007] by considering extremal functions (for instance the indicator of a calibrable set).

Figure 5.2 – *Decomposition of the famous "Barbara" image using A2BC algorithm (top) and TVL1 (bottom).*

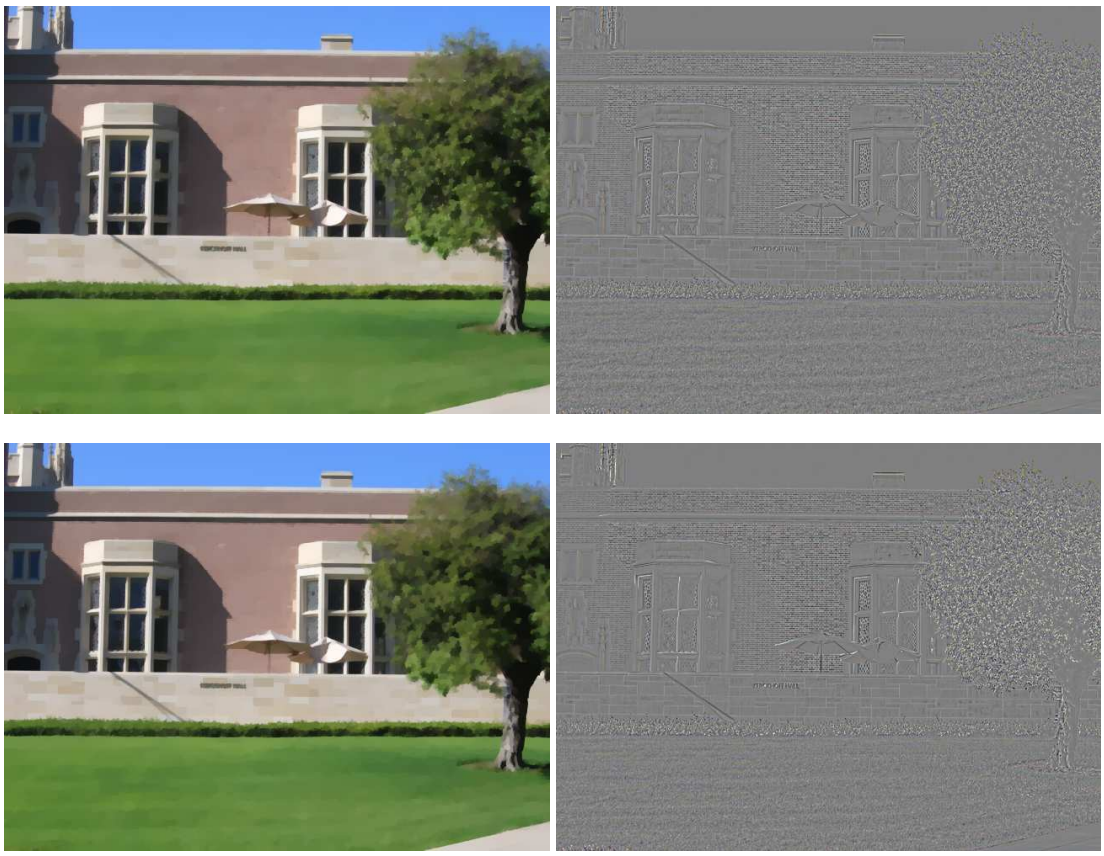


Figure 5.3 – *Cartoon-texture decomposition using A2BC algorithm (upper row) and TVL1 (lower row).*

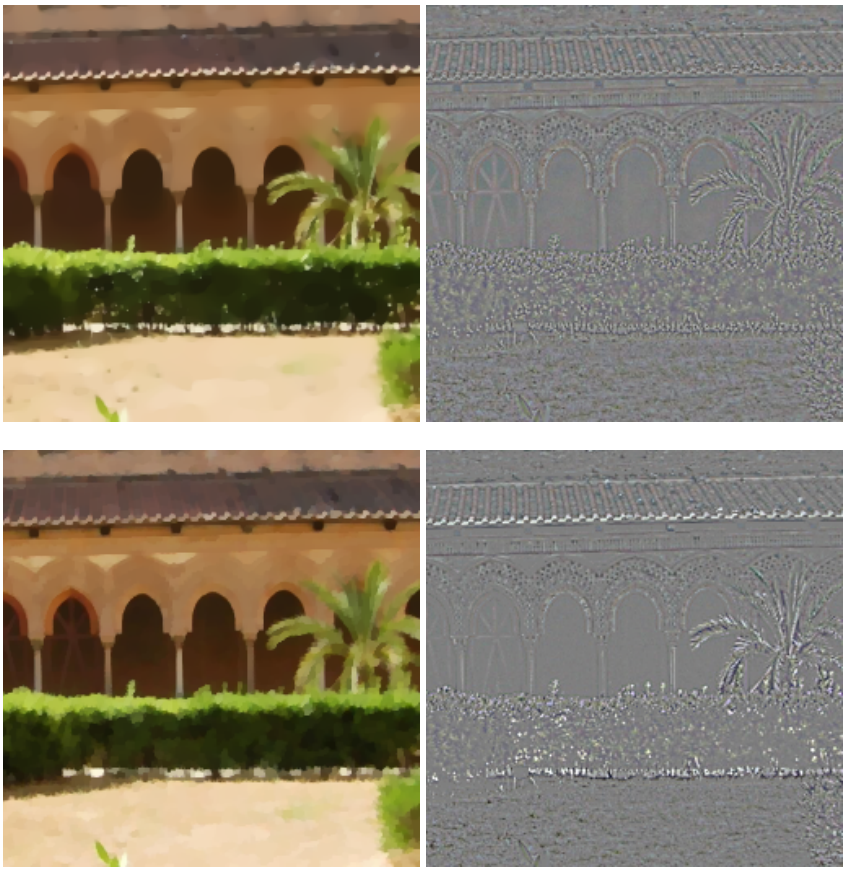


Figure 5.4 – *Cartoon-texture decomposition using the A2BC algorithm (top) and the TVL1 model (bottom).*

5.2.2 First properties

In [Kindermann and Osher, 2005], Kindermann *et al.* give necessary and sufficient conditions on a couple (u, v) where $v = f - u$ to be a solution of Problem (5.7):

Theorem 5.2.5 ([Kindermann and Osher, 2005] Theorem 3.4)**.** *Let $f = u + v$ be a $BV + G$ decomposition of $f \in L^2$. Then the pair (u, v) is optimal if and only if there exists a function $p \in BV(\mathbb{R}^2)$ such that:*

$$\lambda \int up = \int |Du|, \tag{5.8}$$

$$\int pv = \|v\|_G, \tag{5.9}$$

$$\int |Dp| \leq 1, \tag{5.10}$$

$$\|p\|_G \leq \frac{1}{\lambda}. \tag{5.11}$$

Using the framework of Section 1.2 on the Euler-Lagrange equation, this result can be obtained straightforwardly under the following equivalent form:

Figure 5.5 – *Cartoon-texture decomposition using the A2BC algorithm (top) and the TVL1 model (bottom). The parameters were chosen as a compromise so that the results are comparable. However, with the A2BC algorithm, if one wants to send the raised pattern on the left side of the wall to the texture part (as in the TVL1 model), one has to considerably degrade the clouds.*

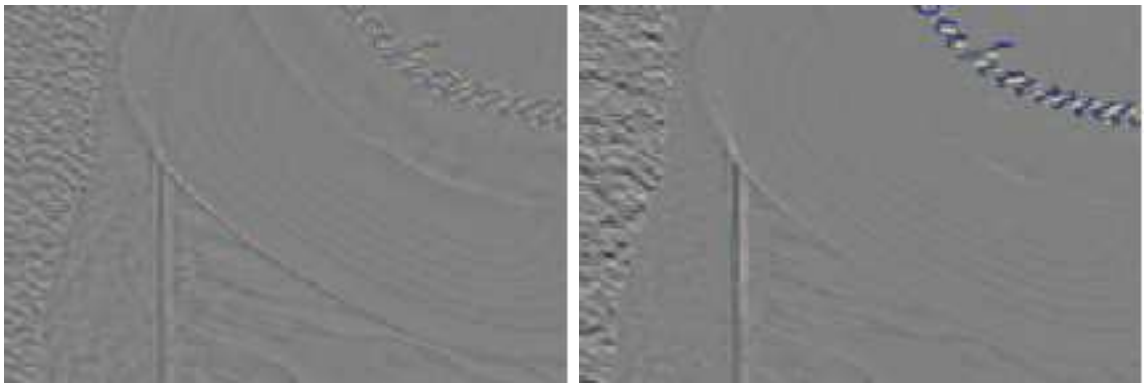


Figure 5.6 – *Detail of the texture parts of Figure 5.5 (left: A2BC, right TVL1). Notice how the edge of the cap appears in the A2BC model.*

Proposition 5.2.6. *The decomposition (u, v) is optimal if and only if:*

$$\exists z \in X_\infty, \quad |z| \leq 1 \text{ and } \int u div\, z = \int |Du|, \tag{5.12}$$

$$\exists p \in BV(\mathbb{R}^2), \quad \int |Dp| \leq 1 \text{ and } \int pv = \|v\|_G, \tag{5.13}$$

$$\text{such that} \quad div\, z + \lambda p = 0. \tag{5.14}$$

In [Haddad and Meyer, 2007], Haddad and Meyer study the stability of the decomposition $f = u+v$ of simple functions using such a model. They also show several interesting properties of decompositions. In particular:

Lemma 5.2.7 ([Haddad and Meyer, 2007], Lemma 4.1)**.** *For $0 < \lambda < 4\pi$ and any $f \in L^2$, the optimal decomposition is unique and it is given by $(u, v) = (0, f)$.*

Proof. From the inequality (5.6), we get:

$$\begin{aligned} \forall u \in L^2, \quad \int |Du| + \lambda \|f - u\|_G &\geq 4\pi \|u\|_G + \lambda \|f - u\|_G \\ &\geq \lambda \left(\|u\|_G + \|f - u\|_G \right) \\ &\geq \lambda \|f\|_G, \end{aligned}$$

with strict inequality when $u \neq 0$. □

Theorem 5.2.8 ([Haddad and Meyer, 2007], Theorem 4.1)**.** *Let $f = u + v$ be an optimal decomposition of $f \in L^2$. Then $u = u + 0$ is an optimal decomposition of u and $v = 0 + v$ is an optimal decomposition of v.*

We have noticed in Remark 3.2.4 that this property also holds for the TVL1 model. In fact, since the proof relies on the triangle inequality, it holds for any model of the form $\inf_u \int |Du| + \lambda \|f - u\|_E$ where E is a normed space.

5.2.3 A toy example

In [Haddad and Meyer, 2007], it is noted that if the inequality is strict in (5.10), then Eq. (5.9) implies that $v = 0$. Hence, let us assume that $v \neq 0$ and that g is a vectorfield adapted to v, i.e. div $g = v$ and $\|g\|_\infty = \|v\|_G$. Then (5.9) is equivalent to :

$$\begin{aligned} -\int (g, Dp) &= \|g\|_\infty \int |Dp| \\ \text{i.e.,} \quad 0 &= \int \left((g, \frac{Dp}{|Dp|}) + \|g\|_\infty \right) d|Dp|. \end{aligned} \tag{5.15}$$

This means that $|Dp|$-almost everywhere, $|g|$ must reach its maximum, and g must be pointing in the opposite direction from Dp.

Before proceeding with the example, let us recall the following remark by Haddad and Meyer :

Theorem 5.2.9 ([Haddad and Meyer, 2007], Theorem 3.2)**.** *Consider a radial function $f \in L^2$, i.e. $f(x) = \tilde{f}(r)$ with $r = |x|_2$, $\tilde{f} \in L^2(rdr)$. Then,*

$$\|f\|_G = \left\| \frac{1}{r} \int_0^r \tilde{f}(s) s ds \right\|_\infty . \tag{5.16}$$

In fact the radial vectorfield $g(x) = \tilde{g}(r)e_r$ with $\tilde{g}(r) = \frac{1}{r}\int_0^r \tilde{f}(s)sds$ is adapted to f.

Using this theorem, Haddad [Haddad, 2007] gives a semi-explicit decomposition of the radial function $\tilde{f}(r) = e^{-r}$. We provide here an explicit example that shows that the superposition of the indicator of a disc and a perturbed sine is perfectly decomposed by the TV-G model:

Proposition 5.2.10. *Let $N \in \mathbb{N}^*$ be an odd number. For $4\pi \leq \lambda \leq 4\pi\left((N-1)(2N+\frac{1}{2}) + \frac{2N}{N+1}\right)$, the function f defined by*

$$\tilde{f}(r) = \mathbb{1}_{[0,1]}(r) - \beta\left(\sin(N\pi r) - \frac{1}{\pi N}\cos(N\pi r)\right)\mathbb{1}_{[1/2,3/2]}(r), \tag{5.17}$$

has an optimal decomposition given by:

$$\begin{aligned} u(x) &= \mathbb{1}_{B(0,1)}(x), \\ \tilde{v}(r) &= -\beta\left(\sin(N\pi r) - \frac{1}{\pi N}\cos(N\pi r)\right)\mathbb{1}_{[1/2,3/2]}(r). \end{aligned}$$

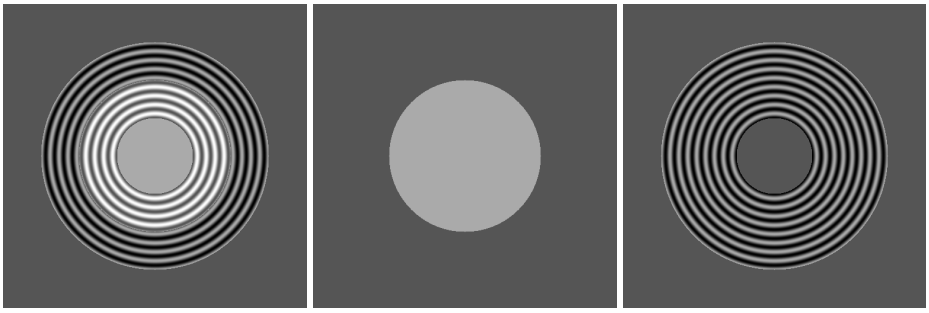


Figure 5.7 – *TV-G decomposition predicted by Proposition 5.2.10: original image f (left), cartoon part $u = \mathbb{1}_{B(0,1)}$ (middle), texture part $v(x) = \left(-\sin(N\pi r) + \frac{1}{\pi N}\cos(N\pi r)\right)\mathbb{1}_{[1/2,3/2]}(r)$ (right).*

Proof. The idea is to check that (u, v) satisfy Equations (5.12)-(5.14).

To this end, consider the vectorfield $\tilde{g}(r) = \frac{1}{N\pi}\cos(N\pi r)\mathbb{1}_{[1/2,3/2]}(r)$. Notice that $\tilde{g}(1) = -\frac{1}{N\pi} = -\|g\|_\infty$, and $g(\frac{1}{2}) = g(\frac{3}{2}) = 0$, so that g is continuous at these points. Then set $v := \text{div } g$. We have $\tilde{v}(r) = \frac{1}{r}\frac{\partial(rg(r))}{\partial r} = \left(-\sin(N\pi r) + \frac{1}{N\pi}\cos(N\pi r)\right)\mathbb{1}_{[1/2,3/2]}(r)$, and $\|v\|_G = \|g\|_\infty$.

- For $\lambda = 4\pi$. Let us define $p = \left(-\frac{2}{\lambda}\right)\mathbb{1}_{B(0,1)}$. We clearly have $\int |Dp| = \frac{4\pi}{\lambda} = 1$, and :

$$\begin{aligned} \int p(x)v(x)dx &= 2\pi\int_0^{+\infty} \tilde{p}(r)\tilde{v}(r)rdr && (5.18)\\ &= 2\pi\int_0^1 \left(\frac{-2}{\lambda}\right)\tilde{v}(r)rdr && (5.19)\\ &= -\frac{4\pi}{\lambda}g(1) = \|v\|_G. && (5.20) \end{aligned}$$

Eventually, we define $z(x) = \begin{cases} re_r \text{ for } r \leq 1 \\ \frac{1}{r}e_r \text{ otherwise.} \end{cases}$. We see that $|z| \leq 1$, div $z = 2 \times \mathbb{1}_{B(0,1)}$, and $\int u\text{div } z = 2\pi = \int |Du|$. Since div $z + \lambda p = 0$, the pair (u, v) is optimal.

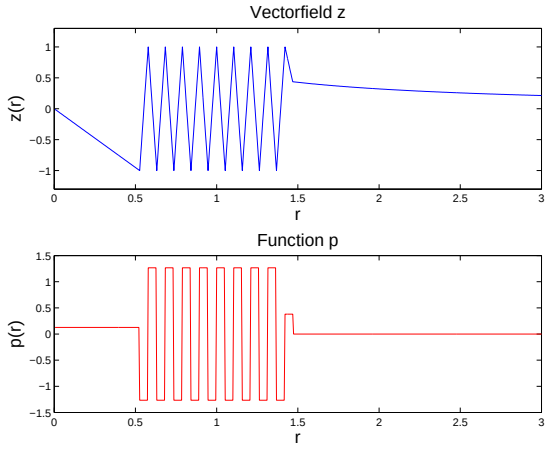


Figure 5.8 – *Construction of the function p and the vectorfield z in Proposition 5.2.10.*

- For $\lambda > 4\pi$, the idea is to distribute the mass $\int |Dp|$ on the extrema of g. Let $k^* \in \{0, \dots, \lfloor \frac{N}{2} \rfloor\}$, $j^* \in \{-\lfloor \frac{N}{2} \rfloor, \dots, 0\}$; we build a piecewise constant function $p(r) = p_k$ for $r \in [r_k, r_{k+1})$, where $r_k = 1 + \frac{k}{N}$, $k \in \{j^*, \dots, 0, 1, \dots, k^*\}$.

  For $r \in [0, r_{j^*})$, we set $p(r) = p_{j^*-1} := (-1)^{j^*} \frac{2}{\lambda r_{j^*}}$, so that $\tilde{z}(r) = (-1)^{j^*+1} \frac{r}{r_{j^*}}$. For $r \geq r_{k^*}$, we set $p(r) = p_{k^*} = 0$.

  Now, we choose each p_k so that $\int |Dp| = \sum_{k=j^*}^{k^*} |Dp|(\{r_k\}) = 1$, with the constraint that $|z| \leq 1$. Since $|Dp|(\{r_k\}) = 2\pi r_k |p_k - p_{k-1}|$, the condition $\frac{Dp}{|Dp|}(\{r_k\}) \cdot g(r_k) < 0$ imposes that $(-1)^{k+1}(p_{k+1} - p_k) \geq 0$.

  If z is a continuous vectorfield such that $\operatorname{div} z = -\lambda p_k$ on $[r_k, r_{k+1})$, then

$$\forall r \in [r_k, r_{k+1}), \quad \tilde{z}(r) = \frac{1}{r} \left(r_k \tilde{z}_k + \frac{\lambda p_k}{2} (r_k^2 - r^2) \right). \tag{5.21}$$

  Plugging this expression in the constraint $|z| \leq 1$, we eventually observe that we may choose[1] :

$$\forall k \in \{j^*, \dots, 0, \dots, k^* - 1\}, \quad p_k = (-1)^k \frac{2}{\lambda(r_{k+1} - r_k)}, \tag{5.22}$$

$$p_{k^*-1} = (-1)^{k^*-1} \frac{2\eta}{\lambda(r_{k^*+1} - r_k^*)}, \tag{5.23}$$

$$p_{j^*-1} = (-1)^{j^*-1} \frac{2}{\lambda r_{j^*}} \tag{5.24}$$

$$\tag{5.25}$$

  where $0 \leq \eta \leq 1$.

  The condition $\int |Dp| = 1$ can be reformulated as:

$$2\pi r_{j^*} |p_{j^*} - p_{j^*-1}| + \sum_{k=j^*+1}^{k^*-1} 2\pi r_k |p_k - p_{k-1}| + 2\pi r_{k^*} |p_{k^*} - p_{k^*-1}| = 1,$$

$$\text{i.e. } 4\pi \left(N + \frac{1}{1 + \frac{j^*}{N}} + 2N \sum_{k=j^*+1}^{k^*-1} (1 + \frac{k}{N}) + \left(1 + \frac{k^*}{N}\right) \eta N \right) = \lambda$$

[1] Notice that the following expression is somewhat optimal, since for each $k \in \{-k^* + 1, \dots, 0, \dots, k^* - 1\}$ the constraint $|z| \leq 1$ is active since $\tilde{z}_k = \tilde{z}(r_k) = (-1)^{k+1}$.

With an appropriate choice of j^*, k^*, η, any value of $\lambda \in \left[4\pi, 4\pi\left((N-1)(2N+\frac{1}{2})+\frac{2N}{N+1}\right)\right]$ may be reached.

□

Notice that the above construction shares similarities with the construction of an extremal function in [Haddad and Meyer, 2007].

More generally, the following result holds:

Proposition 5.2.11. *Let $\tilde{g} \in W^{1,\infty}(\mathbb{R}_+, \mathbb{R})$ with $\tilde{g}(0) = 0$. Assume that there are values $0 < r_{j^*} < \ldots < r_0 < \ldots < r_{k^*}$ with $r_0 = 1$ such that $g(r_k) = (-1)^{k+1}\|\tilde{g}\|_\infty$ for $k \in \{j^*, \ldots 0, \ldots k^*\}$. Then, for:*

$$4\pi \leq \lambda \leq 4\pi \left(r_{j^*}\left(\frac{1}{r_{j^*}} + \frac{1}{r_{j^*+1} - r_{j^*}}\right) + \sum_{k=j^*+1}^{k^*-1} r_k \left(\frac{1}{r_k - r_{k-1}} + \frac{1}{r_{k+1} - r_k}\right) + \frac{r_{k^*}}{r_{k^*} - r_{k^*-1}} \right), \tag{5.26}$$

the radial function f defined by $\tilde{f}(r) = \mathbb{1}_{[0,1]}(r) + \beta\frac{1}{r}\frac{\partial r\tilde{g}(r)}{\partial r}$ has a perfect decomposition (u, v) given by:

$$\begin{cases} \tilde{u}(r) &= \mathbb{1}_{[0,1]}(r) \\ \tilde{v}(r) &= \beta\frac{1}{r}\frac{\partial r\tilde{g}(r)}{\partial r} \end{cases}.$$

It is well-known that the G-norm favors oscillations, but to our knowledge, little is known about the influence of the frequency of the original pattern on the actual solution of the $TV + G$ decomposition. In the above example one may observe that the quicker the oscillations (i.e. in the above example the higher N or the larger $\frac{1}{r_k - r_{k-1}}$), the larger the interval of λ for which the decomposition is perfect. For λ above this interval, we do not know precisely what happens. It is likely that parts of the oscillations are sent to the cartoon part u, as it is observed in the numerical experiments of [Yin et al., 2007a] on one-dimensional signals.

Another interesting point about this toy example is that it shows that the value r_0 where the jump of the cartoon part lies is special. In fact, to build the above example, it is necessary that v should contain some texture around the jump at r_0. The next Section gives more detail about this and explains why the jumps of an image often appear in the v part using TV-G.

5.2.4 Shadow edges in the texture part

In [Duval et al., 2010], we reported a visual artifact that appears with the TV-G decomposition (see the visual experiments of Section 5.1.2), namely a sort of halo in the smooth areas and shadow edges in the texture part (at least more visibly than with the TVL1 model). We gave an elementary explanation to justify this phenomenon. We believe it is not related to a wrong choice of parameters but rather to an inherent property of the G-norm, which is the use of $\|.\|_\infty$ to measure the vectorfield.

Let us recall this argument in a one-dimensional setting (for the sake of simplicity). Notice however that the following argument could be adapted to a radial setting as above. Suppose for instance, that one wants to decompose a signal $f \in L^1 \cap L^2$, i.e. find a pair (u, v) that solves :

$$\min_{u+v=f} \int |Du| + \|v\|_G \tag{5.27}$$

where the functions are defined on $\mathbb{R}$. In dimension 1, the divergence operator reduces to the derivation:

$$\|v\|_G = \inf_{C\in\mathbb{R}} \sup_{t\in\mathbb{R}} \left| \int_{-\infty}^{t} v(s)ds + C \right|. \tag{5.28}$$

Let $V(t) = \int_{-\infty}^{t} v(s)ds$, V is continuous bounded (since $v \in L^1(\mathbb{R})$) and the optimal C is $C = -\frac{\sup V + \inf V}{2}$.

Now, consider a step edge, perturbed with some textures as in Figure 5.9 (a); for instance:

$$f(x) = \mathbb{1}_{(0,1)}(x) + \beta \sin(8p\pi x)\mathbb{1}_{\frac{1}{4}\leq|x|\leq\frac{3}{4}} \tag{5.29}$$

where $\beta > 0$, $p \in \mathbb{N}^*$. In that case $C = 0$ in (5.28). The ideal decomposition one would dream of is a perfect step $u(x) = \mathbb{1}_{(0,1)}(x)$ on the one hand, and the pure oscillation $v(x) = \beta \sin(8p\pi x)\mathbb{1}_{\frac{1}{4}\leq|x|\leq\frac{3}{4}}$ on the other hand (see Figure 5.9 (b) and (c)). The energy of the cartoon part is simply $\int |Du| = u(0+) - u(0-) + u(1-) - u(1+) = 2$, whereas the energy of the texture part is given by $\|v\|_G = \beta \int_{\frac{1}{4}}^{\frac{1}{4}+\frac{1}{8p}} \sin(8p\pi t)dt = \frac{\beta}{4p\pi}$. Yet, replacing u on $[-\frac{1}{4}, \frac{1}{4}]$ with any non-decreasing function u^* with the same limits at $\pm\frac{1}{4}$, say, a ramp $x \mapsto (\frac{1}{2} + \frac{1}{2\eta}x)\mathbb{1}_{[-\eta,\eta]}(x)$ as in Figure 5.9 (d)), one still gets the same cartoon energy $\int |Du^*| = 2$. As for the texture part, we should notice that one extra oscillation is added near the discontinuity of the original function f. But, precisely, the G-norm favors oscillations, so that this change in the texture part will not be penalized. Indeed:

$$\|v^*\|_G = \max\left(\frac{\beta}{4p\pi}, \int_{-\eta}^{0}(\frac{1}{2} + \frac{1}{2\eta}t)dt\right) = \frac{1}{4p\pi} \tag{5.30}$$

for η small enough ($0 \leq \eta \leq \min(\frac{1}{4}, \frac{\beta}{p\pi})$).

On the second discontinuity, a similar argument shows that one may also replace the jump with a ramp without changing the energy.

In a nutshell, we see that given any decomposition with sharp edges, there exists a decomposition with the same energy where shadows of edges appear in the texture part. So it is not surprising to see edges appear in the texture part of our experiments. This phenomenon may also explain the kind of halo that appears in smooth areas: replacing the ramp with a slow gradation, one may alter smooth parts without changing the energy. This is why in Figure 5.5 the caps in the texture part look so "inflated" compared to those of TVL1.

Notice that we do not claim that the ideal decomposition (u, v) of the function in (5.29) is indeed a solution of the TV-G problem. We simply observe that, *if it were the case*, another optimal decomposition would have shadow edges: one may object that depending on luck or on the choice of the algorithm, this artifact may not appear.

The following result suggests that there should *always* be some texture in the v part around the points where u has jumps: to explain the apparition of halos we used the example (5.29) for its simplicity, but in this particular case the proposed ideal decomposition $u = \mathbb{1}_{(0,1)}$, $v = f - u$ *cannot* be a solution of the TV-G problem. In such cases, *the artifact is bound to appear* near edges.

Proposition 5.2.12. *Let $f \in L^2(\mathbb{R}^2)$ be a radial function, $u = \mathbb{1}_{B(0,1)}$ and $v = f - \mathbb{1}_{B(0,1)}$ such that $v \neq 0$. Let $\epsilon > 0$ and $\Omega_{(-\varepsilon,\varepsilon)} = \{x \in \mathbb{R}^2,\ 1-\varepsilon < |x| < 1+\epsilon\}$. If $v_{|\Omega_{(-\varepsilon,\varepsilon)}} = 0$, then the decomposition (u, v) is not optimal.*

Proof. Observe that we may assume that z and p are radial in the Euler-Lagrange equation (5.14). Indeed, if $z = z_r e_r + z_\theta e_\theta$, we replace z with z^* such that $\tilde{z_r}^* = \frac{1}{2\pi}\int_0^{2\pi} \tilde{z_r}(r,\theta)d\theta$, $z_\theta^* = 0$, so that $\|z^*\|_\infty \leq \|z\|_\infty \leq 1$, $z^* \cdot e_r = 1$ for $r = 1$. Since $\operatorname{div} z = \frac{1}{r}\frac{\partial r z_r}{\partial r} + \frac{1}{r}\frac{\partial z_\theta}{\partial \theta}$ in the distribution sense, this leads to $\operatorname{div} z^* = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{div} z d\theta$. On the other hand, we replace p with p^* such that $\tilde{p}^*(r) = \frac{1}{2\pi}\int_0^{2\pi} p(r,\theta)d\theta$. Since $\int |Dp^*| \leq \int |Dp|$, we see that z^* and p^* satisfy Equations (5.12), (5.13) and (5.14). In the following, we assume that z and p are radial. As noticed above, we may also assume that g is radial.

If $v = \operatorname{div} g = 0$ in $\Omega_{(-\varepsilon,\varepsilon)}$, then $\tilde{g}(r) = \frac{C}{r}$ in $\Omega_{(-\varepsilon,\varepsilon)}$, so that:

$$\sup_{x\in\Omega_{(-\frac{\varepsilon}{2},\frac{\varepsilon}{2})}} |g(x)| < \sup_{x\in\Omega_{(-\varepsilon,\varepsilon)}} |g(x)| \tag{5.31}$$

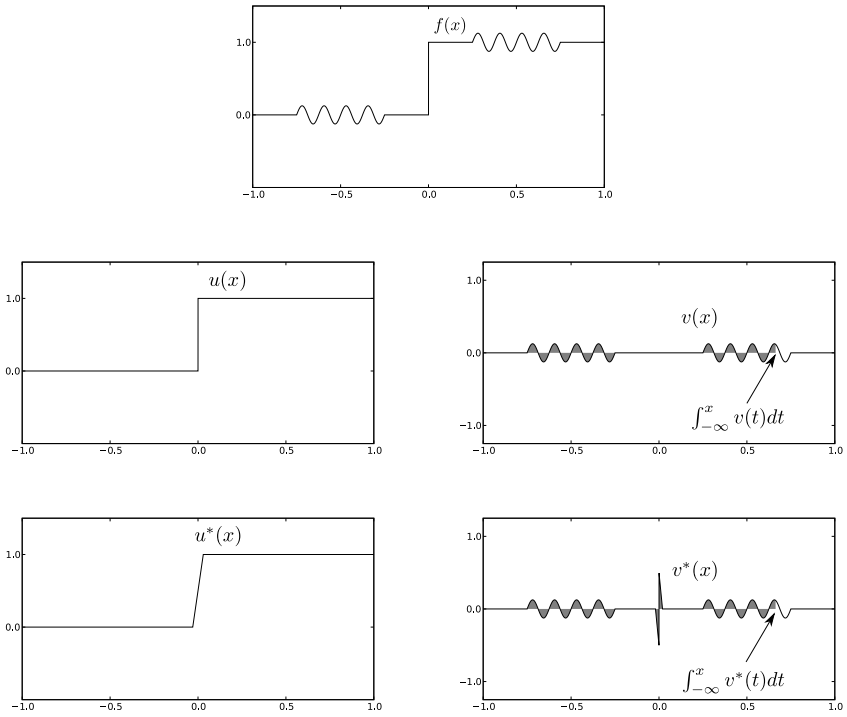


Figure 5.9 – *Decomposition of a 1-dimensional signal. Top: Original signal f. Middle row: ideal decomposition (left: cartoon part, right texture part). Bottom row: another decomposition. The total variation of the cartoon part is equal to 1 in both cases, and by (5.28), the G-norm of each texture part is equal to the maximum area of its "bumps". Therefore both decompositions have the same energy.*

and thus $\left((\frac{Dp}{|Dp|}\cdot g) + \|g\|_\infty\right) > 0$, $|Dp|$-a.e. in $\Omega_{(-\frac{\varepsilon}{2},\frac{\varepsilon}{2})}$. By (5.15), we must have $|Dp|(\Omega_{(-\frac{\varepsilon}{2},\frac{\varepsilon}{2})}) = 0$, so that p is constant on $\Omega_{(-\frac{\varepsilon}{2},\frac{\varepsilon}{2})}$.

Then by (5.14), we see that $\tilde{z}(r) = -\frac{\lambda p}{2} r + \frac{C'}{r}$ in $\Omega_{(-\frac{\varepsilon}{2},\frac{\varepsilon}{2})}$. Since $\tilde{z}(1) = 1$, we have $C' = 1 + \frac{\lambda p}{2}$. Since $|z| \leq 1$, we must have $\tilde{z}'(1) = -\frac{\lambda p}{2} - \left(1 + \frac{\lambda p}{2}\right) = 0$, therefore $p = \frac{-1}{\lambda}$. As a consequence, $\tilde{z}(r) = \frac{1}{2}\left(r + \frac{1}{r}\right) > 1$ for $r \neq 1$. Contradiction. □

The rest of this section is devoted to the generalization of Proposition 5.2.12 to characteristic functions of convex sets. Let $\Omega \subset \mathbb{R}^2$ be a $C^{1,1}$ convex body. For $\varepsilon > 0$, we denote by $\Omega_\varepsilon = \Omega \oplus \varepsilon B(0,1)$ the dilation of C with radius ε, and $\Omega_{-\varepsilon} = \Omega \ominus \varepsilon B(0,1) := \{x \in \Omega,\ x + B(0,1) \subset \Omega\}$ its erosion. We also write $\Omega_{(-\varepsilon,\varepsilon)} := \text{ int } (\Omega_\varepsilon \setminus \Omega_{-\varepsilon})$.

Lemma 5.2.13. *Let $f \in L^2(\mathbb{R}^2)$, $u = \mathbb{1}_\Omega$ and $v = f - \mathbb{1}_\Omega$ such that $v \neq 0$. Assume that the pair (u, v) is optimal, and let $z \in X_\infty, p \in BV(\mathbb{R}^2)$ be as in the Euler-Lagrange Equation (Proposition 5.2.6). Then $|Dp|(\Omega_{(-\varepsilon,\varepsilon)}) > 0$.*

Proof. By contradiction, let us assume that $|Dp|(\Omega_{(-\varepsilon,\varepsilon)}) = 0$, so that $p(x) = p_0 \in \mathbb{R}$ for a.e. $x \in \Omega_{(-\varepsilon,\varepsilon)}$. The area and perimeter of the eroded and dilated Ω are given by Steiner's formula:

$$\forall r > 0, \quad \begin{cases} |\Omega_r| &= |\Omega| + r\text{Per }\Omega + \pi r^2 \\ \text{Per }\Omega &= \text{Per }\Omega + 2\pi r \end{cases}. \tag{5.32}$$

$$\text{For } \frac{1}{r} \geq \operatorname*{ess\,sup}_{x\in\partial\Omega} \kappa_\Omega(x), \quad \begin{cases} |\Omega| &= |\Omega_{-r}| + r\text{Per }\Omega_{-r} + \pi r^2 \\ \text{Per }\Omega &= \text{Per }\Omega_{-r} + 2\pi r \end{cases}. \tag{5.33}$$

Let $0 < r < \varepsilon$.

- We apply the Gauss-Green theorem[2] on int $(\Omega_r \setminus \Omega)$:

$$\begin{aligned}\int_{\partial\Omega_r} z\cdot\nu d\mathcal{H}^1 &= \int_{\partial\Omega} z\cdot\nu d\mathcal{H}^1 + \int_{\Omega_r\setminus\Omega} \text{div } z\\ &= \text{Per }\Omega - \lambda p_0\left(r\text{Per }\Omega + \pi r^2\right)\end{aligned}$$

  Since $\|z\|_\infty \leq 1$, we have $\int_{\partial\Omega_r} z\cdot\nu d\mathcal{H}^1 \leq \mathcal{H}^1(\partial\Omega_r) = \text{Per }\Omega + 2\pi r$, so that $-\lambda p \leq \frac{2\pi}{\text{Per }\Omega+\pi r}$.

- Similarly, we have:

$$\begin{aligned}\int_{\partial\Omega} z\cdot\nu d\mathcal{H}^1 &= \int_{\partial\Omega_{-r}} z\cdot\nu d\mathcal{H}^1 + \int_{\Omega\setminus\Omega_{-r}} \text{div } z,\\ \text{i.e. Per }\Omega &= \int_{\partial\Omega_{-r}} z\cdot\nu d\mathcal{H}^1 - \lambda p_0\left(r\text{Per }\Omega - \pi r^2\right).\end{aligned}$$

  Again, $\int_{\partial\Omega_{-r}} z\cdot\nu d\mathcal{H}^1 \leq \mathcal{H}^1(\partial\Omega_{-r}) = \text{Per }\Omega - 2\pi r$, so that $-\lambda p_0 \geq \frac{2\pi}{\text{Per }\Omega-\pi r}$.

Eventually, we see that for $r > 0$ small enough:

$$\frac{2\pi}{\text{Per }\Omega - \pi r} \leq -\lambda p_0 \leq \frac{2\pi}{\text{Per }\Omega + \pi r}, \tag{5.34}$$

which is absurd. Therefore we must have $|Dp|(\Omega_{(-\varepsilon,\varepsilon)}) > 0$. □

Proposition 5.2.14. *Let $\Omega \subset \mathbb{R}^2$ be a $C^{1,1}$ convex body, $f \in L^2(\mathbb{R}^2)\setminus\{\mathbb{1}_\Omega\}$, $u = \mathbb{1}_\Omega$ and $v = f - \mathbb{1}_\Omega$. Assume that* Spt $v \subset \overline{\Omega}$ *and that the decomposition (u, v) is optimal. Then* Spt $v \cap \partial\Omega \neq \emptyset$.

Proof. By contradiction, assume that $d(\text{Spt } v, \partial\Omega) > 0$, so that there exists $\varepsilon > 0$ such that Spt $v \subset \Omega_{-\varepsilon}$.

As (u, v) is optimal, the Euler-Lagrange equation (5.14) holds.

Since $\frac{\int_{\mathbb{R}^2} pv}{\int |Dp|} = \|v\|_G = \sup_{w\in \text{BV}(\mathbb{R}^2)\setminus 0} \frac{\int_{\mathbb{R}^2} wv}{\int |Dw|}$, p is a non-trivial minimizer of:

$$G(w) := \int |Dw| - \int_{\mathbb{R}^2} \frac{v}{\|v\|_G} w, \tag{5.35}$$

which means that p is a solution of the prescribed mean curvature problem associated to $-\frac{v}{\|v\|_G}$. Now, let $E_t = \{p \geq t\}$ for $t \geq 0$ and $E_t = \{p < t\}$ for $t < 0$. Since $\Omega_{-\varepsilon}$ is convex:

$$\text{Per }\left(E_t \cap \Omega_{-\varepsilon}\right) \leq \text{Per } E_t, \tag{5.36}$$

with strict inequality if $|E_t \setminus \Omega_{-\varepsilon}| > 0$.

Now, we replace p with $p^* = p\mathbb{1}_{\Omega_{-\varepsilon}}$, so that $\{p^* \geq t\} = E_t \cap \Omega_{-\varepsilon}$ for $t \geq 0$, and $\{p^* < t\} = E_t \cap \Omega_{-\varepsilon}$ for $t < 0$.

Since $|Dp|(\Omega_{(-\varepsilon,\varepsilon)}) > 0$, the coarea formula implies that the set $B = \{t > 0,\ |E_t \cap \Omega_{(-\varepsilon,\varepsilon)}| > 0\}$ has positive Lebesgue measure. Therefore, inequality (5.36) is strict for all $t \in B$ so that $\int |Dp^*| < \int |Dp|$. Since $\int p^* v = \int pv$, we see that $G(p^*) < G(p)$, which is a contradiction. □

Proposition 5.2.15. *Let $\Omega \subset \mathbb{R}^2$ be a $C^{1,1}$ convex body, $f \in L^2(\mathbb{R}^2)\setminus\{\mathbb{1}_\Omega\}$, $u = \mathbb{1}_\Omega$ and $v = f - \mathbb{1}_\Omega$. Assume that* Spt $v \subset D_0 \cup D_1$, *where $D_0 \subset \Omega$, $D_1 \subset \mathbb{R}^2 \setminus \Omega$ are such that $\overline{D_0} \cap \overline{D_1} = \emptyset$, $\overline{D_1}$ is strictly convex and*

$$\forall U \subset \mathbb{R}^2 \text{ such that } U \supset D_0 \cup D_1, \quad \text{Per } U \leq \text{Per } D_0 + \text{Per } D_1. \tag{5.37}$$

with strict inequality when $|U \setminus (D_0 \cup D_1)| > 0$.

If (u, v) is an optimal decomposition for the TV-G model, then Spt $v \cap \partial\Omega \neq \emptyset$.

[2] In each integral below, ν points toward the exterior of Ω.

Proof. We proceed as in Proposition 5.2.14 by contradiction by assuming that $d(\text{Spt } v, \partial\Omega) > 0$. First, let us observe that we may replace D_1 with any convex set which is contained in D_1, without losing the property (5.37). Indeed, let $D_2 \subset D_1$ be a convex set. By submodularity of the perimeter, for all $U \subset \mathbb{R}^2$ such that $D_0 \cup D_2 \subset U$:

$$\text{Per } (U \cup D_1) + \underbrace{\text{Per } (U \cap D_1)}_{\geq \text{Per } D_2} \leq \text{Per } U + \text{Per } D_1 \tag{5.38}$$

$$\text{so that Per } D_0 + \text{Per } D_1 + \text{Per } D_2 \leq \text{Per } U + \text{Per } D_1, \tag{5.39}$$

$$\text{Per } D_0 + \text{Per } D_2 \leq \text{Per } U. \tag{5.40}$$

Similarly we may replace D_0 with any convex set which is contained in D_0.

As a consequence, since by assumption $d(\text{Spt } v, \partial\Omega) > 0$, we may assume that $d(D_0, \partial\Omega) > \epsilon$ and $d(D_1, \partial\Omega) > \epsilon$ for some $\epsilon > 0$. Let p be as in the proof of Proposition 5.2.14, and consider $E_t = \{p \geq t\}$ for $t \geq 0$ and $E_t = \{p < t\}$ for $t < 0$. By the submodularity of the perimeter:

$$\text{Per } E_t \cap (D_0 \cup D_1) \leq \text{Per } E_t + \underbrace{\text{Per } D_0 \cup D_1 - \text{Per } (E_t \cup D_0 \cup D_1)}_{\leq 0} \tag{5.41}$$

with strict inequality if $|E_t \setminus (D_0 \cup D_1)| > 0$. Since $|Dp|(\Omega_{(-\varepsilon,\varepsilon)}) > 0$, the coarea formula implies that the set $B = \{t > 0,\ |E_t \cap \Omega_{(-\varepsilon,\varepsilon)}| > 0\}$ has positive Lebesgue measure. Therefore, inequality (5.41) is strict for all $t \in B$ so that if $p^* = p\mathbb{1}_{D_0 \cup D_1}$, $\int |Dp^*| < \int |Dp|$. Since $\int p^* v = \int pv$, we see that $G(p^*) < G(p)$, which is a contradiction. □

We conjecture that, in the above propositions, the assumptions on the geometry of Spt v can be removed. The fact that v vanishes in a neighborhood of $\partial\Omega$ implies that the level lines of p are line segments. On the other hand, since $-\lambda p = \text{div } z$, the value of $-\lambda p$ on $\partial\Omega$ should be equal, at least formally, to the curvature of $\partial\Omega$. Those two behaviors seem difficult to conciliate, but we have not managed to reach a contradiction.

To sum up, our interpretation is that for the TV-G model to provide good decompositions, the original image should have textures in a neighborhood of the edges (as in Figures 5.3 and 5.4). Otherwise, the texture part is modified so that it is not empty in a neighborhood of the edges, and a shadow of the edges or a gradation appears in the texture part.

5.3 The TVL1 model

Several authors [Yin et al., 2007a, Haddad, 2007, Aujol et al., 2006] have noticed the ability of the TVL1 model in the problem of cartoon-texture separation. Relying mainly on empirical observations, they argued that this model could compare favorably with the TV-G model by discriminating objects in function of their scales. While the last section detailed the main properties of the TV-G model for decompositions we now turn to the TVL1 model and we try to understand the results of Section 5.1.2 in light of the results of Chapter 3.

5.3.1 Toy example

Like the TV-G model, the $TVL1$ model is able to decompose some functions of the form "indicator + sine", as the following proposition illustrates.

Proposition 5.3.1. *Let $N \in \mathbb{N}^*$. For $2 \leq \lambda \leq 2N$, the function f defined by*

$$\tilde{f}(r) = \mathbb{1}_{[0,1]}(r) - \beta \sin(N\pi r)\mathbb{1}_{[2,3]}(r), \tag{5.42}$$

has an optimal decomposition given by:

$$u(x) = \mathbb{1}_{B(0,1)}(x),$$
$$\tilde{v}(r) = -\beta \sin(N\pi r)\mathbb{1}_{[2,3]}(r).$$

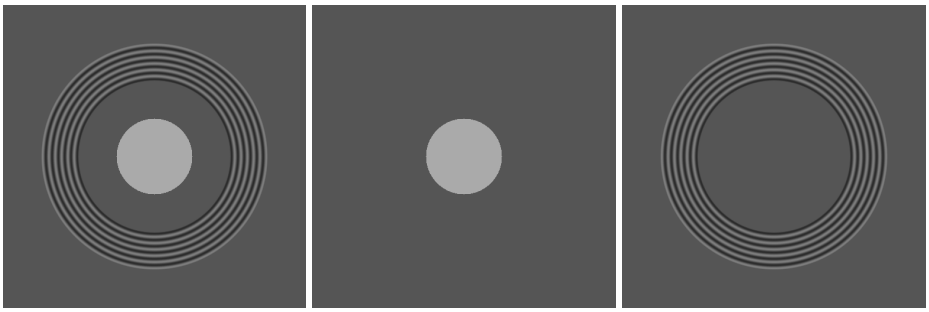


Figure 5.10 – *$TV - L^1$ decomposition predicted by Proposition 5.3.1: original image f (left), cartoon part $u = \mathbb{1}_{B(0,1)}$ (middle), texture part $v(x) = -\sin(N\pi r)\mathbb{1}_{[2,3]}(r)$ (right).*

Proof. Let $k^* \in \{0, 1 \dots, N\}$ and $r_k = 2 + \frac{k}{N}$, for $k \in \{0, 1 \dots, k^*\}$. The Euler-Lagrange equation (2.2) amounts to finding a vectorfield $z \in X_\infty$, a function $p \in L^\infty$ such that:

$$\begin{cases} \tilde{z}(1) = 1, \\ |\tilde{z}(r)| \leq 1 \text{ for } r \geq 0, \\ p_k := \tilde{p}(r) = (-1)^{k+1} \text{ for } r \in [r_k, r_{k+1}) \text{ and } k \in \{0, \dots, N-1\}, \\ |\tilde{p}(r)| \leq 1 \text{ for } r \geq 0 \end{cases} \tag{5.43}$$

$$\text{and div } z + \lambda p = 0. \tag{5.44}$$

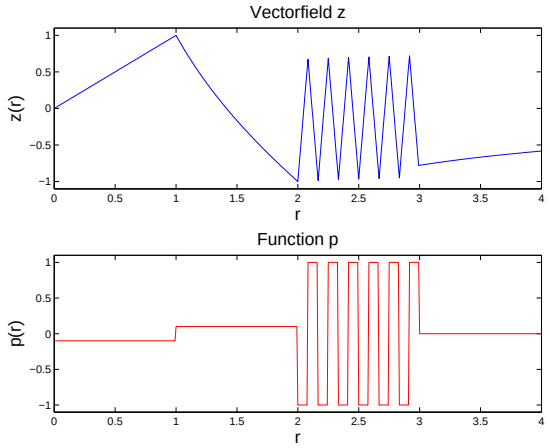


Figure 5.11 – *Construction of the vectorfield z and function p in Proposition 5.3.1.*

Notice that if $\tilde{p}$ is constant on $[\rho, r)$, the integration of div $z = -\lambda p$ leads to:

$$\tilde{z}(r) = \frac{1}{r}\left(\rho\tilde{z}(\rho) - \frac{\lambda\tilde{p}}{2}(r^2 - \rho^2)\right) \tag{5.45}$$

- For $r \in [0, 1)$, we define $\tilde{p}(r) = -\frac{2}{\lambda}$, so that $\tilde{z}(r) = r$.

- For $r \in [1,2)$, we define $\tilde{p}(r) = \frac{2}{\lambda}$, so that $\tilde{z}(r) = \frac{1}{r}(2 - r^2)$.
- For $r \in [r_k, r_{k+1})$, $k \in \{0, \dots, N-1\}$, we define by induction $\tilde{p}(r) = (-1)^{k+1}$ and:

$$\tilde{z}(r) = \frac{1}{r}\left(r_k z_k + \frac{\lambda(-1)^k}{2}(r^2 - r_k^2)\right). \tag{5.46}$$

- For $r \in [r_N, +\infty)$, we set $\tilde{p}(r) = 0$, so that $\tilde{z}(r) = z_N \frac{r_N}{r}$, where z_N is defined below.

For $\lambda \geq 2$, the derivative of (5.46) has the sign of $(-1)^k$. By monotonicity, it is therefore sufficient to check that $|\tilde{z}(r_k)| \leq 1$ for all $k \in \{0, \dots, N\}$ so as to ensure that $|\tilde{z}(r)| \leq 1$ for all $r \geq 0$. If we set $V_k := z_k r_k$, we see with Eq. (5.46) that V_k satisfies the simple recursion:

$$V_{k+1} = V_k + \frac{\lambda(-1)^k}{2}(r_{k+1}^2 - r_k^2). \tag{5.47}$$

Therefore:

$$\underbrace{\sum_{j=0}^{k-1}(V_{j+1} - V_j)}_{V_k - V_0} = \frac{\lambda}{2}\sum_{j=0}^{k-1}(-1)^j\left(r_{j+1}^2 - r_j^2\right) \tag{5.48}$$

$$= \frac{\lambda}{2}\sum_{j=0}^{k-1}(-1)^j\left(\frac{1}{N^2} + \frac{2}{N}\left(2 + \frac{j}{N}\right)\right) \tag{5.49}$$

$$= \frac{\lambda}{2}\left[\left(\frac{1-(-1)^k}{2}\right)\left(\frac{1}{N^2} + \frac{4}{N}\right) - \frac{1}{2N^2}\left((-1)^k(2k-1)+1\right)\right], \tag{5.50}$$

where we have used the identity: $\sum_{j=0}^{k-1} j(-1)^{j-1} = \frac{1}{4}\left((-1)^k(2k-1)+1\right)$. As a consequence, we eventually get:

$$V_k = V_0 + \frac{\lambda}{2}\left[(-1)^{k+1}\left(\frac{2}{N} + \frac{k}{N^2}\right) + \frac{2}{N}\right]. \tag{5.51}$$

The condition $|z| \leq 1$ is equivalent to:

$$-1 \leq \frac{1}{r_k}\left(-2 + \frac{\lambda}{2}\left[(-1)^{k+1}\left(\frac{2}{N} + \frac{k}{N^2}\right) + \frac{2}{N}\right]\right) \leq 1. \tag{5.52}$$

Considering the parity of k, one sees that this condition is in fact equivalent to $\lambda \leq 2N$ since $r_k = 2 + \frac{k}{N}$. By construction, z and p satisfy all the conditions in (5.44). □

Again, we see that the higher the frequency N, the larger the interval of λ for which the decomposition is optimal.

5.3.2 No shadow edges nor halos

From the above example, let us notice that the model does not produce shadow edges or extra-oscillations: it is also clear that the argument in Section 5.2.4 does not hold with the TVL1 model, since any extra-oscillation added in the texture part has a cost in terms of energy. Moreover, there is no need for the texture part to be filled near edges of the cartoon part. From the Figures 5.2 and 5.5, we can see that the texture part does not look inflated compared to the one of TV-G, and that the edges are less visible.

As a consequence, we see that the TVL1 model does not suffer from some of the drawbacks of TV-G. Yet, we still have to explain why it performs so well in cartoon-texture decomposition: we believe it is related to the connection between the TVL1 scale-space and granulometries.

5.3.3 Granulometries

Granulometries were introduced with mathematical morphology by Matheron and Serra to measure the petrographic distribution of oolites and chlorite cement in Lorraine iron ores (see [Matheron and Serra, 2002]). To this end, they have introduced operations like erosions, dilations and openings and computed variograms called granulometries that allow to discriminate objects by their size (see [Matheron, 1964, Haas et al., 1967]). Since then, the use of mathematical morphology and granulometries has spread to various areas of image processing especially in the study of textures in general [Soille, 2003].

Definition 5.3.2. A granulometry is a family of openings $\{\gamma_\mu\}$ depending on a positive parameter μ, that are decreasing functions with respect to μ: $\mu_2 \geq \mu_1 > 0 \Rightarrow \gamma_{\mu_2} \leq \gamma_{\mu_1}$. The cumulative size distribution of a set F is $\mu \mapsto 1 - \frac{|\gamma_\mu F|}{|F|}$. Its derivative is called the granulometric spectrum of F: $-\frac{1}{|F|}\frac{d}{d\mu}|\gamma_\mu F|$.

Figure 5.12 shows the result of openings with increasing radii on a Brodatz texture. The larger the radius, the larger the patterns that are removed. Therefore it is natural to think that granulometries indicate the proportion of objects at each scale within the image. This behavior is illustrated in Figure 5.13 for three Brodatz textures. The profile of the granulometric spectrum gives important information about the characteristic scales of textures: a texture whose granulometric spectrum is concentrated on small values of μ is "rough" (that is, composed of tiny objects) whereas if the granulometric spectrum is concentrated on high values of μ the texture is considered smoother (made with objects at larger scale) (see for instance [Soille, 2003]).

The property of granulometries, which is to progressively destroy tiny objects, has proved its practical efficiency for the discrimination of textures, since granulometric spectra have been used for years in texture analysis (see for instance [Serra, 1988, Vanrell and Vitria, 1993]). Leaving aside the condition on the ratio perimeter/area in Theorem 3.3.9, we notice that the TVL1 model provides a granulometry on the family of convex sets (since $\gamma_\mu F = F^\mu$ if $\mu \leq \frac{1}{\lambda^*}$, $\emptyset$ otherwise). Eventually, the tresholding on the ratio perimeter/area, which gives another information about the scale of the object, reinforces this property.

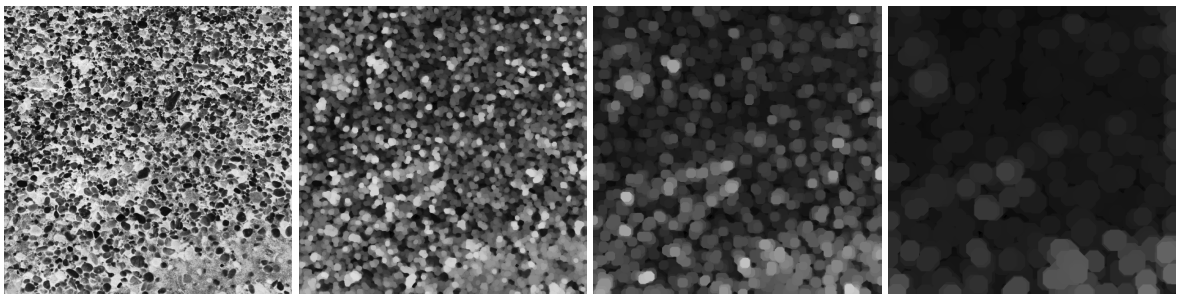


Figure 5.12 – *From left to right: original image, openings of radius 5, 10, 20. As the radius increases, larger patterns are removed.*

In a nutshell, TVL1 acts on images as a granulometry (which kills small details and high curvatures) followed by a thresholding on the ratio perimeter/area. When the fidelity parameter λ varies, objects are altered in function of their granulometry and then vanish when their ratio perimeter/area is too small. The v part contains object with fine granulometry, whereas u contains objects with coarse granulometry (which is what we expect of a cartoon-type image).

It is interesting that, aside from the many works on cartoon-texture decomposition relying on the idea of favoring oscillations in the texture term [Meyer, 2001, Vese and Osher, 2003, Osher et al., 2003, Aujol et al., 2005, Aujol and Chambolle, 2005, Aujol et al., 2006], a norm that does not even notice oscillations provides similar results.

To conclude, let us mention that the granulometric behavior of TVL1 is so striking that granulometries have been independently reinvented in the context of TVL1 by Vixie *et al.* [Vixie et al., 2010]. In

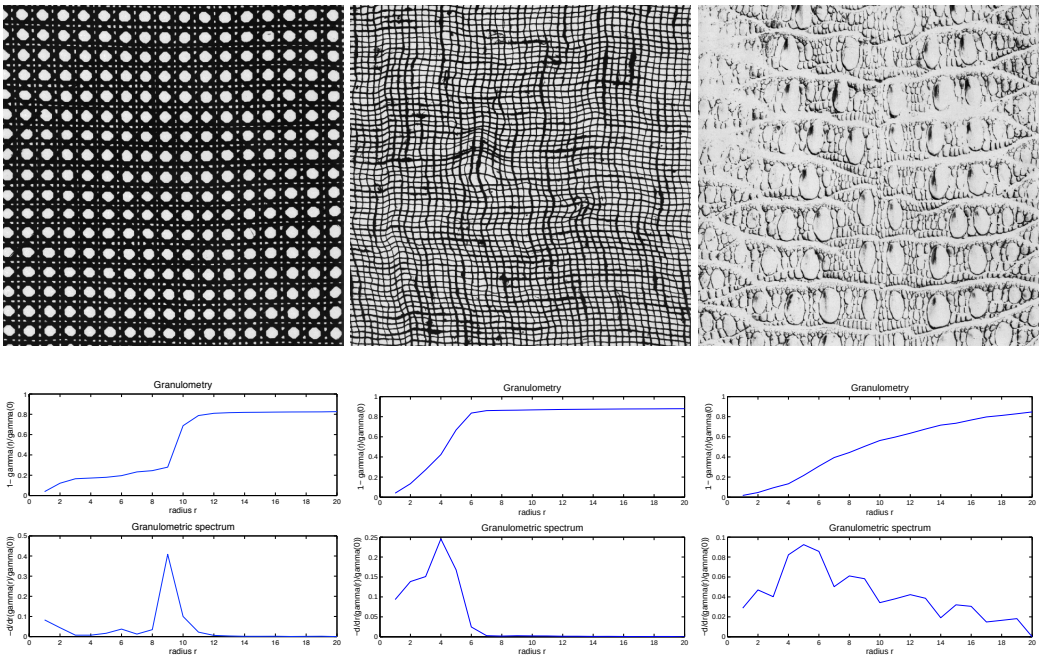


Figure 5.13 – *Examples of granulometries. The granulometric spectrum indicates the characteristic scales of the image.*

that paper, the authors aim at classifying shapes. In order to discriminate between shapes or images, the authors consider (among other signatures) the function:

$$s_F : \lambda \mapsto \frac{|U(\lambda)\Delta F|}{|F|} \tag{5.53}$$

where $U(\lambda)$ is a solution to the (geometric) TVL1 problem associated to $F \subset \mathbb{R}^2$:

$$\min_{U \subset \mathbb{R}^2} \text{Per}\, U + \lambda |U \Delta F|.$$

Then, they plot $\lambda \mapsto s_F(\frac{1}{\lambda})$ and they show that this curve gives information about the scales of the object. In particular, they notice the importance of its jumps and they propose to examine the derivative $\lambda \mapsto \frac{d}{d\lambda} s_F(\frac{1}{\lambda})$ (see Figure 5.15). This is precisely the philosophy of granulometries and granulometric spectra used in [Maragos, 1989] for shape analysis, and Corollary 3.4.3 shows that the examined quantities are in fact almost the same.

(a) Fork 4 (b) Fork 16 (c) Spoon 10

Figure 5.14 – *Example of shapes considered in [Vixie et al., 2010] for classification.*

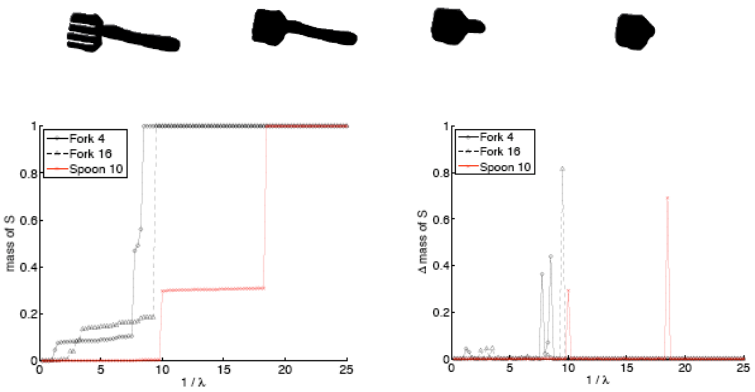


Figure 5.15 – *This Figure is extracted from [Vixie et al., 2010]. Top: Evolution the first fork as λ decreases. Bottom: Cumulated histograms $s_F(\frac{1}{\lambda})$ for different shapes F (left) and their derivatives (right), the "granulometric" spectra $\frac{d}{d\lambda}s_F(\frac{1}{\lambda})$.*

Chapter 6

Spatially adaptive TVL1

In this chapter we focus on a variant of the TVL1 model of the form:

$$\inf_u \int_{\mathbb{R}^2} |Du| + \int_{\mathbb{R}^2} |u-f|\lambda(x)dx, \tag{6.1}$$

in which the fidelity parameter λ can be locally adapted to the content of the image. The motivation for this model comes from two areas. First, following [Gilboa et al., 2006] for the ROF model, one may want to use a higher fidelity parameter in textured regions so as to preserve fine details while smoothing homogeneous regions. Second, Nikolova has shown that the TVL1 model was well-suited to the denoising of images degraded by impulse noise or salt and pepper noise (see [Nikolova, 2002] for a theoretical analysis, [Nikolova, 2004a] for impressive numerical experiments with both types of noise). Recent algorithms like [Chan et al., 2004, Chan et al., 2005, Cai et al., 2010] for the removal of impulse noise consist in first detecting the location of the impulses, then modifying an energy which is close to a modified TVL1 energy that smoothes only the pixels detected as noisy. A way to interpret this model is to consider this as a spatially adaptive TVL1 where $\lambda(x) = 0$ for all detected noisy pixels x. In our framework, we assume for technical reasons that $\lambda(x) \geq c > 0$ for some constant c. When $c \to 0$, the model is close to the one used in [Chan et al., 2004].

This chapter is composed of a theoretical part and an experimental part. The first part shows that there are (locally) two options when using adaptive TVL1 and defining $\lambda(x)$: either one bounds $\lambda(x)$ and gets regular solutions ($\lambda(x)$ giving a bound on the curvature), or one lets λ unbounded and then any shape can be preserved by TVL1, including those with corners. The second part shows numerical experiments with impulse noise.

6.1 Theoretical analysis of the model

In this section, we give elementary properties which show that the model does indeed what it is expected to do: as before, the functional problem is equivalent to a family of geometric problems on the level sets of the image, and the value λ controls the curvature of the level lines. Moreover, setting very large values for λ allows to preserve objects of the original image, even with corners.

6.1.1 Definition of the model

It will sometimes be more convenient to consider the weight $\lambda \in L^1_{loc}$ as the density of a measure Λ which is absolutely continuous with respect to the Lebesgue measure: $\lambda(x)dx = d\Lambda$. We will assume that there exists $c > 0$ such that:

$$\lambda(x) \geq c > 0 \text{ for a.e. } x \in \mathbb{R}^2. \tag{6.2}$$

Notice that under these assumptions, the space of Λ-integrable functions, that we denote by L^p_Λ, enjoys the usual properties of Lebesgue spaces (duality L^p_Λ-$L^{p'}_\Lambda$, density of smooth functions, etc.).

When λ is constant, Problem (6.1) amounts to the regular TVL1 model, in which case we shall say that the problem is stationary .

Proposition 6.1.1 (Existence)**.** *Assume condition (6.2) is true. Then there exists (at least) one solution to Problem (6.1). Let us define:*

$$T : L^1_\lambda \to \mathcal{P}(L^1_\lambda) \tag{6.3}$$

the operator which maps f to the set of solutions $Tf = \{u \in L^1_\lambda, u \text{ is a solution of } (6.1)\}$. Then the set Tf is convex and closed in L^1_λ.

Proof. The proof relies on the direct method of the calculus of variations. Let u_n be a minimizing sequence. Then,

$$\min(c,1)\left(\int |Du_n| + \int_{\mathbb{R}^2} |f(x) - u_n(x)|dx\right) \leq \int |Du_n| + \int_{\mathbb{R}^2} |f - u_n| d\Lambda \leq C \tag{6.4}$$

thus we can extract a subsequence (still denoted as u_n) which converges to $u \in L^1$ in L^1 and $Du_n \rightharpoonup Du$ in the sense of measure convergence. Up to another extraction, we can even assume that the sequence $u_n(x)$ converges to $u(x)$ for almost every $x \in \mathbb{R}^2$. But then by semicontinuity and Fatou's lemma:

$$\liminf_n \left(\int |Du_n| + \int_{\mathbb{R}^2} |f - u_n| d\Lambda\right) \geq \int |Du| + \int_{\mathbb{R}^2} |f - u| d\Lambda \tag{6.5}$$

and u is a solution of Problem 6.1. □

Proposition 6.1.2. *The energy reformulates:*

$$\mathbb{E}(u) = \int_{-\infty}^{+\infty} \text{Per}\,\{x, u(x) > t\} + \Lambda\left(\{x, u(x) > t\}\Delta|\{x, f(x) > t\}\right) dt. \tag{6.6}$$

We are thus led to study the geometric problem:

$$\inf_{U \subset \mathbb{R}^2} \mathbb{E}_G(U) := \text{Per}\, U + \Lambda(U \Delta F) \tag{6.7}$$

for $F \subset \mathbb{R}^2$. To insist on the dependence on the data F, we shall sometimes write $\mathbb{E}_G(U, F)$.

Proof. Notice that:

$$\begin{aligned}
\int_{\mathbb{R}^2} |u - f| d\Lambda &= \int_{\mathbb{R}^2} \left(\int_{f(x)}^{u(x)} dt\right) \mathbb{1}_{f(x)<u(x)} + \left(\int_{u(x)}^{f(x)} dt\right) \mathbb{1}_{u(x)<f(x)} d\Lambda \\
&= \int_{\mathbb{R}^2} \int_{-\infty}^{+\infty} \left(\mathbb{1}_{f(x)\leq t<u(x)} + \mathbb{1}_{u(x)\leq t<f(x)}\right) dt\, d\Lambda
\end{aligned}$$

with

$$\{x, (u(x) > t \geq f(x)) \text{ or } (u(x) \leq t < f(x))\} = \{x, u(x) > t\}\Delta\{x, f(x) > t\}. \tag{6.8}$$

By Fubini's theorem:

$$\begin{aligned}
\int_{\mathbb{R}^2} |f - u| d\Lambda &= \int_{-\infty}^{+\infty} \int_{\mathbb{R}^2} \mathbb{1}_{\{x,u(x)>t\}\Delta\{x,f(x)>t\}} d\Lambda dt \\
&= \int_{-\infty}^{+\infty} \int_{\mathbb{R}^2} \Lambda\left(\{x, u(x) > t\}\Delta\{x, f(x) > t\}\right) dt.
\end{aligned}$$

□

6.1.2 Monotonicity

In this section we adapt the monotonicity result (Theorem 3.1.3). We mainly follow the steps of [Yin et al., 2007b] for the stationary case, though the proofs we give are slightly different.

Lemma 6.1.3. *Let $F \subset G \subset \mathbb{R}^2$ with finite Λ-measure, let U_1, U_2 be corresponding solutions, and $U_\wedge = U_1 \cap U_2$. We have:*

$$\mathbb{E}_G(U_1, F) - \mathbb{E}_G(U_\wedge, F) \geq \mathbb{E}_G(U_1, G) - \mathbb{E}_G(U_\wedge, G) \tag{6.9}$$

where $\mathbb{E}_G(U, F)$ is the geometric energy of U (see Equation (6.7)) with data F.

Proof. Since the perimeter parts of the energies are equal, it is sufficient to prove the following inequality for the fidelity terms:

$$\Lambda(U_1 \Delta F) - \Lambda((U_1 \cap U_2)\Delta F) \geq \Lambda(U_1 \Delta G) - \Lambda((U_1 \cap U_2)\Delta G). \tag{6.10}$$

Let $U \subset \mathbb{R}^2$ with finite Λ-measure. Since G and F have finite measure we can write:

$$\begin{aligned}
\Lambda(U\Delta G) &= \Lambda(U \cup G) - \Lambda(U \cap G) \\
&= \Lambda(U \cup F) + \Lambda((G \setminus (F \cup U)) - (\Lambda(U \cap F) + \Lambda(U \cap (G \setminus F)) \\
&= \Lambda(U\Delta F) + \Lambda((G \setminus F) \cap U^C) - \Lambda(U \cap (G \setminus F))
\end{aligned}$$

Therefore:

$$\Lambda(U\Delta F) - \Lambda(U\Delta G) = \Lambda(U \cap (G \setminus F)) - \Lambda((G \setminus F) \cap U^C) \tag{6.11}$$

Notice that the right side of the last equation is a non-decreasing function of U. Therefore its value for $U = U_1$ is larger than for $U = U_1 \cap U_2$, hence the result. □

Lemma 6.1.4. *The mapping $U \mapsto \Lambda(U\Delta F)$ is submodular, i.e.:*

$$\Lambda\left((U \cap V)\Delta F\right) + \Lambda\left((U \cup V)\Delta F\right) \leq \Lambda(U\Delta F) + \Lambda(V\Delta F). \tag{6.12}$$

Proof. Let us write: $u = \mathbb{1}_U$, $v = \mathbb{1}_V$, $f = \mathbb{1}_F$ and notice that $u^2 = u$, $v^2 = v$.

We have:

$$|uv - f| = |(u - f)v + (1 - v)(v - f)| \leq v|u - f| + (1 - v)|v - f|, \tag{6.13}$$

$$|u + v - uv - f| = |(u - f)(1 - v) - vf + v^2| \leq (1 - v)|u - f| + v|v - f|. \tag{6.14}$$

Therefore: $|uv - f| + |u + v - uv - f| \leq |u - f| + |v - f|$. Integration with respect to the measure Λ gives the result. □

Theorem 6.1.5 (Pseudo-monotonicity)**.** *Let $S_1 \subset S_2 \subset \mathbb{R}^2$, and $U_i, i = 1, 2$ be a solution of the adaptive geometric problem associated to S_i. Then*

- *$U_\wedge = U_1 \cap U_2$ is a solution associated to S_1*
- *$U_\vee = U_1 \cup U_2$ is a solution associated to S_2.*

Proof. We proceed as in [Yin et al., 2007b].

$$\begin{aligned}
0 &\geq \mathbb{E}_G(U_1, S_1) - \mathbb{E}_G(U_\wedge, S_1) \text{ by optimality of } U_1 \\
&\geq \mathbb{E}_G(U_1, S_2) - \mathbb{E}_G(U_\wedge, S_2) \text{ by Lemma 6.1.3} \\
&\geq \mathbb{E}_G(U_\vee, S_2) - \mathbb{E}_G(U_2, S_2) \text{ (see below)} \\
&\geq 0 \text{ by optimality of } \mathbb{E}_G(U_2, S_2).
\end{aligned}$$

The third inequality is a consequence of the submodularity of the perimeter and of the fidelity term. As a result, all inequalities are in fact equalities and $U_\wedge$ (resp. $U_\vee$) is a minimizer associated to S_1 (resp. S_2). □

A consequence of this monotonicity property is the geometric equivalence theorem.

Theorem 6.1.6 (Geometric equivalence)**.** *Let $f \in L^1_\lambda$.The following assertions are equivalent:*

(i) u is solution of (6.1)

(ii) Almost every level set U_ρ of u is solution of (6.7).

Proposition 6.1.7 (Contrast invariance)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be a Lipschitz homomorphism. Then, for all $f \in L^1_\lambda$:*

$$T_\lambda(g \circ f) = g(T_\lambda f). \tag{6.15}$$

6.1.3 Smoothness of the boundary

The following results explain that, as in the stationary case, when λ is bounded, the boundary of the solutions is smooth.

Proposition 6.1.8. *Assume λ is essentially bounded. Let $F \subset \mathbb{R}^2$, and $U \subset \mathbb{R}^2$ be a solution of the TVL1 problem associated to F. Then ∂U coincides with the reduced boundary of U, and it is a $C^{1,1}$ hypersurface.*

Proof. It is the same as the stationary case. Let $M \geq \operatorname{ess\,sup} |\lambda|$. Since $TVL1$ is idempotent, U is a strong M-minimizer of the perimeter:

$$\text{Per}\,(U, \Omega) \leq \text{Per}\,(U', \Omega) + \int_{U \Delta U'} \lambda(x) dx \leq \text{Per}\,(U', \Omega) + M|U \Delta U'|, \tag{6.16}$$

for every open set Ω and every U' with locally finite perimeter such that $U \Delta U' \subset\subset \Omega$. We get the result by applying the regularity theorem 4.7.4 in [Ambrosio, 1997]). □

Notice that, if we assume λ to be only locally bounded, the result holds locally.

In the following proposition, we assume that the weight has bounded variation, which ensures that it has a well-defined interior trace $T^{(i)}\lambda$ and exterior trace $T^{(e)}\lambda$ on the boundary of the solution U. This trace controls the curvature of the solution.

Theorem 6.1.9. *Let us assume that $\lambda \in BV_{loc} \cap L^\infty_{loc}$. Let $U \subset \mathbb{R}^2$ be a bounded open set solution of the TVL1 problem for some $F \subset \mathbb{R}^2$. Then:*

$$-T^{(e)}\lambda(x) \leq \kappa_{\partial U}(x) \leq T^{(i)}\lambda(x)\ \mathcal{H}^1 a.e. x \in \partial U. \tag{6.17}$$

Proof. Let $V = D \times I \subset \mathbb{R}^2$ be an open set intersecting $\partial^* U$, such that, up to an $\mathcal{H}^1$ negligible set, $U \cap V$ is the hypograph of a C^1 function, and $h \in C^1_0(V)$. Let $\alpha > 0$ small, and $\psi_\alpha(x) := x + \alpha h(x)\nu(x)$, where $\nu \in C^1(V, \mathbb{R}^2)$ is a vector field such that $|\nu| = 1$ on U, and coincides with the outer normal of U on $\partial^* U \cap V$. Let us extend ψ_α by $\psi_\alpha(x) = x$ for $x \notin U$. Notice that when α is small enough, ψ_α is a C^1 diffeomorphism of $\mathbb{R}^2$ in $\mathbb{R}^2$, such that $\psi_\alpha(U \cap V) \subset U \cap V$. Let us write $U_\alpha := \psi_\alpha(U)$. Then, if h is non-positive, since U and U_α coincide outside V and $U_\alpha \cap V \subset U \cap V$:

$$0 \leq \frac{\mathbb{E}_G(U_\alpha) - \mathbb{E}_G(U)}{\alpha} = \frac{\text{Per}\, U_\alpha - \text{Per}\, U}{\alpha} + \frac{\Lambda(U) - \Lambda(U_\alpha)}{\alpha} \tag{6.18}$$

(Note that $\mathbb{E}_G$ refers here to the geometric energy associated with data U.)

The first term tends to $\int_{\partial^* U} \kappa_{\partial^* U} h d\mathcal{H}^1$ for $\alpha \to 0$, so let us focus on the second one. Since:

$$\int_{U_\alpha} \lambda(y) dy = \int_U \lambda(x + \alpha h(x)\nu(x)) \det(I + \alpha D(h\nu)) dx \tag{6.19}$$

we have:

$$\frac{\Lambda(U_\alpha)-\Lambda(U)}{\alpha}=\int_U \frac{\lambda(x+\alpha\nu h)-\lambda(x)}{\alpha}\det(I+\alpha Dw)dx+\int_U \lambda(x)\left(\frac{\det(I+\alpha D(h\nu))-1}{\alpha}\right)dx \tag{6.20}$$

The limit of this expression for $\alpha\to 0^+$ (see Lemmas 6.1.11 and 6.1.12) is:

$$\int_U (D\lambda,h\nu)+\int_U \lambda\text{div }(h\nu)=\int_{\partial^* U} T^{(i)}\lambda h|\nu|^2 d\mathcal{H}^1 \tag{6.21}$$

Eventually, for all $h\le 0$ with support in V:

$$0\le\int_{\partial^* U}(\kappa_{\partial^* U}-T^{(i)}\lambda)hd\mathcal{H}^1 \tag{6.22}$$

thus $\kappa(x)\le T^{(i)}\lambda(x),\ \ \mathcal{H}^1 a.e. x\in V\cap\partial U$.

Proceeding similarly with U^C, we get the other inequality. □

Remark 6.1.10. It is easy to see (consider for instance a disc), that even in the convex case, the condition ($\frac{\text{Per } C}{\Lambda(C)}\le 1$, $\kappa_{\partial C}(x)\le\lambda(x)\ \mathcal{H}^1 a.e. x\in\partial C$) of Theorem 3.3.4 is necessary but not sufficient for a general weight λ.

However, it is possible to compare to the stationary case to get sufficient conditions. If:

$$|\kappa_{\partial C}(x)|\le \operatorname*{ess\,inf}_{y\in C}\lambda(y) \text{ and } \frac{\text{Per } C}{|C|}\le \operatorname*{ess\,inf}_{y\in C}\lambda(y), \tag{6.23}$$

then C is solution[1].

Below are two lemmas that justify the passage to the limit in the proof of Proposition 6.1.9. We give them for the convenience of the reader, but they are quite standard. We shall write $w(x):=h(x)\nu(x)$ (we have $w\in C^1_c(\mathbb{R}^2,\mathbb{R}^2)$, with Spt $w\subset\subset V$).

Lemma 6.1.11. *Under the hypothesis of Theorem 6.1.9:*

$$\lim_{\alpha\to 0^+}\int_U\lambda(x)\left(\frac{\det(I+\alpha D(w))-1}{\alpha}\right)=\int_U\lambda\,\text{Tr}(Dw)h \tag{6.24}$$

Proof. Recall that when A and H are two matrices:

$$\lim_{t\to 0}\frac{\det(A+tH)-\det A}{t}=(\det A)\,\text{Tr}(A^{-1}H) \tag{6.25}$$

Therefore $\frac{d}{d\alpha}\det(I+\alpha Dw)=\det(I+\alpha Dw)\,\text{Tr}\left((I+\alpha Dw)^{-1}Dw\right)$, and by the mean value theorem:

$$\frac{\det(I+\alpha Dw)-1}{\alpha}\le\sup_{\beta\in[0,\alpha]}\det(I+\beta Dw)\,\text{Tr}\left((I+\beta Dw)^{-1}Dw\right) \tag{6.26}$$

which is uniformly bounded on U.

As a consequence, the dominated convergence theorem applies, and the result follows. □

Lemma 6.1.12. *Under the hypothesis of Theorem 6.1.9:*

$$\lim_{\alpha\to 0^+}\int_U\frac{\lambda(x+\alpha w(x))-\lambda(x)}{\alpha}\det(I+\alpha Dw)dx=\int_U(D\lambda,w) \tag{6.27}$$

[1] It is sufficient to write: Per $U+\int_\emptyset d\Lambda=$ Per $U+c\int_\emptyset dx\le$ Per $V+c\int_{C\setminus V}dx\le$ Per $V+\int_{C\setminus V}d\lambda$

Proof. Let U_1, U_2 be two open sets, such that $U_2 \subset\subset U_1 \subset\subset U$ and $|D\lambda|(U\setminus U_2) \leq \epsilon$ (which is possible since U is open). Let us choose $\varphi \in C_c^1$, $0 \leq \varphi \leq 1$, such that $\varphi(x) = 1$ for $x \in U_1$, Spt $\varphi \subset U$, and $\left|\langle (D\lambda, w), \varphi\rangle - \int_U (D\lambda, w)\right| \leq \epsilon$. Notice that for α small enough, $\psi_\alpha(U \setminus U_1) \subset U \setminus U_2$, and we can even assume that $1/2 \leq \det D\psi_\alpha \leq 2$ uniformly on U.

The result will follow from the inequality:

$$\begin{aligned}\left|\int_U \tfrac{\lambda(x+\alpha w)-\lambda(x)}{\alpha} \det D\psi_\alpha dx - \int_U (D\lambda, w)\right| \leq \quad & \left|\int_U \tfrac{\lambda(x+\alpha w)-\lambda(x)}{\alpha} (\det D\psi_\alpha - \varphi)\, dx\right| \\ & + \left|\int_U \tfrac{\lambda(x+\alpha w)-\lambda(x)}{\alpha} \varphi dx - \int_U (D\lambda, w)\right|\end{aligned}$$

- Let us show that:

$$\left|\int_U \frac{\lambda(x+\alpha w)-\lambda(x)}{\alpha} (\det D\psi_\alpha - \varphi)\, dx\right| \leq 6|D\lambda|(U\setminus U_2)\, \|w\|_\infty \tag{6.28}$$

Since $\varphi \equiv \det D\psi_\alpha \equiv 1$ on U_1, and $|\det D\psi_\alpha - \varphi| \leq 3$ in $U \setminus U_1$, it suffices to prove that:

$$\int_{U\setminus U_1} \left|\frac{\lambda(x+\alpha w)-\lambda(x)}{\alpha}\right| dx \leq 2|D\lambda|(U\setminus U_2)\, \|w\|_\infty \tag{6.29}$$

We proceed by approximation: let $\lambda_n \in C^\infty(U)$ converging to λ in L^1 and such that $|D\lambda_n|(U \setminus U_2) \to |D\lambda|(U\setminus U_2)$ (such a sequence exists as we may assume by choice of U_2 that $|D\lambda|(\partial U_2) = 0$, see for instance [Andreu-Vaillo et al., 2002, Theorem B.4]).

$$\begin{aligned}\int_{U\setminus U_1} |\lambda_n(x+\alpha w) - \lambda_n(x)| dx &\leq \alpha \int_{U\setminus U_1}\int_0^1 |\langle \nabla\lambda_n(x+\alpha w s), w\rangle| ds dx \\ &\leq \alpha \int_{U\setminus U_1}\int_0^1 |\nabla\lambda_n(x+\alpha w s)| ds dx \|w\|_\infty \\ &\leq \alpha \int_0^1 \int_{U\setminus U_1} |\nabla\lambda_n(x+\alpha w s)| ds dx \|w\|_\infty \\ &\leq \alpha \int_0^1 \int_{U\setminus U_2} |\nabla\lambda_n((y))| \det D\psi_{s\alpha}^{-1} dy \|w\|_\infty \text{ since } \psi_{s\alpha}(U\setminus U_1) \subset U\setminus U_2 \\ &\leq 2\alpha \|\nabla\lambda_n\|_{L^1(U\setminus U_2)} \|w\|_\infty = 2\alpha |D\lambda_n|(U\setminus U_2)\|w\|_\infty\end{aligned}$$

We pass to the limit $n \to +\infty$ to get the inequality.

- On the other hand, we have:

$$\int_U \frac{\lambda(x+\alpha w(x)) - \lambda(x)}{\alpha} \varphi(x) dx = \frac{1}{\alpha}\int_{U_\alpha} \lambda(y)\varphi(\psi_\alpha^{-1}(y))|\det D\psi_\alpha^{-1}| dy - \frac{1}{\alpha}\int_U \lambda(x)\varphi(x) dx \tag{6.30}$$

But for $y \in U \setminus U_\alpha$, $\psi_\alpha^{-1}(y) \notin U$ thus $\varphi(\psi_\alpha^{-1}(y)) = 0$. Therefore:

$$\begin{aligned}\int_U \tfrac{\lambda(x+\alpha w(x)-\lambda(x))}{\alpha} \varphi(x) dx = \quad & \int_U \lambda(y) \left(\underbrace{\frac{\varphi(\psi_\alpha^{-1}(y)) - \varphi(y)}{\alpha}}_{\to -\nabla\varphi\cdot w}\right) |\underbrace{\det D\psi_\alpha^{-1}}_{\to 1}| dy \\ & + \int_U \lambda(y)\varphi(y) \underbrace{\frac{\det D\psi_\alpha^{-1} - 1}{\alpha}}_{\to -\operatorname{Tr} Dw} dy\end{aligned}$$

The quotients are uniformly bounded (by the mean value theorem) and converge pointwise. Since $\lambda \in L^1$, we can apply Lebesgue's dominated convergence theorem, which gives us the following limit:

$$-\int_U \lambda(y)\nabla\varphi(y)\cdot w - \int_U \lambda\varphi \text{div } w = \langle (D\lambda, w), \varphi\rangle \tag{6.31}$$

- Finally:

$$\begin{aligned}\limsup_{\alpha\to 0}\left|\int_U \frac{\lambda(x+\alpha w)-\lambda(x)}{\alpha}\det D\psi_\alpha dx - \int_U (D\lambda, w)\right| &\leq \epsilon + 6|D\lambda|(U\setminus U_2)\,\|w\|_\infty \\ &\leq (1+6\|w\|_\infty)\epsilon\end{aligned}$$

and this is true for every $\epsilon > 0$, hence the result.

□

6.1.4 Corners with adaptive TVL1

The last subsection showed that in regions where λ is bounded, the boundary of the solution is regular. In the stationary case it is well-known that one cannot preserve sharp corners while minimizing the total variation in the ROF model (see for instance [Meyer, 2001]) or in the TVL1 model (Proposition 3.2.5). On the contrary, one might want, in certain regions of an image, to preserve corners or tiny details.

In [Barozzi et al., 1987], it was proved that any bounded set $F \subset \mathbb{R}^2$ with finite perimeter has a variational curvature, i.e. there exists a non-negative weight $\lambda(x)$ such that F is solution of the problem:

$$\inf_{U\subset F} \text{Per } U - \int_U \lambda(x)dx. \tag{6.32}$$

In view of the geometrical analysis of this paper, this clearly implies that when the data F is convex (for instance a square), there exists a weight $\lambda(x)$ such that TVL1 preserves this data. The idea in [Barozzi et al., 1987] is to consider an increasing sequence $\lambda_n \to +\infty$ and consider the associated solutions of the stationary problem (which gives a non-decreasing sequence of sets E_i). Then define $\lambda(x) = \lambda_i$ if $x \in E_i \setminus E_{i-1}$ (see Figure 6.1). We see that no condition is imposed on the weight λ outside of the convex set since $E_i \subset F$.

It seems difficult to generalize this result to non convex shapes. However, in view of image processing applications, we can extend it to polygons.

Proposition 6.1.13. *Let $F \subset \mathbb{R}^2$ be a simple polygon (not necessarily convex). Then there exists a weight $\lambda \in L^1$ such that F is invariant by adaptive TVL1. Moreover, given a distance $r > 0$, one can assume that $\lambda(x) = 0$ for dist$(x, F) > r$.*

Let us remind the reader that a simple polygon is a polygon whose boundary does not cross itself. The last part of the proposition means that the choice of λ is almost local, and that one can combine weights associated with different objects provided they are far enough from one another.

Proof. Let $X = \{(x_1, y_1), \dots, (x_N, y_N)\}$ denote the set of vertices of the polygon F. Set $0 < \alpha < \min\left(\{|x_i - x_j|, /x_i \neq x_j|\} \cup \{|y_i - y_j|, /y_i \neq y_j|\}\right)$ and define a grid $\alpha\mathbb{Z}^2$. Define its open cells by:

$$\forall (i,j) \in \mathbb{Z}^2,\ V_{i,j} = \{(x,y) \in \mathbb{R}^2, \alpha i < x < \alpha(i+1),\ \alpha j < y < \alpha(j+1)\}. \tag{6.33}$$

Since the number of segments is finite, we can assume, up to a rotation of the axes, that no segment is parallel with the axes (therefore the edges of F intersect with the grid only at isolated points). Moreover, by choice of α the closure of each $V_{i,j}$ contains at most one vertex, and either $F \cap V_{i,j}$ is convex or

$F^C \cap V_{i,j}$ is convex (since it is the intersection of the triangle defined by the vertex and its two neighbors with the convex cell $V_{i,j}$), see Figure 6.1.

In each cell $V_{i,j}$ such that $F \cap V_{i,j} \neq \emptyset$, consider the restricted TVL1 problem (we assume for clarity that $F \cap V_{i,j}$ is convex, but by self-duality of the model it deals with the other case as well):

$$\inf_{U \subset F \cap V_{i,j}} \text{Per}\, U - \int_U \lambda_{i,j}(x)dx \tag{6.34}$$

By the result of [Barozzi et al., 1987], one can find a weight $\lambda_{i,j}$ such that $F \cap V_{i,j}$ is a solution, and $\lambda_{i,j}$ vanishes outside $F \cap V_{i,j}$.

Now we collect the weights and define a function λ such that $\lambda \equiv \lambda_{i,j}$ on $V_{i,j}$ for all $(i,j) \in \mathbf{Z}^2$. Then, since by assumption the edges of F intersect the grid at most at isolated points:

$$|D\mathbb{1}_F|(\mathbb{R}^2) = \sum_{(i,j)\in\mathbb{Z}^2} |D\mathbb{1}_F|(V_{i,j}) \tag{6.35}$$

By summing the inequalities on cells, for any $U \subset \mathbb{R}^2$:

$$|D\mathbb{1}_F|(\mathbb{R}^2) = \sum_{(i,j)\in\mathbb{Z}^2} \left(|D\mathbb{1}_U|(V_{i,j}) + \int_{(F\cap V_{i,j})\Delta(U\cap V_{i,j})} \lambda(x)dx \right) \leq |D\mathbb{1}_U|(\mathbb{R}^2) + \int_{(F\Delta U)} \lambda(x)dx \tag{6.36}$$

Therefore F is a solution with such λ. □

To be coherent with our framework (Equation 6.2), once such λ is found, one can even replace λ by $\lambda + c$.

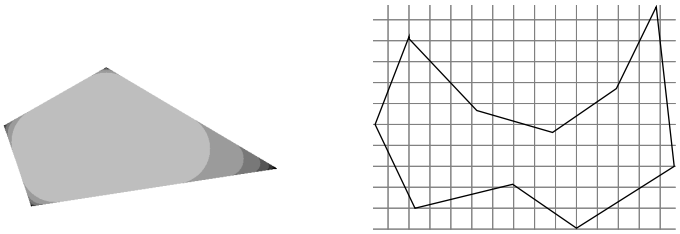


Figure 6.1 – *(Left) A convex polygon and its solution E_i with TVL1 using an increasing sequence λ_i (the darker the greater value of λ). In [Barozzi et al., 1987], it is proved that using precisely the weights λ_i in the corresponding sets leads to a perfect preservation of the convex set. (Right) Generalization to non convex polygons. If the boundary is simple, one can define a grid such that for each cell $V_{i,j}$, $F \cap V_{i,j}$ is convex or $F^C \cap V_{i,j}$is convex.*

6.1.5 Zoom and the generalized Cheeger problem

In the stationary case, it is equivalent to say that a disc suddenly vanishes when λ varies or that it suddenly vanishes when its radius varies. For the spatially adaptive TVL1, we may similarly ask what happens when the image is zoomed (simultaneously with the fidelity map).

Such a zoom consists in replacing $F \subset \mathbb{R}^2$ with $\alpha F = \{\alpha x, x \in F\}$ and $\lambda(x)$ with $\lambda(\frac{x}{\alpha})$ for $\alpha > 0$. Since the following two problems are equivalent:

$$\min_{U' \subset \mathbb{R}^2} \text{Per}\, U' + \int_{\mathbb{R}^2} \mathbb{1}_{U'\Delta(\alpha F)}(x)\lambda\left(\frac{x}{\alpha}\right) dx \tag{6.37}$$

and

$$\min_{U \subset \mathbb{R}^2} \text{Per}\, U + \int_{\mathbb{R}^2} \mathbb{1}_{U\Delta F}(y)\alpha\lambda(y)dy \tag{6.38}$$

(the correspondence is given by $U' = \alpha U$), a zoom of factor α of the image amounts to changing the density λ with $\alpha\lambda$. The question of which set suddenly vanishes when the image is zoomed is therefore related to the notion of generalized Cheeger set studied in [Carlier and Comte, 2007], [Buttazzo et al., 2007], and [Carlier et al., 2009].

Given a nonempty open bounded set $F \subset \mathbb{R}^2$, determine:

$$h_F := \inf_{X \subset F} \frac{\int_F |D\mathbb{1}_X|(x)}{\int_X \lambda(x)dx} = \frac{\text{Per } X}{\Lambda(X)} \tag{6.39}$$

Any minimizer X of this problem is called a generalized Cheeger set of F. If F itself is a minimizer, we say that F is Cheeger in itself.

Notice that F is Cheeger in itself if and only if it is solution of the problem:

$$\min_{X \subset F} \text{Per } X - \lambda_F \Lambda(X) \ . \tag{6.40}$$

with $\lambda_F = \frac{\text{Per } F}{\Lambda(F)}$.

We can also extend the notion of calibrable sets to this framework.

Definition 6.1.14. Let E be a bounded set with finite perimeter. We say that E is Λ-calibrable if there exists a vector field $\xi \in X_\infty$ with $\|\xi\|_\infty \leq 1$ such that $-\text{div } \xi = \lambda_E \mathbb{1}_E \lambda(x)$ in $\mathcal{D}'(\mathbb{R}^2)$ for some λ_E, and $(\xi, D\mathbb{1}_E) = |D\mathbb{1}_E|$ as measures in $\mathbb{R}^2$.

With this definition, Propositions 1.3.2 and 1.3.3 are true when replacing the Lebesgue measure with Λ.

As a consequence, a Λ-calibrable set F is (Λ-)Cheeger in itself, and the converse is true when F is convex.

Now the answer to the initial question is clear: replacing Λ with $\alpha\Lambda$ ($\alpha > 0$), every Λ-calibrable set suddenly vanishes, and every suddenly vanishing set is Λ-Cheeger in itself. In the convex case, those implications are in fact equivalences, but one should be careful that at the critical point (when $\alpha = \lambda_F$) the set of solutions may strictly contain $(\{\emptyset, F\})$, since uniqueness may not hold in the generalized Cheeger problem.

6.1.6 *Shouldn't we change the regularization term instead?*

Another choice for a spatially adaptive TVL1 is to change the regularization term instead of the fidelity term. In a Bayesian setting, this makes sense if for instance we focus on the preservation of the texture or small details: the prior (i.e. the total variation) should change rather than the likelihood (the fidelity term). Yet, we do not choose this approach because for our application it is much more complicated (although several authors have obtained excellent results with it, notably in segmentation [Bresson et al., 2007, Koko and Jehan-Besson, 2010]).

The first (minor) issue one has to face is related to the discretization of the model. Assume that we want to solve:

$$\min_{u \in \mathbb{R}^{N \times N}} \sum_{1 \leq i,j \leq N} g_{i,j} \sqrt{(\nabla_x u)_{i,j}^2 + (\nabla_y u)_{i,j}^2} + \lambda \sum_{1 \leq i,j \leq N} |u_{i,j} - f_{i,j}|,$$

where the gi, j are chosen in function of the observed regularity of the region or the detection of the impulse noise. The traditional forward discretization of the gradient:

$$(\nabla_x u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } 1 \leq i \leq N-1 \\ 0 \text{ if } i = N \end{cases}, \quad (\nabla_y u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } 1 \leq j \leq N-1 \\ 0 \text{ if } j = N \end{cases}. \tag{6.41}$$

raises the issue of the choice of α. Assume for instance that the pixel (i_0, j_0) should be smoothed (because it is in a smooth region or it is detected as noisy). A natural idea is to set $g_{i_0,j_0} \gg 1$, but

this would only force u_{i_0,j_0} to be close to u_{i_0+1,j_0} and u_{i_0,j_0+1}. To have u_{i_0,j_0} close to u_{i_0-1,j_0} or u_{i_0,j_0-1}, other coefficients should be increased. In [Koko and Jehan-Besson, 2010], the authors bypass this difficulty by choosing a large value α_{i_0,j_0} and then they convolve α with a Gaussian with small standard deviation. Another way to solve this problem could be to use a centered scheme of the form $(\nabla_x u)_{i,j} = \frac{1}{2}(u_{i+1,j} - u_{i-1,j})$, $(\nabla_y u)_{i,j} = \frac{1}{2}(u_{i,j+1} - u_{i,j-1})$ (this would then require an adaptation of the divergence for the minimization scheme).

In the continuous case, a related issue may be observed through the Euler Lagrange Equation[2]. Considering the $TV_g + L^1$ problem:

$$\min_u \int_{\mathbb{R}^2} g(x)|\nabla u(x)|dx + \int_{\mathbb{R}^2} |f(x) - u(x)|dx, \tag{6.42}$$

we formally get:

$$-\text{ div } \left(\frac{g(x)\nabla u(x)}{|\nabla u|(x)}\right) + \text{ sign } (u(x) - f(x)) = 0, \tag{6.43}$$

which means:

$$-g(x)\text{div } \left(\frac{\nabla u(x)}{|\nabla u|(x)}\right) - \nabla g(x) \cdot \frac{\nabla u(x)}{|\nabla u|} + \text{ sign } (u(x) - f(x)) = 0. \tag{6.44}$$

As a consequence the $TV_g + L^1$ problem is not simply a weighted TVL1 model because of the apparition of the middle term. The impact of this middle term seems difficult to predict. In segmentation, it is interpreted as a shock term that enhances the detection of edges. Indeed, Bresson *et al.* [Bresson et al., 2007] have shown that the $TV_g + L^1$ problem amounts to the geodesic active contour [Caselles et al., 1997] on the level sets of the image.

Therefore, another way to investigate the model is to consider the geodesics of the associated geometric problem:

$$\min_U \int_{\partial U} g(x)d\mathcal{H}^1(x) + \lambda|U\Delta F|. \tag{6.45}$$

Figure 6.2 shows the difference between the $TV_g + L^1$ with convolution of g and the spatially varying λ. Using a spatially adaptive λ amounts to interpolating the level lines in the noisy regions with line segments. This should avoid too oscillating boundaries in the resulting image. On the contrary, the adaptivity on g favors oscillating boundaries. This might be the reason why the authors of [Koko and Jehan-Besson, 2010] propose to postprocess the result of the $TV_g + L^1$ model with a deinterlacer (EDDI) [De Haan and Lodder, 2002].

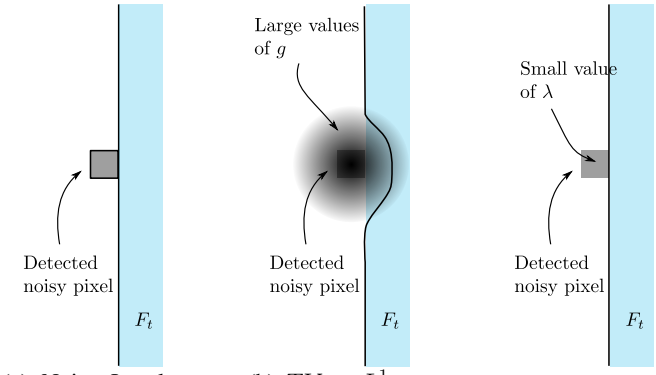
As a conclusion, the adaptivity on the regularization term is useful in several problems like segmentation, but in our case it seems to mostly introduce difficulties.

6.2 Application to impulse noise removal

In this section, we show an application of the spatially adaptive TVL1: impulse noise removal. By combining a noise detector with spatially adaptive TVL1, we show that performance close to state-of-the-art methods can be obtained. Notice that we do not claim that the proposed method is a breakthrough: its purpose is mainly to illustrate an application of the model. We consider discrete images defined on a grid $\Omega = [\![1, N]\!]^2$.

Impulse noise comes from bit errors in transmission, wrong pixels or faulty memory locations in hardware. It can be divided in two categories:

[2]This remark is inspired from the one of Tshcumperlï¿½ and Deriche on anistropic diffusion with divergence form expressions [Tschumperlé and Deriche, 2005].



(a) Noisy Level set (b) $TV_g + L^1$ + convolution of g (c) Adaptive λ

Figure 6.2 – *Left: a level set contaminated by an impulse noise. Middle: with the $TV_g + L^1$ method with Gaussian filtering of g [Koko and Jehan-Besson, 2010], changing the weight of the total variation changes the geodesics. Right: with the adaptive λ the geodesics are not modified.*

.

- Salt-and-Pepper noise: each pixel x is assigned the value

$$f(x) = \begin{cases} \alpha_{max} & \text{with probability } p/2 \\ \alpha_{min} & \text{with probability } p/2 \\ I_0(x) & \text{with probability } 1-p \end{cases} \tag{6.46}$$

  where I_0 is the uncorrupted image, and α_{max} (resp α_{min}) is its maximum (resp. minimum) value.

- Random-valued impulse noise: each pixel x is assigned the value

$$f(x) = \begin{cases} n(x) & \text{with probability } p \\ I_0(x) & \text{with probability } 1-p \end{cases} \tag{6.47}$$

  where $\{n(x)\}_{x\in\Omega}$ are identically uniformly distributed random variables in $[\![\alpha_{min}, \alpha_{max}]\!]$.

In [Nikolova, 2004a], a method called Detail Preserving Variational Method (DPVM) is proposed to remove salt-and-pepper noise as well as random-valued impulse noise. The idea is to solve the minimization problem:

$$\min_{u\in\mathbb{R}^{N\times N}} \sum_{i,j} \varphi\left((\nabla u)_{i,j}\right) + \lambda \sum_{i,j} |f_{i,j} - u_{i,j}|. \tag{6.48}$$

Different choices for the function φ are possible. It may be the ℓ^1 norm: $\varphi(p_x, p_y) = |px| + |p_y|$ (in which case this amounts to the TVL1 model) or the regularized total variation, or the Huber function. The author uses $\varphi(p_x, p_y) = |p_x|^{1.3} + |p_y|^{1.3}$ in her numerical experiments.

As noted in [Chan et al., 2005], this method has the drawback of smoothing the pixels that are regarded as non-corrupted. In [Chan et al., 2004, Chan et al., 2005], the method is therefore modified to include a detection step (based on the adaptive median filter (AM) in the case of salt-and-pepper noise, or the Adaptive Center Weighted Median Filter (ACWMF) [Chen and Wu, 2001a] for the random-valued impulse noise) which gives a map $\mathcal{N}$ of the (supposedly) noisy pixels. Then, they minimize the following

energy [Chan et al., 2005]:

$$\min_{u\in\mathbb{R}^{N\times N}} \sum_{(i,j)\in\mathcal{N}} |u_{i,j}-f_{i,j}| + \beta \sum_{(i,j)\in\mathcal{N}} \left(\sum_{\substack{(m,n)\sim(i,j),\\ (m,n)\in\mathcal{N}}} \varphi(u_{i,j}-u_{m,n}) + 2 \sum_{\substack{(m,n)\sim(i,j),\\ (m,n)\notin\mathcal{N}}} \varphi(u_{i,j}-f_{m,n}) \right), \tag{6.49}$$

where $(m,n)\sim(i,j)$ means that (m,n) is a neighbor of (i,j).

A similar energy can be found in [Chan et al., 2004] for impulse noise removal and in [Dong et al., 2007] where it is adapted to the iterating of the procedure. Let us also mention that a related challenging problem has been considered in [Cai et al., 2008, Cai et al., 2010], where the authors restore images contaminated by blur and impulse noise. We do not consider blurred images in the following.

Now let us give more details about the detection step.

6.2.1 Noise detectors: a short review

Impulse noise removal algorithms have been relying on a detection step for almost two decades. Before focussing on detectors for the random-valued impulse noise, let us mention that problem (6.47) is much more difficult than problem (6.46). Indeed, very often in natural images, only few pixels reach the extremal values. With salt-and-pepper noise, a first guess for the set of corrupted pixels is therefore the pixels for which $f(x) = \alpha_{min}$ or α_{max} (see [Sun and Neuvo, 1994, Koko and Jehan-Besson, 2010]). A more sophisticated procedure is used in [Chan et al., 2005]. Both methods seem to lead to excellent detection rates.

For the random-valued impulse noise, things are not that easy, since corrupted pixels may take any value between 0 and 255. Many detection procedures have been proposed. Some of them rely on the median or similar ideas: switching median (SM) filter [Sun and Neuvo, 1994], multi-state median (MSM) filter [Chen and Wu, 2001b], adaptive center-weighted median filter (ACWMF) [Chen and Wu, 2001a], peak-and valley filter [Windyga, 2001] conditional signal-adaptive median (CSAM) filter [Pok et al., 2003], pixel-wise Median Absolute Deviation (PWMAD) [Crnojevic et al., 2004], etc. More generally, many filters rely on the order of the pixel values, or the order of their differences: Signal Dependent Rank Order Median [Abreu et al., 1996], Rank Ordered Absolute Differences (ROAD) [Garnett et al., 2005] or the recent Rank-Ordered Logarithmic Difference (ROLD) [Dong et al., 2007] Let us also mention the modified threshold Boolean filter (TBF) [Aizenberg et al., 2005], the Jarque-Bera test based median (JM) [Besdok and Yüksel, 2005]. filter, and fuzzy-based methods like [Van De Ville et al., 2003, Xu et al., 2004].

Since the recent state-of-the art methods related to TVL1 like ACWMF-EPR [Chan et al., 2004] and ROLD-EPR [Dong et al., 2007] are based respectively on the ACWMF and ROLD detectors, we focus on the ACWMF, ROAD and ROLD detectors.

6.2.1.1 Adaptive Center-Weighted Median Filter (ACWMF)

The Adaptive Center-Weighted Median Filter is proposed [Chen and Wu, 2001a] to detect impulse noise.

Let $f_{i,j}$ be the values of the noisy image. For $w \in \mathbb{N}$, the weighted median $m_w(f)_{i,j}$ of f is :

$$m_w(f)_{i,j} = \operatorname{med}\left(\{f_{k,l}, |k-i|\le h, |l-j|\le h\} \cup \{w \diamond f_{i,j}\}\right) \tag{6.50}$$

where $w \diamond f_{i,j}$ stands for w copies of the value $f_{i,j}$, and $h \in \mathbb{N}$ is a window size. For $w = 0$ the weighted median amounts to the standard median, and for $w \ge h(2h+1)$ it amounts to the identity filter.

Now, let $L = 2h(2h+1)$, and define the distance $d_k = |m_{2k+1}(f)_{i,j} - f_{i,j}|$ for $k \in \{0,1,\dots,L-1\}$. It is clear that $d_K \le d_{k-1}$ for $k \ge 1$. The idea of the ACWMF detector is that if one of the distances d_k for $k \in \{0,1,\dots,L-1\}$ is above a certain threshold T_k, the pixel should be regarded as noisy.

In practice, the authors consider 3×3 windows ($h = 1$), and they propose the following thresholds:

$$T_k = s \cdot MAD + \delta_k \tag{6.51}$$

where $[\delta_0, \delta_1, \delta_2, \delta_3] = [40, 25, 10, 5]$, and MAD is the median of the absolute deviations from the median:

$$MAD = \text{med}\{|f_{k,l} - (\text{med}\, f)_{i,j}|,\ |k-i| \leq 1, |l-j| \leq 1\}. \tag{6.52}$$

This detector yields satisfying detection rates, it is used in [Chen and Wu, 2001a] to precess a median filter, and in an iterated fashion in [Chan et al., 2004, Cai et al., 2008, Cai et al., 2010] to precess a variational method.

6.2.1.2 Rank Ordered Absolute/Logarithmic Differences (ROAD and ROLD)

The Rank Ordered Absolute Differences detector was proposed in [Garnett et al., 2005]. In the original paper, it is followed with a "trilateral filter" which removes the noisy pixel in an efficient way. For $x = (i, j)$, the idea is to consider the distances with the neighbors $\{|f_{k,l} - f_{i,j}|,\ |k-i| \leq h, |l-j| \leq h, (i,j) \neq (k,l)\}$ and to sort them by increasing order: $r_1(x) \leq r_2(x) \ldots \leq r_{(2h+1)^2-1}(x)$. The ROAD indicator is then:

$$ROAD_m(x) = \sum_{n=1}^{m} r_n(x), \tag{6.53}$$

where $2 \leq m \leq (2h+1)^2 - 2$. In practice, the authors use $h = 1$, $m = 4$. To detect noisy pixels, one selects a threshold $T > 0$, and the noise candidate pixels x are such that $ROAD_m(x) > T$.

In [Dong et al., 2007], a variant was proposed using the logarithm of the differences:

$$ROLD_m(x) = \sum_{n=1}^{m} \tilde{r}_n(x), \tag{6.54}$$

where $\tilde{r}_n = 1 + \max\{\log_2(r_n), -5\}/5$.

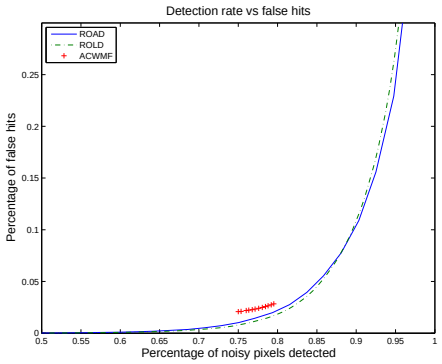


Figure 6.3 – *Detection versus percentage of false hits for the ROAD, ROLD and ACWMF detectors.*

.

Figure 6.3 illustrates the detection rate versus the percentage of false hits on the Lena image as T varies. The ACWMF is also shown, where s is the varying parameter, which takes values between 0.05 and 0.6 as recommended in [Chen and Wu, 2001a, Chan et al., 2004]. It is clear that ROAD and ROLD are better detectors than the ACWMF. The difference between ROAD and ROLD is less obvious. For less than 0.88% of the noisy pixels detected, ROLD has less false hits than ROAD. Since it is usually below this value that one uses the detector (because above 0.88% the corresponding false hit rate is not acceptable), ROLD is more interesting, and the experiments of [Dong et al., 2007] show that the mean square error is lower when applying ROLD instead of ROAD before the Edge Preserving Regularization (EPR).

6.2.2 The importance of iterating the detection

A first denoising experiment is illustrated in Figures 6.4 and 6.5. We have used the spatially adaptive TVL1 to denoise the image using the map of outliers given by ACWMF, ROAD and ROLD. We set $\lambda_{i,j} = 0.001$ for all pixels detected as noisy and $\lambda_{i,j} = 80$ for the others. The corresponding functional is very similar to the one used in [Chan et al., 2004, Chan et al., 2005], but as we will see, its interpretation is much simpler. We have used the Chambolle-Pock algorithm (see Chapter 4) with 800 iterations. The test image is lena with 30% random-valued impulse noise. The quality is measured using the PSNR, given by the formula:

$$\mathrm{PSNR}(u, I_0) = 10 \log_{10} \frac{255^2}{\frac{1}{N^2}\sum_x (I_0(x) - u(x))^2}. \tag{6.55}$$

This experiment confirms that the ACWMF is less efficient than ROAD and ROLD, since many spots are not detected with ACWMF. In fact, with ROLD and ROAD, the quantity of undetected noisy pixels is not negligible but their intensity is usually close to the original value of each pixel, so that this is not visually disturbing.

However, this is not always true with ROLD, as illustrated in Figure 6.5, where ROLD leaves a noisy white spot. Although ROLD provides a better PSNR than ROAD, it is less sensitive to extreme outliers, because of the concavity of the logarithm. As a result, it may happen that ROLD misses visibly noisy pixels. For this reason, we will not use ROLD in the following but ROAD.

Yet, it is clear that the remaining noisy pixel is now very easy to detect using ROAD or ROLD, so that another iteration of the detection and denoising steps allows to detect missed noisy pixels. The natural procedure is therefore to first use a very high threshold to detect and remove the extreme outliers. Then the ordered distances should be more robust and a smaller threshold allows to detect more subtle outliers without many false detections. This iterative approach is used in [Chan et al., 2004, Dong et al., 2007]. We have adapted it using ROAD and TVL1 and different thresholds from those used in [Dong et al., 2007] (we use 2.3, 1.1 and 0.7, corresponding respectively to the detection of 50%, 75% and 82% of the noisy pixels on the Lena image). The quantitative justification of iterating the detections is given in Figure 6.6: we see that the at the second iteration the false hit rate is lower. The gain of the third iteration is much smaller, and there is no point in iterating more.

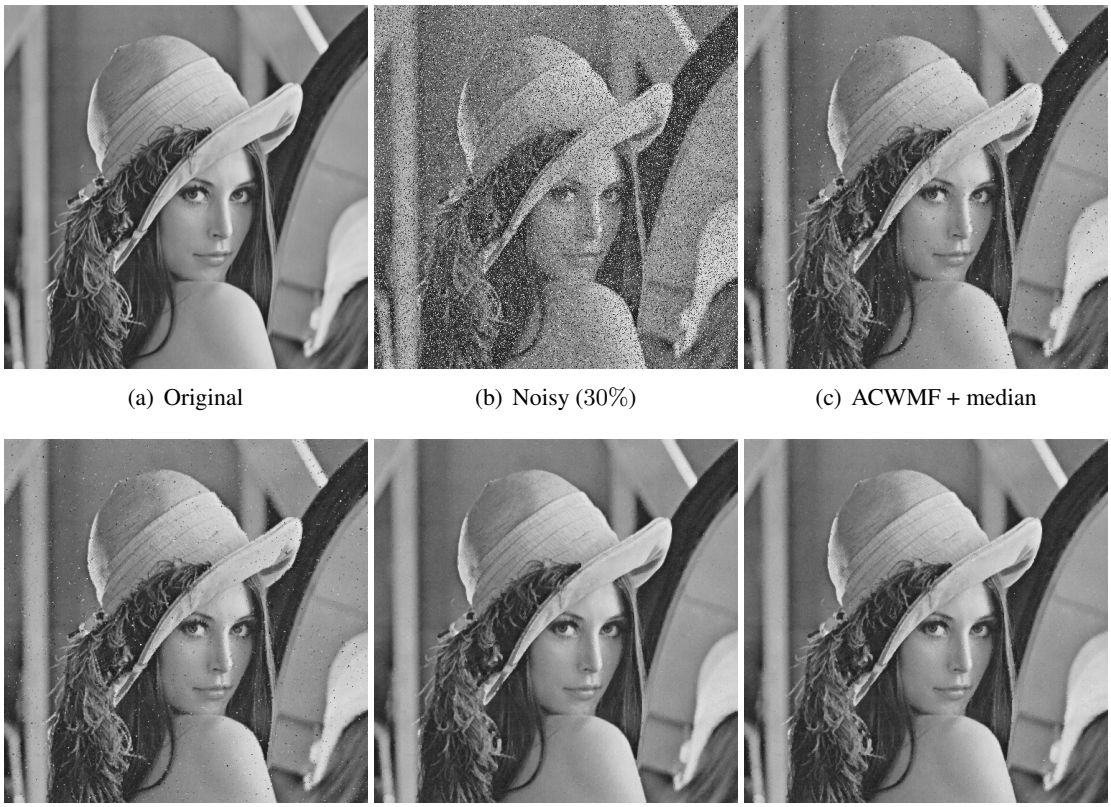
Figure 6.7 illustrates the result of each iteration on the Goldhill image. The first iteration has a deliberately high threshold, so that the false detection rate is very low but only the most obvious outliers are detected. Then, as the thresholds increase, the visual quality is improved.

Table 6.1 describes the PSNR obtained by the exposed denoising methods. As in [Chan et al., 2004], we give the highest value obtained by letting the parameters vary. Results from [Chan et al., 2004, Dong et al., 2007] with comparable noise intensities are also reported, but comparison is difficult since the experiment is different. In any case, this table shows that the performance of the exposed method is neither worse nor better than the state-of-the art methods, and that the noise removal method has in fact little impact on the final result compared to the noise detection.

6.2.3 Conclusion

The spatially adaptive TVL1 with ROAD provides similar performance as [Chan et al., 2004, Dong et al., 2007]. In fact, its main advantage is its simplicity. The algorithms for ACWMF-EPR are summarized in Figure 6.8 and 6.9. The pseudo-code for ROAD-TVL1 is more compact, and basically the iterated spatially adaptive TVL1 and EPR methods are almost identical (leaving aside the factor 2 in the regularization).

In fact, we think that the real issue with random-valued impulse noise lies in the detection step. The way to "fill the gaps" does not matter much provided one uses a reasonable method. As an extreme illustration of this, we show in Figure 6.10 two results based on the ROAD detection on the Lena image with 30% random-valued impulse noise. One is reconstructed using spatially adaptive TVL1, the other



(a) Original (b) Noisy (30%) (c) ACWMF + median

(d) ACWMF + TVL1 (e) ROAD + TVL1 (f) ROLD + TVL1

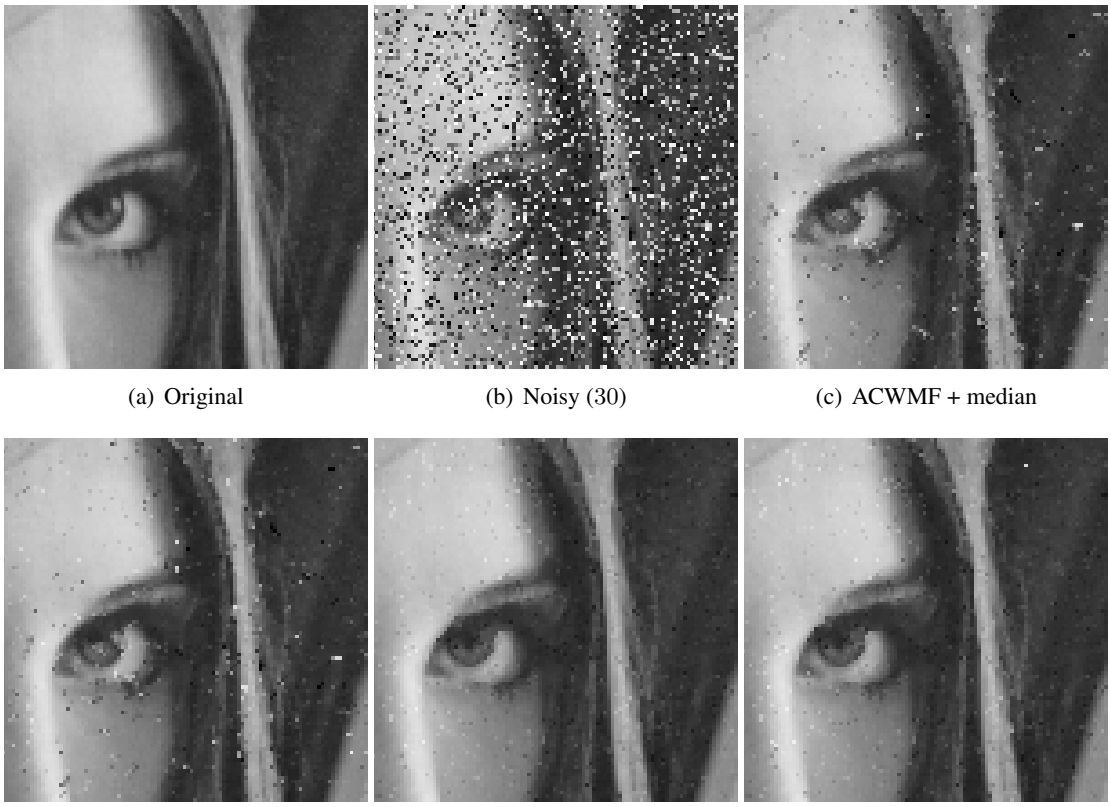
Figure 6.4 – *Experiment using a noise detector + a local denoiser (TVL1 or median). The PSNR are (c) 27.61 dB, (d) 28.75 dB, (e) 30.84 dB, (f) 31.05 dB.*

Table 6.1 – *Numerical result for* 30% *noise removal. For the algorithms with a star, the given values are taken from [Chan et al., 2004, Dong et al., 2007] when available for the same image and comparable noise intensity. They do not reflect the same experiment.*

	Lena	Bridge	Cameraman	Goldhill
ACWMF+med	27.94	23.72	23.27	26.90
DPVM *	27.29	22.44	24.72	27.13
ACWMF-EPR (4 iter.) *	28.33	22.76	25.08	27.52
ACWMF-TVL1 (1 iter.)	28.75	23.91	23.61	27.01
ROAD-TVL1 (1 iter.)	30.84	24.53	23.32	28.95
ROLD-TVL1 (1 iter.)	31.05	24.73	23.63	29.33
ROAD-TVL1 (3 iter.)	31.29	24.74	23.95	29.75
ROLD-EPR (4 iter.) *	32.32	24.79	-	-

one is given by "an oracle" and consists in replacing the noisy pixels with the values of the noise free image. The PSNR in the first case is 30.05 dB, in the second one it is 36.6 dB. Yet, leaving aside a few geometric details[3], the overall visual impression is the same, with many slight spots of noise.

[3]To be precise, the lower part of the iris is not well reconstructed. It seems here that, to minimize the perimeter, the algorithm has cut the iris at the boundary of the spot. Maybe a curvature penalizing energy would be more adequate in this case (see for instance [Masnou and Morel, 1998, Schoenemann et al., 2011]).



(a) Original (b) Noisy (30) (c) ACWMF + median

(d) ACWMF + TVL1 (e) ROAD + TVL1 (f) ROLD + TVL1

Figure 6.5 – *Zoom of Figure 6.4. Notice that ROLD misses a white noisy pixel in the top right-hand corner.*

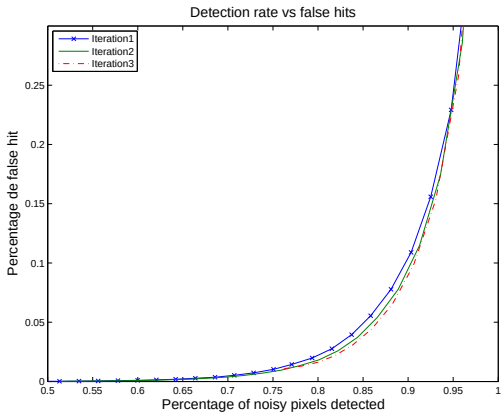


Figure 6.6 – *Detection versus percentage of false hits for the ROAD filter when iterating the detection and denoising steps. A second iteration reduces the false hit rate, but more iterations hardly provide any gain.*

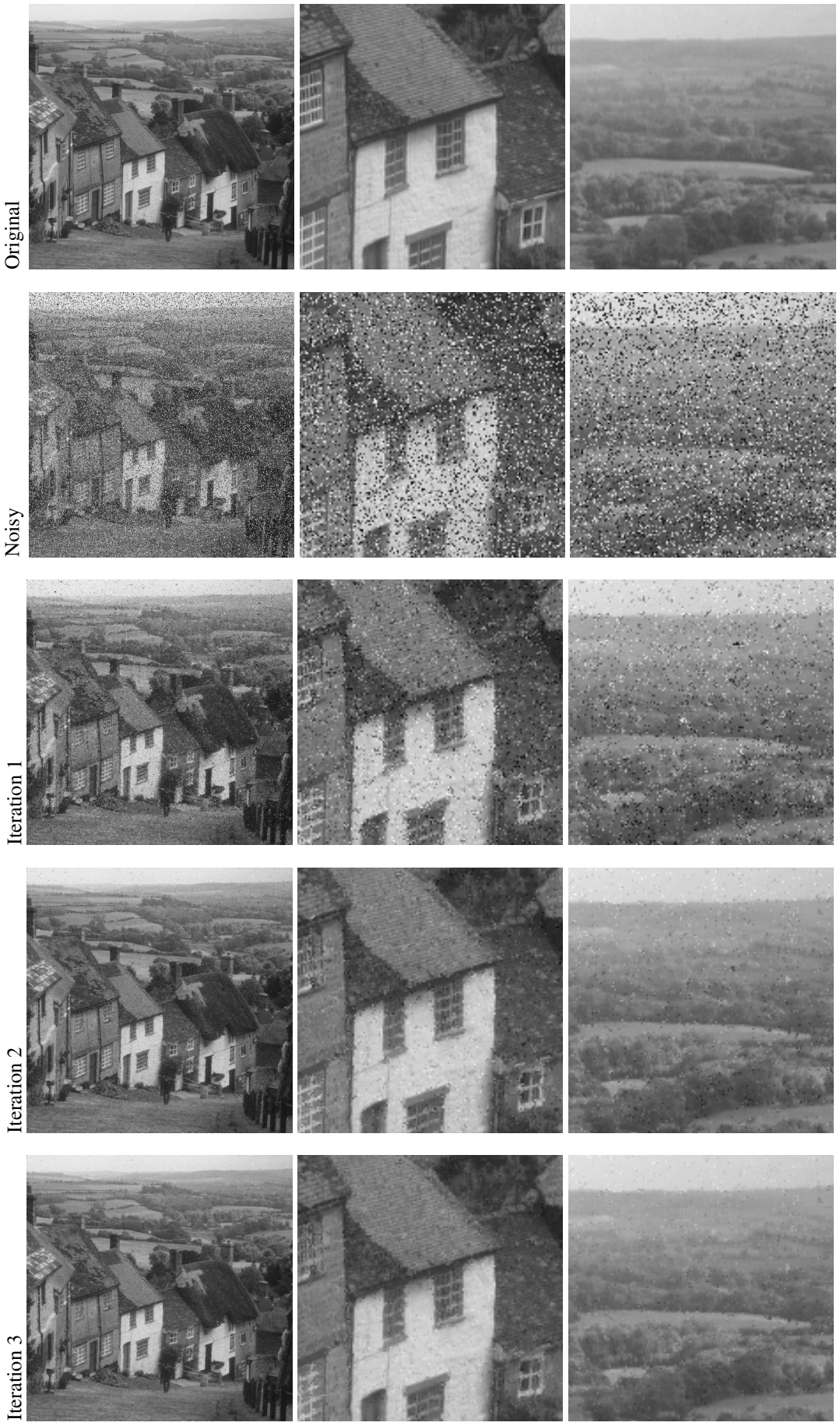


Figure 6.7 – *Evolution of the result with the iterations using ROAD-TVL1. The PSNR for iterations 1,2,3 is 21.80 dB, 27.97 dB, 29.75 dB.*

Algorithm ACWMF-EPR [Chan et al., 2004]

Inputs: Noisy Image f
Output: Denoised image $u^{(4)}$

Set $u^{(0)} := f$.
for all $r \in \{0, 1, 2, 3\}$ **do**
Apply ACWMF with the thresholds $T_k^{(r)} = s \cdot MAD + \delta_k + 20 \times (3 - r)$ to get the noise set $\mathcal{M}^{(r)}$.
Set $\mathcal{N}^{(r)} = \bigcup_{l=0}^{r} \mathcal{M}^{(r)}$.
Set $u_{i,j}^{(r+1)} := u_{i,j}^{(r)}$ for $(i,j) \notin \mathcal{N}^{(r)}$.
Set $u_{i,j}^{(r+1)}$ for $(i,j) \in \mathcal{N}^{(r)}$ such that $u^{(r+1)}$ solves:

$$\min_{u \in \mathbb{R}^{N \times N}} \sum_{(i,j) \in \mathcal{N}} |u_{i,j} - u_{i,j}^{(r)}| + \beta \sum_{(i,j) \in \mathcal{N}} \left(\sum_{\substack{(m,n) \sim (i,j), \\ (m,n) \in \mathcal{N}}} \varphi(u_{i,j} - u_{m,n}) + 2 \sum_{\substack{(m,n) \sim (i,j), \\ (m,n) \notin \mathcal{N}}} \varphi(u_{i,j} - u_{m,n}^{(r)}) \right).$$

end for

Figure 6.8 – *The ACWMF-EPR algorithm*

Algorithm ROAD-TVL1

Inputs: Image f
Output: Denoised image $u^{(3)}$

Set $u^{(0)} := f$ (noisy image).
for all $r \in \{0, 1, 2\}$ **do**
Apply ROAD on $u^{(r)}$ with the thresholds $T^{(r)}$ to get the noise set $\mathcal{M}^{(r)}$.
Set $\lambda_{i,j}^{(r)} := 0.001$ for $(i,j) \in \mathcal{M}^{(r)}$, $\lambda_{i,j}^{(r)} := 80$ otherwise.
Set $u^{(r+1)}$ a minimizer of :

$$\min_{u \in \mathbb{R}^{N \times N}} \sum_{(i,j)} |(\nabla u)_{i,j}| + \sum_{(i,j)} \lambda_{i,j} |u_{i,j} - u_{i,j}^{(r)}|.$$

end for

Figure 6.9 – *The iterated ROAD-TVL1 algorithm*

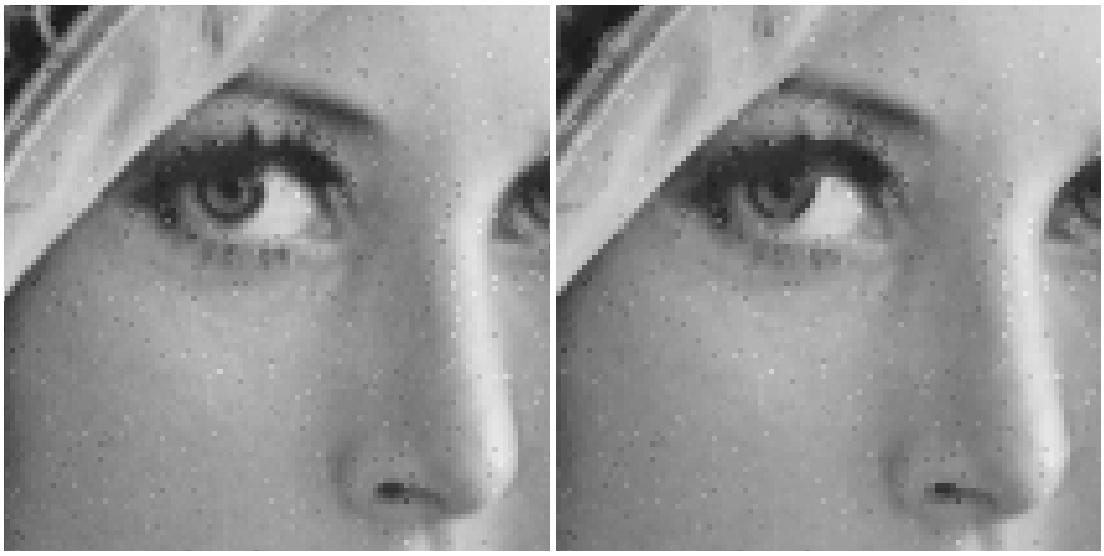


Figure 6.10 – *In both images, the noise (*30%*) was detected using ROAD. One of them was reconstructed with spatially adaptive TVL1, the other using the "oracle", i.e. the values of the noise free image. Can the reader guess which one was restored using TVL1?*

Answer: the right one.

Chapter 7

Extension to the anisotropic case

In this chapter, our aim is to justify the empirical observations of Chapter 4 concerning the Darbon-Sigelle algorithm [Darbon and Sigelle, 2006] with 4-connectivity, namely that the solution of the TVL1 geometric problem for convex sets seems to be given by openings with squares instead of balls. As we mentioned in Chapter 4, in that case the numerical scheme for the energy is:

$$\mathbb{E}(u) := \left(\sum_{k=0}^{N-1} \sum_{l=0}^{N} |u_{k+1,l} - u_{k,l}| + \sum_{k=0}^{N} \sum_{l=0}^{N-1} |u_{k,l+1} - u_{k,l}| \right) + \sum_{1 \leq k,l \leq N} |u_{k,l} - f_{k,l}|. \tag{7.1}$$

This energy is anisotropic, and when the pixel size goes to zero, one can show that this model Γ-converges[1] to the following continuous model (see for instance [Chambolle et al., 2010]):

$$\int |Du|_1 + \int_{\mathbb{R}^2} |f(x) - u(x)| dx. \tag{7.2}$$

where for smooth functions u the integrand in the left term is the ℓ^1 norm of the gradient.

This is a particular case of anisotropic total variation, which is the subject of intensive research [Bellettini et al., 2001, Kawohl and Novaga, 2008, Bellettini et al., 2009, Caselles et al., 2008]. Invoking a result by Bellettini *et al.* [Bellettini et al., 2001], we explain in this chapter that given a crystalline anisotropic total variation, the minimizer of the TVL1 geometric problem with convex input is given by an opening with the Wulff shape followed by a thresholding on the ratio anisotropic perimeter/area. The Wulff shape is the set which minimizes the perimeter for a given volume.

We begin by recalling some basic facts about anisotropic total variation, then we state the basic properties of anisotropic TVL1 before explaining the result of [Bellettini et al., 2001].

7.1 Anisotropic total variation

7.1.1 Introduction

The total variation is used to model surface tension in physical problems where interfaces between substances (fluid or solid) occur. Notice that it may also model the interface between one substance and a vacuum. Since the atoms, molecules or ions on the interface have a different environment than those inside the substance, a surface energy arises. However, as the atoms are organized in a lattice, the interactions between them may crucially depend on the orientation. Therefore, in many problems the surface tension is anisotropic, and the notion of perimeter used in the models should be anisotropic.

The corresponding evolution, observed in crystal growth [Taylor et al., 1992], or statistical physics [Spohn, 1993] is the anisotropic mean curvature motion. In the mathematical community, the study

[1] The Γ-convergence is a convergence for functionals which implies the convergence of the minimizers. See for instance [Braides, 2002].

of such problems was initiated by Taylor in [Taylor, 1978]. A peculiarity of the anisotropic mean curvature motion is that some facets of the evolving object may bend or split in several pieces. In [Bellettini et al., 2001], which provides the key result of this chapter, Bellettini *et al.* characterize the convex facets that evolve without bending or splitting: such facets are called calibrable. We refer the reader to [Bellettini, 2004] for a didactic introduction to the anisotropic mean curvature and many useful references.

As far as image processing is concerned, the anisotropic total variation appears for instance in [Esedoglu and Osher, 2004] where the anisotropic version of the Rudin-Osher-Fatemi is studied, and in [S. Setzer and Teuber, 2008, Steidl and Teuber, 2009] where the dual norm of the anisotropy is the composition of the ℓ^1 norm and an affine transform: the authors use such energies to preserve parallelograms in images. In [Grasmair and Lenzen, 2010], a linear transform is introduced before the computation of the ℓ^2 norm of the gradient, and this transform is chosen locally, depending on the structure of the image. The authors show that this modification yields sharper images than usual Rudin-Osher-Fatemi model. Moreover, anisotropies appear naturally when working with graph-cut methods. As these methods have become standard in computer vision[Greig et al., 1989, Hochbaum, 2001, Boykov and Kolmogorov, 2004, Chambolle, 2005], the study of anisotropic total variation in image processing is far from being anecdotal.

Closely related to our problem are the papers about the total variation flow and Cheeger sets: in [Moll, 2005], Moll shows the well-posedness of the anisotropic total variation flow, and in particular he characterizes the subdifferential of the total variation. In [Kawohl and Novaga, 2008], the anisotropic Cheeger problem is studied as the limit of the p-Laplace eigenvalue problems for $p \to 1$. The uniqueness of the anisotropic Cheeger set of a convex set in the plane is also shown. In [Caselles et al., 2008], Caselles *et al.* characterize by their curvature convex calibrable sets with respect to anisotropic norms in dimension N (hence generalizing the result of [Bellettini et al., 2001] in dimension 2), with the additional assumption that the convex set satisfies a ball condition. As with the isotropic case, the connection with the ROF problem plays a crucial role in the study of dependence of the solutions of $\min_{X \subset C} \mathrm{Per}_\phi\, X - \lambda |X|$ on the parameter λ. In [Caselles et al., 2009], applications of the anisotropic Cheeger problem are given in connection with the 3D segmentation problem and color diffusion. Notice that in this chapter we only consider uniform anisotropies whereas for instance [Moll, 2005, Caselles et al., 2009] deal with the more general case of spatially varying anisotropies.

7.1.2 Anisotropies and total variation

In this section, we consider an anisotropy, i.e. a convex function $\phi : \mathbb{R}^2 \mapsto \mathbb{R}_+$ such that:

$$\exists c > 0, \quad \forall x \in \mathbb{R}^2, \quad \phi(x) \geq c|x|, \tag{7.3}$$

and such that ϕ is 1-homogeneous:

$$\forall t \in \mathbb{R}, \quad \phi(tx) = |t|\phi(x). \tag{7.4}$$

Notice that this implies that there exists $C > 0$ such that $\phi(x) \leq C|x|$ for all $x \in \mathbb{R}^2$. The unit ball W_ϕ for this norm will be called the *Wulff shape*:

$$W_\phi = \{y \in \mathbb{R}^2, \phi(y) \leq 1\}. \tag{7.5}$$

It is a convex body (i.e. a bounded convex set which contains the origin in its interior).

Definition 7.1.1. Let $\phi : \mathbb{R}^2 \to \mathbb{R}_+$ be an anisotropy. We say that:

- The metric ϕ is crystalline if W_ϕ is a polytope.
- The dual norm of ϕ is the anisotropy ϕ^o defined by:

$$\forall x \in \mathbb{R}^2, \quad \phi^o(x) = \sup\{\langle x, y\rangle,\ y \in W_\phi\}. \tag{7.6}$$

For instance $\phi : x \mapsto |x|_1$ and $\phi : x \mapsto |x|_\infty$ are crystalline, and they are dual to each other. Notice that in general, if ϕ is an anisotropy, $(\phi^o)^o = \phi$.

Given an anisotropy ϕ, the ϕ-total variation is defined as:

$$\int |Du|_\phi = \sup \left\{ \int_{\mathbb{R}^2} u \operatorname{div} \varphi,\ \varphi \in C_c^1(\mathbb{R}^2, W_\phi) \right\}. \tag{7.7}$$

Similarly, the anisotropic perimeter of a measurable set $E \subset \mathbb{R}^2$ is:

$$\operatorname{Per}_\phi(E) = \sup \left\{ \int_E \operatorname{div} \varphi,\ \varphi \in C_c^1(\mathbb{R}^2, W_\phi) \right\}. \tag{7.8}$$

For u and E smooth enough, these quantities are respectively equal to:

$$\int |Du|_\phi = \int_{\mathbb{R}^2} \phi^o(\nabla u(x))dx, \text{ and } \operatorname{Per}_\phi(E) = \int_{\partial E} \phi^o(\nu) d\mathcal{H}^1, \tag{7.9}$$

where ν refers to the outer unit normal of E. In particular, we see that if we take $\phi = |\cdot|_\infty$, we obtain the smoothing term used in the anisotropic TVL1 (7.2).

7.1.3 Basic properties

Most properties of Chapter 1 for the isotropic total variation remain true with the anisotropic one. We state them without proof. The reader may consult [Bellettini et al., 1999, Moll, 2005, Caselles et al., 2008, Caselles et al., 2009] for more details.

For convenience, we write:

$$J_\phi : \begin{array}{ccc} L^1 & \longrightarrow & \mathbb{R}_+ \cup \{+\infty\}, \\ u & \longmapsto & \int |Du|_\phi \end{array}. \tag{7.10}$$

As J_ϕ is the supremum of convex, lower semi-continuous functions, we have:

Proposition 7.1.2. *The total variation J_ϕ is convex, lower semi-continuous with respect to L^1 convergence.*

Bellettini *et al.* have proved an extension of the coarea in [Bellettini et al., 1999]:

Proposition 7.1.3 (Coarea formula)**.** *Let $u \in BV(\mathbb{R}^2)$. Then:*

$$\int |Du|_\phi = \int_{-\infty}^{\infty} \operatorname{Per}_\phi \{x \in \mathbb{R}^2,\ u(x) \geq t\} dt. \tag{7.11}$$

If the right member is finite, then $u \in BV(\mathbb{R}^2)$.

The subdifferential of the anisotropic total variation is used in the definition of the anisotropic total variation flow [Moll, 2005], but also in the study the Cheeger problem in connection with the Rudin-Osher-Fatemi problem [Caselles et al., 2008, Caselles et al., 2009]:

Proposition 7.1.4 (Subdifferential of the total variation)**.** *Let $p \in L^\infty(\mathbb{R}^2, \mathbb{R}^2)$. Then $p \in \partial J(u)$ if and only if:*

$$\exists z \in L^\infty, \phi(z(x)) \leq 1 \text{ a.e., } \operatorname{div} z = p \text{ and } -\int (z, Du) = \int |Du|_\phi. \tag{7.12}$$

The following three results concern the anisotropic perimeter.

Proposition 7.1.5 (Submodularity of the anisotropic perimeter)**.** *Let $E, F \subset \mathbb{R}^2$ be measurable sets. Then:*

$$\operatorname{Per}_\phi(E \cap F) + \operatorname{Per}_\phi(E \cup F) \leq \operatorname{Per}_\phi(E) + \operatorname{Per}_\phi(F). \tag{7.13}$$

The next result can be found in [Caselles et al., 2008]:

Proposition 7.1.6. *Let $C \subset \mathbb{R}^2$ be a closed convex set. For all $E \subset \mathbb{R}^2$ with finite perimeter:*

$$\text{Per}_\phi\,(E \cap C) \leq \text{Per}_\phi\, E. \tag{7.14}$$

Eventually, the isoperimetric inequality is proved in [Fonseca, 1991]:

Proposition 7.1.7 (Isoperimetric inequality for anisotropies)**.** *There is a constant $C > 0$ such that for all $E \subset \mathbb{R}^2$ with finite perimeter,*

$$C|E| \leq \text{Per}_\phi\, E, \tag{7.15}$$

with equality if and only if $E = x + W_\phi$ for some $x \in \mathbb{R}^2$.

7.2 Anisotropic TVL1 and openings

In this section, we state without proof the results that can be obtained by following the same pattern as in Chapter 3. We eventually obtain the result for openings of convex sets as a consequence of [Bellettini et al., 2001]. In this section we assume that the anisotropy is crystalline.

7.2.1 The geometric problem

The anisotropic TVL1 problem is given by:

$$\min_{u \in \text{BV}(\mathbb{R}^2)} \int |Du|_\phi + \lambda \int_{\mathbb{R}^2} |f - u|. \tag{7.16}$$

Existence of solutions for (7.16) follows from the compactness theorem and the semicontinuity result (Proposition 7.1.2). The coarea formula (7.11) implies that the problem reformulates as:

$$\min_{u \in \text{BV}(\mathbb{R}^2)} \int_{-\infty}^{+\infty} (\text{Per}_\phi\, U_t + \lambda |U_t \Delta F_t|)\, dt, \tag{7.17}$$

where U_t and F_t denote the upper level sets of respectively u and f.

As in the isotropic case, this problem is equivalent to the family of geometric problems, for almost every $t \in \mathbb{R}$:

$$\min_{U \subset \mathbb{R}^2} \text{Per}_\phi\, U + \lambda |U \Delta F_t|, \tag{7.18}$$

The solutions of this problem are "smooth" in a certain sense: assume for instance that $|F| < +\infty$, then U is a minimizer of

$$\text{Per}_\phi\, U + \lambda \int_U (1 - 2 \cdot \mathbb{1}_F(x)) dx, \tag{7.19}$$

and the following regularity result for quasi-minimizers of the crystalline perimeter[Ambrosio et al., 2002, Theorem 6.19] applies.

Theorem 7.2.1 ([Ambrosio et al., 2002])**.** *There exists $\Sigma(U) \subset \mathbb{R}^2$ with $\mathcal{H}^1(\Sigma(U)) = 0$ such that for all $x_0 \in \partial U \setminus \Sigma(U)$, there exists a neighborhood V of x_0 where ∂U is the graph of a Lipschitz function.*

7.2.2 The case of convex sets

By Proposition 7.1.6, in case $C = F_t$ or $C = \mathbb{R}^2 \setminus F_t$ is convex, the geometric problem associated with $C \subset \mathbb{R}^2$ is equivalent to:

$$\min_{U \subset C} \text{Per}_\phi\, U - \lambda |U|. \tag{7.20}$$

Eventually, the solutions of this problem are obtained from the results of [Bellettini et al., 2001]:

Proposition 7.2.2 (Openings with the Wulff shape)**.** *Let ϕ be a crystalline anisotropy, and $C \subset \mathbb{R}^2$ be a bounded convex set. Let C_r be the opening of C with rW_ϕ, and consider the geometric problem (7.20).*

Then:

- *if* $\frac{\text{Per}_\phi\, C_{1/\lambda}}{|C_{1/\lambda}|} < \lambda$, *the solution is* $C_{1/\lambda}$,
- *if* $\frac{\text{Per}_\phi\, C_{1/\lambda}}{|C_{1/\lambda}|} = \lambda$, *the solutions are* $\emptyset$ *and* $C_{1/\lambda}$,
- *if* $\frac{\text{Per}_\phi\, C_{1/\lambda}}{|C_{1/\lambda}|} > \lambda$, *the solution is* $\emptyset$.

Let us recall that $C_r = \bigcup \{B,\ B \subset C \text{ and } \exists y \in C,\ B = y + rW_\phi\}$.

Proof. In [Bellettini et al., 2001, Theorem 5.2], Bellettini *et al.* characterize the minimizers of problem (7.20) by their anisotropic curvature when C is a convex facet. The proof amounts to a "calibration argument" similar to those exposed in Chapter 2. They show that for a general anisotropy ϕ, the sets:

$$\Omega_\lambda^C = \{x \in \text{ int } (C), \kappa_\phi < \lambda\}, \text{ and } \Theta_\lambda^C = \{x \in \text{ int } (C), \kappa_\phi \leq \lambda\}, \tag{7.21}$$

are solutions of (7.20), and that any other minimizer $\tilde{U}$ satisfies $\Omega_\lambda^C \subset \tilde{U} \subset \Theta_\lambda^C$. In the case where the anisotropy ϕ is crystalline, the meaning of the anisotropic curvature κ_ϕ above is not straightforward, we refer the reader to [Bellettini, 2004, Caselles et al., 2008] for more details.

The point with crystalline anisotropies is that Bellettini *et al.* have characterized the sets Ω_λ^C and Θ_λ^C as openings. Writing $\kappa_{min} = \text{ ess inf}_C\, \kappa_\phi$, they obtain that:

$$\forall \lambda > \kappa_{min},\ \text{int}\, \Omega_\lambda^C = \bigcup \left\{ B, B \subset \text{ int } (C) \text{ and } \exists y \in C,\ B = y + \frac{1}{\lambda} W_\phi \right\}, \tag{7.22}$$

$$\forall \lambda \geq \kappa_{min}, \overline{\Theta_\lambda^C} = \bigcup \left\{ B, B \subset \overline{C} \text{ and } \exists y \in C,\ B = y + \frac{1}{\lambda} W_\phi \right\}. \tag{7.23}$$

For $\lambda > \kappa_{min}$, since C is convex, the sets Ω_λ^C and Θ_λ^C can be identified with $C_{1/\lambda}$, thus the solution of the geometric problem is given by an opening. As κ_{min} is the value such that Ω_λ^C vanishes, i.e. the maximum value of λ such that:

$$\forall U \subset C, 0 = \text{Per}_\phi\, \emptyset - \lambda |\emptyset| \leq \text{Per}_\phi\, U - \lambda |U|, \tag{7.24}$$

the test on the ratio perimeter/area follows.

The last point is to check that only $\emptyset$ and $C_{1/\lambda}$ are solutions for $\lambda = \kappa_{min}$. This amounts to the uniqueness of the anisotropic Cheeger set which was proved in [Kawohl and Novaga, 2008, Theorem 5.1]. □

7.3 Illustrations

From the previous section, we know that when the anisotropy is crystalline, the result of the geometric problem associated with a convex set is given by an opening with the Wulff shape followed by a thresholding on the ratio perimeter/area. Let us emphasize that this result holds for all total variations of the form:

$$\int_{\mathbb{R}^2} \phi^o(Du), \text{ with } \phi(v) \text{ or } \phi^o(v) = \max_{1\leq i\leq k} |g_i \cdot v|, \tag{7.25}$$

and Span$\{g_1, \dots g_k\} = \mathbb{R}^2$.

In particular, for $\phi^o = |\cdot|_1$, the structuring element is the square

$$W_\phi = \{(x_1, x_2) \in \mathbb{R}^2, \max(|x_1|, |x_2|) \leq 1\}. \tag{7.26}$$

In a dual way, if $\phi^o = |\cdot|_\infty$, the structuring element is the rotated square

$$W_\phi = \{(x_1, x_2) \in \mathbb{R}^2, |x_1| + |x_2| \leq 1\}. \tag{7.27}$$

In Figure 7.2, we have used the following anisotropic total variation. Let $e_+ = \frac{1}{\sqrt{2}}(1,1)$, $e_- = \frac{1}{\sqrt{2}}(-1,1)$). We set $W_\phi = \left\{(x_1, x_2) \in \mathbb{R}^2,\ |\langle x, e_+\rangle| \leq 2\sqrt{2}, |\langle x, e_-\rangle| \leq \frac{\sqrt{2}}{2}\right\}$, which corresponds to:

$$\phi(y) = 2\sqrt{2}|\langle y, e_+\rangle| + \frac{\sqrt{2}}{2}|\langle y, e_-\rangle|, \tag{7.28}$$

$$\phi^o(x) = \max\left(\frac{1}{2\sqrt{2}}|\langle x, e_+\rangle|, \sqrt{2}|\langle x, e_-\rangle|\right). \tag{7.29}$$

The projection on W_ϕ is straightforward:

$$P_{W_\phi}(x) = \frac{\langle x, e_+\rangle}{\max\left(1, \frac{\langle x, e_+\rangle}{2\sqrt{2}}\right)} e_+ + \frac{\langle x, e_-\rangle}{\max\left(1, \langle x, e_-\rangle\sqrt{2}\right)} e_-, \tag{7.30}$$

and we inject this expression in (4.10) to replace the ℓ^2 projection, so as to adapt the Pock-Chambolle algorithm [Chambolle and Pock, 2011] to our total variation.

The Wulff shape is displayed on Figure 7.1. To be coherent with the Matlab simulation, we have used the Matlab convention so that the vertical axis is oriented downwards. The result of the adapted Pock-Chambolle algorithm is given on Figure 7.2. As λ decreases, the Wulff shape appears more clearly.

Let us now consider more traditional anisotropies. Figure 7.3 compares the results given by the Darbon-Sigelle algorithm [Darbon and Sigelle, 2006] using 4-connectivity (i.e. $\phi_0 = |\cdot|_1$), and the FLST algorithm with opening with squares. For these simple shapes the results are almost identical. For more complicated shapes however, the result may differ in a more patent way, as we have seen in Chapter 4. Notice that the evolution is reminiscent of the anisotropic mean curvature motion experiments (see for instance [Chambolle and Darbon, 2009]).

Figure 7.4 shows the result given by the Darbon-Sigelle algorithm [Darbon and Sigelle, 2006] with 8-connectivity. As the fidelity parameter decreases, the Wulff shape progressively appears. It is a regular octagon.

7.4 Conclusion

In this chapter we have extended the core result of this part of the thesis to the case of anisotropic TVL1: the geometric problem associated with convex sets is solved by an opening with the Wulff shape followed by a test on the ratio anisotropic perimeter/area. As in the isotropic case, this leads to a precise approximation of the solutions of TVL1 in the non-convex case when the level-set is not too complicated.

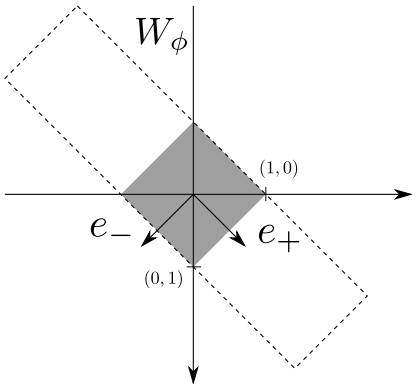


Figure 7.1 – *Anisotropic Wulff shape used in the experiment of Figure 7.2. The vertical axis is oriented using the convention adopted in Matlab.*

This concludes the part devoted to the TVL1 model: we have given more insight to the observation by Darbon [Darbon, 2005] that TVL1 is a morphological filter by highlighting the connection with openings for convex sets. This connection also allows to interpret the behavior of the filter for cartoon-texture decomposition as a granulometry. Incidentally we have compared the TVL1 and the TV+G models, and we have shown that the former gives satisfying decompositions, whereas the latter yields artifacts in the texture part around edges. Then we have investigated the case of a spatially varying fidelity parameter, and we have proposed a framework which encompasses the core ideas of [Chan et al., 2004] while being very simple.

Figure 7.2 – *Result of the Pock-Chambolle algorithm [Chambolle and Pock, 2011] with 8-connectivity. As λ decreases the Wulff shape appears: it is a rectangle as in 7.1.*

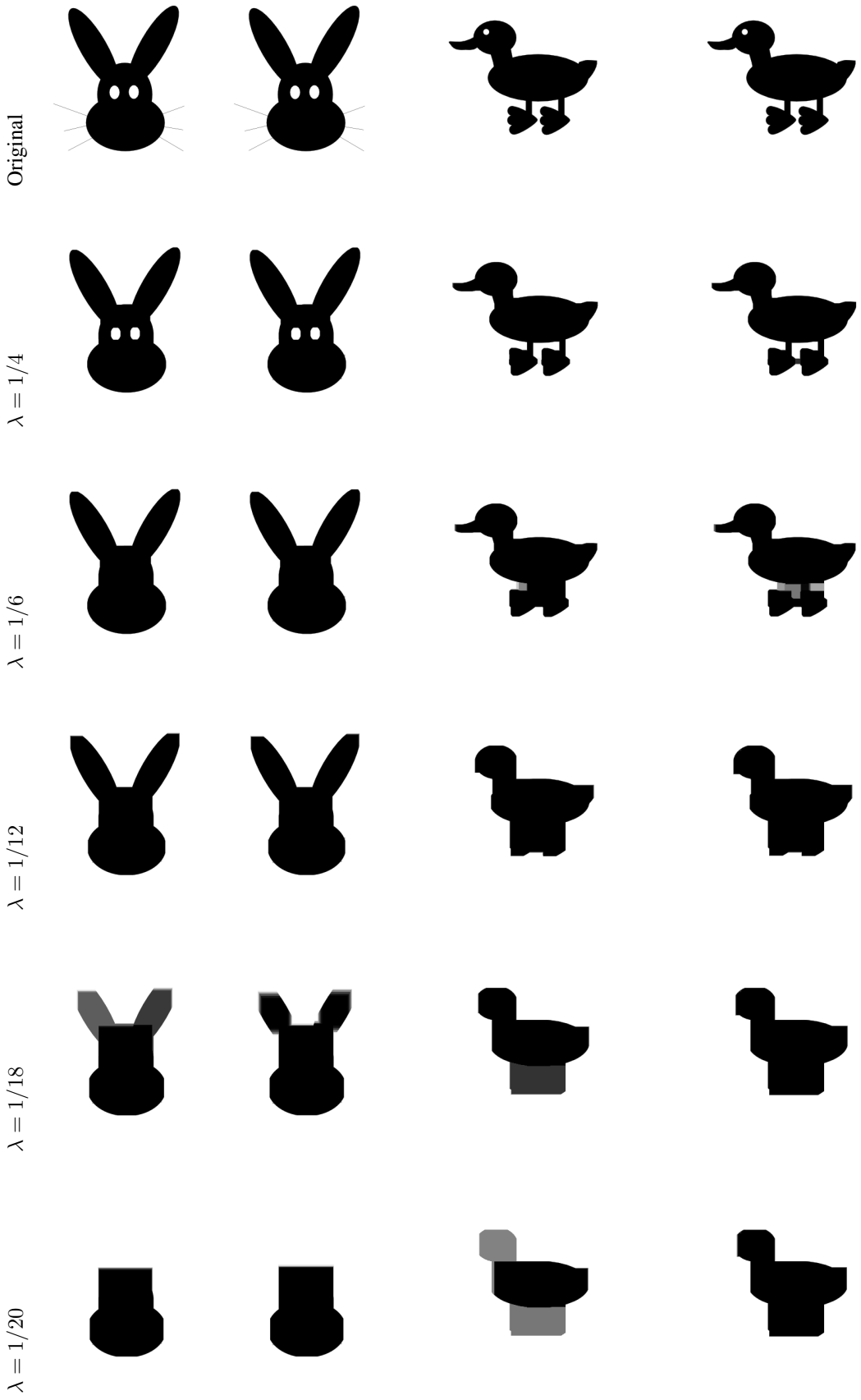


Figure 7.3 – *Comparison of the Darbon-Sigelle algorithm [Darbon and Sigelle, 2006] and the FLST-based algorithm. As the images are not exactly binary (the edge pixels are interpolated), the solutions for the geometric problem may differ depending on the intensity value, hence the apparition of grey zones. No animals were injured during this experiment.*

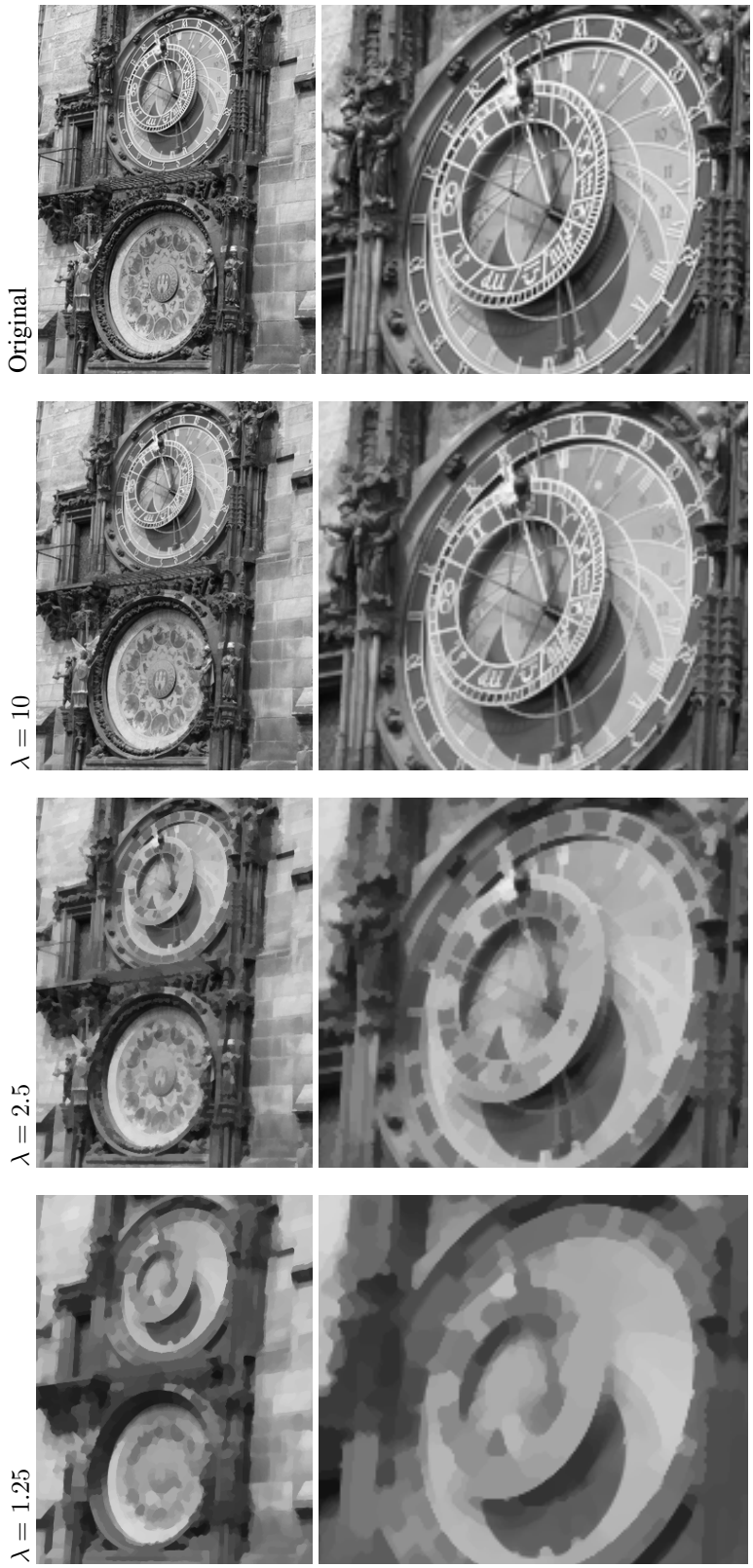


Figure 7.4 – *Result of the Darbon-Sigelle algorithm [Darbon and Sigelle, 2006] with 8-connectivity. As λ decreases the Wulff shape appears: it is a regular octagon.*

Part II

Non-local Methods for Image Restoration

Chapter 8

Introduction to the Non Local Means

This part is devoted to non local algorithms for image denoising. In this chapter, we recall the definition of the Non-Local Means filter (NLM) and, using simple toy models, we illustrate some of its properties. Topics such as the patch size, the bandwidth or the search window are discussed. The framework is a bias-variance trade-off.

In this part of the thesis, we assume that we given a noisy image $\mathbf{u}$ which is a noisy version of an unobserved deterministic image $\mathbf{f}$ corrupted by an Additive White Gaussian Noise (AWGN). This time we work in a discrete setting: let $\Omega \subset \mathbb{Z}^2$ be the (bounded) set indexing the pixels. For any pixel x in the grid Ω, the model has the following formulation:

$$\mathbf{u}(x) = \mathbf{f}(x) + \boldsymbol{\varepsilon}(x)\,, \tag{8.1}$$

where $\boldsymbol{\varepsilon}$ is a centered Gaussian noise with known variance σ^2 and the noise components $\boldsymbol{\varepsilon}(x)$ are independent.

8.1 Introduction

8.1.1 The Non-Local Means and their descendants

Whereas the total variation is a good prior to model cartoon images, it is less adequate to denoise textures. Around 2005, a breakthrough in image denoising was made with the simultaneous introduction of the Non-Local Means [Buades et al., 2005] and the UINTA filters [Awate and Whitaker, 2006]. Both filters rely on the idea of taking advantage of self-similarities of images. When averaging patterns that are repeated across the image, one may reduce drastically the variance of the noise without blurring the details (see Figure 8.1). These similarities may appear at large distances, hence the term "non-local" to qualify the algorithm.

The idea of averaging pixels depending on their similarity had been proposed before the paper by Buades, Coll and Morel, and the authors highlight the similarity between the Non-Local Means and the Yaroslavsky filter [Yaroslavsky, 1985]:

$$YNF_{h,\rho}\mathbf{u}(x) = \frac{1}{C(x)} \sum_{y \in B(x,\rho)} e^{-\frac{|\mathbf{u}(x)-\mathbf{u}(y)|}{h^2}} \mathbf{u}(y). \tag{8.2}$$

where $C(x)$ is a normalization constant, and $B(x,\rho)$ is a window with radius ρ centered around x. However, this filter, as well as the similar sigma filter introduced in [Lee, 1983] is meant to be local: the typical window $B(x,\rho)$ is 5×5 or 7×7. Similarly, the bilateral filter [Tomasi and Manduchi, 1998]:

$$BL_{h,\rho}\mathbf{u}(x) = \frac{1}{C(x)} \sum_{y \in \Omega} e^{-\frac{|\mathbf{u}(x)-\mathbf{u}(y)|^2}{2h^2}} e^{-\frac{|x-y|^2}{2\rho^2}} \mathbf{u}(y), \tag{8.3}$$

Figure 8.1 – *Most natural images have self-similarities. Averaging similar patterns allows to reduce the noise without smoothing out the details. This is the principle of the Non-Local Means. This picture is taken from the Kodak Image Suite.*

gives a strong emphasis to the neighboring pixels, although all pixels in the image may have a (small) contribution.

In fact, it is the combination of the two major innovations of the filter, the claimed non-locality and the use of patches, that makes the denoising so efficient. The non-locality is a bit controversial since the performance of the filter is in fact better when the search for similar pixels is reduced to a (large) sliding window, and several authors [Brox et al., 2008, Gilboa and Osher, 2007, Salmon and Strozecki, 2010b] prefer to use the term "Semi-Local Means" to underline this point. We will come back to this issue in Section 8.3.3. Still, the used window size (between 11×11 and 25×25) is much larger than its equivalent in other filtering methods, and this allows to reduce the variance of the noise by a larger amount than with local filters.

The use of patches, on the contrary, has been widely adopted by the image denoising community. Buades, Morel and Coll were inspired by the work of Efros and Leung [Efros and Leung, 1999] (and more recently [Criminisi et al., 2004]) in texture synthesis, where the use of patches allows to compare neighborhoods in a Markovian framework. In image denoising, the interest of using patches to compare neighborhood comes from their robustness to noise. As a result, several authors have proposed adaptations or variants of existing methods to the patch framework with a substantial improvement of the denoising performance.

In [Kindermann et al., 2005, Gilboa and Osher, 2007, Gilboa and Osher, 2008], the total variation is replaced with a non-local total variation which favors self-similarities within the image. An interesting point with these variational methods is that they also allow to deblur images. A variational framework was also chosen by Brox *et al.* [Brox and Cremers, 2007] and Azzabou *et al.* [Azzabou et al., 2007c]. All these methods interpret the Non-Local Means (or its variant) as one step of a gradient descent for the minimization of a certain energy.

A close point of view is the study of the Non-Local Means in the patch space: in [Szlam, 2006] and [Tschumperlé and Brun, 2009], the Non-Local Means are interpreted as one step of a heat equation in the patch space and the connection with classical diffusion-based denoising algorithms is established. The analysis of the non-local heat equation is carried further in [Peyré, 2008, Peyré, 2009]: Peyré argues that natural images lie on a manifold in the patch space, and the Non-Local Means consist in a diffusion on this manifold. His argumentation is strengthened by the study of different image models (e.g. smooth or cartoon) for which the author gives an explicit parametrization. Let us also mention the work of Singer *et al.* [Singer et al., 2009] where the link with the Fokker-Planck equation is established for various non-local neighborhood filters (including the Non-Local Means).

Some authors adopt a more statistical framework. In [Kervrann and Boulanger, 2008], Kervrann and Boulanger provide a patch-based variant of Lepski's method. Their iterative filter automatically and locally selects the search window by considering a bias-variance trade-off. In [Goossens et al., 2008], the Non-Local Means are connected to robust estimation problem. This leads the authors to question the

choice of the exponential weight for the Non-Local Means and to propose a series of weight functions (Section 8.3.3 will support their conclusions). Interestingly enough, Elad established a similar connection between the bilateral filter and robust estimation in [Elad, 2002], which brought him to the same conclusions about the exponential weight of the bilateral filter. In [Azzabou et al., 2007b], the authors propose a denoising algorithm based on Marginal Posterior Modes estimation. This allows them to select locally the bandwidth of the filter. Let us also mention the work of Doré and Cheriet [Dore and Cheriet, 2009] where the bandwidth is also selected locally, this time relying on the C_p statistic. Whereas the Non-Local Means are designed for additive Gaussian white noise, many applications require to tackle different kinds of noise. Let us mention the extension of the Non Local Means to colored Gaussian noise by Goossens *et al.* [Goossens et al., 2008], and the work of Deledalle *et al.* [Deledalle et al., 2009], who propose an iterative algorithm based on maximum likelihood estimation which allows to deal with different noise models, e.g. multiplicative speckle noise. This is especially important when processing Synthetic Aperture Radar (SAR) images (see [Deledalle et al., 2010a]). The authors also focussed on Poisson noise in [Deledalle et al., 2010b]. The statistical framework also allows to interpret the performance of the Non-Local Means: Salmon and Le Pennec interpret the filter as a PAC-bayesian aggregation of estimators of each patch [Salmon and Le Pennec, 2009a, Salmon and Le Pennec, 2009b]. This leads to oracle inequalities to bound the performance of the Non-Local Means.

Another interesting statistical result was given in [Van De Ville and Kocher, 2009] by Van De Ville and Kocher: they gave a closed form expression for Stein's Unbiased Risk Estimate (SURE) in the case of the Non-Local Means. We will use this result in Chapters 9 and 10.

Many works deal with direct improvements of the Non-Local Means, whether regarding the visual quality or the computation time: Louchet and Moisan propose in [Louchet and Moisan, 2010] a combination of a (local) total variation based filter and the Non-Local Means. The parameter of the total variation filter is locally chosen so that the Non-Local Means find enough similarities in the smoothed image. They show that this helps reducing the visual artifacts of the filter. Brox *et al.* [Brox et al., 2008] refine the Non-Local Means both in terms of visual quality (their algorithm is designed for texture denoising) and computation speed (their filter is iterative, so that the computation time becomes a critical issue). Other methods to accelerate the Non-Local Means have been proposed: identifying a convolution in the patch distances, Wang *et al.* [Wang et al., 2006] a use a Fast Fourier Transform to accelerate the computation of the patch distances (which is the most time consuming step of the algorithm). In [Darbon et al., 2008], Darbon *et al.* use summed area tables (like Viola and Jones [Viola and Jones, 2001] for object detection) to speed up the computation of the distances. Both methods have complexity that is independent of the size of the patch, contrary to the original algorithm.

This list is far from being exhaustive, but it is hopeless to list all the works that have been inspired by the Non-Local Means. Let us conclude this introduction by mentioning two state-of-the art methods which are both patch-based. The first one, called Learned Simultaneous Sparse Coding (LSSC), is based on dictionary learning techniques, well-suited for overcomplete representation (based for instance on ℓ_1 regularization or greedy algorithms). It was proposed by Mairal *et al.* [Mairal et al., 2008, Mairal et al., 2009] following the work of Aharon *et al.* [Aharon et al., 2006]. It has to be noted that this algorithm incorporates a prior on images by initializing a dictionary of patches on a large clean dataset. Another state-of-the-art method in denoising is BM3D [Dabov et al., 2007]. It also relies on patches and consists of a smart combination of classical filtering techniques, such as wavelet denoising and Wiener Filter, applied in the space of patches.

In this thesis, we do not claim to compete with these methods. Our aim is to provide more insight on the Non-Local Means, and, without changing the philosophy of the algorithm, to try to improve its performance.

8.1.2 Notations

We consider a noisy image $\mathbf{u}$ defined on the domain $\Omega \subset \mathbb{Z}^2$. Given an odd number s and a pixel $x \in \Omega$, we define the square patch $U(x)$ of width s centered at x as the s^2-dimensional vector whose coordinates

are the gray values of the pixels in a square neighborhood of x with side s :

$$U(x) = (\mathbf{u}(x+j))_{|j|_\infty \leq \frac{s-1}{2}}. \tag{8.4}$$

The Non-Local Means filter (NLM) compares patches in the image $\mathbf{u}$ to restore the value at pixel x according to the following formula:

$$\mathrm{NL}\,\mathbf{u}(x) = \frac{\sum_{y\in\Omega} e^{-\frac{\|U(x)-U(y)\|^2}{2h^2}} \mathbf{u}(y)}{\sum_{y'\in\Omega} e^{-\frac{\|U(x)-U(y')\|^2}{2h^2}}} = \frac{1}{C(x)} \sum_{y\in\Omega} \omega(x,y)\mathbf{u}(y), \tag{8.5}$$

where $\omega(x,y) = e^{-\frac{\|U(x)-U(y)\|^2}{2h^2}}$ and $C(x)$ is a normalizing factor.

In Chapter 9 and in most of the present chapter, we use the following normalized ℓ^2 norm to compare the patches:

$$\|U(x) - U(y)\|^2 = \frac{1}{s^2} \sum_{|j|\leq \frac{s-1}{2}} (\mathbf{u}(x+j) - \mathbf{u}(y+j))^2,$$

so that h is homogeneous to a gray level. However, in Chapter 10, we will consider more general weighted ℓ^2 distances, which encode some geometry. A shape S is a collection of non-negative values $(S_i)_{i\in[\![-\frac{s-1}{2},\frac{s-1}{2}]\!]^2}$ such that $\sum_j S_j = 1$, and the associated distance is:

$$\|U(x) - U(y)\|_S^2 = \sum_{|j|\leq \frac{s-1}{2}} S_j \cdot (\mathbf{u}(x+j) - \mathbf{u}(y+j))^2.$$

Sometimes, when the context makes it clear, we may write $d^2_{\boldsymbol{S}}(x,y)$ instead of $\|U(x) - U(y)\|_S^2$. Notice that in the original paper [Buades et al., 2005], the ℓ^2 distance is weighted by a Gaussian function. For simplicity, and since this does not affect much performance, most authors do not keep this Gaussian weighting.

Since the common habit is to restrict the above search for patches to a search window of side-length $\mathcal{W}$ around x, the sums in (8.5) may be replaced by sums over all $y \in \Omega$ such that $|x-y|_\infty \leq \frac{\mathcal{W}-1}{2}$.

As a consequence the user of the NLM has to set three parameters to denoise an image:

Choice of the bandwidth h. Many authors (for instance [Kervrann and Boulanger, 2008, Louchet and Moisan, 2010]) use a χ^2 test to set the parameter h. They only accept patches that are likely to be exact replicas of the one they want to denoise: for instance they choose the smallest h such that 99% of exact replicas contaminated by the noise are accepted. This leads to a linear relation between h and σ, and the experiments reported in [Buades et al., 2005, Tasdizen, 2009, Van De Ville and Kocher, 2009] confirm that, in terms of PSNR, the best value of h is roughly proportional to σ. Still, we prove in this thesis that there is interest in choosing the parameter h depending on the image. First, the rule of selecting $h = C\sigma$ is too rough: the visual difference between the results with the optimal h and the predicted value $C\sigma$ may be noticeable. Second, the optimal bandwidth widely varies between the different regions of an image (see Section 9).

In fact, an exact replica is not always available (e.g. along contrasted curved edges or on isolated details), and different options are available. A first possibility is to smooth the image nonetheless by imposing a large value of h. This is the approach used in [Dore and Cheriet, 2009, Duval et al., 2011], and we will describe it in Chapter 9. Another approach is to change the similarity criterion by replacing patches with other shapes, in order to take into account the local geometry. This idea was introduced by Salmon and Strozecki [Salmon and Strozecki, 2010a] by shifting the patches in order to have more similar patches near edges. With Charles Deledalle and Joseph Salmon, we took this approach to a step further by replacing shifted patches with general shapes. This is the subject of Chapter 10.

Choice of the search window $\mathcal{W}$. In the seminal paper [Buades et al., 2005], the theoretical NLM is exposed as a fully non-local filter, i.e. the search for similar pixels is done in the whole image. The complexity of this algorithm is then $O(|\Omega|^2 \cdot s^2)$, which is far too large for most applications. As a consequence the authors propose to reduce the search to a window around x (with typical size 11×11 to 25×25). The complexity becomes $O(|\Omega| \cdot \mathcal{W}^2 \cdot s^2)$ with $\mathcal{W}^2 \ll |\Omega|$. However, the computation time is not the only reason to reduce the search window, since several authors have noticed that a large search window often degrades the quality of the denoising [Kervrann and Boulanger, 2006, Kervrann and Boulanger, 2008, Gilboa and Osher, 2007]. The most striking example is the approach chosen in [Kervrann and Boulanger, 2006, Kervrann and Boulanger, 2008] where the size of the search window is made following bias-variance considerations. Like most authors, we will only consider a fixed search window. Section 8.3.3 shows that the influence of the search window is considerably reduced when the exponential function in the weights is replaced with a function with compact support.

Choice of the patch size s^2. In the literature, this parameter is typically set to 5×5, 7×7 or 9×9. A value of 1×1 yields the Yaroslavsky filter (8.2). The interest of using large patches is to have a robust estimation of the similarity. The drawback is that the measure is less adapted to quick transitions in the image: as a result, details are more blurred (see Section 8.2.2.2) and there are more noisy halos since it is more difficult to find similar pixels near contrasted edges (see 8.3.4). Few authors have tried to mix different sizes of patches: [Mairal et al., 2008] in the context of learning with patches, and [Salmon and Strozecki, 2010a] in a variant of NLM.

To give more intuition about these parameters, the rest of this chapter is made of various examples that highlight the interaction between them, and the artifacts that they induce.

8.2 The importance of the bias of a method

8.2.1 Bias-variance trade-off

Like any other denoising method, the NLM is based on several assumptions on the images to process. As noted in [Szlam, 2008], these are mainly:

- There are similar patches (H_1).
- Similar patches have similar central pixels (H_2).

Roughly speaking, the stronger the assumptions a method makes, the less variance it has, but the more biased it is in case the images do not really satisfy these assumptions. In the case of NLM, the estimation of the variance of the result is much easier than the estimation of the bias. In this thesis, our goal is to take the bias into account when making a decision (whether on the choice of the parameter or on the choice of the shape) so as to minimize the sum of the bias and the variance (i.e. the quadratic error).

Now, let us give more detail about this bias-variance tradeoff. To fix ideas, we momentarily assume that the weights are computed on the noise free image (and thus *deterministic*). If the variance σ^2 of the noise is small and the patch size is large[1], this approximation makes sense. Indeed, if $U(z)$ denotes the patches of the noisy image $\mathbf{u} = \mathbf{f} + \varepsilon$, with ε Gaussian White Noise, we have with high probability[2]: $\|U(x) - U(y)\|^2 \approx \mathbb{E}\|U(x) - U(y)\|^2 = \|F(x) - F(y)\|^2 + 2\sigma^2$.

The risk of denoising the pixel x is given by:

$$\begin{aligned}\mathbb{E}|\operatorname{NL}\mathbf{u}(x) - \mathbf{f}(x)|^2 = \mathbb{E}|\operatorname{NL}\mathbf{u}(x) - \operatorname{NL}\mathbf{f}(x)|^2 + \mathbb{E}|\operatorname{NL}\mathbf{f}(x) - \mathbf{f}(x)|^2 \\ + 2\mathbb{E}\left((\operatorname{NL}\mathbf{u} - \operatorname{NL}\mathbf{f}(x))(\operatorname{NL}\mathbf{f}(x) - \mathbf{f}(x))\right).\end{aligned}$$

[1] Roughly, we assume that $\frac{\sigma^2}{s}$ is small compared to the typical square distance $\|F(x) - F(y)\|^2$, which is typically 10^2 as illustrated in Figure 8.6

[2] This approximation is not valid for $x = y$. However, the same qualitative conclusions can be drawn by slightly adapting the following discussion. We skip this for the sake of clarity.

The last term vanishes since $\mathbb{E}(\mathrm{NL}\,\mathbf{u}(x) - \mathrm{NL}\,\mathbf{f}(x)) = \frac{1}{C(x)} \sum_y \mathbb{E}(\varepsilon(y))\omega(x,y) = 0$. The first term is the variance term: it is small when the smoothing parameter h is large. The second one is the bias term: it is small when h is small. Thus, the optimal choice of h is a trade-off between bias and variance.

An image is all the easier to denoise with the NLM filter as its bias term stays low for large intervals of h. We call such images *patch regular*. This property corresponds to the assumption (H_2). This *patch regularity* will be made precise in the rest of the thesis. Notice that a similar, information theoretic formulation can be found in [Awate and Whitaker, 2006], stating that conditionally to the rest of the patch, the entropy of the law of the central pixel is very low.

In Chapters 9 and 10, we estimate the risk of NLM using SURE (recently introduced in [Van De Ville and Kocher, 2009] in the context of NLM), so that we need not assume that the weights are deterministic.

8.2.2 The bias on simple examples

To give more intuition about the bias of the NLM, we now study three toy models on a noise free signal (so that $\mathbf{u} = \mathbf{f}$). As the variance term vanishes, we may observe the consequences of the assumptions of the NLM, namely:

1. Even periodic images are altered.
2. The search window has an impact on the visual quality of the result.
3. A large patch size tends to blur objects.
4. There is a loss of contrast depending on the occurrence of each pattern
5. A weight with compact support instead of an exponential allows to reduce the bias.
6. The less contrasted the details, the more they are degraded, and this relation is highly non linear.

We do not claim that that these properties are new. We simply propose to highlight them on simple examples.

8.2.2.1 Periodic crenel

The reader who is not familiar with the NLM might think that it is able to restore arbitrarily well any periodic signal. However, it can be shown that the only functions that are invariant by NLM are constant functions (a direct proof is given in [Gilboa and Osher, 2007]), so that even periodic signals must be altered. To illustrate this, we consider a quickly oscillating texture modelled by a periodic crenel with period T and intensity α. For simplicity, we assume that:

- The periodic signal is a crenel, with period $T = 2p$ (where p is an odd number) and intensity α.
- The patch size is $s = \left(k + \frac{1}{2}\right) T$, for $k \in \mathbb{N}$ (i.e. the period of the signal is small compared to the patch size).
- The size of the search window is infinite, or it is a multiple of T.

An illustration of the following discussion is given in Figure 8.2.

We set the origin at the center of a crenel, and we consider one pixel x. Let us first compute the patch distances to x. For $-\frac{T}{2} + 1 \leq j \leq \frac{T}{2}$, a translation by j pixels leads to a difference of $(2k+1)|j|$ pixels, so that :

$$\|F(x) - F(x_j)\|^2 = \frac{(2k+1)|j|\alpha^2}{s} = \frac{2}{T}|j|\alpha^2$$

where $F(x)$ and $F(x_j)$ denote the patches centered respectively at x and x_j, and $x_j = x + j$. The distances for $j > \frac{T}{2}$ and $j < -\frac{T}{2} + 1$ are obtained by periodicity, and the resulting distance function is a periodic triangle (see Figure 8.2).

Now, we assume for instance that the pixel x belongs to the crenel, i.e. there exist $j_1, j_2 \geq 0$ such that $\mathbf{f}(x_j) = \alpha$ for $-j_2 \leq j \pmod T \leq j_1$, with $j_1 + j_2 + 1 = \frac{T}{2}$ and $\mathbf{f}(x_j) = 0$ otherwise.

If we set $r = \frac{1}{T}\frac{\alpha^2}{h^2}$, we have by periodicity:

$$\begin{aligned}
\mathrm{NL}\,\mathbf{f}(x) &= \frac{\sum_{-\frac{T}{2}<j\leq\frac{T}{2}} e^{-\frac{\|F(x)-F(x_j)\|^2}{2h^2}}\mathbf{f}(j)}{\sum_{-\frac{T}{2}<j\leq\frac{T}{2}} e^{-\frac{\|F(x)-F(x_j)\|^2}{2h^2}}}, \\
&= \alpha\frac{\sum_{j=0}^{j_1} e^{-rj} - 1 + \sum_{j=0}^{j_2} e^{-rj}}{2\sum_{j=0}^{\frac{T}{2}-1} e^{-rj} - 1 + e^{-r\frac{T}{2}}}, \\
&= \alpha\frac{\frac{1}{e^{-r}-1}\left(2e^{-(\frac{j_1+j_2}{2}+1)r}\cosh(\frac{j_1-j_2}{2}r) - (e^{-r}-1)\right)}{\frac{1}{e^{-r}-1}(e^{-r\frac{T}{2}}-1)(1+e^{-r})}.
\end{aligned}$$

Recalling that $j_1 = \frac{T}{2} - 1 - j_2$, we get:

$$\mathrm{NL}\,\mathbf{f}(x) = \frac{\alpha}{(1-e^{-r\frac{T}{2}})(1+e^{-r})}\left(1 - e^{-r} - 2e^{-\frac{1}{2}(\frac{T}{2}+1)r}\cosh rx\right). \tag{8.6}$$

The expression for the other parts of the signal are obtained by replacing $\mathrm{NL}\,\mathbf{f}(x)$ with $\alpha - \mathrm{NL}\,\mathbf{f}(x)$ and translating the signal by $\frac{T}{2}$.

The corresponding curve is a piecewise catenary that is all the farther from the crenel as r is small (see Figure 8.2). This shows that even periodic signals may suffer from bias (*point 1.*).

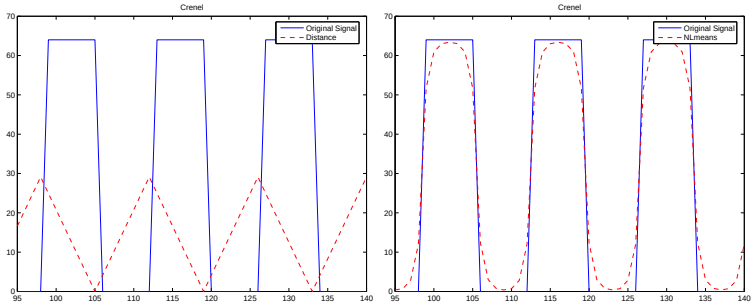


Figure 8.2 – *Effect of NLM on a periodic crenel. Left: extract of the original signal and the patch distance related to pixel 118. Right: the result is a piecewise catenary (see (8.6)).*

8.2.2.2 Isolated detail

Let us consider an isolated pattern modeled by a single crenel (as before, its size is $\frac{T}{2}$ and its intensity is α, but this time, we assume moreover that the patch size is $s = T/2$). If the signal has length $N \geq T$ and if we neglect boundary effects, we have inside the crenel :

$$\mathrm{NL}\,\mathbf{f}(x) = \alpha\frac{1 - e^{-r} - 2e^{-\frac{1}{2}(\frac{T}{2}+1)r}\cosh rx}{(1-e^{-r})\left(2\sum_{j=0}^{\frac{T}{2}} e^{-rj} - 1 + (N-T-1)e^{-r\frac{T}{2}}\right)}, \tag{8.7}$$

where $r = 2\frac{\alpha^2}{sh^2}$ and $x \in \left[-\frac{T}{4} + \frac{1}{2}, \frac{T}{4} - \frac{1}{2}\right]$.

This is again a catenary. Observe that, contrary to total variation based methods, the result when denoising an object depends on the size of the whole image: in the case of an infinite signal (or image) ($N \to +\infty$) we see that $\mathrm{NL}\,\mathbf{f}(x) = 0$ for all x! An interpretation of this phenomenon is that NLM has the following a-priori : *good images are images with repetitive patterns*. The larger the background, the less the isolated detail is "repeated" compared to the background. The phenomenon also appears when denoising lines in an image (see Figure 8.4). If a line contains $O(\sqrt{N})$ pixels in an image with N pixels, the output has intensity $O(\frac{1}{\sqrt{N}})$.

In practice, one usually restricts the set of patches to a search window around x, so that this dependence on N is changed into a dependence on $\mathcal{W}$ (*hence point 2.*). The dependence on $\mathcal{W}$ was noted in [Kervrann and Boulanger, 2008, Dore and Cheriet, 2009] and we believe this example strengthens their argument. The idea is that, when denoising a small detail, pixels are averaged with *any other*. Because of the exponential function, the weights assigned to the wrong patches are small, but they are nonzero. If these patches are overwhelming they will have a strong influence.

Two remedies have been proposed: use a small search window $\mathcal{W}$ or replace the exponential weights with functions with compact support (i.e. that vanish for $\|F(x) - F(y)\|^2/(2h^2)$ large enough) so that $e^{-r\frac{T}{2}}$ in (8.7) is replaced with zero (hence *point 5*, see also [Goossens et al., 2008, Louchet and Moisan, 2010, Salmon, 2010b]). The connection between these methods is discussed in Section 8.3.3. Let us also mention [Kervrann and Boulanger, 2008] where the bias is locally controlled by choosing $\mathcal{W}$ at each pixel.

Incidentally, notice that the larger the patch size, the larger the impact of $\mathcal{W}$ on the blurring of the detail (see *point 3.* and Figure 8.3).

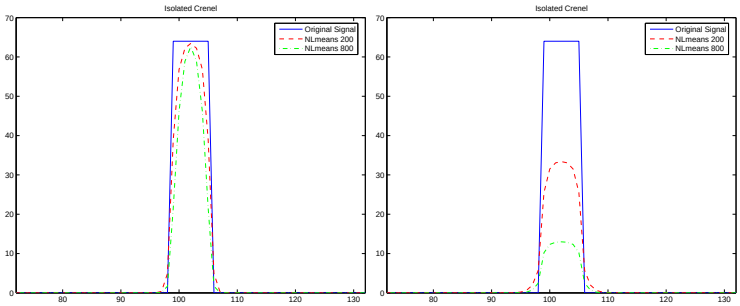


Figure 8.3 – *Loss of isolated details. Left: an extract of a synthetic input signal and the result provided by the NLM filter, it is a catenary inside the crenel (see (8.7)). The size of the crenel is 7, its intensity $\alpha = 64$, the patch size is $s = 7$, $h = 20$. Depending on the total size of the signal ($N = 200$ or $N = 800$), the result does not vary much since $e^{-r\frac{T}{2}} \ll 1$. Right: same experiment, using a patch size $s = 15$. Since the patch size is larger than the pattern, the size of the signal has a large impact on the bias ($e^{-r\frac{T}{2}} \approx 1$).*

Figure 8.4 – *Loss of lines and isolated details. Left: Boat image with little noise ($\sigma = 5$). Middle: Result of the NLM filter ($h = 6$, $s^2 = 7 \times 7$, search window 11×11). Right: same experiment, using a search window 61×61. Notice that several ropes vanish when the size of the search window increases (this should be seen on a computer screen).*

8.2.2.3 Step edge

As a last example, let us consider two regions (one with θN black pixels, the other with $(1-\theta)N$ pixels with intensity α, and $\theta \in (0,1)$) delimited by a step edge. Taking the limit $N \to +\infty$, one may compute the asymptotic output of the filter (see Figure 8.2):

$$\mathrm{NL}\,\mathbf{f}(x) = \frac{\alpha(1-\theta)}{(1-\theta) + \theta e^{(2b(x)-s)r}}, \tag{8.8}$$

where $b(x)$ is the number of black pixels in the patch centered at x (e.g. $b(x) = s$ if the patch of center x lies completely inside the black region). Observe that the gray levels of the two regions are shifted differently, depending on their number of pixels (*point 4.*). Eventually, the transition width between the two regions is proportional to the patch size.

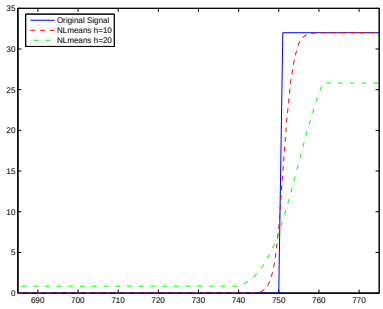


Figure 8.5 – *Effect of NLM on a step signal. The proportion of black pixels is $\theta = \frac{3}{4}$, and the size of the signal is $N = 1000$. The parameters are $s = 21$, $h = 10$ (dashed line) and $h = 20$ (dash-dot line). The length of the transition is equal to the patch size. The bias at each pixel is all the more important as h is large and the pixel value is rare (see (8.8)).*

The previous three examples also show the nonlinear behavior of the filter with respect to the contrast (*point 6.*). Two regions with the same geometry but different contrast are handled differently by the filter (notably for the rare patch effect exhibited in Section 9.3.2).

8.3 The bias-variance tradeoff in the patch space

As we have seen, the NLM filter introduces bias even on simple signals that *apparently* satisfy assumptions (H_1) and (H_2). As with many filters, a first way to control this effect is to tune the bandwidth h so as to accept a little bias in order to reduce the variance while minimizing their sum. A second approach, if images do not satisfy assumptions (H_1) or (H_2) (i.e. if there are no similar patches, or if similar patches do not have similar central pixels), is to *change the similarity measure*.

The present section gives some insight of the two approaches in the patch space.

8.3.1 Regularity in the patch space

The aim of this paragraph is to interpret the easiness with which one may solve the bias-variance trade-off as a regularity property of the image that can be read in the patch space. Several authors [Szlam, 2006, Peyré, 2008, Peyré, 2009, Tschumperlé and Brun, 2008, Tschumperlé and Brun, 2009, Singer et al., 2009] interpret the behavior of the algorithm as a diffusion on a manifold in the patch space. We agree with this model but we wish to stress the importance of considering a manifold *with density* (or a current).

The patch application[3]:

$$F : \begin{array}{ccc} \Omega & \longrightarrow & \mathcal{P} \subset \mathbb{R}^{s^2} \\ x & \mapsto & (\mathbf{f}(y), |y-x|_\infty \leq \frac{s-1}{2}) \end{array}, \tag{8.9}$$

which maps every pixel x of the domain to the patch of center x in the patch space $\mathcal{P}$, gives a parametrization of a surface in the patch space (provided $\mathbf{f}$ is smooth enough).

However, the geometry of such surfaces is complicated as there are many self-intersections. Moreover, the geometry of the patch manifold does not take into account redundancies in the image. For instance, a closed surface may represent either a single pattern or a periodic one, while this difference is crucial for the NLM. Thus, inasmuch as we are interested in the bias-variance trade-off in the patch space, we focus on the mass distribution of the patch set rather than on its geometry, in a more elementary framework.

A measure in the patch space may be defined by pushing forward the Lebesgue measure of the spatial domain:

$$\forall A \in \mathcal{B}(\mathbb{R}^{s^2}),\ m(A) = \mathcal{L}^2(F^{-1}(A)).$$

If the search window is the whole image, the weights $\omega(x,y) = e^{-\frac{\|F(x)-F(y)\|^2}{2h^2}}$ depend only on the patch value $F(x)$ rather than on the pixel position x itself, and so do the normalization factors $C(x)$. Therefore we can denote them by $W(F(x),F(y)) = \frac{1}{C(F(x))}e^{-\frac{\|F(x)-F(y)\|^2}{h^2}}$ and $C(F(x))$, and we may write:

$$\mathrm{NL}\,\mathbf{f}(x) = \frac{1}{C(x)}\int_\Omega \mathbf{f}(y)\omega(x,y)dy = \frac{1}{C(F(x))}\int_{\mathcal{P}} c(P)W(F(x),P)dm(P), \tag{8.10}$$

where c is the application that maps a patch to the value of its central pixel (i.e. $c(F(x)) = \mathbf{f}(x)$).

Because of the normalization of the weights, the measure $\frac{1}{C(F(x))}W(F(x),\cdot)m$ has total mass 1. The "bias" of the filter can therefore be expressed as:

$$\mathrm{NL}\,\mathbf{f}(x) - \mathbf{f}(x) = \frac{1}{C(F(x))}\int_{\mathcal{P}} \left(c(P) - c(F(x))\right) W(F(x),P)dm(P). \tag{8.11}$$

As a result the bias can be expressed as the moment of the measure $\frac{1}{C(F(x))}W(F(x),\cdot)m$ along the axis corresponding to the central pixel. For instance, if an image has a measure m widely spread along the central pixel axis in a neighborhood of $F(x)$, the estimation of $\mathbf{f}(x)$ might be considerably biased.

The patch regularity assumption (H_2) may be reformulated by asking that the measure m makes the above integral small.

Lipschitz regularity A natural assumption to control the bias in (8.11) is to assume a Lipschitz regularity of the center: the support of $W(F(x),\cdot)m$ is contained in the set $\{P \in \mathcal{P}, |c(P) - \mathbf{f}(x)| \leq k\|\dot{P} - \dot{F}(x)\|\}$, where $\dot{P} \in \mathbb{R}^{s^2-1}$ denotes the patch omitting its central pixel (i.e. $P = (c(P), \dot{P})$), and $k > 0$ is a constant. This provides the upper bound:

$$|\,\mathrm{NL}\,\mathbf{f}(x) - \mathbf{f}(x)| \leq \frac{k}{\sqrt{1+k^2}}\frac{1}{C(F(x))}\int_{\mathcal{P}} \|P - F(x)\|W(F(x),P)dm(P).$$

However, this bound is not satisfactory since it may be arbitrarily close to $\frac{k}{\sqrt{1+k^2}}255$ (or any positive constant, depending on $\sup_\Omega u$): consider the mass defined by the crenel signal of Section 8.2.2.2, with $\alpha = 255$ and $N \to +\infty$ Again, since the exponential weights never cancel, the effect of overwhelming patches is out of control. On the contrary, if one uses a kernel φ with compact support as proposed in [Goossens et al., 2008] (e.g. $\varphi(x) = 0$ for $x \geq 1$), the bias is bounded by $\frac{\sqrt{2}k}{\sqrt{1+k^2}}h$.

[3] In this paragraph, to be coherent with the manifold model, we assume that images are defined on a continuous domain. Otherwise, the Lebesgue measure may be replaced with the counting measure.

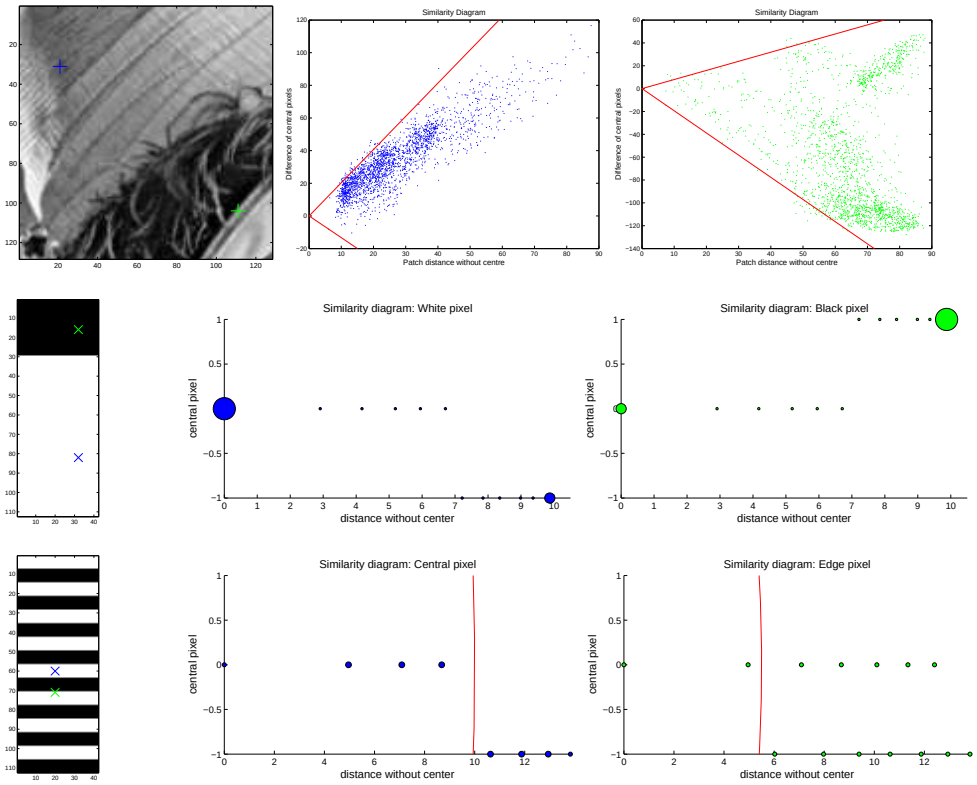


Figure 8.6 – *Similarity diagrams for the pixels indicated by a blue or green cross. Top: A natural image. Middle: Two regions. It is more difficult to reduce the variance without introducing bias for the minority pixels than for the majority. Bottom: Stripes. Pixels near edges suffer from more bias than pixels at the center of each stripe. The Lipschitz constant k is larger near edges.*

Similarity diagram A visual way to examine the regularity at a point is to plot the patch set as the couples: $(\|\dot{P} - \dot{F}(x)\|, c(P) - \mathbf{f}(x))$. This representation is inspired from the one in [Singer et al., 2009] where, in dimension 1, patches of size 2 are represented as couples $(\mathbf{f}(y), \mathbf{f}(y+1))$ for all y. It allows to see if the estimation of each pixel will be much biased and it contains exactly the information NLM needs to compute $\mathrm{NL}\,\mathbf{f}(x)$. Such *similarity diagrams* are shown in Figure 8.6 in the case of a natural image in the case of the examples of Sections 8.2.2.1 and 8.2.2.3. This shows that the regularity constant k varies in the image, and it is especially high near edges.

Replacing patches with shapes Replacing square patches with a shape S amounts to changing the metric in the patch space:

$$\|F(x) - F(y)\|^2 = \sum_j \frac{1}{s^2} \mathbb{1}_{[\![-\frac{s-1}{2}, \frac{s-1}{2}]\!]^2}(j)(\mathbf{f}(x+j) - \mathbf{f}(y+j))^2, \tag{8.12}$$

with :

$$\|F(x) - F(y)\|_S^2 = \sum_j S(j)(\mathbf{f}(x+j) - \mathbf{f}(y+j))^2. \tag{8.13}$$

This is equivalent to replacing the level lines of $W(F(x), \cdot)$, which are balls, with ellipsoids (which may be degenerate depending on which coefficients vanish). In a similarity diagram, this generally changes the order of the pixels, so that the regularity of an image depends on the chosen shape (see Chapter 10).

8.3.2 Bias-variance trade-off in the patch space

The bias-variance dilemma above may be translated to the patch space. Let $R_1 := \mathbb{E}|\operatorname{NL}\mathbf{u}(x) - \operatorname{NL}\mathbf{f}(x)|^2$ and $R_2 := \mathbb{E}|\operatorname{NL}\mathbf{f}(x) - \mathbf{f}(x)|^2$. Since[4]:

$$R_1 = \mathbb{E}\left(\frac{1}{C(x)}\sum_y \varepsilon(y)\omega(x,y)\right)^2 = \sigma^2\frac{1}{C(x)^2}\sum_y(\omega(x,y))^2 = \frac{1}{C(F(x))^2}\sigma^2\int_{\mathcal{P}}(W(F(x),P))^2\,dm(P), \tag{8.14}$$

the risk is given by:

$$R_1 + R_2 = \sigma^2\int_{\mathcal{P}}\frac{(W(F(x),P))^2}{C(F(x))^2}dm(P) + \left(\int_{\mathcal{P}}(c(P)-c(F(x)))\frac{W(F(x),P)}{C(F(x))}dm(P)\right)^2. \tag{8.15}$$

By Cauchy-Schwarz's inequality, the first term is minimal when the weights are uniform (i.e. $h \to +\infty$). On the contrary, if the image is patch regular, the second one is small for small values of h (since patches with a different central pixel are allowed, but they should be very far in the patch space - see Figure 8.6). As expected, the best choice for h is thus a trade-off between reducing the variance by taking a large number of pixels in the average, and not averaging pixels that belong to very different patches.

The interpretation of this quantity is even clearer when the kernel is not Gaussian but given by an indicator function: $\omega(x,y) = \mathbb{1}_{\|F(x)-F(y)\|^2 \le h^2}$. In the patch space, the weights can be written: $W(P,Q) = \mathbb{1}_{B_h(P)}(Q)$, where $B_h(P)$ denotes a ball with radius h and center P, and $C(P) = m(B_h(P))$ is its m-measure, i.e. the number of patches within distance h of P. Then, similar computations show that:

$$R_1 + R_2 = \frac{\sigma^2}{m(B_{h'}(F(x)))} + \left(\frac{1}{m(B_{h'}(F(x)))}\int_{B_{h'}(F(x))}(c(P)-c(F(x)))dm(P)\right)^2, \tag{8.16}$$

where $h' = \sqrt{h^2 - 2\sigma^2}$ corresponds to a threshold on the noise free image. The minimum value of $R_1 + R_2$ is all the lower as the image is regular (e.g. the lipschitz constant k is small). Figure 8.6 shows the trade-off for different pixels. In the middle row, a threshold $h' = 7$ allows to reduce the variance without introducing bias. However the variance will decrease more for the majority pixels than for the minority ones. In the bottom row, it is quickly necessary to introduce bias near edges in order to reduce the variance, contrary to the center of stripes.

8.3.3 The choice of the search window seen from the patch space

We have seen in Sections 8.2.2.2 and 8.3.1 that the bias term depends on the size of the search window. Using a small search window is the common use and the reason invoked is that, besides the speed-up, the result is visually better. However, on the theoretical ground, it is very different to restrict the search of patches to a window for computational reasons than to do this because we know that it will produce better results: the second approach contradicts the "non local philosophy" introduced in [Buades et al., 2005]. If a similar object appears at the other side of the image there is no reason not to use it to denoise a pattern. On the other hand, we have seen that kernels with compact support allow to control the bias. The following experiment determines to what extent.

To this end, we need to distinguish *two different kinds of pixels that may introduce bias* when denoising a pixel x (see Figure 8.7). *First*, pixels that belong to patches that are similar but have very different central values ($\|\dot{F}(x) - \dot{F}(y)\|$ is small but $|\mathbf{f}(x) - \mathbf{f}(y)|$ is large): these patches are cause of non patch regularity. We have seen that this kind of pixels arise for instance at edges. *Second*, pixels that belong to very different patches (both $\|\dot{F}(x) - \dot{F}(y)\|$ and $|\mathbf{f}(x) - \mathbf{f}(y)|$ are large). These pixels have a small weight but, as in Section 8.2.2.2, it is non-zero. Contrary to the pixels of the first type, we can get rid of them by using a truncated kernel.

[4] Again, we assume in this section that the weights are computed on the noise-free image. We do not make this assumption in the next two chapters.

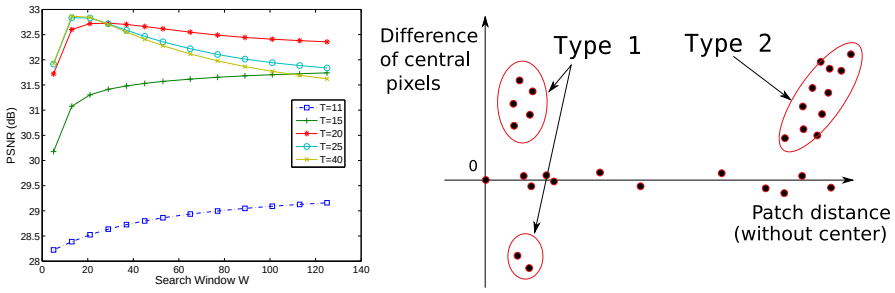


Figure 8.7 – *Left: Evolution of the PSNR as the size of the search window increases, for different values of the threshold T, on the boat image. Using a good threshold (e.g. $T = 20$) for the weights makes the algorithm more robust to changes of the search window. The parameters are: $h = 10$, $s = 7$, $\sigma = 10$. Right: the different kinds of irrelevant pixels in a similarity diagram.*

We run the NLM filter on several images degraded by a noise with $\sigma = 10$, and look at the evolution of the PSNR when the size of the search window increases. However, we truncate the weights in the following way:

$$\omega(x,y) = \begin{cases} e^{-\frac{\|U(x)-U(y)\|^2}{2h^2}} \text{ if } \|U(x)-U(y)\| \leq T, \\ 0 \text{ otherwise.} \end{cases} \tag{8.17}$$

where the parameter h is set to 10. Since the patch distance is normalized, for a threshold value $T = 255$, the algorithm is equivalent to the usual NLM filter. Typical curves for several values of T are shown in Figure 8.7. The PSNR first increases with $\mathcal{W}$, since the variance decreases and the bias introduced on small neighborhoods is very low. When $\mathcal{W}$ gets large enough, many pixels in the search window are irrelevant to denoise x and the bias increases much: the PSNR drops.

The interesting point in this experiment is that when imposing a small threshold value T, the filter is almost insensitive to the increase of the search window. This shows that the pixels of the first kind are not prominent when the search window increases, and therefore the loss of PSNR without thresholding is due to the bias induced by pixels of the second kind.

As a consequence, images are mostly patch regular and truncated weights make the algorithm more robust to the choice of the search window. In [Goossens et al., 2008], it is proposed to use kernels with compact support and the authors show that this allows to preserve textures better (which can also be understood in light of Section 8.2.2.2). Let us stress the fact that it also makes the algorithm more robust to the choice of the search window.

Remark 8.3.1. It is shown in [Tasdizen, 2009] that projecting the patches on their principal components before computing the distances reduces the decay of the PSNR when $\mathcal{W}$ increases. Since the main advantage of this modified ℓ^2 distance is its improved robustness to noise, this suggests that some of the error committed when increasing $\mathcal{W}$ is due to the uncertainty on the distances (for large σ). However, the PCA does not solve the problem of non-zero weights (as in Section 8.2.2.2) and we have performed several experiments highlighting the same bias problem for large $\mathcal{W}$ when using PCA.

8.3.4 The choice of the patch size seen from the patch space

In the literature ([Tasdizen, 2009, Van De Ville and Kocher, 2009, Mairal et al., 2009]), the best results with strong noise are obtained with large patches. As noted in [Singer et al., 2009], using a large patch allows a more robust discrimination between areas that are not actually similar, which is interesting in the presence of noise. Let us illustrate this in an experiment displayed in Figure 8.8. We consider an image with mainly two kinds of patches, say two regions (with intensities α and 0) and a noise with standard deviation σ such that the two noisy regions are hard to distinguish. This time we compute the similarity diagram *on the noisy image*: the two regions cannot be discriminated with a 3×3 patch, but with size 15×15, two different clouds clearly appear.

The shapes of the similarity diagrams are explained by the Bienaymé-Tchebychev inequality. Indeed, let P (resp. Q) be a perfect grey (resp. black) patch and $U(x)$ be a noisy patch completely included in the grey region. Then

$$\forall \epsilon > 0,\ \mathbb{P}\left(\left|\|U(x) - P\|^2 - \sigma^2\right| > \epsilon\right) \leq \frac{2\sigma^4}{s^2\epsilon^2},$$

so that for a large patch size s^2, most patches of the grey (resp. black) regions lie near the sphere of radius σ and center P (resp. Q).

Since the two clusters are separated with a large patch size, a good threshold value h allows to average the pixels of the first cluster only. This does not introduce bias if the first cluster actually corresponds to the pixels of the grey region, i.e. $\|U(x) - P\|^2 \leq \|U(x) - Q\|^2$. But if $\varepsilon_i \sim \mathcal{N}(0,1)$ i.i.d. for $i \in \{1, \dots, s^2\}$:

$$\begin{aligned}\mathbb{P}\left(\|U(x) - Q\|^2 < \|U(x) - P\|^2\right) &= \mathbb{P}\left(\frac{1}{s^2}\sum_{i=1}^{s^2}(\alpha - \sigma\varepsilon_i)^2 < \frac{1}{s^2}\sum_{i=1}^{s^2}(\sigma\varepsilon_i)^2\right) \\ &= \mathbb{P}\left(\frac{\alpha s}{2\sigma} < \frac{1}{s}\sum_{i=1}^{s^2}\varepsilon_i\right) = \int_{\frac{\alpha s}{2\sigma}}^{+\infty} e^{-t^2/2}\frac{dt}{\sqrt{2\pi}},\end{aligned} \tag{8.18}$$

so that when s is large, most patches are averaged with the correct cluster.

In the case of a general shape S, similar computations show that:

$$\forall \epsilon > 0,\ \mathbb{P}\left(\left|\|U(x) - P\|_S^2 - \sigma^2\right| > \epsilon\right) \leq \frac{2\sigma^4}{\epsilon^2}\sum_i S_i^2,$$

and

$$\mathbb{P}\left(\|U(x) - Q\|_S^2 < \|U(x) - P\|_S^2\right) = \int_{\frac{\alpha}{2\sigma\sqrt{\sum_i S_i^2}}}^{+\infty} e^{-t^2/2}\frac{dt}{\sqrt{2\pi}}. \tag{8.19}$$

Since $\sum_i S_i = 1$, it is more robust to use large binary shapes (but there is no prescription on their geometry).

The same holds for natural images: large patches are needed for their robustness to noise. An example is shown in Figure 8.9 where a small patch size induces a mottling effect. However, large patches make the algorithm more exposed to both bias and variance. First, as in Section 8.2.2.2 using a large patch size reduces the importance of little contrasted small details, so that they are more blurred. Second, if the image has textures with highly contrasted transitions or curved and contrasted edges, using a too large patch will prevent the algorithm from finding redundancies and the variance will stay large (see Figure 8.10). The next two chapters explain how to overcome both issues using very different strategies. The first one consists in selecting the smoothing parameter h locally: when choosing the useful patches, be more selective when denoising a small fading out detail and more lenient when it becomes difficult to find similar patches. The second approach is to change the shape of the patch, in order to adapt to the local geometry. This amounts to changing the metric in the patch space, so that the similarity diagrams are changed.

8.4 Conclusion

In this chapter we have studied the classical NLM filter, and we have highlighted some of its properties on simple examples. In particular, as the search window gets larger, the bias of the filter increases, but this effect can be limited with the use of weights with compact support. Large patches are robust to noise but they tend to blur the details that are not contrasted, and they create noise halos around details or edges

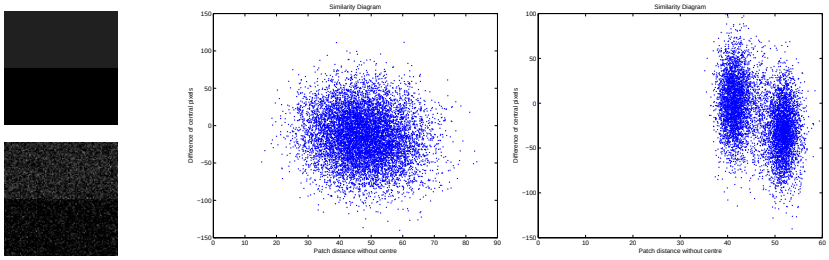


Figure 8.8 – *Left: original (intensity* 0 *and* $\alpha = 32$*) and noisy image (*$\sigma = 30$*). Middle and right: similarity diagram for a pixel at the center of the grey region (the patch size is respectively* 3×3 *and* 15×15*). With a large patch size one sees two clusters. Question: Does the closest cluster correspond to patches which were originally grey as well? When* s *is large, the answer is yes with high probability: in strong noise, a large patch size discriminates better the two regions than a small one.*

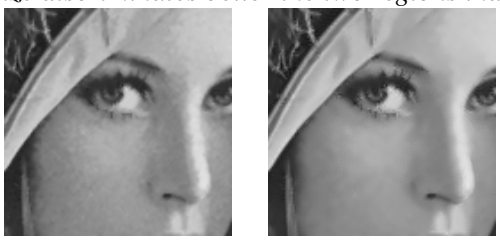
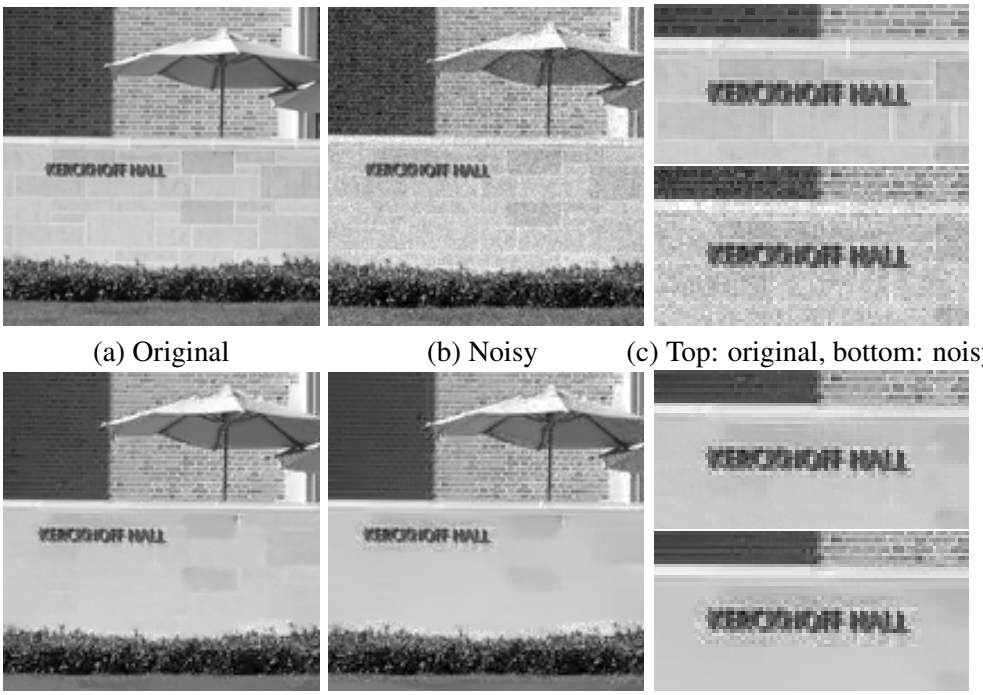


Figure 8.9 – *Left and right: extract of the result of the NLM filter with* $s^2 = 3 \times 3$ *and* $s^2 = 5 \times 5$ *on a noisy image (*$\sigma = 10$*,* h *optimized for PSNR). Using a too small patch size makes the algorithm less robust to noise and Lena's skin looks mottled. It looks smoother with patch size* 5×5 *but visual artifacts appear in the eye.*

that have a high contrast. For each exhibited artifact, we have seen that an adequate choice of parameters could reduce it.

Concerning the rare patch artifact, notice that many other tricks have been proposed to reduce it. Some authors propose to modify the "self-similarity" of x, $\omega(x,x) = 1$, by replacing it with either $\max_{y \neq x} \exp(-\frac{\|U(x)-U(y)\|^2}{2h^2})$ [Buades et al., 2005] or $\exp(-\frac{\sigma^2}{h^2})$ [Salmon, 2010b]. This trick is generally not enough and more stringent methods are sometimes used. In [Gilboa and Osher, 2007], and later in [Brox et al., 2008], all weights are set to zero except the five largest and those of the four spatial neighbors of the pixel to denoise. In [Brox and Cremers, 2007] Brox and Cremers use uniform weights and they impose a minimum of at least 20 similar patches. In [Louchet and Moisan, 2010], Louchet and Moisan propose to prefilter the image using a localized Rudin-Osher-Fatemi model. At each pixel, they smooth the image so that the NLM finds enough similarities. Let us also mention the work of Salmon and Strozecki [Salmon and Strozecki, 2010a], where shifted patches replace the traditional centered patch, which allows to find more similarities near edges.

Most of those methods are efficient, but some of them move the method away from the simplicity of the original Non-Local Means. In the next chapter, we design a variant of the NLM which sticks to the original spirit, except that its parameters vary locally. The choice of these parameters is made according to a bias variance tradeoff.



(a) Original (b) Noisy (c) Top: original, bottom: noisy

(d) Patch size 5×5 (e) Patch size 9×9 (f) Top: 5×5, bottom: 9×9

Figure 8.10 – *Choice of the patch size (bis): original image (a), noisy image (b) ($\sigma = 10$), NLM filter with patch size $s^2 = 5 \times 5$ (d), and $s^2 = 9 \times 9$ (e). Zoomed versions are shown in (c) and (f). Around the letters it is very difficult to find similar patches, and a noisy halo appears. A smaller patch size reduces the spread of the halo since it allows to find similar patches for the furthest pixels. Another solution would have been to force a high smoothing parameter h (those used here were chosen to maximize the PSNR).*

Chapter 9

Spatially adaptive choice of the bandwidth h

In this chapter, we propose an algorithm that locally selects the smoothing parameter h of the NLM, so as to reduce the artifacts exhibited in the last chapter. To handle this problem as a bias-variance trade-off, we need an estimate of the quadratic risk. The first section studies an oracle estimate to describe the optimal behavior of a local selection of the parameter h. Then, we use Stein's Unbiased Risk Estimate to estimate the local quadratic error and to select the right parameter.

As the previous chapter advocates for kernels with compact support (see also [Goossens et al., 2008]), we consider in the following general filters of the form:

$$\mathrm{NL}\,\mathbf{u}(x) = \frac{\sum_{y\in\Omega}\omega(x,y)\mathbf{u}(y)}{\sum_{y'\in\Omega}\omega(x,y')} = \frac{\sum_{y\in\Omega}\varphi\left(\frac{\|U(x)-U(y)\|^2}{2h^2}\right)\mathbf{u}(y)}{\sum_{y'\in\Omega}\varphi\left(\frac{\|U(x)-U(y')\|^2}{2h^2}\right)}, \tag{9.1}$$

with φ non-negative and non-decreasing. Since with a kernel with compact support, the filter is more robust to the choice of $\mathcal{W}$ than it is with a global one, so that in the following, we fix $\mathcal{W}$ in advance and then we choose locally h. This approach is dual to the one proposed by Kervrann and Boulanger [Kervrann and Boulanger, 2008] who fix h and then control the bias and variance of the filter by choosing the size of the search window.

In this chapter and in the following, NL refers to the NLM as an operator, whereas $\hat{\mathbf{f}}$ refers to the corresponding estimator of $\mathbf{f}$, i.e.:

$$\hat{\mathbf{f}}(x) = \mathrm{NL}\,\mathbf{u}(x). \tag{9.2}$$

The performance of each method is evaluated using the Peak Signal to Noise Ratio (PSNR) defined by

$$\mathrm{PSNR}(\hat{f}, f) = 10\log_{10}\frac{255^2}{\frac{1}{|\Omega|}\sum_{x\in\Omega}(\mathbf{f}(x)-\hat{\mathbf{f}}(x))^2}, \tag{9.3}$$

and the Structural SIMilarity (SSIM) defined in [Wang et al., 2004]. The SSIM is between 0 and 1 and a value closed to 1 means that the estimated image has a similar structure to the noise-free image.

9.1 Oracle estimation

To show that the behavior of the SURE-based method of Section 9.3 is an approximation of the optimal one, we first build an oracle which has access to the *real* local squared error. Although this algorithm is not usable in practice, it gives an idea of what parameters should be used in each region and what can be expected in terms of visual quality.

For the oracle estimation only, we have used the indicator kernel $\varphi(x) = \mathbb{1}_{[0,1/2]}(x)$. This kernel has compact support, and it does not suffer from the overestimation of the *self-weights* $\omega(x,x)$ pointed out in [Salmon, 2010b]. Let us recall that for $y \neq x$, the distance between noisy patches $\|U(x) - U(y)\|^2 \approx \|F(x) - F(y)\|^2 + 2\sigma^2$ is increased by the noise level, whereas the *self-distance* $\|U(x) - U(x)\|^2$ is always zero. Therefore, with Gaussian weights, the weight $\frac{1}{C(x)}\omega(x,x) = \frac{1}{C(x)}$ is proportionally $e^{\frac{2\sigma^2}{2h^2}}$ times larger in presence of noise than it would be without noise. As a consequence, some authors set $\omega(x,x)$ to $e^{-\frac{\delta^2}{2h^2}}$, where $\delta^2 = \min_{y \neq x} \|U(x) - U(y)\|^2$, or replace it with $e^{-\frac{\sigma^2}{h^2}}$ (see [Salmon, 2010b]). On the contrary, the indicator weights do not behave this way, and we need not give a special value to $\omega(x,x)$. Notice that the indicator weights have also been used in [Brox and Cremers, 2007, Salmon and Strozecki, 2010a] for similar reasons.

With this kernel, minimizing the bound (8.16) over the radius h' amounts to finding the optimal number of pixels $n_x := m(B_{h'}(U(x)))$ when denoising x. To build our oracle estimate, we compute the risk on the *noise free image* u for each integer, and we keep the minimizing value n_x. Then we can estimate $\mathbf{f}(x)$ from $\mathbf{u}$ by averaging the centers of the n_x patches $U(y)$ that are the nearest to $U(x)$ in euclidean distance. In a nutshell, we use the oracle to define a map n_x and then compute the non-local filter on the noisy image $\tilde{u}$, keeping only the best n_x patches. This roughly amounts to one iteration of [Brox and Cremers, 2007] where the number of similar patches n is selected locally using an oracle.

Figure 9.1 shows the number of pixels n_x recommended by the oracle and the associated smoothing parameter h_x (that is, we display the norm $\|U(x) - U(y)\|$ where y is the last pixel taken into account). As expected, in very smooth regions the oracle selects as many pixels as possible whereas in regions where the image is not patch regular (i.e. near edges) the oracle recommends to use very few pixels. The case of textures is in-between. More surprising is the map of the corresponding h_x: the values prescribed near edges are much higher than in smooth regions or textured regions. In fact, even though the oracle prescribes very few pixels to reduce the variance term, one has to go very far in the patch space to gather enough pixels. This is illustrated in Figure 9.2 (notice that here these similarity diagrams are computed on the noisy image). Therefore, *one should use much higher values of h near edges*. Let us stress that this problem is *not* related with the overestimation of the self-weight mentioned above, since the indicator kernel does not suffer from this drawback.

Figure 9.3 shows the PSNR as a function of the size of search window (to be compared with Figure 8.7). This time, the global trend of the PSNR is non decreasing with the size of the search window. The conclusion is that there is no harm in computing *fully non-local* means, provided that we choose well the pixels in the means[1], but there is also little gain in doing so unless the image has large smooth areas (like the *mountain* image). In general, the PSNR tends to stabilize for a side-length $\mathcal{W}$ greater than 25. Slights oscillations of the PSNR are due to the following balance: when increasing the search window, one adds to the mean several relevant pixels that help reduce the variance, but also a few pixels of the first kind (see Section 8.3.3) which perturb the estimation of $\mathbf{u}(x)$. However they are very few, so that on the overall the PSNR is stabilized.

9.2 Stein's Unbiased Estimation of the Risk (SURE) for the NLM

This section shows that the error committed by the NLM can be estimated (without the noise free image!) by using Stein's Unbiased Estimate of the Risk (SURE). This is a way to handle the bias-variance dilemma of Section 8.3 without making the assumption that the weights are deterministic.

[1] Put it this way, this statement looks as a tautology, but one should remember that the expression (9.1) imposes a structure on the choice of the pixels, and it would not be true if pixels of the first kind (see Section 8.3.3) were overwhelming in natural images.

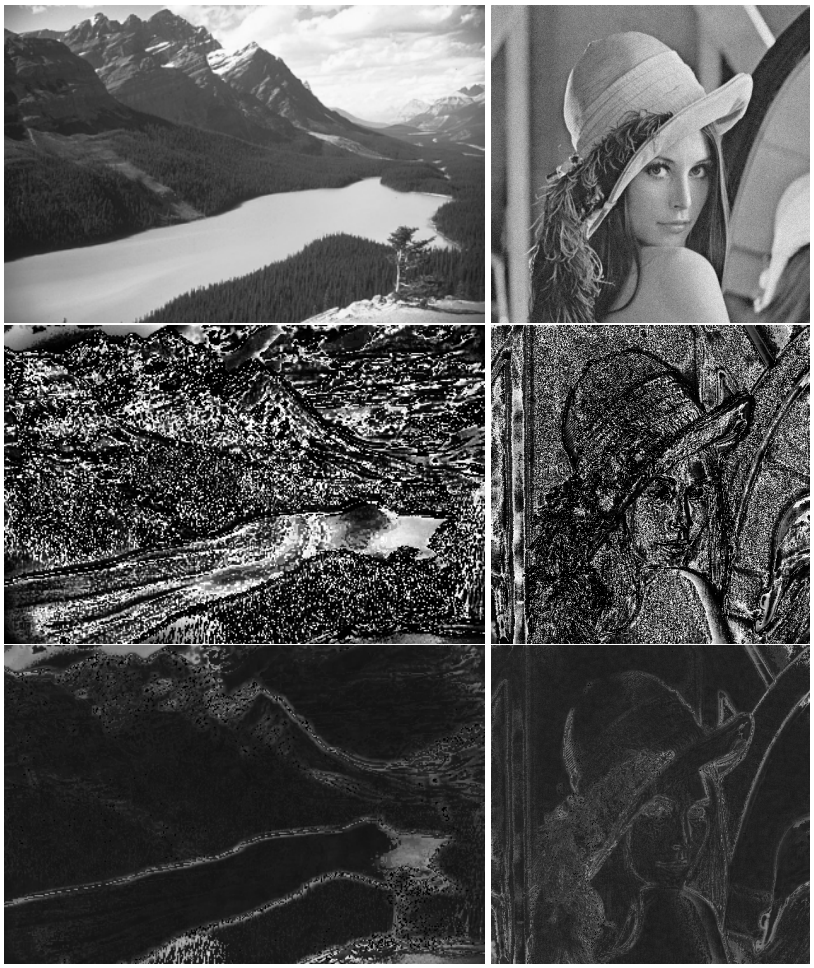
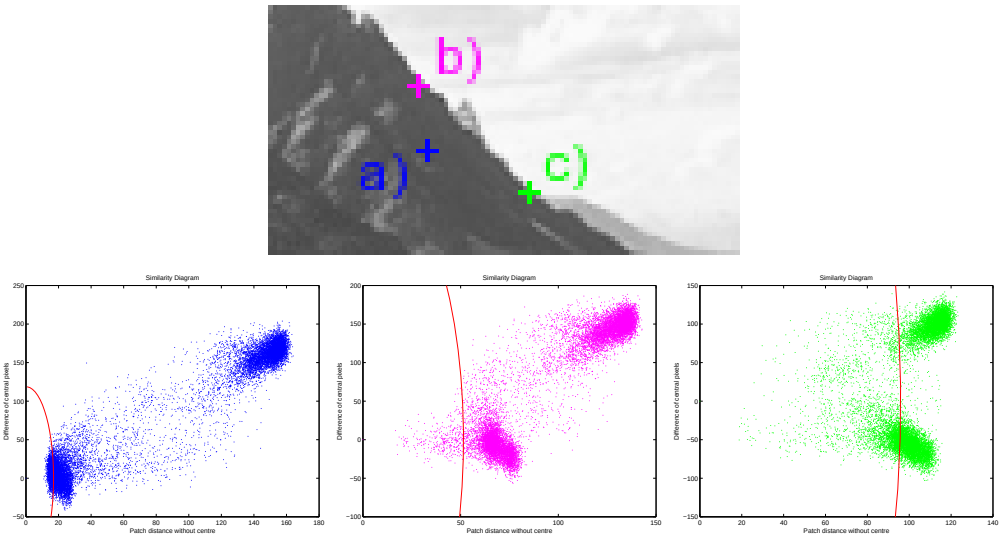


Figure 9.1 – *Map of parameters prescribed by the oracle. The original images were degraded with* $\sigma = 10$*, the patch size is* $s = 7$*. From top to bottom: original images (without noise), map of the number of pixels in the mean, map of the corresponding* h *parameter. In the second row, the white regions represent a number of pixels* $n_x \approx 3000$*. In the third row, first image, the parameter* h *is approximately equal to 14 on the lake, while it ranges from 15 to 20 in the forest, and from 40 to 75 along the edges of the mountains and the lake. On the rocks it is around 30. Although the denoising of edges should be performed with very few pixels, the corresponding parameter* h *should be very large.*

9.2.1 Estimation of the risk

The first applications of SURE date from the 1980's, when it was used to choose the smoothing parameter in families of linear estimates [Li, 1985] such as for model selection, ridge regression, smoothing splines, etc. It was then widely used in the wavelet community [Zhang and Luo, 1999, Combettes and Pesquet, 2004] after the introduction of the SURE-Shrink algorithm [Donoho and Johnstone, 1995] . Solo [Solo, 1996] gave a general form of SURE for an estimator defined as a minimizer of a regular energy, especially for least square regression regularized by a Sobolev norm or the Total Variation. More recently, Blu and Luisier [Blu and Luisier, 2007] proposed to build linear combinations of estimates and to select the coefficients using SURE. Moreover, Ramani *et al.* [Ramani et al., 2008] have proposed a Monte Carlo approach to evaluate SURE when a closed-form expression is not available or too computer-intensive. In the case of the NLM, it was recently shown by Van De Ville and Kocher [Van De Ville and Kocher, 2009] that SURE had a closed-form expression. We had come to the same conclusion when we discovered their paper.

Let us first recall the result by Stein (see [Stein, 1981]):



(a) (b) (c)

Figure 9.2 – *Similarity diagram for three different pixels. The red line indicates the ball of radius h_x prescribed by the oracle (the pixels at the right of this line are not taken into account into the mean). In the interior of homogeneous regions (a), a small radius ($h_x \approx 17$) is sufficient to reduce the variance of the noise. Near edges (b and c), one has to look very far in the patch space to find enough pixels to reduce the variance ($h_x \approx 50$ and 95 respectively). The threshold is especially large in the third case since the compensation of the darker and brighter pixels yields a very small bias.*

Proposition 9.2.1 (Stein). *Let $\mathbf{f}(x) \in \mathbb{R}$, $\boldsymbol{\varepsilon}(x) \sim \mathcal{N}(0, \sigma^2)$, and $\mathbf{u}(x) = \mathbf{f}(x) + \boldsymbol{\varepsilon}(x)$. Denote by $\gamma(\mathbf{u}(x))$ an estimate of $\mathbf{f}(x)$ and assume that $\gamma : \mathbb{R} \to \mathbb{R}$ is absolutely continuous, and*

i) $\lim_{|z|\to\infty} \gamma(\mathbf{f}(x) + z) e^{-\frac{z^2}{2\sigma^2}} = 0$,

ii) $\mathbb{E}(\gamma(\mathbf{f}(x) + \boldsymbol{\varepsilon}(x)))^2 < +\infty$, *and*

iii) $\mathbb{E}|\gamma'(\mathbf{f}(x) + \boldsymbol{\varepsilon}(x))| < +\infty$.

Then the following relation holds:

$$\mathbb{E}|\gamma(\mathbf{u}(x)) - \mathbf{f}(x)|^2 = \mathbb{E}\left((\gamma(\mathbf{u}(x)) - \mathbf{u}(x))^2 + 2\sigma^2 \frac{\partial \gamma(\mathbf{u}(x))}{\partial \boldsymbol{\varepsilon}(x)} \right) - \sigma^2 \,. \tag{9.4}$$

The proof relies on an integration by parts. Let $\mathbf{u}$ be the noisy image, $\hat{\mathbf{f}} = \mathrm{NL}\,\mathbf{u}$ the result of NLM applied to the noisy image using the noisy weights. Then:

$$\boldsymbol{r}(x) = (\hat{\mathbf{f}}(x) - \mathbf{u}(x))^2 + 2\sigma^2 \frac{\partial \hat{\mathbf{f}}(x)}{\partial \boldsymbol{\varepsilon}(x)} - \sigma^2 \,, \tag{9.5}$$

is an unbiased estimate of the risk at pixel x, i.e. $\mathbb{E}(\boldsymbol{r}(x)) = \mathbb{E}|\hat{\mathbf{f}}(x) - \mathbf{f}(x)|^2$. In [Van De Ville and Kocher, 2009], an analytic expression of $\boldsymbol{r}$ is given in the case of the Gaussian weights. The authors show that this estimator yields a very robust estimation of the global mean square error.

In the general case of a kernel φ with compact support, the middle term rewrites:

$$\frac{\partial \hat{\mathbf{f}}(x)}{\partial \boldsymbol{\varepsilon}(x)} = \frac{\varphi(0)}{C_x} + \frac{1}{C_x} \sum_{x'} \mathbf{u}(x') \frac{\partial \omega(x, x')}{\partial \boldsymbol{\varepsilon}(x)} - \left(\frac{1}{C_x} \sum_{x'} \mathbf{u}(x') \omega(x, x') \right) \left(\frac{1}{C_x} \sum_{x''} \frac{\partial \omega(x, x'')}{\partial \boldsymbol{\varepsilon}(x)} \right) , \tag{9.6}$$

where $C_x = \sum_{x'} \omega(x, x')$ is the normalization constant, and

$$\frac{\partial \omega(x, x')}{\partial \boldsymbol{\varepsilon}(x)} = \frac{1}{h^2} \varphi' \left(\frac{d^2_{\boldsymbol{S}}(x, x')}{2h^2} \right) \left(\boldsymbol{S}\,(0) \left[\mathbf{u}(x) - \mathbf{u}(x') \right] + \boldsymbol{S}\,(x - x') \left[\mathbf{u}(x) - \mathbf{u}(2x - x') \right] \right) . \tag{9.7}$$

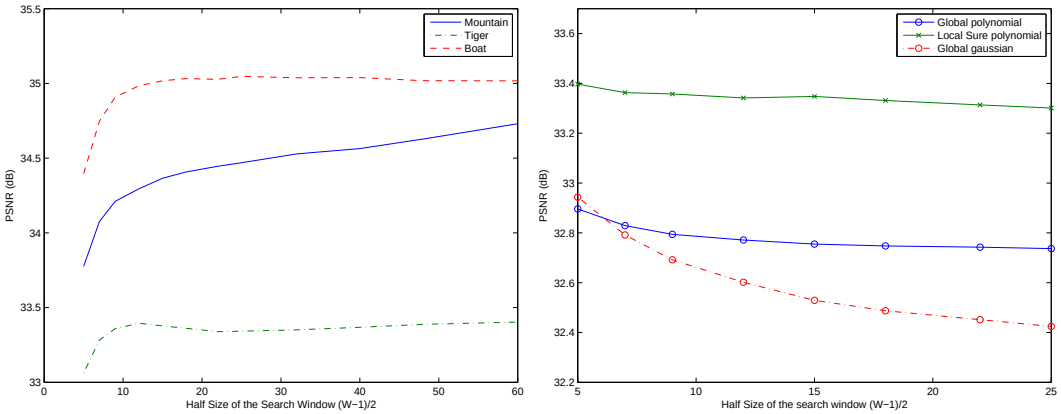


Figure 9.3 – *Evolution of the PSNR as the size of the search window $\mathcal{W}$ increases. Left: for the oracle with indicator weights on different images. The patch size is $s = 7$, the noise level is $\sigma = 10$. With a locally defined smoothing parameter h, there is no or very little loss when using a large window, so that the choice of $\mathcal{W}$ is not a real issue, contrary to Figure 8.7. Right: for the Local Minimizer of the Risk Estimate (MRE) on the image "mountain". The local MRE filter, the NLM with polynomial weight and the classical NLM are displayed (for the last two filters, the parameter h was optimized for PSNR for each size of the search window). The local SURE filter is robust to the size of the search window, mainly because of the compact support of the weights.*

Here, for convenience, we have used the notation for shapes:

$$\boldsymbol{S}(\tau) = \begin{cases} \frac{1}{s^2}, & \text{if } \|\tau\|_\infty \leq \frac{s-1}{2}, \\ 0, & \text{otherwise,} \end{cases} \tag{9.8}$$

so that the term $\boldsymbol{S}(x - x')$ of (9.7) vanishes when x' does not belong to the patch centered at x (i.e. $\|x - x'\|_\infty \leq \frac{s-1}{2}$).

As with the Gaussian weights, this procedure yields a very reliable estimation of the (global) mean square error when it is summed over all pixel x in the image. Notice that it is necessary to compute the NLM for each parameter to estimate the corresponding risk, and that Eq. (9.6) adapts straightforwardly to the trick of replacing $\varphi(0)$ with $\exp(-\frac{\sigma^2}{h^2})$ in the self-weight.

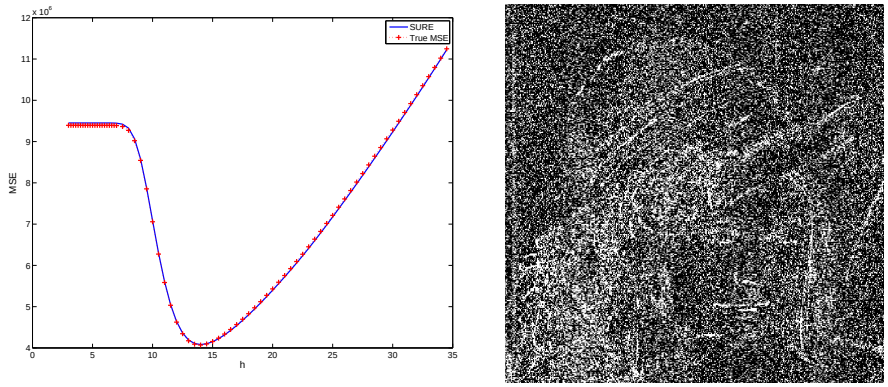


Figure 9.4 – *Left: Evolution of the global Mean Square Error (MSE) as the bandwidth h increases, and its estimation using SURE. Right: map of the estimated risk on the Lena image ($h = 29$).*

9.2.2 Is it legal to average the SURE map?

As illustrated in Figure 9.4, SURE gives a reliable estimate of the global mean square error, but its pointwise values are very oscillating. As a result it cannot be directly used to find the best local parameter.

To circumvent this difficulty, we locally average values of the estimated risk so as to make it more robust.

In [Luisier, 2010], Luisier computes $\mathbb{E}\left(\boldsymbol{r}(x) - (\gamma(\mathbf{u}(x)) - \mathbf{f}(x))^2\right)^2$ for a general estimator. He shows that for a shrinkage operator, this quantity vanishes as the size of the blocks on which the MSE is computed goes to infinity. In the case of NLM, a theoretical proof is not so straightforward, but our experiments confirm that the larger the size of the block, the more reliable the estimation of the MSE (this is why the global SURE is so reliable).

As a consequence, in this chapter and the following we will average the estimated risk on small neighborhoods to have a measure that is as robust and local as possible. The assumptions behind this are that for each pixel:

- $\sum_{i \in \mathcal{N}(x)} \alpha_i \boldsymbol{r}(x+i)$ is a reliable estimator of $\mathbb{E}\left(\sum_{i \in \mathcal{N}(x)} \alpha_i (\hat{\mathbf{f}}(x+i) - \mathbf{f}(x+i))^2\right)$.
- The local MSE, $\mathbb{E}\left[\sum_{i \in N(x)} \alpha_i (\hat{\mathbf{f}}(x+i) - \mathbf{u}(x+i))^2\right]$ is a good approximation of $\mathbb{E}\left(\hat{\mathbf{f}}(x) - \mathbf{u}(x)\right)^2$.

Again we have to face a tradeoff between locality and robustness.

9.3 A local parameter choice for the NLM

9.3.1 Algorithm

In order to select local parameters, we use the estimation (9.5) to minimize the risk depending on the local content of the image (textured areas, smooth regions, ...). Since the pointwise estimation of the risk is not robust, we need to locally average the estimations. The underlying assumption is that the risk is roughly homogeneous within each region (smooth/textured). One should find the right balance between having enough samples to estimate the risk, and keeping a local estimation.

Considering a set of values $\{h_1, h_2, \dots h_n\}$ for the smoothing parameter, we compute for each value the output of the filter $(\mathrm{NL}_i\, \mathbf{u})_{i=1..n}$ and the associated risk map $(\boldsymbol{r}_i)_{i=1..n}$. We convolve each risk map $\boldsymbol{r}_i$ with a disk indicator or a Gaussian with small radius to have a more robust estimation of the local risk. Then we choose for each pixel x the value $i(x)$ that minimizes the convolved risk at pixel x, $\hat{\boldsymbol{r}}_i(x)$, and we retain the corresponding estimation $\mathrm{NL}_{i(x)}\, \mathbf{u}(x)$. We call this selection procedure Local Bandwidth Minimizer of the Risk Estimate (LBMRE) for h, for reasons that will be clear in the next chapter.

Implementation The procedure we propose, LBMRE, is described in Figure 9.5. It is necessary to compute many NLM filters, but this procedure is simpler than several methods proposed in the literature inasmuch as it is *not* iterative. The expensive computations of the patch distances need be performed only once (since all the filters work with the same input image), and as the other computations are independent, they can be parallelized. As an indication our code takes 26 s to execute lines 1 to 15 (the rest is negligible) on a 256×384 image using a search window $\mathcal{W}$ of 23×23 and 64 values of h, on an Intel Core2 Duo 2.5GHz and 4Gb RAM. And the speed can still be improved, since our C code (which uses SSE instructions to vectorize the computations) uses only one of the two cores. Additional tricks could be added like taking advantage of the fact that for each pixel, if a weight is zero for some value h_1, it is necessarily zero for all $h_2 \leq h_1$. Moreover, several approaches proposed in the literature (like the use of integral images to compute the patch distances in [Darbon et al., 2008], or the cluster tree in [Brox et al., 2008] to accelerate the NLM) could be adapted.

9.3.2 Experimental results

In this section, we illustrate the differences between the NLM filter with optimal global parameter (*i.e.* using the value of h that minimizes the true MSE) and the NLM with local parameter, estimated us-

Non-Local Means with Local bandwidth (LBMRE)

for all pixel x **do**
  for all translation $k \in \mathbb{Z}^2$, $|k|_\infty \leq \frac{\mathcal{W}-1}{2}$ **do**
    dist $\leftarrow \frac{\|U(x)-U(x+k)\|^2}{2}$
    for i=1 to n **do**
      $(\sum u\varphi)_i \leftarrow (\sum u\varphi)_i + \mathbf{u}(x+k)\varphi(\frac{\text{dist}}{h_i^2})$
      $(\sum \varphi)_i \leftarrow (\sum \varphi)_i + \varphi(\frac{\text{dist}}{h_i^2})$
      $(\sum u\varphi')_i \leftarrow (\sum u\varphi')_i + \mathbf{u}(x+k)\varphi'(\frac{\text{dist}}{h_i^2})$
      $(\sum \varphi')_i \leftarrow (\sum \varphi')_i + \varphi'(\frac{\text{dist}}{h_i^2})$
    end for
  end for
  for i=1 to n **do**
    $\text{NL}_i\,\mathbf{u}(x) \leftarrow (\sum u\varphi)_i \,/\, (\sum \varphi)_i$
    $\boldsymbol{r}_i(x) \leftarrow \ldots$ (Eq. 9.5)
  end for
end for
for i=1 to n **do**
  $\hat{\boldsymbol{r}}_i \leftarrow \boldsymbol{r}_i * G_\rho$
end for
for all pixel x **do**
  $\text{LBMRE}(x) \leftarrow \text{NL}_{i(x)}\,\mathbf{u}(x)$ where $i(x) = \arg\min_i \boldsymbol{r}_i(x)$
end for

Figure 9.5 – *Local Bandwidth Minimizer of the Risk Estimate (LBMRE) algorithm.*

ing SURE (LBMRE). The indicator oracle (Section 9.1) is also shown. The NLM filters [2] used with SURE are given by the polynomial kernel: $\varphi(x) = \mathbb{1}_{[0,1]}(x)\left(1-(10x^6-24x^5+15x^4)\right)$. Observe that this kernel is smooth enough to apply Proposition 9.2.1. In the following, the local and global versions share the same values for the parameters that are not locally selected and unless otherwise stated, the size of the search window is set to 29×29, the patch size is 7×7. Notice that the use of compactly supported weights already gives a better result than the original Gaussian ones, as explained in [Goossens et al., 2008]. When the noise level is $\sigma = 10$, we take $\{h_1, h_2, \ldots, h_n\} = \{3, 3.5, \ldots, 34.5\}$ (see the previous paragraph) and we scale these values proportionally when σ varies.

Original[3] and noisy images are displayed on Figure 9.6. On Figure 9.7, it is shown that the local selection of h enables one to get rid of the *rare patch* effect, responsible for noisy halos around edges. The advantage of using the local approach is further illustrated in Figure 9.8, where one observes in particular that the macro-texture made of the tiger stripes is better preserved using a local selection of h. Both the PSNR and SSIM are improved in these experiments, as illustrated by Tables 9.1 and 9.2.

Next, we display in Figure 9.9 the maps of prescribed values for h, first using the (ideal) oracle, then using the approach described in Section 9.3.1. Of course, the values prescribed by the oracle are spatially more accurate than those using LBMRE. The corresponding denoising result is also better. It is therefore tempting to try to improve the estimation of h. We performed various attempts in this direction, for example using some non-local regularization of the risk map, relying on weights computed on the noisy image. This did not improve the PSNR but, instead, yielded some visible artifacts and we did not

[2] We do not replace the self weight $\omega(x,x)$ by the maximum of the weights $\omega(x,y)$. This would slightly reduce the rare patch effect described here but it would not solve it, because this effect is due to the configuration of the patch cloud (see Figure 9.2) and even the indicator weights suffer from it. This trick would also favor the loss of details. Moreover, its relevance is questioned in [Salmon, 2010b].

[3] These images are taken from the Berkeley Database [Martin et al., 2001], and so are the mountain image on Figure 9.1 and the country house image in Figure 9.16

pursue further in this direction. Observe also that some abnormally high h values are present on the map obtained from local SURE. They may be explained by very flat risk curves in these areas, as illustrated on the same figure, and therefore do not impair significantly the denoising process.

In the experiments of Figures 9.10 and 9.11, it is shown that the optimal local value of h near an edge strongly depends on its contrast. As a consequence there is no way to globally set h to efficiently denoise edges.

The experiment of Figure 9.12 shows that the LBMRE approach permits the reduction of noise halos and artifacts that appear when the patch size s is increased. Figure 9.13 displays an example where the local approach yields a better preservation of fine details. Last, it is illustrated in Figures 9.15 and 9.14, that this ability to increase the value of s enables to avoid the *mottling effect* (local intensity fluctuation due to a non-robustness of small patch sizes), without washing out textures.

Eventually, we can also locally adapt the patch size using the estimation given by SURE. We will come back to this issue in the next chapter. The gain in letting both s and h vary locally is visually small compared to varying h only, or s only. The explanation is that one degree of freedom already allows to limit the artifacts of the NLM. There is nothing to more to gain in having two degrees of freedom.

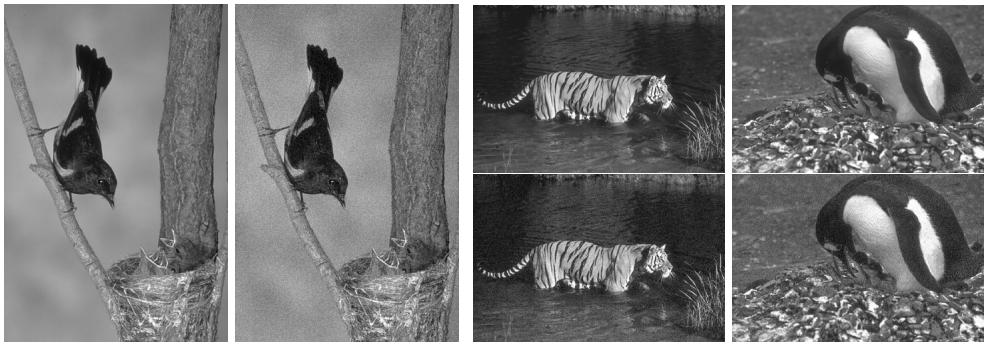


Figure 9.6 – *Original and noisy images ($\sigma = 10$).*

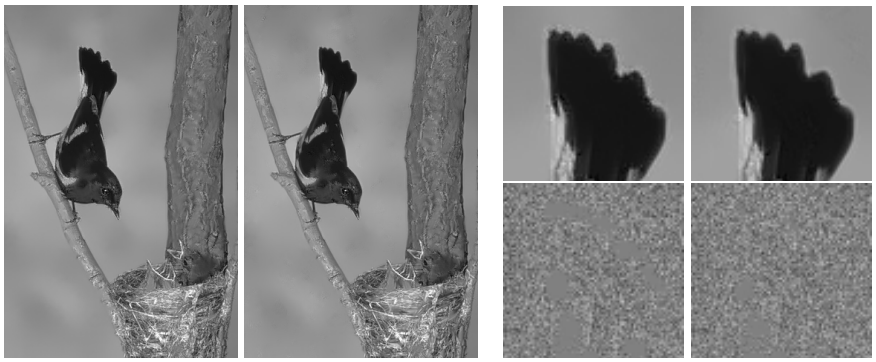


Figure 9.7 – *Bird image. From left to right: NLM with global smoothing parameter h optimized for PSNR, NLM with local h using LBMRE, Zoom of the NLM with global h, local h , and their respective method noise ($\mathbf{u} - \mathrm{NL}\,\mathbf{u}$). Along contrasted edges, the global NLM leaves a noisy halo. This "rare patch effect" is all the stronger as the edge is winding. The local choice of the smoothing parameter corrects this shortcoming.*

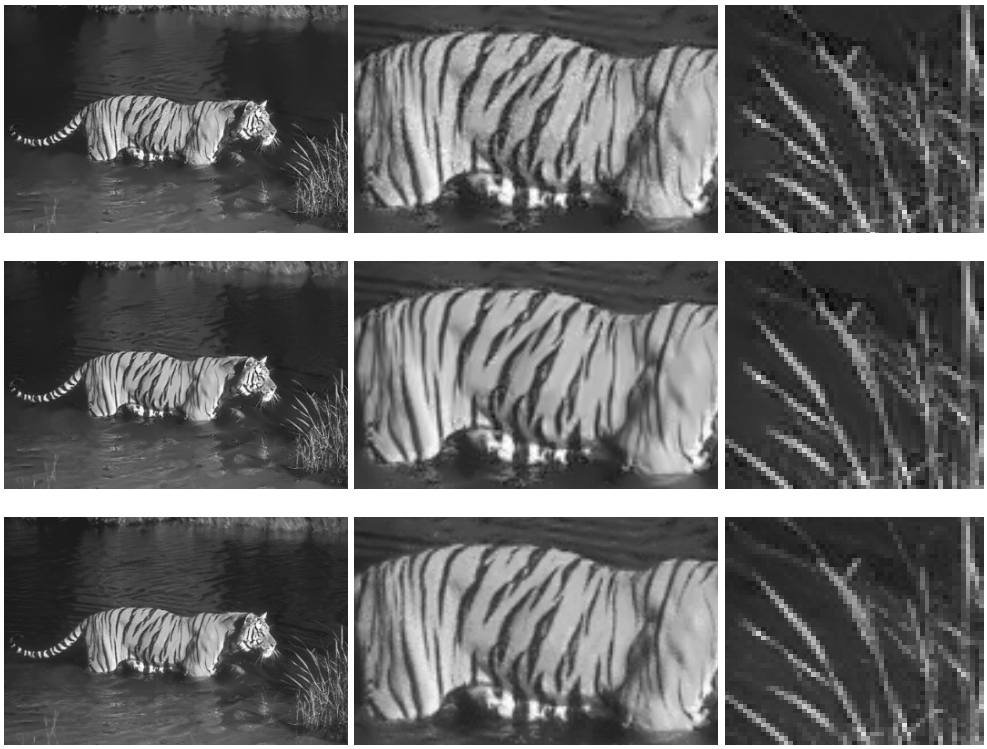


Figure 9.8 – *Tiger image. From top to bottom: NLM with global parameter h optimized for PSNR, NLM with local h (LBMRE), oracle.*

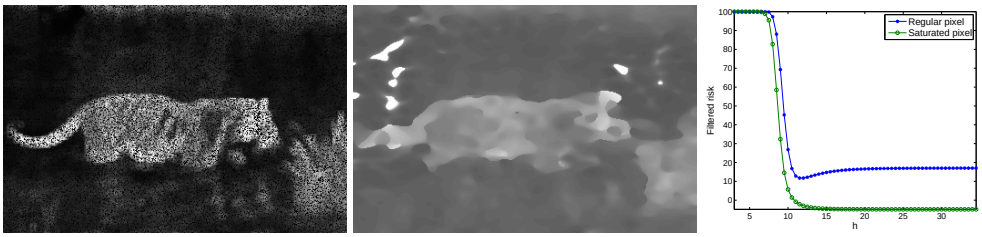


Figure 9.9 – *Map of the prescribed value of h using the oracle (left) and the filtered SURE (middle). The middle map is rough but it shows the same general behavior as the left one. In some areas, the chosen h is very high since the filtered SURE is flat when h varies (right). It has no visual consequence.*

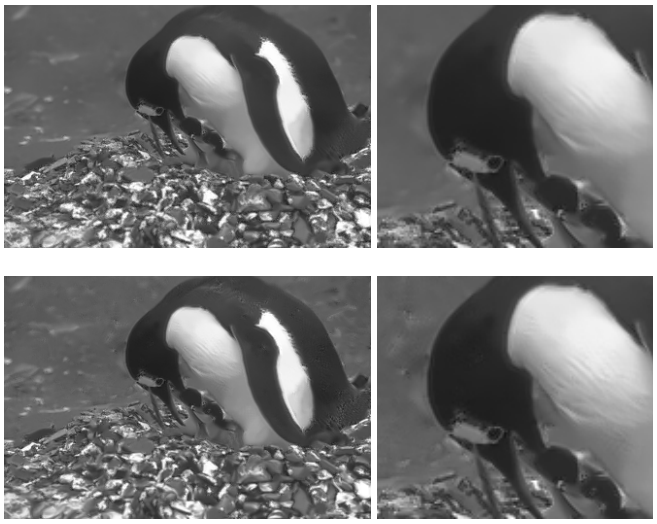


Figure 9.10 – *Penguin image. Top: NLM with global parameter optimized for PSNR. Bottom: local h using LBMRE.*

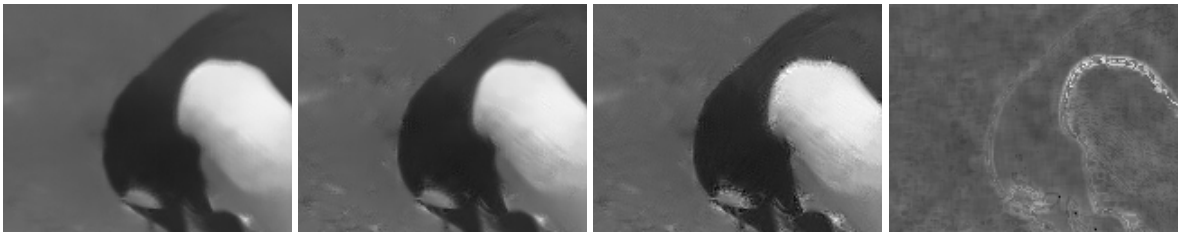
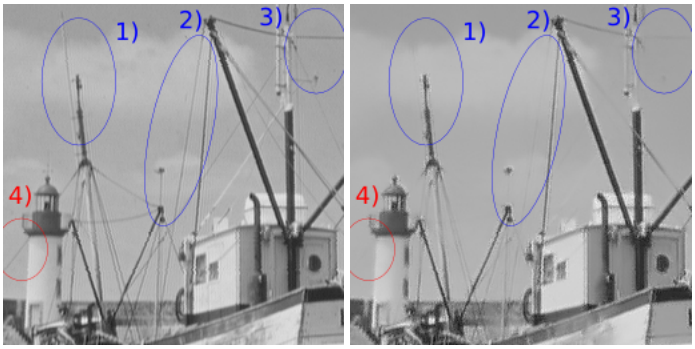


Figure 9.11 – *Experiment with $\sigma = 20$. Left: NLM filter ($h = 30$ optimized for the PSNR). The global optimal parameter is too high for the least contrasted edges, so that, as in (8.8), they are blurred. Middle left: local h using LBMRE. Along the least contrasted edge, the chosen value of h is about 24. If we set the global parameter to 24 (middle right), these edges become sharp but the more contrasted edges become noisy. Right: map of h prescribed by the indicator oracle. The more contrasted the edge, the higher h should be.*

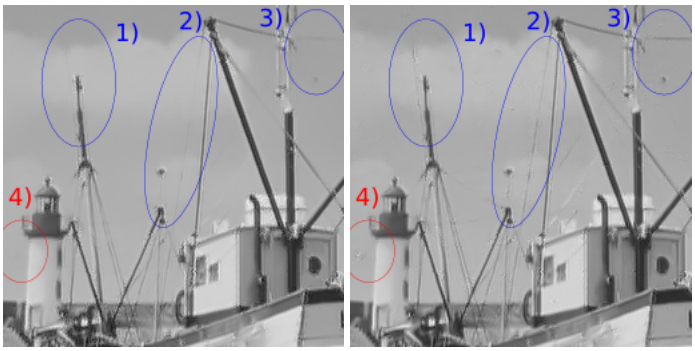
(a) Original (b) Global h, $s^2 = 5 \times 5$ (c) Global h, $s^2 = 7 \times 7$

(d) Noisy $\sigma = 10$ (e) Local h, $s^2 = 5 \times 5$ (f) Local h, $s^2 = 7 \times 7$

Figure 9.12 – *Couple image (only an extract is shown). Top: extract of the original image (a), NLM with global parameter h: the patch size is $s^2 = 5 \times 5$ in (b) (PSNR 32.46 dB), and $s^2 = 7 \times 7$ in (c) (32.14 dB). In both cases, h was chosen to maximize the PSNR. Bottom: noisy image (c) ($\sigma = 10$), NLM with local h using LBMRE: the patch size is $s^2 = 5 \times 5$ in (e) (32.77 dB), and $s^2 = 7 \times 7$ in (f) (32.70 dB). Notice how the face, the tie and the shoulder are smoother with the local SURE (the reader should zoom on this picture) with both patch sizes. Yet the contrast of the wall is not lost.*

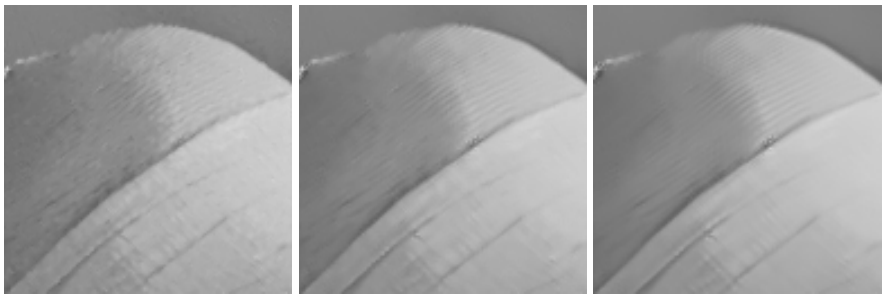


(a)Original image (without noise). (b) Gaussian weights, global h.

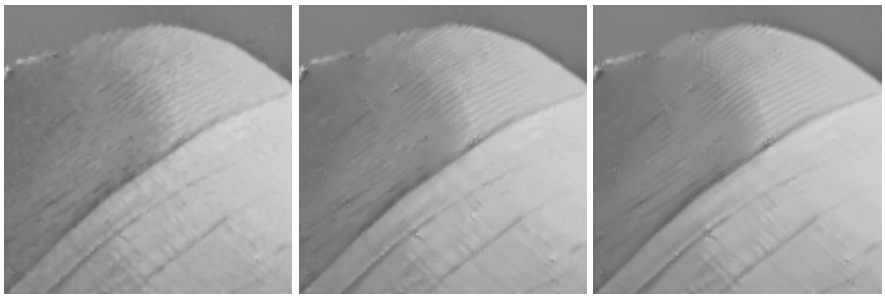


(c) Polynomial weights, global h. (d) Polynomial weights, local h.

Figure 9.13 – *Comparison of the NLM on a noisy image ($\sigma = 10$). In (b) and (c), the parameter h is optimized for PSNR. The difference between them is barely visible. In regions 1),2) and 3), the adaptivity using SURE (d) allows to reconstruct fine structures such as ropes and antennas. However, the filter leaves a noisy spot in region 4) when trying to preserve a fine rope.*

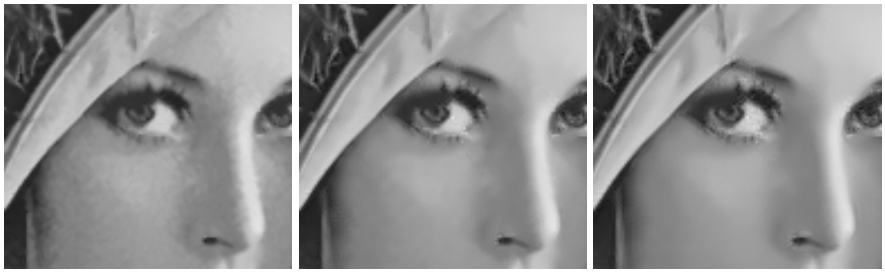


(a) Global h, $s^2 = 3 \times 3$ (b) Global h, $s^2 = 5 \times 5$ (c) Global h, $s^2 = 7 \times 7$

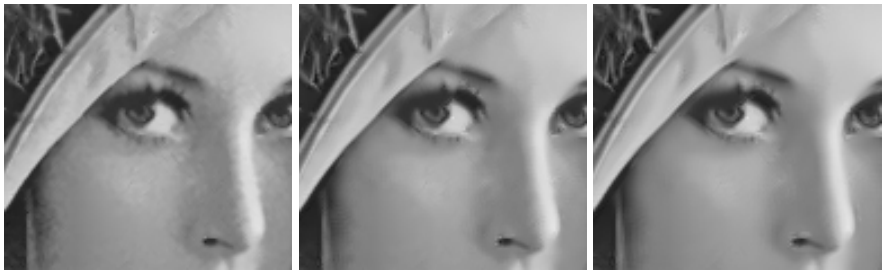


(d) Local h, $s^2 = 3 \times 3$ (e) Local h, $s^2 = 5 \times 5$ (f) Local h, $s^2 = 7 \times 7$

Figure 9.14 – *Patch size and textures. Top: NLM with different patch sizes s^2 using a global parameter h optimized for PSNR. Bottom: Local h using LBMRE. The least contrasted textures are better preserved with a small patch size. However, what looks like texture with patch size 3×3 might as well be the mottling artifact (see below). With the local parameter h, the textures are preserved even with a large patch size.*



(a) Global h, $s^2 = 3 \times 3$ (b) Global h, $s^2 = 5 \times 5$ (c) Global h, $s^2 = 7 \times 7$



(d) Local h, $s^2 = 3 \times 3$ (e) Local h, $s^2 = 5 \times 5$ (f) Local h, $s^2 = 7 \times 7$

Figure 9.15 – *Patch size and robustness to noise (same experiment as in Figure 9.14). With a too small patch size, the algorithm leaves too much noise: Lena's skin looks mottled. As the patch size increases, this effect reduces but the rare patch artifact appears. With a local h, the rare patch effect is reduced, which allows to use large patches.*

9.3.3 A variant using aggregation

The experiments above show that the proposed algorithm allows to preserve some details with low contrast. However, it sometimes leaves a spot of noise where the risk is wrongly estimated (see Figure 9.13). A way to trust less the estimation of the risk is to use the Exponential Aggregation detailed in Chapter 10 At the end of Algorithm 9.5, instead of taking:

$$\mathrm{LBMRE}(x) = \mathrm{NL}_{i(x)}\,\mathbf{u}(x) \text{ where } i(x) = \arg\min_i \hat{\boldsymbol{r}}_i(x) \tag{9.9}$$

we take:

$$\mathrm{LBEWA}(x) = \frac{\sum_{i=1}^{n} e^{-\frac{\hat{\boldsymbol{r}}_i(x)}{T}}\,\mathrm{NL}_i\,\mathbf{u}(x)}{\sum_{i=1}^{n} e^{-\frac{\hat{\boldsymbol{r}}_{h_i}(x)}{T}}}. \tag{9.10}$$

We have used this aggregation procedure with a temperature $T = 0.5\sigma^2$, and we refer to it as Local Bandwidth with Exponentially Weighted Aggregation (LBEWA). Tables 9.1 and 9.2 compare the PSNR/SSIM of the algorithm 9.5 and its variant using aggregation on several images, the noise standard deviation is respectively $\sigma = 10$ and $\sigma = 20$. It shows that a higher PSNR may be reached by averaging different estimators instead of taking the (supposedly) best one at each pixel. However this is not always true, and the result is sometimes not as good as the local MRE for h. Another point to notice is that the BM3D algorithm [Dabov et al., 2007] challenges the oracle in terms of PSNR. This indicates that no matter how well we estimate the risk (with SURE or something else), a pixelwise NLM with local h cannot beat BM3D.

The most notable difference between LBMRE and LBEWA is in fact visual. The aggregation procedure yields images that look slightly more noisy, but more pleasant than the output of the NLM filter (whether with global optimal bandwidth or LBMRE). In textured areas, the NLM tends to blur all the patterns that have little contrast and to do absolutely nothing as soon as the contrast increases. This sharp transition is visually annoying. The LBMRE generally reduces this artifact near edges, but it might fail in smooth regions. The aggregation procedure makes the transition between these regions less shocking since the algorithm leaves a little noise everywhere on the one hand and on the other hand the contrasted regions are less noisy: the aggregation contains smoothed estimations that reduce the effect of the rough ones. In images with textures with little contrast, this usually allows to preserve some more details. As the comparison with BM3D [Dabov et al., 2007] shows, the textures are less blurred (still, on the overall BM3D yield a better visual and numerical quality).

The interest of averaging different estimators in denoising has been known for years (e.g. [Leung and Barron, 2006, Coifman and Donoho, 1995]). In the specific case of NLM, let us mention the aggregation of shifted NLM in [Salmon and Strozecki, 2010a], and the combination of global NLM parameters in [Van De Ville and Kocher, 2011] in the spirit of the SURE-let framework.

Another connection with the EWA aggregation for h is the common practice of taking a convex combination of the denoised image and the noisy one [Buades et al., 2005, Van De Ville and Kocher, 2011]. Indeed, the estimators corresponding to the small values of h are close to the noisy image. However, the estimators corresponding to large values of h correspond to highly smoothed images, so that the EWA aggregation reduces the variance of the noise in regions with high contrast.

Table 9.1 – *Comparison of the PSNR (dB) / SSIM [Wang et al., 2004] for the aggregation* ($\sigma = 10$).

	NLM Gaussian weights		NLM polynomial weights			BM3D	NLM ind. w.
	Global h	Local h (LBMRE)	Global h	Local h (LBMRE)	Local h (LBEWA)		Local h Oracle
barbara	33.08/0.961	33.71/0.968	33.28/0.962	33.84/0.970	33.51/0.966	34.94/0.977	35.21/0.978
boat	31.84/0.938	32.59/0.952	32.13/0.943	32.76/0.955	32.55/0.949	33.86/0.966	33.95/0.963
bird	32.60/0.916	33.11/0.923	32.60/0.917	32.92/0.920	33.07/0.922	33.83/0.931	33.68/0.923
bridge	29.50/0.879	29.98/0.893	29.42/0.879	29.73/0.890	29.86/0.896	30.79/0.911	30.68/0.916
cameraman	32.35/0.910	33.26/0.922	32.52/0.912	33.07/0.921	33.14/0.920	34.08/0.932	34.19/0.925
city	31.46/0.938	32.12/0.949	31.56/0.940	32.08/0.950	32.05/0.943	33.19/0.958	33.67/0.962
country house	30.81/0.823	31.38/0.848	30.78/0.822	31.21/0.841	31.36/0.848	32.02/0.851	32.35/0.880
couple	31.92/0.931	32.58/0.951	32.23/0.941	32.81/0.955	32.47/0.946	34.00/0.967	33.99/0.965
fingerprint	30.34/0.979	30.76/0.987	30.37/0.983	30.69/0.987	30.69/0.989	32.48/0.991	31.94/0.989
flinstones	31.27/0.971	31.93/0.978	31.14/0.971	31.50/0.977	32.04/0.979	32.48/0.980	33.71/0.984
hill	30.55/0.854	31.08/0.868	30.53/0.847	30.90/0.863	31.01/0.870	31.81/0.883	32.13/0.902
house	34.55/0.879	35.20/0.896	34.88/0.887	35.46/0.900	35.10/0.891	36.59/0.918	36.52/0.915
lake	32.14/0.926	32.65/0.936	32.19/0.927	32.59/0.937	32.58/0.930	33.63/0.949	33.60/0.948
lena	34.00/0.947	34.56/0.958	34.26/0.952	34.82/0.960	34.35/0.953	35.86/0.969	35.77/0.967
man	32.08/0.936	32.74/0.949	32.32/0.941	32.87/0.952	32.59/0.946	33.94/0.963	33.99/0.963
mandrill	30.28/0.931	31.02/0.949	30.51/0.930	31.33/0.951	30.98/0.950	33.18/0.966	32.18/0.963
peppers	32.89/0.903	33.54/0.912	33.17/0.906	33.67/0.914	33.41/0.908	34.72/0.927	35.18/0.936
tiger	31.59/0.846	32.28/0.855	31.72/0.846	32.31/0.858	32.17/0.850	33.42/0.887	33.41/0.886
ucla	30.43/0.928	31.03/0.940	30.48/0.932	30.78/0.937	30.93/0.939	31.63/0.948	31.68/0.952
windmill	34.11/0.942	35.13/0.956	34.31/0.946	35.04/0.958	34.83/0.947	35.81/0.966	36.87/0.958

Table 9.2 – *Comparison of the PSNR (dB) / SSIM [Wang et al., 2004] for the aggregation* ($\sigma = 20$).

	NLM Gaussian weights		NLM polynomial weights			BM3D	NLM ind. w.
	Global h	Local h (LBMRE)	Global h	Local h (LBMRE)	Local h (LBEWA)		Local h Oracle
barbara	29.44/0.917	29.81/0.929	30.11/0.929	30.52/0.938	29.12/0.912	31.67/0.952	31.20/0.946
bird	28.79/0.813	29.28/0.836	28.91/0.826	29.36/0.837	28.95/0.813	30.18/0.859	29.85/0.825
boat	28.55/0.860	29.07/0.887	29.03/0.871	29.59/0.897	28.70/0.865	30.80/0.926	30.21/0.915
bridge	25.56/0.729	25.99/0.750	25.58/0.714	25.92/0.741	25.85/0.750	26.83/0.782	26.97/0.809
cameraman	28.73/0.830	29.47/0.852	28.96/0.841	29.57/0.854	29.22/0.833	30.52/0.878	30.28/0.844
city	26.80/0.868	27.44/0.884	27.25/0.882	27.76/0.895	27.36/0.866	29.06/0.916	29.05/0.903
couple	28.32/0.867	28.68/0.884	28.85/0.884	29.29/0.897	28.26/0.864	30.73/0.929	29.85/0.915
country house	27.71/0.703	28.06/0.724	27.77/0.702	28.12/0.724	27.93/0.716	28.88/0.743	28.88/0.772
fingerprint	26.59/0.938	26.86/0.953	27.02/0.951	27.21/0.957	26.52/0.948	28.86/0.972	28.38/0.969
flinstones	27.59/0.941	28.47/0.958	28.17/0.952	28.87/0.961	28.31/0.954	29.60/0.966	30.06/0.967
hill	27.13/0.724	27.48/0.739	27.35/0.721	27.70/0.743	27.16/0.725	28.54/0.778	28.53/0.800
house	31.60/0.841	31.87/0.844	32.32/0.851	32.56/0.851	31.30/0.826	33.77/0.869	32.81/0.838
lake	27.90/0.855	28.33/0.865	28.01/0.863	28.37/0.870	28.13/0.843	29.37/0.893	29.30/0.874
lena	30.93/0.907	31.22/0.916	31.44/0.912	31.70/0.922	30.62/0.902	32.94/0.940	31.89/0.928
man	28.79/0.862	29.18/0.881	29.13/0.867	29.52/0.888	28.68/0.858	30.56/0.917	30.16/0.913
mandrill	26.59/0.840	26.90/0.863	26.91/0.836	27.36/0.872	26.67/0.847	29.11/0.912	28.18/0.906
peppers	29.38/0.845	29.83/0.851	29.83/0.854	30.26/0.858	29.33/0.830	31.38/0.885	31.04/0.867
tiger	28.29/0.741	28.82/0.758	28.46/0.742	29.02/0.764	28.63/0.746	30.01/0.806	29.58/0.776
ucla	26.81/0.824	27.28/0.855	26.73/0.829	27.19/0.855	27.08/0.844	27.94/0.875	28.12/0.892
windmill	30.12/0.899	30.88/0.908	30.69/0.910	31.32/0.915	30.41/0.881	32.05/0.934	32.42/0.890

(a) Original image

(b) Noisy Image ($\sigma = 10$).

(c) Global h.

(d) Local h, minimizer of the risk (MRE)

(e) Local h, EWA aggregation

(f) BM3D

Figure 9.16 – *Comparison on the country house image ($\sigma = 10$).*

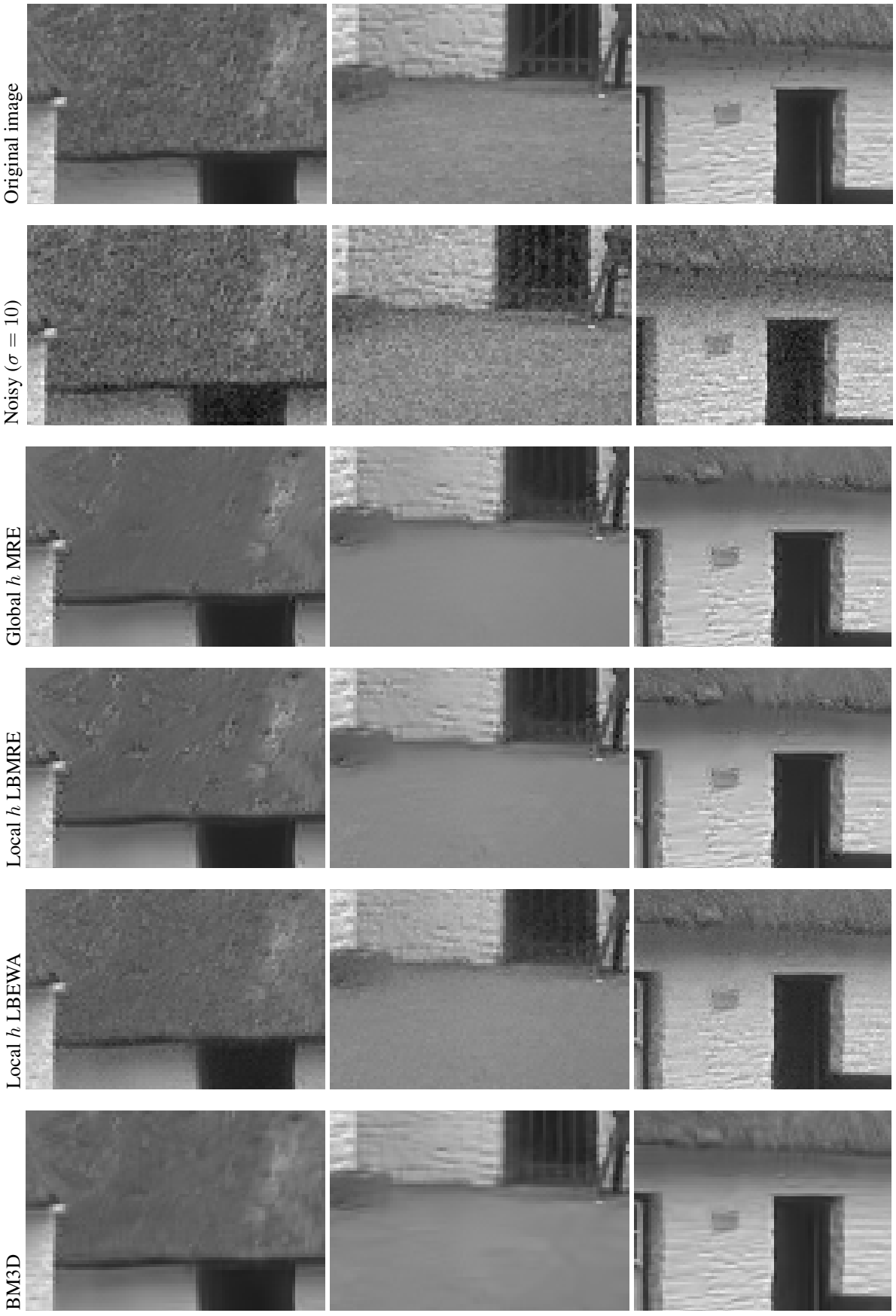


Figure 9.17 – *Zoom of Figure 9.16.*

(a) Original image (b) Noisy Image ($\sigma = 10$).

(c) Global h. (d) Local h, minimizer of the risk (MRE)

(e) Local h, EWA aggregation (f) BM3D

Figure 9.18 – *Comparison on the UCLA image.*

(a) Original image (b) Local h, minimizer of the risk (MRE)

(c) Local h, EWA aggregation (d) BM3D

Figure 9.19 – *Comparison on the UCLA image.*

(a) Original image (b) Local h, minimizer of the risk (MRE)

(c) Local h, EWA aggregation (d) BM3D

Figure 9.20 – *Comparison on the UCLA image.*

Chapter 10

From patches to shapes: NLM-SAP

In the previous chapter, we have introduced an algorithm to locally set the bandwidth parameter h. Although it was originally designed to prevent the degradation of textured areas, the main benefit of this locality is to reduce the noisy halo around rare contrasted areas. Given a similarity diagram (as defined in Chapter 8), the choice of h amounts to defining a cut: in order to reduce the halo one accepts to introduce some bias. The philosophy is not very different from the one of [Brox and Cremers, 2007] who impose a minimum number of similar patches, so that even rare patches are smoothed enough.

On a totally different basis, another method to reduce the noisy halo has been proposed by Salmon and Strozecki [Salmon and Strozecki, 2010a]: in some cases, for a given pixel, it is sufficient to shift the patch used in the computation of the similarities to provide enough similar pixels (see Figure 10.1). This approach is very efficient to remove the noisy halo (when used with a smart aggregation procedure, e.g. the Weighted Average Reprojection described below), but it is constrained by the square shape of the patch: at every location where a square patch cannot fit, the halo effect is likely to appear. A way to circumvent this limitation is to replace the square patches with more general shapes: in narrow regions, one may fit elongated shapes to look for similarities.

The object of this chapter is to describe an extension of the method of Salmon and Strozecki to general shapes.

This chapter relates a joint work with Charles Deledalle and Joseph Salmon[1].

10.1 Non-Local Means with an arbitrary shape

10.1.1 Shapes: a generalization of patches

As we have seen in the last two chapters, the original NLM algorithm sometimes suffers from a *noise halo* around edges because of an abrupt lack of redundancy in the image (see Fig. 10.2). This phenomenon is also called the *rare patch effect*. To reduce the variance of the noise in these areas, we propose to replace the simple square patches with more general shapes (see Fig. 10.3).

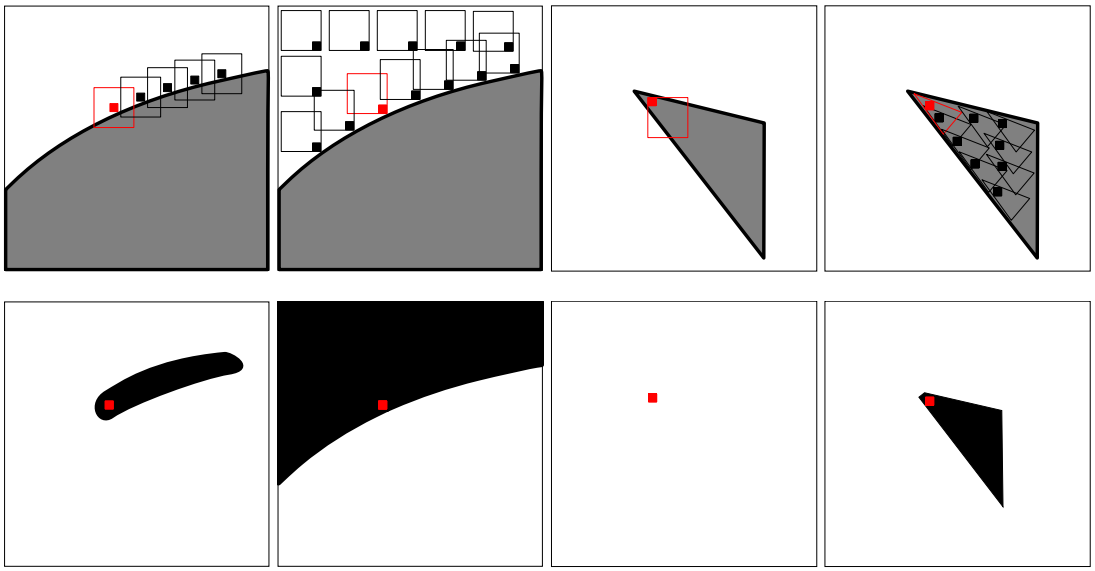
More precisely, the (dis)similarity between pixels is usually measured using a Euclidean norm:

$$d^2(x,x') = \sum_{|\tau|_\infty \leq \frac{s-1}{2}} \frac{1}{s^2} (\mathbf{u}(x+\tau) - \mathbf{u}(x'+\tau))^2 \,, \tag{10.1}$$

or using the weighted Euclidean distance $\|\cdot\|_{2,a}$ as in [Buades et al., 2005]. In fact, to encode a general shape using the metric, we may replace the usual distance with the following weighted ℓ^2 distance :

$$d^2_{\boldsymbol{S}}(x,x') = \sum_{\tau \in \Omega} \boldsymbol{S}(\tau) (\mathbf{u}(x+\tau) - \mathbf{u}(x'+\tau))^2 \,, \tag{10.2}$$

[1] a little anisotropic diffusion of this work towards the last two chapters might even be noticed!



(a) Centered square patch (b) Shifted square patch (c) Shifted square patch (d) Adapted shape

Figure 10.1 – *Similarities using patches/shapes. Top: the geometric configuration and the similar patches. Bottom: the associated weights. If the patch is centered (a) fewer similar patches candidates are found than if the patch is decentered (b). Yet in narrow areas, even decentered patches cannot fit (c) and one should use more complicated shapes to find enough neighbors (d).*

Figure 10.2 – *Illustration of the noise halo appearing around high contrasted edges on images denoised by the NLM. The input noisy images are corrupted versions of the noise-free images presented on Fig. 10.6 damaged by AWGN with standard deviation $\sigma = 20$. Noise halo arises from an abrupt lack of redundancy around edges sometimes referred as the rare patch effect.*

where for all τ, $0 \leq \boldsymbol{S}(\tau)$ and $\sum_{\tau \in \Omega} \boldsymbol{S}(\tau) = 1$. As before Ω refers to the whole image. In this chapter, since we use the Discrete Fourier Transform, the image is assumed to be periodic, so that in the above sum $\mathbf{u}(x+\tau)$ makes sense for all $\tau \in \Omega$.

With this notation, we can easily rewrite the original NLM with a simple shape $\boldsymbol{S}$ by choosing:

$$\boldsymbol{S}(\tau) = \begin{cases} \frac{1}{s^2}, & \text{if } \|\tau\|_\infty \leq \frac{s-1}{2}, \\ 0, & \text{otherwise,} \end{cases} \tag{10.3}$$

The interest of this notation is that $\boldsymbol{S}$ may encode a more complicated geometry than a simple square.

Changing the shape amounts to changing the metric in the patch space and therefore the similarity diagram. In the patch space, the modification of the coefficients of $\boldsymbol{S}$ changes the level sets of the distance from balls to ellipsoids. Notice that the axes of the ellipsoid will always coincide with the canonical axes of the patch space. This is clear since the weighted ℓ^2 distance corresponds to the usual Euclidean distance after a diagonal map in the patch space. To consider ellipsoids with different axes,

one should apply a more general linear map before computing the ℓ^2 norm. This is out of the scope of this work. Let us mention that recently Van De Ville and Kocher [Van De Ville and Kocher, 2011] have computed an expression of SURE when the distance is computed after a general linear map.

The main point with the framework of shapes is that in the next Section we present an algorithm which is able to compute very efficiently the NLM with shapes, using the Fast Fourier Transform (FFT) to compute the distances in Eq. (10.2). Our implementation is independent of the shape, and can thus be applied with different shapes (see Section 10.1.2 for details). As soon as we consider the use of anisotropic shapes, and not just squares or disks centered on the pixel of interest, two questions emerge. The first one is how to choose the collection of shapes to consider. The second issue is to propose a way to combine the estimators provided by each shape.

The collection of shapes should be composed of more than one shape to locally take into account the geometrical properties of natural images. The collection should therefore be diversified enough to identify directional features (see Fig. 10.8 for a visual illustration). At the same time, it should remain small enough so that the algorithm is not computationally intensive.

10.1.2 Fast algorithm to handle shapes

In this section, we present a fast way to compute the NL-Means weights for general shapes, based on the 2D-FFT. It is inspired from works initiated in [Wang et al., 2006] and [Darbon et al., 2008] to speed up the NL-Means algorithm. However, contrary to these approaches, ours can deal with non-square and/or non-binary patches, i.e., with general shapes $\boldsymbol{S}$. Like them, our method is independent of the shape size. Let us also mention that other fast implementations of the NL-Means have been proposed [Mahmoudi and Sapiro, 2005, Bilcu and Vehvilainen, 2008, Pang et al., 2009]: such methods use a pre-selection of the patches based either on statistical tests or comparisons of the gradient. We should however emphasize that the final estimates with those approaches are approximated solutions of the original NL-Means. Our method does not rely on such tricks and computes the exact NL-Means in the case of a square shape in Eq. (10.2).

In [Wang et al., 2006] and [Darbon et al., 2008], the authors propose to compute the Euclidean distances using "Summed Area Tables" [Crow, 1984] (also called "Integral Images" [Viola and Jones, 2001]). This allows them to reduce the computational cost of the NL-Means from $O(|W|\cdot|\Omega|\cdot|\mathsf{P}|)$ to $O(|W|\cdot|\Omega|)$, where $|W| = \ell^2$ is the number of pixels in the search window, $|\Omega|$ is the image domain size and $|\mathsf{P}|$ is the patch size (we refer to [Darbon et al., 2008] for more details). To compute these integral images, the authors change the original algorithm by swapping the two "for" loops: instead of considering all the shifts for each pixel, they consider all the pixels for each shift.

We use basically the same swapping trick. Notice that Equation (10.2) can be reformulated for any translation parameter δ (i.e., taking $x' = x + \delta$) as a discrete convolution:

$$\begin{aligned} d^2_{\boldsymbol{S}}(x, x+\delta) &= \sum_{\tau\in\Omega} \boldsymbol{S}(\tau)\,(\mathbf{u}(x+\tau) - \mathbf{u}(x+\delta+\tau))^2 \\ &= (\check{\boldsymbol{S}} \star \boldsymbol{\Delta}_\delta)(x)\,, \end{aligned} \tag{10.4}$$

where $(\check{\boldsymbol{S}}(\tau) = \boldsymbol{S}(-\tau)$, $\boldsymbol{\Delta}_\delta(x) = (\mathbf{u}(x) - \mathbf{u}(x+\delta))^2$ and $\star$ is the convolution operator. This term can be interpreted as the correlation between the shape $\boldsymbol{S}$ and the square difference of the observe image and the δ-shifted version. The convolution $\check{\boldsymbol{S}} \star \boldsymbol{\Delta}_\delta$ can be computed quickly thanks to following relation:

$$\check{\boldsymbol{S}} \star \boldsymbol{\Delta}_\delta = \mathcal{F}^{-1}(\mathcal{F}(\check{\boldsymbol{S}})\mathcal{F}(\boldsymbol{\Delta}_\delta)) = \mathcal{F}^{-1}(\overline{\mathcal{F}(\boldsymbol{S})}\mathcal{F}(\boldsymbol{\Delta}_\delta))\,, \tag{10.5}$$

where $\mathcal{F}$ is the 2D discrete Fourier transform (2D-FFT) and $\mathcal{F}^{-1}$ is its inverse transform. According to Equation (10.5), and given a translation δ, we only need to perform one term by term multiplication in Fourier domain and two 2D-FFT (note that $\mathcal{F}(\boldsymbol{S})$ can be computed off-line). The repetition of this procedure for every translation δ covering the search window, leads to an algorithm (whose pseudo-code is detailed in Fig. 10.4) with a complexity of $O(|W|\cdot|\Omega|\cdot\log(|\Omega|))$. After submitting this work to a

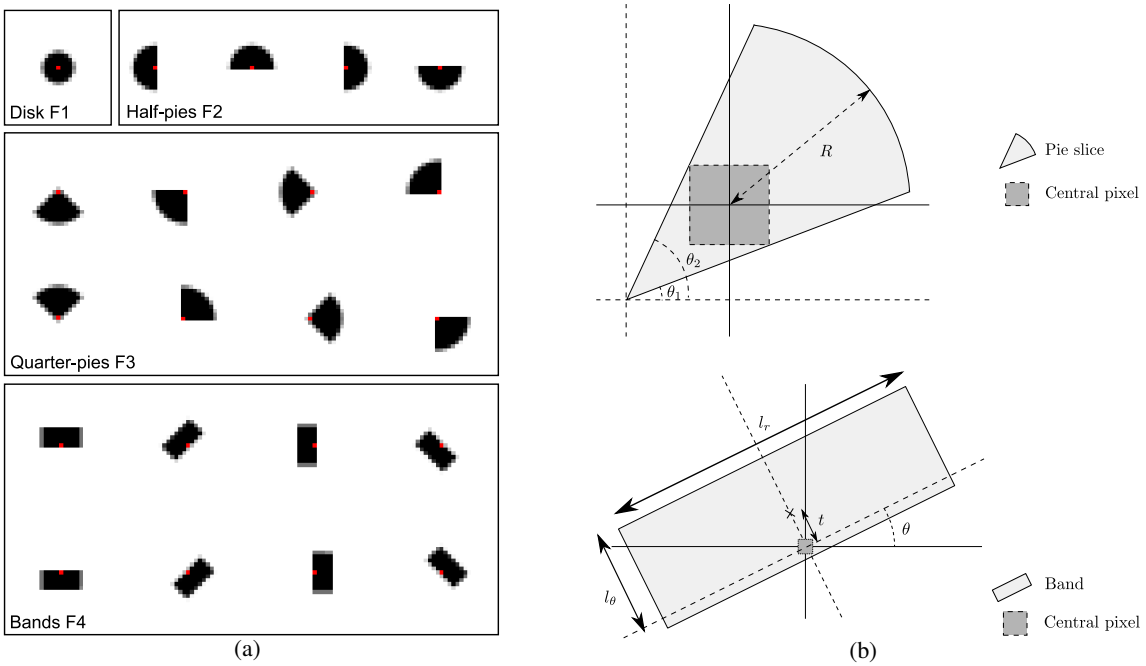


Figure 10.3 – *(a) Examples of shapes considered. The "central" pixel is shown in red, dark pixels illustrate high weights. Shapes are grouped in four categories: F1. the isotropic disk family, F2. the half-pies family, F3. the quarter-pies family and F4. the bands family. (b) Parametrization of the pie slices and bands.*

journal we learnt that Laurent Condat had simultaneously proposed a similar algorithm using recursive filters [Condat, 2010].

10.1.3 Families of shapes

The main purpose of this chapter is to show that the use of different shapes allows to reduce the *rare patch effect*. This point of view is a generalization of the NLM based on square patches with the reprojection studied in [Salmon and Strozecki, 2010a], since each translated patch can be regarded as a de-centered shape. Here, h is fixed and the challenge is to find shapes with enough similar candidates in the search window to reduce the noise.

We now present several types of families that we have considered. The first collections consist of classical squares and disks shapes. Then, we propose more directional shapes such as *pie slices* and *bands* displayed in Fig. 10.3.

Squares: To begin with, we apply our framework to the most commonly used shapes, i.e., the square shapes of odd length (so that the squares have a central pixel). For instance, choosing:

$$\boldsymbol{S}(\tau) = \begin{cases} \frac{1}{s^2}, & \text{if } \|\tau\|_\infty \leq \frac{s-1}{2}, \\ 0, & \text{otherwise,} \end{cases} \tag{10.6}$$

leads to the classical (simplified) NLM definition with square patches of size $s \times s$ and distance between patches measured by the Euclidean norm.

Gaussian: The original, but less common choice, is to set:

$$\boldsymbol{S}(\tau) = \begin{cases} \frac{1}{Z}\exp(-(\tau_1^2+\tau_2^2)/2a^2), & \text{if } \|\tau\|_\infty \leq \frac{s-1}{2}, \\ 0, & \text{otherwise,} \end{cases} \tag{10.7}$$

Algorithm 2D-FFT NL-Means for an arbitrary shape

Inputs: noisy image $\mathbf{u}$, 2D-FFT of the shape $\mathcal{F}(\boldsymbol{S})$
Parameters: search window W, bandwidth h
Output: estimated image $\hat{\mathbf{f}}$
Initialize accumulator images $\mathbf{A}$ and $\mathbf{B}$ to zero
for all shift vector δ in the search window W **do**
Compute the square difference image $\boldsymbol{\Delta}_\delta$

$$\boldsymbol{\Delta}_\delta(x) := (\mathbf{u}(x) - \mathbf{u}(x+\delta))^2 \text{ for all pixels } x \text{ in } \Omega$$

Compute the 2D-FFT $\mathcal{F}(\boldsymbol{\Delta}_\delta)$
Perform the convolution of $\boldsymbol{\Delta}_\delta$ by the shape $\check{\boldsymbol{S}}$

$$d^2_{\boldsymbol{S}}(\cdot, \cdot + \delta) \leftarrow \left(\mathcal{F}^{-1}\left(\overline{\mathcal{F}(\boldsymbol{S})}\mathcal{F}(\boldsymbol{\Delta}_\delta)\right)\right)(\cdot)$$

▷ $O(|\Omega| \cdot \log|\Omega|)$ operations using 2D-FFT

for all pixels x in Ω **do**
Compute the weights

$$\omega(x, x+\delta) = \varphi\left(\frac{d^2_{\boldsymbol{S}}(x, x+\delta)}{2h^2}\right)$$

Update the accumulators

$$\mathbf{A}(x) \leftarrow \mathbf{A}(x) + \omega(x, x+\delta)\mathbf{u}(x+\delta)$$
$$\mathbf{B}(x) \leftarrow \mathbf{B}(x) + \omega(x, x+\delta)$$

end for
end for
Final (normalized) estimator $\hat{\mathbf{f}}(x) = \frac{\mathbf{A}(x)}{\mathbf{B}(x)}$ for all pixel x
Note: the central pixel ($\delta = 0$) is treated as a special case

▷ see Section 10.3 for details

Figure 10.4 – *NL-Means pseudo-code for an arbitrary patch shape $\boldsymbol{S}$. Pre-computations (based on 2D-FFT) of distances between shapes from the noisy image and shapes from its shifted version leads to a smaller complexity of $O(|W| \cdot |\Omega| \cdot \log|\Omega|)$, independent of the shape $\boldsymbol{S}$.*

where Z is a normalizing factor. Equation (10.7) means that the norm $\|\cdot\|_{2,a}$ is used to measure the distance between patches. This limits the influence of square patches corners and leads to a more isotropic comparison between patches.

Disks: Disk shapes are defined in the same way, using the Euclidean norm instead:

$$\boldsymbol{S}(\tau) = \begin{cases} \frac{1}{Z}, & \text{if } \|\tau\|_2 \leq \frac{p-1}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{10.8}$$

A non-binary version may also be defined for pixels crossed by the boundary.

Pie slices: We study a family of shapes, denoted as "pie", whose elements are defined with three parameters: two angles and a radius. These shapes represent a portion of a disk delimited by two lines and

surrounding the discrete central pixel.

Bands: This family of shapes is simply composed of rectangles, potentially rotated and decentered with respect to the pixel of interest.

10.2 Aggregation of shape-based estimates

In the previous chapter, we have considered many NLM estimators, and at each pixel we have kept the estimator with the smallest estimated risk. However, SURE does not yield a perfect estimation of the pointwise square error, therefore it is sometimes better to combine the different estimators instead of keeping a single one. This is called the aggregation problem.

In this section we investigate several ways to aggregate the NLM estimators based on different shapes of "patches". We have extended the standard square shape to other shapes such as *disks*, *pies* or *bands* (see Section 10.1.3). Thus, the new goal in this context is to determine how to locally take the most of each proposed denoiser.

Assume that for any pixel x in the image, we have built a collection of K pixel estimators $\hat{\mathbf{f}}_1(x), \cdots, \hat{\mathbf{f}}_K(x)$ based on different shapes, and that estimates of their corresponding performances are available. With this information at hand, we can address different aggregation/optimization problems. Indeed, the ways of combining the estimators at hand may depend on the theoretical aggregation problem we aim to solve (as described by [Nemirovski, 2000] and [Tsybakov, 2003]):

(S) The selection problem: finding the best estimator among $\hat{\mathbf{f}}_1(x), \cdots, \hat{\mathbf{f}}_K(x)$.

(L) The linear problem: finding the best linear combination of $\hat{\mathbf{f}}_1(x), \cdots, \hat{\mathbf{f}}_K(x)$.

(C) The convex problem: finding the best convex combination of $\hat{\mathbf{f}}_1(x), \cdots, \hat{\mathbf{f}}_K(x)$.

Before looking into these problems, let us recall that the risk should be filtered before considering its pointwise value, as in the previous chapter.

10.2.1 Uniformly weighted aggregation (UWA)

This is the simplest form of aggregation that we consider. The idea is to give the same weight to any shape-based estimator. With few shapes it is already an improvement in practice (both in term of PSNR and SSIM, see Table 10.3), but as the number of shapes increases, we can take into account positions that are completely irrelevant. It is in essence, the one proposed in [Buades et al., 2005] and referred as blockwise NL-means in [Kervrann and Boulanger, 2006], with shapes being decentered square patches. Such an aggregation procedure still suffers from the *rare patch effect* (see Fig. 10.12, third line, for a visual illustration). We can express for any pixel x the uniform estimate $\hat{\mathbf{f}}_{\text{UWA}}(x)$ as the mean of the shape-based estimates:

$$\hat{\mathbf{f}}_{\text{UWA}}(x) := \frac{1}{K} \sum_{k=1}^{K} \hat{\mathbf{f}}_k(x)\,. \tag{10.9}$$

In our experiments, this method can be considered as the benchmark to be improved when using more complex aggregating strategies.

10.2.2 Variance-based decision, Weighted Average (WAV)

A possible way to limit the halo effect is to adapt WAV-reprojection [Salmon and Strozecki, 2010a] to general shapes. The idea, also proposed by Dabov *et al.* [Dabov et al., 2007] in a different context, is to perform a weighted average of the estimates $\hat{\mathbf{f}}_1(x), \cdots, \hat{\mathbf{f}}_K(x)$. To limit the *noise halo*, each weight

should be chosen inversely proportional to the (approximate) variance of the corresponding estimator. In the context of NLM, this approximate variance can be obtained in closed-form in the same way as in [Kervrann and Boulanger, 2006], assuming that the coefficients $\omega(x, x')$ can be treated as deterministic, as in Chapter 8. It is the sum of the square of the weights, as shown in Eq. (8.15). Measuring the performance of the estimators in term of variance is well justified since the halo effect results from the high variance of our estimators around the edges (see [Salmon and Strozecki, 2010a]). However, it tends to over-smooth the edges and the thin details since it does not consider the bias of each estimator.

10.2.3 SURE-based decisions

A way to take the bias into account is to consider the risk estimate rather than the variance to locally attribute more weight to the estimators with small risks. Stein's Unbiased Risk Estimate (SURE) is described in Section 9.2.1. In [Van De Ville and Kocher, 2009] and in the previous chapter, the aim is to select globally the best bandwidth for a given image. Here, we employ SURE to choose the shape locally (i.e., for each pixel), since it is very unlikely that a single shape should be optimal for a whole natural image.

Let us recall that by Proposition 9.2.1, the following quantity is an unbiased estimator for the k-th shape at pixel x:

$$\boldsymbol{r}_k(x) = (\hat{\mathbf{f}}_k(x) - \mathbf{u}(x))^2 + 2\sigma^2 \frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \boldsymbol{\varepsilon}(x)} - \sigma^2. \tag{10.10}$$

Equivalently: $\mathbb{E}(\boldsymbol{r}_k(x)) = \mathbb{E}|\hat{\mathbf{f}}_k(x) - \mathbf{f}(x)|^2$.

As in the previous chapter, the expression is:

$$\frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \boldsymbol{\varepsilon}(x)} = \frac{\varphi(0)}{C_x} + \frac{1}{C_x}\sum_{x'} \mathbf{u}(x') \frac{\partial \omega(x,x')}{\partial \boldsymbol{\varepsilon}(x)} - \left(\frac{1}{C_x}\sum_{x'} \mathbf{u}(x')\omega(x,x')\right)\left(\frac{1}{C_x}\sum_{x''} \frac{\partial \omega(x,x'')}{\partial \boldsymbol{\varepsilon}(x)}\right) \tag{10.11}$$

where $C_x = \sum_{x'} \omega(x, x')$ is a normalization constant. The main difference is that the distance is now the metric associated with each shape $\boldsymbol{S}$. In particular:

$$\frac{\partial \omega(x,x')}{\partial \boldsymbol{\varepsilon}(x)} = \frac{1}{h^2}\varphi'\left(\frac{d^2_{\boldsymbol{S}}(x,x')}{2h^2}\right)\left(\boldsymbol{S}(0)\left[\mathbf{u}(x) - \mathbf{u}(x')\right] + \boldsymbol{S}(x-x')\left[\mathbf{u}(x) - \mathbf{u}(2x-x')\right]\right) \tag{10.12}$$

As we are using K shapes, we get $\boldsymbol{r}_1(x), \cdots, \boldsymbol{r}_K(x)$, unbiased risk estimates respectively for the shape-based estimators $\hat{\mathbf{f}}_1(x), \cdots, \hat{\mathbf{f}}_K(x)$.

10.2.3.1 Regularizing the risk maps

In practice, the estimation of the risk given by (10.10) is too noisy to guide a local choice of the shape (see Figure 10.5). To make it more robust, it is necessary to locally regularize the risk maps, so as to approximate at each pixel the expectations used in Eq. (10.10): our aim is to find estimates $\hat{\boldsymbol{r}}_k(x)$ close to the true risks $\mathbb{E}[(\hat{\mathbf{f}}_k(x) - \mathbf{f}(x))^2]$. These latter quantities will be referred to as the oracle risks since in our simulations we will compute them using the true image.

In the previous chapter, the convolution of the risk map was an efficient way to estimate the local risk in view of setting h. Indeed, for small values of h, the risk is typically high on both sides of edges, whereas it is low for large values of h. As a consequence, there is no harm in diffusing the risk across edges, the resulting algorithm behaves as expected, i.e. uses large values of the bandwidth near edges.

Here, the problem is different since we want to choose a shape. The anisotropy of shapes implies that on one side of an edge the risk may be low whereas it may be very high on the other side. A convolution diffuses the risks across the edges, and any comparison of the risks associated with each shape might become unstable.

In order to average the risks on each side of edges separately, we have adopted a variant of the Yaroslavsky filter [Yaroslavsky, 1985]. A natural idea is to use the noisy image $\mathbf{u}(x)$ to compute the similarities. Yet, we have found a slightly better way. Up to a constant, the estimator of the risk (10.10) can be decomposed in two terms: the square of the method noise [Buades et al., 2005], and a divergence term $\mathbf{D}_k(x) = 2\sigma^2 \frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \boldsymbol{\varepsilon}(x)}$. This divergence term has little variance compared to the noisy image and to the method noise (see Figure 10.5) and contains all the needed information to guide the averaging process: it is uniformly high in the regions where the halo effect is likely to appear, whereas it is low in smooth regions. As a consequence, we obtain better results by using the following regularized version of the risk:

$$\hat{\boldsymbol{r}}_k(x) = \frac{1}{C(x)} \sum_{x'} \mathbb{1}_{\{|\mathbf{D}_k(x) - \mathbf{D}_k(x')| \leq \kappa\}} \, \boldsymbol{r}_k(x'), \tag{10.13}$$

where the sum is taken over a small neighborhood of x, $C(x)$ is a normalizing constant and κ is a bandwidth parameter. Fig. 10.5 shows that this regularization procedure provides smooth risk maps, following edges of the underlying noise-free image, and finer than without regularization. We have displayed the oracle risk map and the estimated risk map provided by SURE and the Yaroslavsky regularization. For illustration purpose, we also show the decomposition of SURE as the sum of the square of the method noise and the divergence term.

Other attempts to regularize the risk map have been performed (median filter, variants of Perona-Malik diffusion and NL-means). Yaroslavsky regularization provides the best trade-off between computing time, visual and numerical results, and we have adopted this approach.

In the following, we assume that the risk maps $\boldsymbol{r}_1(x), \cdots, \boldsymbol{r}_K(x)$ have been filtered, so that we are allowed to consider pointwise values.

10.2.3.2 Minimizer of the risk estimates (MRE)

With our measure of performance, the most natural way to address the selection problem (S) is to select the shape that minimizes the local risk estimate we have at hand:

$$\hat{\mathbf{f}}_{\text{MRE}}(x) := \hat{\mathbf{f}}_{k^*}(x) \quad \text{where} \quad k^* = \arg\min_k \boldsymbol{r}_k(x)\,. \tag{10.14}$$

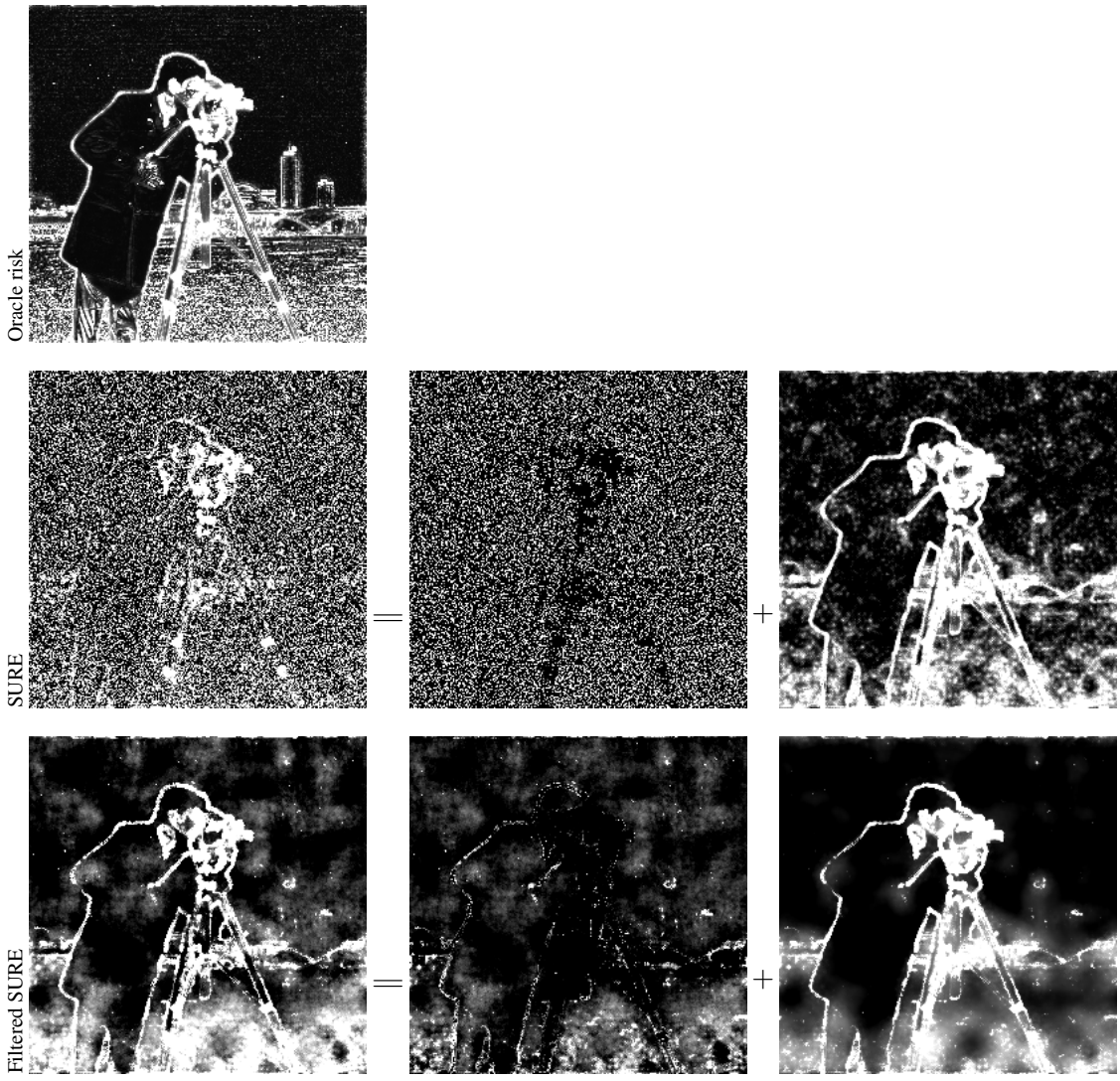
Such a selection rule is all the more relevant as the shapes are really different. In most cases, selecting the locally optimal shape yields satisfying results, but it is sometimes more appropriate to combine some of the best performing estimators as in the next paragraph.

10.2.3.3 Exponentially weighted aggregation (EWA)

In many cases, it might be better to combine several estimators rather than just selecting one. Especially, it happens to be relevant if the best estimators (in term of evaluated risk) are diversified enough or if the risk of the MRE was wrongly under-estimated. Thus, we have also used the statistical method of Exponentially Weighted Aggregation as introduced by Leung and Barron [Leung and Barron, 2006]. This method has been theoretically studied in [Dalalyan and Tsybakov, 2007, Dalalyan and Tsybakov, 2008] and adapted for patch-based denoising in [Salmon and Le Pennec, 2009b]. It consists in aggregating the estimators by performing a weighted average with weights based on the confidence attributed to each estimator, measured in term of the risk. More precisely:

$$\hat{\mathbf{f}}_{\text{EWA}}(x) := \sum_{k=1}^{K} \alpha_k \hat{\mathbf{f}}_k(x)\,, \tag{10.15}$$

$$\text{with} \qquad \alpha_k = \frac{\exp(-\boldsymbol{r}_k(x)/T)}{\sum_{k'=1}^{K} \exp(-\boldsymbol{r}_{k'}(x)/T)}\,.$$



(a) Risk (b) Method noise (c) Divergence

Figure 10.5 – *Maps of the risk associated to a circular shape: (first line) the oracle risk map, (second line) the SURE map without regularization and (third line) the SURE map with Yaroslavsky regularization. Second and third rows correspond to the decomposition* (10.10).

The temperature parameter $T > 0$ is a smoothing parameter, that controls the confidence attributed to the risk estimates. If $T \to \infty$, then the EWA is simply the uniform aggregate $\hat{\mathbf{f}}_{\text{UWA}}$ defined before. Conversely, when $T \to 0$, then $\hat{\mathbf{f}}_{\text{EWA}} \to \hat{\mathbf{f}}_{\text{MRE}}$. Most theoretical works about EWA (see [Leung and Barron, 2006, Dalalyan and Tsybakov, 2007, Dalalyan and Tsybakov, 2008]) recommend a large temperature parameter $T = 4\sigma^2$ under few assumptions (like independence) on the estimators $\hat{\mathbf{f}}_1(x), \cdots, \hat{\mathbf{f}}_K(x)$. In practice, since assumptions on the estimators family may not be satisfied, we have used a smaller value, $T = 0.4\sigma^2$.

10.2.3.4 Minimizing the risk of linear combinations

Suppose we want to solve the linear problem of aggregation. A natural way to find a good linear combination is to solve the following problem:

$$\hat{\mathbf{f}}_{\text{LIN}}(x) := \sum_{k=1}^{K} \alpha_k^* \hat{\mathbf{f}}_k(x)\,, \tag{10.16}$$

$$\text{where} \qquad \alpha^* := \arg\min_{\alpha \in \mathbb{R}^K} \mathbb{E}\left(\sum_{k=1}^{K} \alpha_k \hat{\mathbf{f}}_k(x) - \mathbf{f}(x) \right)^2 .$$

Here, the linearity of the combination allows to use Stein's Lemma (Proposition 9.2.1). It provides an unbiased estimate of the risk of the weighted average estimate:

$$\mathbb{E}\left(\sum_{k=1}^{K} \alpha_k \hat{\mathbf{f}}_k(x) - \mathbf{f}(x) \right)^2 = \mathbb{E}\left(\left(\sum_{k=1}^{K} \alpha_k \hat{\mathbf{f}}_k(x) - \mathbf{u}(x) \right)^2 + 2\sigma^2 \sum_{k=1}^{K} \alpha_k \frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \boldsymbol{\varepsilon}(x)} \right) - \sigma^2.$$

Therefore, neglecting the term that does not depend on α in the last equation, the optimal weights α^* can be obtained by solving the following:

$$\alpha^*(x) = \arg\min_{\alpha \in \mathbb{R}^K} \sum_{k,k'} \alpha_k \alpha_{k'} \mathbb{E}(\hat{\mathbf{f}}_k(x) \hat{\mathbf{f}}_{k'}(x)) + \sum_{k=1}^{K} \alpha_k \left(2\sigma^2 \mathbb{E}\left(\frac{\partial \hat{\mathbf{f}}_k(x)}{\partial \boldsymbol{\varepsilon}(x)} \right) - 2\mathbb{E}\big(\mathbf{u}(x) \hat{\mathbf{f}}_k(x)\big) \right).$$

Using the first order optimality conditions, the last problem amounts to solving a linear system in α. This type of estimator is known to perform quite well for wavelet thresholding estimation where it is referred to as the SURE-LET (for SURE-Linear Expansion of Thresholds, see [Blu and Luisier, 2007]). However, in our framework, this method is not applicable. Indeed, if we work pixelwise, we have only one sample of the SURE for each pixel. This leads to an ill-conditioned system to solve, and the calculated α^* behaves poorly. The reason is that the matrix $[\hat{\mathbf{f}}_k(x)\hat{\mathbf{f}}_{k'}(x)]_{k,k'=1,\dots,K}$ is of rank one, and thus the system to solve is ill-conditioned. If we work on small blocks, these blocks should at least contain K pixels, where K is the number of shapes. However, this is in practice not sufficient since, depending on the noise or on the behavior of the NLM, nothing guarantees that the matrix has full rank or is well-conditioned.

In a global setting (i.e. if the block is the whole image), the matrix is well conditioned with high probability, and Vandeville and Kocher apply this method to globally combine different NLM [Van De Ville and Kocher, 2011] with success. Our attempts to regularize the problem to make it local (e.g. with Tikhonov regularization) did not yield satisfying results.

As a consequence, in the numerical experiments, we only consider aggregation using the Minimizer of the Risk Estimate (MRE), the Weighted Average based on Variance (WAV) and the Exponentially Weighted Aggregation (EWA), which correspond to the selection problem (S) and to the convex problem (C).

10.3 Numerical and Visual Results

This section presents quantitative and qualitative results obtained on four images synthetically corrupted by AWGN.

In all the experiments, **unless otherwise specified**, the NLM-SAP is used with the following default parameters:

- search window: width $\ell = 11$ px,
- shape family: 15 shapes from families F1 and F2 on Fig. 10.3.a with shape areas of 12.5, 25 and 50 px^2,

(a) *Cameraman* (b) *City* (c) *Windmill* (d) *Lake*

Figure 10.6 – *Several* 256×256 *noise-free images for our experiments. These images present edges with high contrast on which the rare patch effect appears.*

- aggregation: EWA with $T = 0.4\sigma^2$,
- risk regularization: Yaroslavsky regularization with search window of size 11×11 px^2 and κ is proportional to the estimated standard deviation of the divergence map.

The computation time is proportional to the number K of shapes used in the NLM-SAP algorithm: we need to perform one NLM-like algorithm per shape. Thanks to the FFT acceleration, the computing time required for one shape, whatever the shape, is of about $2s$ for a 256×256 image with a Matlab implementation on an Intel Pentium 64-bit, 3.00 GHz. For comparison, the naive Matlab implementation of NL-Means takes about $100s$, for square patches of area 7×7 px^2. The computation of one local SURE map, using Yaroslavsky filtering, takes about $0.2s$ per shape. To sum up, NLM-SAP using 15 shapes leads to a computing time of about $32s$ which is still less than the naive Matlab implementation of NLM.

Remark 10.3.1. The focus of this chapter is on the choice of the patch shape to reduce artifacts. To avoid heavy computations, we do not try to find simultaneously an optimal bandwidth h and a combination of shapes. As a consequence, we use the rule proposed in [Buades et al., 2009] for h, $h^2 = 2\sqrt{\frac{8\sigma^4}{|\boldsymbol{S}|}}$ (where $|\boldsymbol{S}|$ is the equivalent size of the shape $|\boldsymbol{S}| = \frac{(\sum \boldsymbol{S}(\tau))^2}{\sum \boldsymbol{S}(\tau)^2}$), and the associated trapezoidal kernel:

$$\omega(x,x') = \begin{cases} 1 & \text{if } \frac{d^2(x,x')}{2h^2} \leq \frac{\sigma^2}{h^2} \\ 1 - \frac{d^2(x,x') - 2\sigma^2}{2h^2} & \text{for } \frac{\sigma^2}{h^2} \leq \frac{d^2(x,x')}{2h^2} \leq \frac{\sigma^2}{h^2} + 1 \\ 0 & \text{for } \frac{d^2(x,x')}{2h^2} > \frac{\sigma^2}{h^2} + 1 \end{cases} . \tag{10.17}$$

Unless otherwise specified, this choice of h and kernel is also the one used in all the numerical illustrations of this chapter.

10.3.1 Behavior of NLM-SAP

In this section, we will study the behavior of NLM-SAP according to some parameters such as the type of shape family, the type of aggregation and the type of risk regularization. Each noisy image is corrupted by AWGN with standard deviation $\sigma = 20$.

Unless otherwise specified, the corrupted images are obtained from four 256×256 noise-free images presented on Fig. 10.6: the famous *cameraman* image and *city*, *windmill* and *lake*[2]. These images are particularly interesting in the study of the Non-Local Means with Adaptive Patch Shapes (NLM-SAP) since they present edges with high contrast for which the classical NLM suffer from the *rare patch effect* (see Fig. 10.2).

[2] three sub-images extracted from Laurent Condat's database: http://www.greyc.ensicaen.fr/~lcondat/

Table 10.1 – *Gain by using multi-scale isotropic shapes in terms of PSNR and SSIM values (PSNR/SSIM). Circular patch shapes respectively with area of 12.5, 25 and 50 px^2 are used and their results are compared to the one obtained when using a combination of these three scales of patches.*

	Cameraman	City	Lake	Windmill
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.456	22.13/0.385
Patch shape area of 12.5 px^2	**29.59**/0.822	28.11/0.873	28.68/0.849	30.91/0.879
Patch shape area of 25 px^2	29.38/0.828	27.94/0.880	28.46/0.855	30.72/0.895
Patch shape area of 50 px^2	29.06/0.825	27.59/0.879	28.33/0.857	30.35/0.899
Combination of these three scales	29.58/**0.844**	**28.32/0.897**	**28.93/0.878**	**31.08/0.912**

(a) Area 12.5 px^2 (b) Area 50 px^2 (c) Combination

Figure 10.7 – *Results obtained with circular shapes of different scales (*12.5 *px*^{2}*,* 50 *px*^{2}*, combination of* 12.5*,* 25 *and* 50 *px*^{2}*). The smallest patch size provide best PSNR but has more artifacts in smooth regions, whereas larger ones suffer from the noise halo. Combination of scales limits those two issues.*

10.3.1.1 Multiscale isotropic shapes

The first experiment consists in choosing locally the patch shape. Table 10.1 and Figure 10.7 illustrate the gain of performance to use multi-scale patch shapes instead of using only one fixed size. In this experiment, we consider three circular shapes of areas: 12.5, 25 and 50 px^2. Comparatively, for the original version of NL-means, the authors suggest to use square patches of fixed size $7 \times 7 = 49$ px^2. In this experiment, the smallest shapes provide the best PSNR. Indeed, we have seen in the previous two chapters that large patches induce bias and *noise halos*. Yet, small patches tend to leave some noise in homogeneous region (in the previous chapter we exhibited a mottling artifact, on Figure 10.7 it is more like a chessboard artifact). The aggregation of these three scales of shapes with our NLM-SAP methods improves slightly the PSNR. Visually speaking, combining different scales of isotropic patch shapes leads to a diminution of both the residual noise and the halo effect.

This is well reflected in Tab. 10.1 by the gain in term of the SSIM criterion which provides quality measurements closer to our perception system. Yet a slight halo still subsists above the hair of the cameraman. We will see in the following that the results can still be improved by considering both multi-scale and anisotropic patch shapes.

10.3.1.2 Families of anisotropic shapes

Table 10.2 gives numerical results obtained by using different families. The compared families are the ones presented on Fig. 10.3.a, i.e., the *disks*, the *half-pies*, the *quarter-pies* and the *bands*. Combinations of these families are also studied. Our experiments show that most suitable shape families, both in terms of PSNR and SSIM, have to contain isotropic shapes, directional shapes and various scales of shapes.

Table 10.2 – *Gain by using anisotropic or mixture of isotropic and anisotropic shapes in terms of PSNR and SSIM values (PSNR/SSIM). The studied patch shapes are the isotropic disks, the half-pies, the quarter-pies, the bands and some combination of them (see Fig. 10.3.a).*

	Cameraman	City	Lake	Windmill
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
Patches with *disk* shapes (family F1)	29.58/**0.844**	28.32/**0.897**	28.93/**0.878**	31.08/0.912
Patches with *half-pie* shapes (family F2)	**29.72**/0.843	28.48/0.896	29.00/0.877	31.29/0.912
Patches with *quarter-pie* shapes (family F3)	29.64/0.842	28.27/0.891	28.89/0.875	31.24/0.912
Patches with *band* shapes (family F4)	29.72/0.841	28.45/0.894	28.98/0.875	31.36/0.912
Combination of shape families F1 and F2	29.74/**0.844**	**28.53/0.897**	**29.04/0.878**	31.32/**0.913**
Combination of shape families F1, F2, F3 and F4	**29.75**/0.842	28.49/0.895	29.02/0.876	**31.40/0.913**

Increasing the number of shapes does not necessarily improve the denoising quality. Using 15 shapes from families F1 and F2 with the three different scales, seems to be a good trade-off between computing time and denoising quality. Figure 10.8 illustrates why using directional shapes is important to reduce the *rare patch effect*. Indeed, each oriented patch shape enables the restoration of edges in the target direction but is inappropriate in the other directions. Then a fine aggregation of them leads to high quality restoration of edges in all directions. Figure 10.9 displays weight maps induced by using patches with only one fixed square shape (i.e., classical NLM) compared to patches with adaptive scales and orientations (i.e., NLM-SAP). For the classical NLM, all the weights are concentrated around the target pixel: this is the *rare patch effect*. For NLM-SAP, the weights are more spread, and other similar pixels are detected thanks to multi-scale and anisotropic patch shapes. It is clear that the limitation of the *rare patch effect* around edges with high contrast leads to a good reduction of the *noise halo*.

10.3.1.3 Aggregation procedure

Table 10.3 presents the numerical performances associated with four aggregation procedures: Uniformly Weighted Aggregation (UWA), Weighted Average based on Variance (WAV), Minimizer of the Risk Estimate (MRE) and Exponentially Weighted Aggregation (EWA). As expected, EWA provides best results, in terms of PSNR and SSIM, since compared to the other three it combines estimates with the best bias-variance trade-off. The local behaviors of NLM-SAP for WAV, MRE and EWA are presented on Fig. 10.10. The average areas and the average orientations of the selected shapes are given for the *cameraman* image. It summarizes for all pixels the information of the average shape as given in Fig. 10.9.d. In smooth regions, anisotropic shapes are not necessarily worse than isotropic ones (like disks or squares). In fact all shapes with the same size should perform equally: weights provided by WAV and EWA are close to uniform distributions.

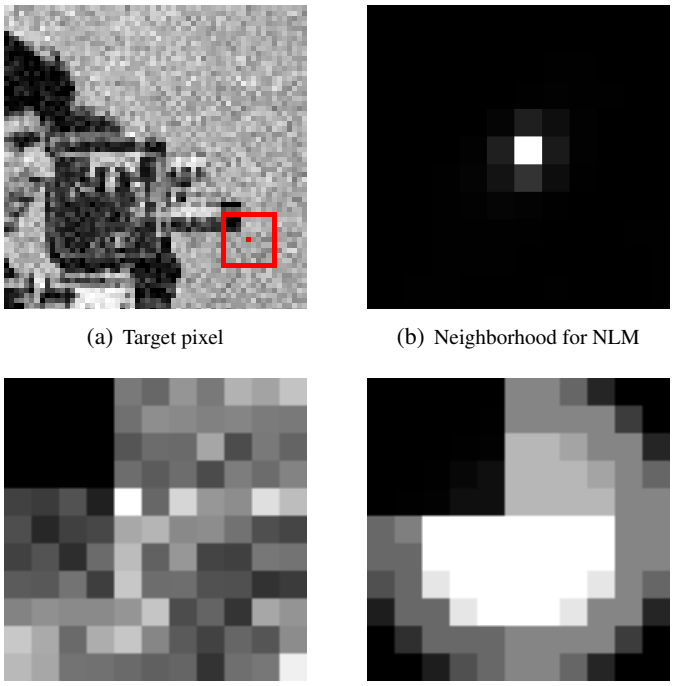
We note that:

- The WAV aggregation scheme is particularly adapted to the reduction of the noise halo, and it gives the most appealing orientation maps. However, this scheme is based on variance and it does not take the bias into account, so that it tends to blur small details with little contrast. This is really minor however.

- The MRE produces very chaotic orientation maps. As in the previous chapter, a slight variation of the risk may induce a brutal variation of the parameter. As a consequence, compared to EWA, MRE suffers from brutal transitions, since it selects only one shape per pixel, while EWA evolves in a smoother way due to the weighted combination of shapes for each pixel.

- The EWA scheme produces orientation maps that are roughly similar to the WAV. In both cases, the selected patch shapes clearly adapt to the local scale and orientation of the image geometry. The chosen sizes of the shapes are smaller around edges and textured areas than in homogeneous areas. The chosen orientation follows the orthogonal orientation of the shapes which is consistent with the remarks given in Section 10.1.1.

Figure 10.8 – *Eight denoised images obtained for different oriented patch shapes. The proposed final aggregate is in the center. Each denoiser provides good performances in a specific target direction but suffers from noise halos in the other directions. The final (central) aggregate takes advantage of every oriented-denoiser to provide high quality restored edges. The patch shape used is indicated in white.*



(a) Target pixel (b) Neighborhood for NLM

(c) Neigh. for NLM-SAP (d) Avg. shape for NLM-SAP

Figure 10.9 – *Illustration of the diminution of the rare patch effect. (a) The noisy image with a highlighted target pixel and its neighborhood. (b) The associated maps of weights obtained by using only square patches of fixed size (i.e., classical NLM). (c) The associated maps of weights with multi-scale and anisotropic patch shapes (i.e., NLM-SAP with F1 and F2). (d) Weighted average of the shapes combined by NLM-SAP.*

Table 10.3 – *Comparisons of different aggregation procedures in terms of PSNR and SSIM values (PSNR/SSIM). The compared aggregation types are UWA, WAV, MRE and EWA.*

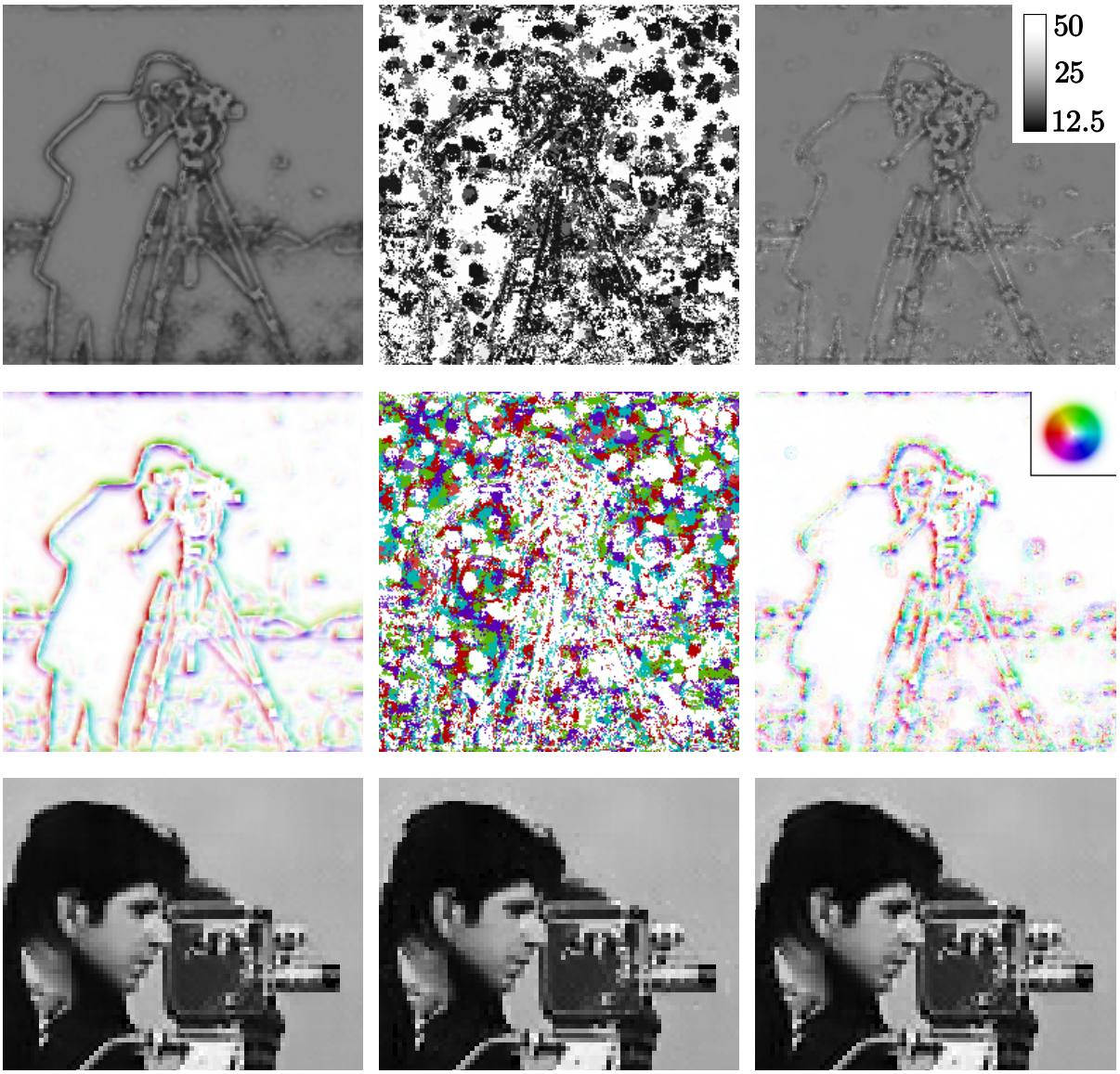
	Cameraman	City	Lake	Windmill
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
Uniform Weighted Average (UWA)	29.60/0.843	28.29/0.895	29.01/**0.879**	31.04/0.910
Weighted Average based on Variance (WAV)	29.64/0.841	28.15/0.887	28.69/0.868	31.10/0.910
Minimizer of the Risk Estimate (MRE)	29.40/0.838	27.99/0.889	28.34/0.864	30.84/0.907
Exponentially Weighted Average (EWA)	**29.74/0.844**	**28.53/0.897**	**29.04**/0.878	**31.32/0.912**

The difference between WAV and EWA is small. It can be noticed on small details with little contrast, like the ear of the cameraman or the top of the camera in Fig. 10.10. The price to pay for EWA is the addition of three parameters: the temperature, and the two parameters of the modified Yaroslavsky filter. Fortunately, we have noticed that these parameters could be set once and for all (as detailed at the beginning of Section 10.3).

10.3.1.4 Regularization method

Finally, we have studied the influence of the regularization of the risk maps on the aggregation results. Three methodologies are compared: aggregation using the noisy risk maps (i.e., SURE maps), the convolved risk maps (using a disk kernel of radius 4) and the risk maps obtained by Yaroslavsky regularization. Table 10.4 gives the corresponding numerical performances and Fig. 10.11 illustrates the behavior of each type of risk map regularization.

The first point to notice is that there is clearly a need for a regularization of the risk map. The aggregation based on the noisy (i.e. not filtered) risk leads to an erratic choice of the shape, which results



(a) WAV (b) MRE (c) EWA

Figure 10.10 – *Aggregation method. Average areas (top) and average orientations (middle line) of selected shapes for different aggregation procedures on a noisy realization of the cameraman image. The average areas and the average orientations are represented using colors whose legends are given on the top right corners. Bottom: Corresponding results focused on the neck of the cameraman. From left to right: WAV, MRE and EWA aggregations. Notice that the ear of the cameraman loses its contrast with the WAV aggregation.*

in a noise halo around edges.

Then comes the question of the choice of the regularization. The risk maps based on convolution and Yaroslavky of SURE maps provide comparable results in terms of PSNR and SSIM (which are far above those of the noisy risk map). However, the choice of the local sizes and orientations of the patch shapes is more relevant in the maps obtained by Yaroslavsky regularization, in terms of scale adaptivity, feature directions and spatial coherency. Using Yaroslavsky filtering, the NLM-SAP acts as expected by selecting big sizes of shapes, even around edges, since the shape orientations have been chosen properly to reduce the *rare patch effect*. By comparison, the convolution forces the size of shapes to be small around edges since it cannot select properly the suitable orientations.

Table 10.4 – *Comparisons of regularization procedures of the risk maps in terms of PSNR and SSIM values (PSNR/SSIM). The compared regularization procedures are the ones using the noisy risk maps directly (i.e., SURE maps), the convolved risk maps and the risk maps obtained by Yaroslavsky regularization.*

	Cameraman	City	Lake	Windmill
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.456	22.13/0.385
Noisy risk maps (SURE maps)	29.13/0.817	27.41/0.865	28.38/0.846	30.15/0.872
Convolved risk maps	29.71/**0.845**	28.49/0.898	**29.13**/0.881	31.26/**0.912**
Yaroslavsky risk maps	**29.74**/0.844	**28.53/0.897**	29.04/0.878	**31.32/0.912**
True risk maps (MSE maps provided by an oracle)	32.09/0.880	32.31/0.938	32.27/0.922	34.43/0.935

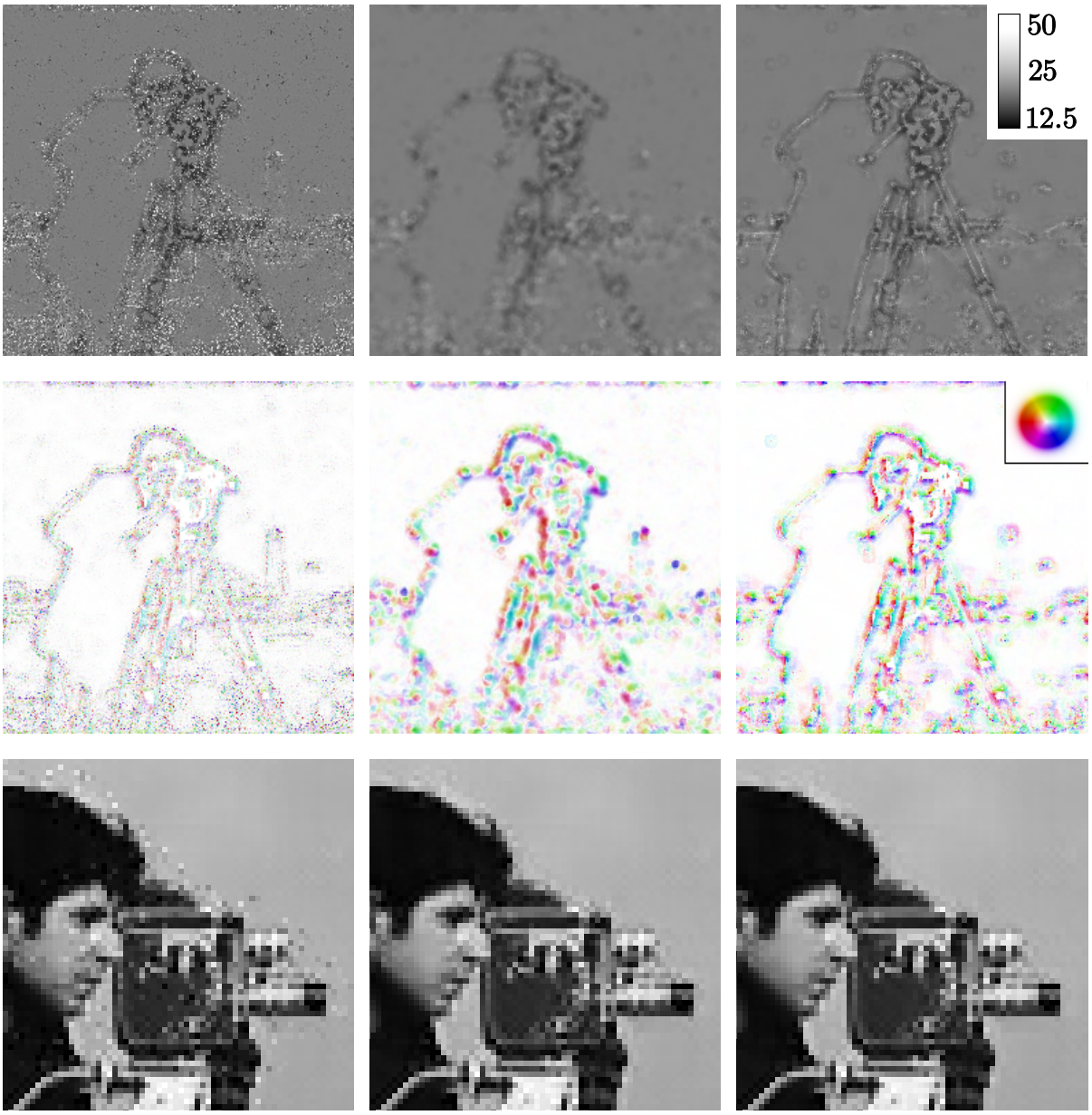
To sum up, the anisotropic regularization is necessary to find a good orientation. What "saves" the convolution is the presence of small shapes in the family of shapes. Visually, on the overall, the differences are slight. Some differences may be noticed around the camera in Figure 10.11.

We have also investigated other regularization strategies (median filter, NL-Means,anisotropic diffusion) but we have not found striking improvements. However, Tab. 10.4 shows that there is still a gap of numerical performance between regularizations of the risk maps and an “oracle risk map” defined as the image of local square errors associated to each shape: $(\hat{\mathbf{f}}_k(x) - \mathbf{f}(x))^2$.

10.3.2 Comparisons with state-of-the-art methods

In this section, the NLM-SAP is compared with *state-of-the-art* denoising methods. The comparisons have been made between the classical (pixelwise) NLM with 9×9 patch size, the blockwise NLM using either UWA or WAV reprojection (with square patches of a single scale), the pixelwise NLM local MRE for h proposed in chapter 9, a refinement of the NLM by Goossens *et al.* [Goossens et al., 2008], the Block-Matching and 3D filtering (BM3D) denoiser [Dabov et al., 2007], and the proposed NLM-SAP approach. Notice that the classical NLM and the NLM with UWA or WAV reprojection, as well as the NLM-SAP use a trapezoidal kernel, whereas the local MRE for h and the NLM by Goossens *et al.* use their respective original kernels.

Table 10.5 shows that NLM-SAP brings a gain of PSNR of about 1 dB compared to the classical NLM (for $\sigma \leq 20$). The SSIM is also usually increased. The BM3D approach leads to better numerical results than all Non Local based approaches. Figure 10.12 gives the visual results. While the blockwise NLM using UWA aggregation and the refinement of the NLM [Goossens et al., 2008] illustrate the *rare patch effect* by the presence of *noise halos*, BM3D and NLM-SAP have reduced a lot this phenomenon. The NLM-SAP provides smooth results with accurate details, such as the head of the cameraman, the house windows, the windmill blades, the tree-trunk and the car. Visually, the quality of images obtained with NLM-SAP challenges those obtained with BM3D.



(a) Noisy risk (b) Convolved risk (c) Yaroslavsky

Figure 10.11 – *Regularization of the risk. Average areas (top) and average orientations (middle line) of selected shapes for different risk maps on a noisy realization of the cameraman image. The average areas and the average orientations are represented using colors whose legends are given on the top right corners. Bottom: Corresponding results focused on the neck of the cameraman. From left to right, results using the noisy risk maps (i.e., SURE maps), the convolved risk maps and the risk maps obtained by Yaroslavsky regularization.*

Table 10.5 – *Numerical performance in terms of PSNR and SSIM values (PSNR/SSIM). The compared methods are the classical (pixelwise) NLM [Buades et al., 2005], the blockwise NLM using UWA reprojection for square patches, the blockwise NLM using WAV reprojection for square patches, the local Bandwidth MRE for h of Chapter 9, BM3D [Dabov et al., 2007], and NLM-SAP.*

	NL-Means	UWA NLM	WAV NLM	LBMRE	Goossens *et al.*	BM3D	NLM-SAP
	$\sigma = 5$						
barbara	36.04/0.986	36.28/0.987	36.48/0.986	37.06/0.986	37.95/0.988	**38.15/0.989**	36.94/0.986
boat	35.41/0.983	35.61/0.983	35.76/0.983	36.13/0.982	36.92/0.985	**37.15/0.986**	36.34/0.984
bridge	34.33/0.958	34.37/0.958	34.43/0.959	34.68/0.961	35.21/0.965	**35.31/0.967**	34.76/0.963
cameraman	36.71/0.949	36.97/0.952	37.29/0.956	37.44/0.956	37.73/0.958	**38.14/0.961**	37.67/0.957
city	35.65/0.963	35.92/0.965	36.32/0.971	36.69/0.974	36.43/0.971	**37.45/0.978**	37.14/0.976
couple	35.73/0.983	35.98/0.984	36.21/0.984	36.47/0.983	37.03/0.985	**37.35/0.986**	36.70/0.983
fingerprint	34.15/0.997	34.16/0.997	34.16/0.997	34.81/0.996	**36.48**/0.997	36.38/**0.997**	34.41/0.997
flinstones	34.85/0.990	35.00/0.990	35.15/0.990	35.21/0.990	35.74/0.991	**36.05/0.991**	35.62/0.991
hill	34.74/0.940	34.83/0.942	34.94/0.943	35.11/0.943	35.73/0.950	**35.96/0.952**	35.43/0.948
lake	36.46/0.960	36.78/0.963	37.10/0.969	37.19/0.970	37.53/0.971	**38.12/0.976**	37.77/0.974
lena	37.07/0.981	37.36/0.982	37.56/0.982	37.77/0.981	38.36/0.983	**38.52/0.984**	37.90/0.981
man	35.87/0.982	36.13/0.983	36.40/0.982	36.76/0.982	37.31/0.984	**37.68/0.985**	36.98/0.982
mandril	34.85/0.986	34.94/0.987	35.02/0.986	35.65/0.983	**37.50**/0.988	37.39/**0.989**	35.41/0.985
windmill	37.75/0.970	38.15/0.972	38.75/0.978	39.27/0.979	38.75/0.978	**39.87/0.983**	39.43/0.980
	$\sigma = 10$						
barbara	32.23/0.969	32.61/0.971	32.89/0.970	33.85/0.970	34.29/0.973	**34.90/0.977**	33.69/0.970
boat	32.00/0.956	32.41/0.958	32.61/0.955	32.77/0.955	33.21/0.962	**33.85/0.967**	32.99/0.953
bridge	29.08/0.884	29.23/0.889	29.40/0.891	29.70/0.887	30.46/0.904	**30.66/0.906**	30.03/0.896
cameraman	32.13/0.909	32.47/0.913	32.85/0.921	33.11/0.920	33.52/0.926	**34.05/0.930**	33.50/0.923
city	30.60/0.922	31.00/0.926	31.49/0.941	32.07/0.948	32.01/0.943	**33.14/0.955**	32.73/0.952
couple	31.99/0.952	32.41/0.954	32.67/0.952	32.81/0.955	33.25/0.959	**33.93/0.967**	33.07/0.948
fingerprint	28.77/0.988	28.83/0.988	28.95/0.988	30.66/0.986	32.14/0.990	**32.41/0.990**	30.44/0.988
flinstones	30.33/0.976	30.71/0.978	31.07/0.977	31.50/0.977	31.68/0.978	**32.40/0.980**	31.85/0.978
hill	30.32/0.859	30.66/0.869	30.96/0.871	30.93/0.863	31.43/0.877	**31.85/0.883**	31.49/0.871
lake	31.64/0.919	32.10/0.924	32.42/0.936	32.56/0.936	32.95/0.940	**33.62/0.949**	33.22/0.943
lena	34.08/0.962	34.47/0.964	34.65/0.962	34.81/0.961	35.34/0.965	**35.79/0.969**	35.00/0.959
man	32.14/0.953	32.53/0.955	32.75/0.951	32.87/0.951	33.34/0.958	**33.90/0.963**	33.20/0.949
mandril	30.11/0.954	30.26/0.955	30.39/0.952	31.29/0.950	32.73/0.960	**33.09/0.966**	31.11/0.948
windmill	33.00/0.938	33.48/0.941	34.06/0.953	35.05/0.957	34.62/0.958	**35.81/0.966**	35.24/0.958
	$\sigma = 20$						
barbara	29.87/0.936	30.30/0.939	30.31/0.937	30.62/0.939	30.95/0.946	**31.76/0.953**	30.41/0.930
boat	29.29/0.892	29.63/0.893	29.55/0.886	29.59/0.897	29.92/0.902	**30.81/0.927**	29.67/0.877
bridge	25.68/0.739	26.11/0.756	26.17/0.743	25.89/0.738	26.20/0.761	**26.76/0.775**	26.24/0.728
cameraman	28.59/0.823	29.01/0.831	29.23/0.838	29.58/0.856	29.49/0.852	**30.34/0.871**	29.74/0.844
city	26.58/0.863	27.07/0.868	27.30/0.877	27.85/0.893	28.00/0.893	**29.06/0.912**	28.53/0.897
couple	29.03/0.892	29.42/0.895	29.41/0.889	29.25/0.893	29.82/0.903	**30.67/0.927**	29.37/0.877
fingerprint	26.48/0.958	26.94/0.960	27.16/0.958	27.20/0.957	27.75/0.965	**28.80/0.972**	27.45/0.951
flinstones	27.19/0.958	27.70/0.959	28.03/0.955	28.92/0.961	28.54/0.962	**29.55/0.966**	29.04/0.960
hill	27.50/0.733	27.86/0.745	27.78/0.735	27.62/0.741	27.99/0.756	**28.51/0.779**	27.83/0.724
lake	27.78/0.854	28.23/0.861	28.18/0.865	28.48/0.872	28.82/0.881	**29.38/0.894**	29.04/0.878
lena	31.61/0.926	31.99/0.928	31.95/0.924	31.67/0.922	32.37/0.932	**32.98/0.940**	31.92/0.918
man	29.34/0.886	29.64/0.887	29.54/0.880	29.55/0.889	29.81/0.895	**30.52/0.915**	29.62/0.872
mandril	27.02/0.869	27.31/0.872	27.35/0.864	27.33/0.867	28.22/0.884	**29.04/0.910**	27.45/0.846
windmill	29.36/0.883	29.94/0.890	30.18/0.901	31.44/0.917	30.85/0.917	**32.06/0.935**	31.32/0.912
	$\sigma = 40$						
barbara	26.65/0.855	26.78/0.858	26.67/0.856	26.86/0.866	27.49/0.879	**28.04/0.896**	26.04/0.831
boat	26.26/0.781	26.38/0.784	26.21/0.779	26.27/0.789	26.74/0.800	**27.64/0.848**	26.06/0.762
bridge	23.01/0.529	23.14/0.540	23.02/0.533	23.16/0.568	23.15/0.552	**23.98/0.615**	22.91/0.511
cameraman	25.89/0.707	26.11/0.724	25.88/0.724	26.29/0.768	26.62/0.766	**27.26/0.801**	26.28/0.749
city	23.50/0.733	23.75/0.746	23.42/0.736	23.77/0.766	24.04/0.779	**25.25/0.829**	23.70/0.747
couple	25.73/0.769	25.87/0.772	25.75/0.767	25.67/0.775	26.33/0.792	**27.43/0.849**	25.39/0.737
fingerprint	23.45/0.855	23.78/0.859	23.72/0.856	23.93/0.885	24.08/0.878	**25.27/0.926**	23.07/0.817
flinstones	24.76/0.903	24.86/0.898	24.63/0.890	25.29/0.915	25.35/0.916	**26.07/0.933**	25.02/0.888
hill	24.57/0.555	24.73/0.568	24.67/0.564	24.63/0.583	24.97/0.590	**25.87/0.659**	24.43/0.540
lake	24.71/0.715	24.88/0.732	24.67/0.727	24.99/0.759	25.25/0.774	**25.86/0.805**	24.79/0.742
lena	28.31/0.846	28.51/0.850	28.42/0.848	28.37/0.853	29.14/0.868	**29.77/0.883**	28.22/0.839
man	26.42/0.779	26.52/0.781	26.45/0.778	26.46/0.785	26.83/0.792	**27.57/0.833**	26.30/0.761
mandril	23.75/0.678	23.83/0.677	23.74/0.671	23.98/0.707	24.16/0.705	**25.22/0.789**	23.39/0.633
windmill	26.41/0.769	26.59/0.787	26.36/0.789	27.46/0.837	27.23/0.839	**28.04/0.878**	26.65/0.815

Figure 10.12 – *Comparisons of the visual denoising performance of the proposed NLM-SAP and other state-of-the-art methods ($\sigma = 20$). From top to bottom, the original and noisy images, the results obtained by the classical (pixelwise) NLM [Buades et al., 2005], the blockwise NLM using UWA reprojection for square patches the BM3D denoiser [Dabov et al., 2007], and the proposed NLM-SAP.*

Table 10.6 – *Comparisons of the proposed approaches for $\sigma = 10$ in terms of PSNR and SSIM values (PSNR/SSIM). All the original images have been taken from the Berkeley database [Martin et al., 2001], except the lighthouse which comes from the Kodak Image Suite.*

	NLM global h	LBMRE	LBEWA	NLM-SAP
Tiger	31.76/0.847	32.36/0.860	32.20/0.850	32.47/0.850
Country house	30.83/0.823	31.19/0.840	31.39/0.849	31.54/0.837
Bear	29.36/0.893	29.90/0.913	30.06/0.917	30.02/0.918
Church	36.51/0.962	36.93/0.963	36.87/0.957	37.47/0.965
Lighthouse	32.45/0.934	32.88/0.946	32.79/0.941	33.24/0.945
Pyramid	35.66/0.902	35.90/0.903	35.34/0.894	36.08/0.901

10.3.3 Comparison with the spatially adaptive choice of h

To close this chapter, let us comment a few visual results which compare the three proposed methods of this thesis: LBMRE (local h, chosen as the minimizer of the risk estimate), LBEWA (local h, using Exponentially Weighted Aggregation) and NLMSAP (global h but local choice of the shape using EWA aggregation).

Figure 10.13 shows the result of a denoising on the image of a bear. Whereas the fur is washed out by the global algorithm, both LBMRE and NLMSAP preserve snippets of this texture, but some other parts are a bit smoothed. It seems that the contrast is slightly better with LBMRE than with NLMSAP. The best result is provided with LBEWA, as all the details of the fur are preserved.

Another example is shown in Figure 10.14 on a church image. Again, the global parameter h optimized for PSNR leads to images that are too smooth on the overall but too noisy around details with high contrast. The three proposed methods reduce the noisy halo, but the cleanest result is provided by NLMSAP. With LBEWA, there is still a little variance around edges with contrast, but as the overall image has a slightly grainy aspect, this is not visually annoying. On the contrary, the contrast of the texture on the roof (especially the glint) seems better preserved with LBMRE and LBEWA than with NLMSAP.

Figure 10.15 shows results on an image of lighthouse. It is clear that all the methods involving locality (whether for h or for the shape) preserve the contrast better than the global one. Although LBMRE reduces the halo, it leaves glitches of noise in the clouds. On the contrary, the result of LBEWA is homogeneously grainy and no noticeable artifact appears. Again, NLMSAP provides smooth images, and leads to the cleanest result in terms of noise halo.

Eventually, the results on the image of a pyramid is shown in Figure 10.16. The same conclusions as in the previous figures hold. This time, one may find that the result of LBEWA is a bit too noisy: in images with very smooth variations and little texture, there is no advantage in introducing very low smoothing parameters in the average, so that the result is unnecessarily noisy. Notice that for coherence, we have used the same parameters for LBEWA in all these experiments. Tuning the temperature parameter would certainly help in this case.

As for the numerical results, we already know from the previous chapter that LBEWA is slightly below LBMRE in terms of PSNR, and from the present chapter that LBMRE is not as good as NLM-SAP. Nevertheless, Table 10.6 gives a few numerical values for noise intensity $\sigma = 10$.

It is is clear that among the three proposed methods, NLM-SAP provides the best numerical performance as well as images with high visual quality. The resulting images are globally smooth, artifact-free, with very clean transitions along edges with high contrast. Yet, we especially like the grainy aspect produced by LBEWA. In several cases, it allows to preserve textures better than with NLM-SAP, and the overall result looks more natural.

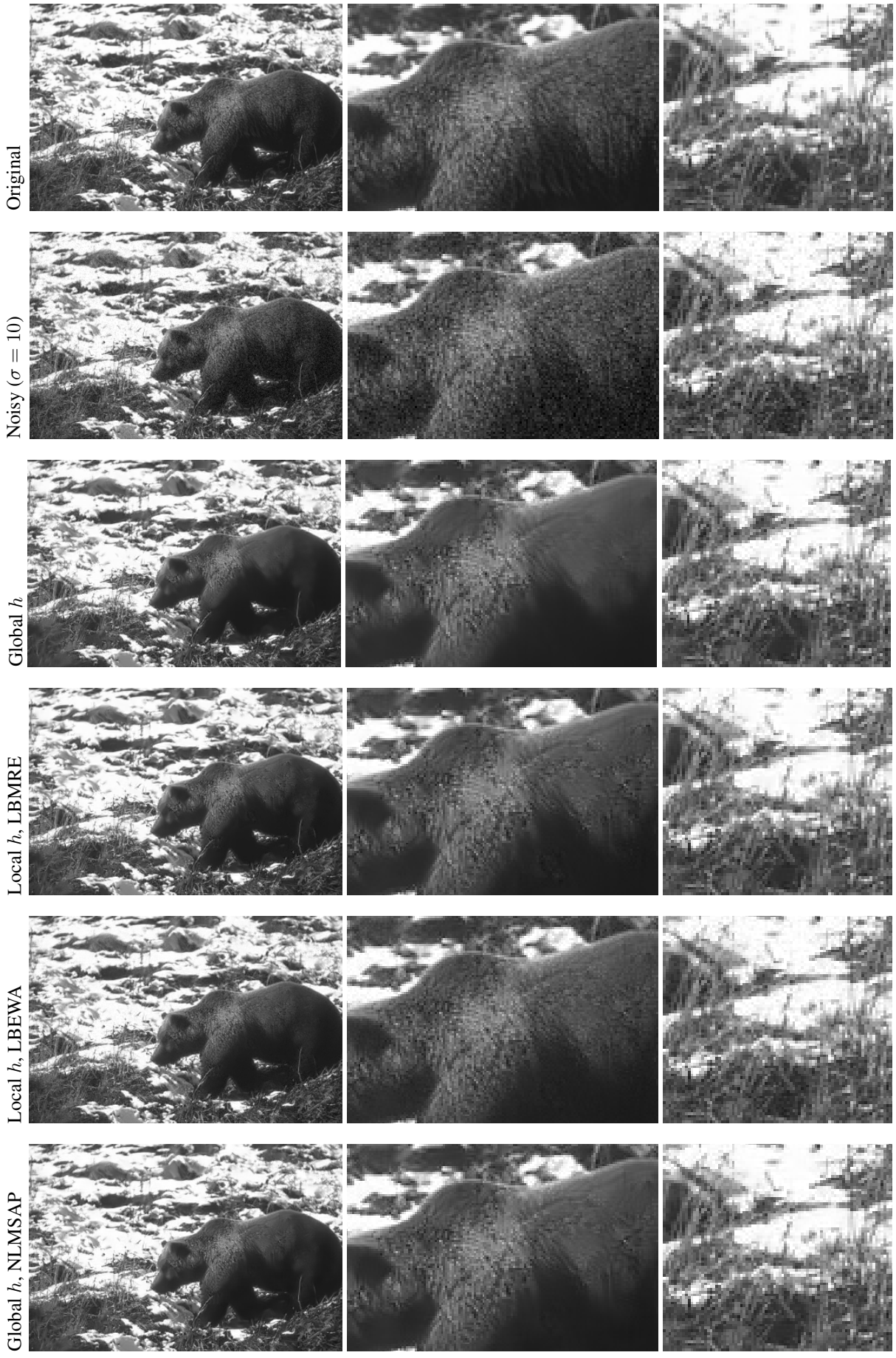


Figure 10.13 – *Comparison of the proposed methods on the bear image.*

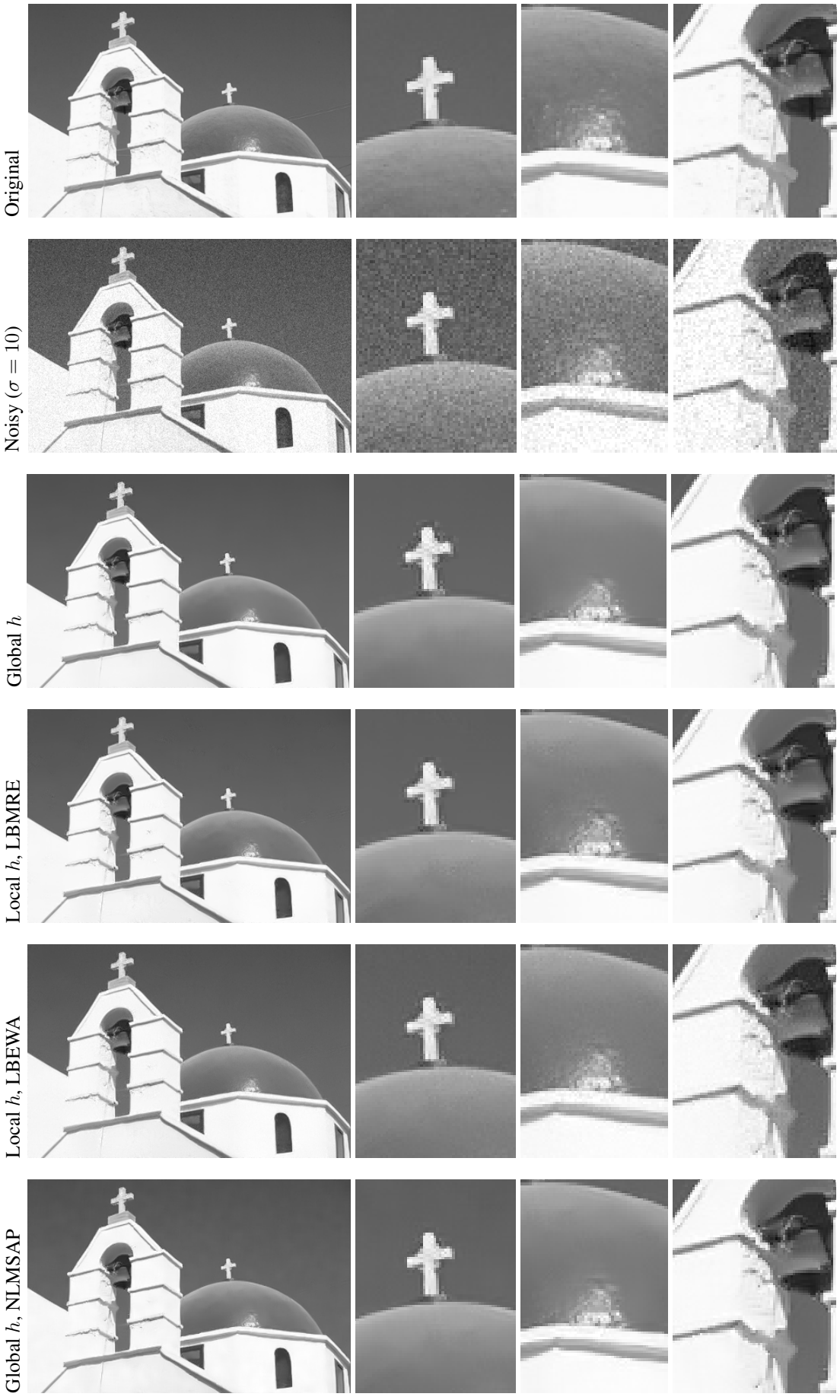


Figure 10.14 – *Comparison of the proposed methods on the church image.*

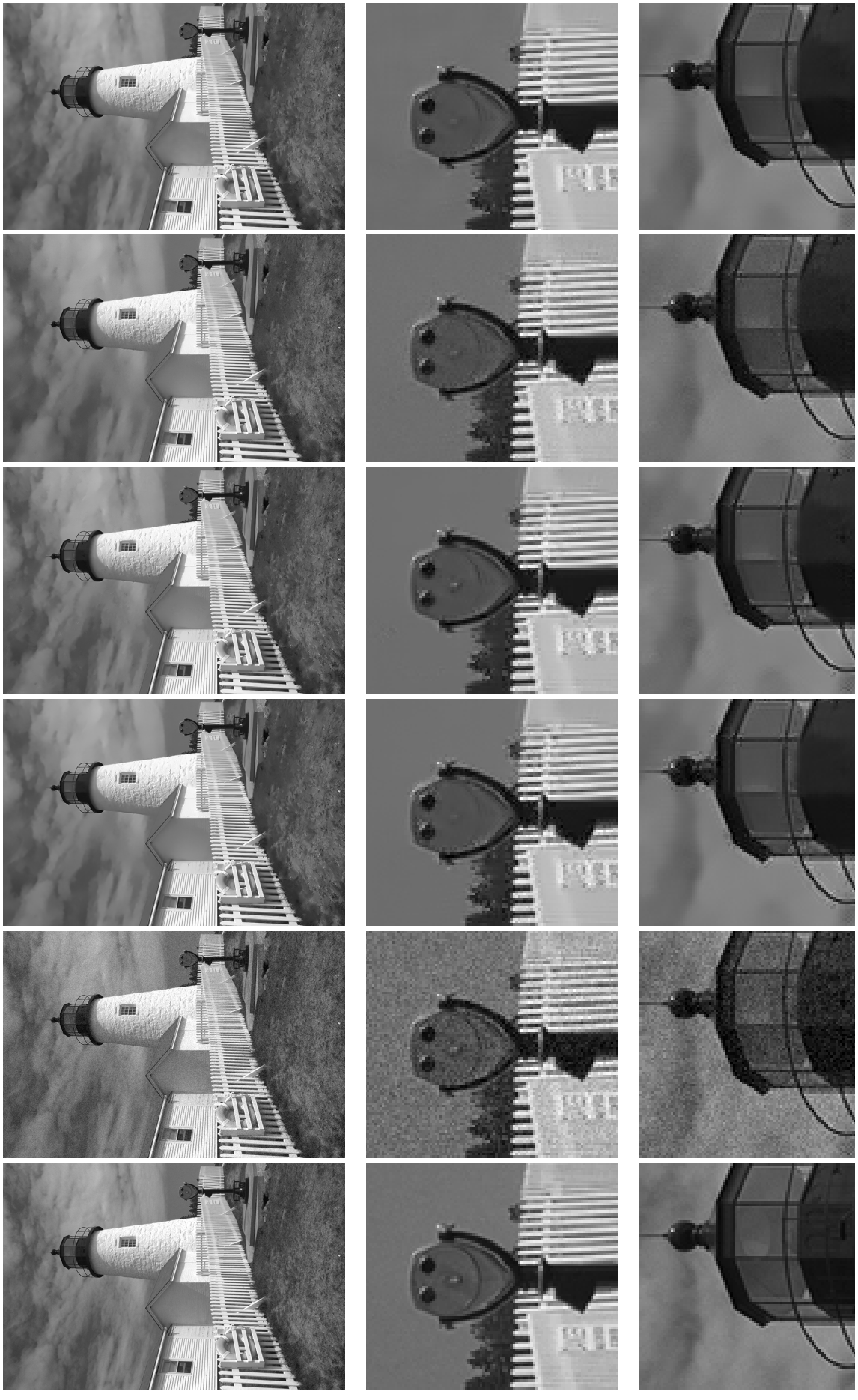


Figure 10.15 – *Comparison of the proposed methods on the lighthouse image.*

10.4 Conclusion

We have pushed the philosophy of using directional shapes to reduce the halo as far as we could. By construction, the aggregation based on Weighted Average Variance (WAV) is well adapted to the problem of reducing the halo. It has the (minor) drawback of slightly blurring details with little contrast.

To reduce this bias, it is necessary to estimate the local risk rather than the variance only. As in the previous chapter, the pointwise values of SURE are not reliable, and it is necessary to regularize them. The convolution of the risk gives good results, but since the filtered risk is not sharp around edges, the algorithm is not able to find the right orientations. As a result it chooses small shapes near edges, which already limits the halo. To reduce the halo further and to comply with our initial philosophy, an anisotropic filtering of the risk is necessary.

Once the risk is filtered, the Exponentially Weighted Aggregation (EWA) produces better results than the Minimizer of the Risk Estimate (MRE), since the decisions of latter have too brutal variations.

The final algorithm is better than those exposed in the previous chapter, and, in terms of visual quality, it is competitive with state-of-the-art methods like BM3D. Yet, a simplified version where the Yaroslavsky filtering and the EWA step are replaced with a WAV aggregation would yield very similar results. It is arguable whether the slight improvement brought by considering the risk and EWA is worth the effort, and we recommend the WAV version as much as the EWA version.

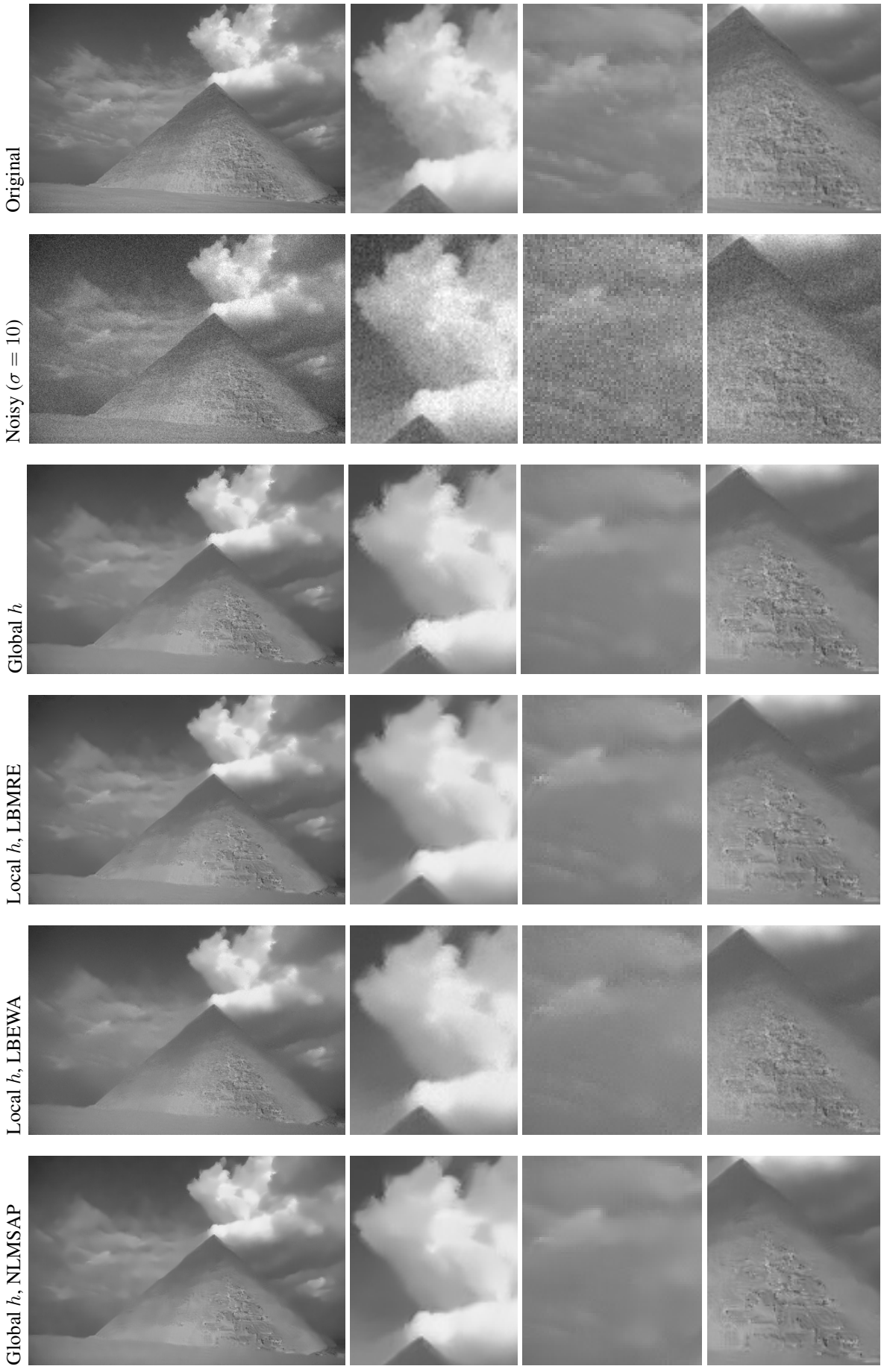


Figure 10.16 – *Comparison of the proposed methods on the pyramid image.*

Conclusion and Perspectives

Conclusion

In this thesis, we have studied two image restoration models: TVL1 and the Non-Local Means. For each model, we have provided a theoretical study, and we have proposed extensions.

In the first part, we have explored the link between the TVL1 model and mathematical morphology. As TVL1 is a morphological filter, it is equivalent to study it as a functional operator or a set operator. Choosing the latter approach, we have shown that the level sets of the solutions have smooth boundary, and that in the case where the original level set is convex, the solution is given by an opening followed by a thresholding. Using this link with mathematical morphology, we have related the good behavior of the model in the cartoon-texture decomposition problem to granulometries, which were introduced in the early years of mathematical morphology. We have compared the TVL1 model to the TV+G decomposition model, and we have exhibited artifacts in the latter which do note appear with the former. In applications like the denoising of textured images, or impulse noise removal, it is important to define a spatially varying fidelity parameter. We have defined such a framework, and we have applied it to impulse noise removal, obtaining comparable results with recent algorithms.

In the second part, we have studied the Non-Local Means filter. Using basic models, we have highlighted its properties and explained its artifacts. Then, we have proposed two algorithms (LBMRE and LBEWA) which locally set the bandwidth parameter. Both rely on Stein's Unbiased Risk Estimate (SURE), and they clearly outperform the Non-Local Means with global parameter: textures and small details are better preserved, whereas the artifacts which appears along edges with high contrast are removed. As LBEWA is a convex combination of estimators, it is more robust to errors in the estimation of the risk, and it takes less brutal decisions than LBMRE. A third algorithm was proposed in collaboration with Charles Deledalle and Joseph Salmon, NLM-SAP, which uses arbitrary shapes instead of patches, in order to find as many similarities as possible. Its performance is better than the first two algorithms, and it is even an interesting alternative to state-of-the art methods.

Perspectives

A first way to extend this work is to characterize more precisely the apparition of edges in the texture part provided by the TV+G model: aside from the simple cases given here, does it always happen?

The spatially adaptive TVL1 model could be applied to problems arising in medical imaging, where one has a mask which indicates areas that should not be smoothed (*e.g.* because they are crucial for diagnosis).

Another question of interest concerns impulse noise removal. As we have seen, the most important step in the denoising algorithm is the detection step. Yet, in case of large spots of noise, the adaptive TVL1 model interpolates the level lines with approximately line segments, which is not what the Gestalt theory would recommend. As the problem of filling holes due to large spots of noise is close to the inpainting problem, an approach based on the curvature of the level lines could be interesting [Masnou and Morel, 1998, Schoenemann et al., 2011]. For instance, the geometric problem on each

level set could be replaced with the following:

$$\min_{U \subset \mathbb{R}^2} \int_{\partial^* U} (1 + \kappa^b_{\partial^* U}) d\mathcal{H}^1 + \int_{U \Delta F} \lambda(x) dx, \tag{10.18}$$

where $b > 0$ and, as in Chapter 6, $\lambda(x) \ll 1$ in noisy regions, $\lambda(x) \gg 1$ otherwise.

Concerning the Non-Local Means, we particularly like the results provided by LBEWA, and we find that its properties should be studied further. We have used a fixed rule for the temperature which does not seem to be optimal, as on smooth images the result looks a bit too noisy. Nevertheless, this approach gives promising visual results and a better control of its behavior would yield a method which removes most of the noise without degrading textures. Moreover, its results confirm the importance of aggregating estimators in image restoration, as claimed in [Salmon, 2010a].

The NLM-SAP is also interesting, and it could be adapted to more general kinds of noise, by adapting results given by [Hudson, 1978, Raphan and Simoncelli, 2007] on unbiased risk estimates for general exponential families, into algorithms like [Deledalle et al., 2009].

Bibliography

[Abreu et al., 1996] Abreu, E., Lightstone, M., Mitra, S., and Arakawa, K. (1996). A new efficient approach for the removal of impulse noise from highly corrupted images. *Image Processing, IEEE Transactions on*, 5(6):1012 –1025.

[Aharon et al., 2006] Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322.

[Aizenberg et al., 2005] Aizenberg, I., Butakoff, C., and Paliy, D. (2005). Impulsive noise removal using threshold boolean filtering based on the impulse detecting functions. *Signal Processing Letters, IEEE*, 12(1):63 – 66.

[Alberti et al., 2003] Alberti, G., Bouchitté, G., and Dal Maso, G. (2003). The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calculus of Variations and Partial Differential Equations*, 16(3):299–333.

[Allard, 2007] Allard, W. (2007). Total variation regularization for image denoising I: Geometric theory. *SIAM Journal on Mathematical Analysis*, 39(4):1150–1190.

[Allard, 2008] Allard, W. (2008). Total variation regularization for image denoising II: Examples. *SIAM Journal on Imaging Sciences*, 1(4):400–417.

[Allard, 2009] Allard, W. (2009). Total variation regularization for image denoising III: Examples. *SIAM Journal on Imaging Sciences*, 2(2):532–568.

[Alliney, 1992] Alliney, S. (1992). Digital filters as absolute norm regularizers. *IEEE Transactions on Signal Processing*, 40(6):1548–1562.

[Alliney, 1996] Alliney, S. (1996). Recursive median filters of increasing order : a variational approach. *IEEE Transactions on Signal Processing*, 44(6):1346–1354.

[Alliney, 1997] Alliney, S. (1997). A property of the minimum vectors of regularizing functional defined by means of the absolute norm. *IEEE Transactions on Signal Processing*, 45(4):913–917.

[Almgren et al., 1993] Almgren, F., Taylor, J. E., and Wang, L.-H. (1993). Curvature-driven flows: a variational approach. *SIAM J. Control Optim.*, 31(2):387–438.

[Alter and Caselles, 2009] Alter, F. and Caselles, V. (2009). Uniqueness of the cheeger set of a convex body. *Nonlinear Analysis: Theory, Methods & Applications*, 70(1):32–44.

[Alter et al., 2005a] Alter, F., Caselles, V., and Chambolle, A. (2005a). A characterization of convex calibrable sets in $\mathbb{R}^N$. *Math. Ann.*, 7:29–53.

[Alter et al., 2005b] Alter, F., Caselles, V., and Chambolle, A. (2005b). Evolution of convex sets in the plane by minimizing the total variation flow. *Interfaces and Free Boundaries*, 332:329–366.

[Alvarez et al., 1993] Alvarez, L., Guichard, F., Lions, P.-L., and Morel, J.-M. (1993). Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis*, 123:199–257. 10.1007/BF00375127.

[Ambrosio, 1994] Ambrosio, L. (1994). Minimizing movements. Lecture notes.

[Ambrosio, 1997] Ambrosio, L. (1997). *Corso introduttivo alla teoria geometrica della misura ed alle superfici minimime*. Quad. Scuola Norm. Sup. Pisa. Pantograph, Genova.

[Ambrosio et al., 2000] Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford mathematical monographs. Oxford University Press.

[Ambrosio et al., 2002] Ambrosio, L., Novaga, M., and Paolini, E. (2002). Some regularity results for minimal crystals. *ESAIM: COCV volume dedicated to J. L. Lions*, 8(69–103).

[Ambrosio and Paolini, 1998] Ambrosio, L. and Paolini, E. (1998). Partial regularity for quasi-minimizers of perimeter. *Ricerce di Mat.*, 48:167–186.

[Andreu et al., 2001a] Andreu, F., Ballester, C. Caselles, V., and Mazón, J. M. (2001a). The dirichlet problem for the total variation flow. *Journal of Functional Analysis*, 180(2):347–403.

[Andreu et al., 2001b] Andreu, F., Ballester, C. Caselles, V., and Mazón, J. M. (2001b). Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360.

[Andreu et al., 2002] Andreu, F., Caselles, V., Diaz, J. I., and Mazón, J. M. (2002). Some qualitative properties for the total variation flow. *Journal of Functional Analysis*, 188(2):516 – 547.

[Andreu-Vaillo et al., 2002] Andreu-Vaillo, F., Caselles, V., and Mazón, J. M. (2002). *Parabolic quasilinear equations minimizing linear growth functionals*, volume 223 of *Progress in Mathematics*. Birkhauser.

[Anzellotti, 1983] Anzellotti, G. (1983). Pairing between measures and bounded functions and compensated compactness. *Ann. di Matematica Pura ed Appl.*, IV(135):293–318.

[Aubert and Aujol, 2005] Aubert, G. and Aujol, J. (2005). Modeling very oscillating signals. Application to image processing. *Applied Mathematics and Optimization*, 51(2).

[Aujol et al., 2005] Aujol, J.-F., Aubert, G., Blanc-Féraud, L., and Chambolle, A. (2005). Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1):71–88.

[Aujol and Chambolle, 2005] Aujol, J.-F. and Chambolle, A. (2005). Dual norms and image decomposition models. *International Journal of Computer Vision*, 63(1):85–104.

[Aujol and Gilboa, 2006] Aujol, J.-F. and Gilboa, G. (2006). Constrained and SNR-based solutions for TV-Hilbert space image denoising. *Journal of Mathematical Imaging and Vision*, 26(1-2):217–237.

[Aujol et al., 2006] Aujol, J.-F., Gilboa, G., Chan, T., and Osher, S. (2006). Structure-texture image decomposition - modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136.

[Aujol and Kang, 2006] Aujol, J.-F. and Kang, S. H. (2006). Color image decomposition and restoration. *Journal of Visual Communication and Image Representation*, 17(4):916–928.

[Awate and Whitaker, 2006] Awate, S. P. and Whitaker, R. T. (2006). Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):364–376.

[Azzabou et al., 2007a] Azzabou, N., Paragios, N., and Guichard, F. (2007a). Image denoising based on adapted dictionary computation. In *ICIP*, pages 109–112.

[Azzabou et al., 2007b] Azzabou, N., Paragios, N., and Guichard, F. (2007b). Uniform and textured regions separation in natural images towards MPM adaptive denoising. In *SSVM*, pages 418–429.

[Azzabou et al., 2007c] Azzabou, N., Paragios, N., Guichard, F., and Cao, F. (2007c). Variable bandwidth image denoising using image-based noise. In *CVPR*.

[Babaud et al., 1986] Babaud, J., Witkin, A. P., Baudin, M., and Duda, R. (1986). Uniqueness of the gaussian kernel for scale-space filtering. *IEEE Trans. Patt. Anal. Mach. Intell.*, 8.

[Bar et al., 2006a] Bar, L., Sochen, N., and Kiryati, N. (2006a). Image deblurring in the presence of impulsive noise. *International Journal of Computer Vision*, 70(3):279–298.

[Bar et al., 2006b] Bar, L., Sochen, N., and Kiryati, N. (2006b). Semi-blind image restoration via mumford-shah regularization. *IEEE Trans. on Image Processing*, 15(2):483–493.

[Barozzi et al., 1987] Barozzi, E., Gonzales, E., and Tamanini, I. (1987). The mean curvature of a set of finite perimeter. *AMS*, 99:313–316.

[Bellettini, 2004] Bellettini, G. (2004). Anisotropic and crystalline mean curvature flow, a sampler of Riemann-Finsler geometry.

[Bellettini et al., 1999] Bellettini, G., Bouchitté, G., and Fragalà, G. (1999). BV functions with respect to a measure and relaxation of metric integral functionals. *Journal of Convex Analysis*, 6(2):349–366.

[Bellettini et al., 2009] Bellettini, G., Caselles, V., Chambolle, A., and Novaga, M. (2009). The volume preserving crystalline mean curvature flow of convex sets in $\mathbb{R}^N$. *Journal de Mathématiques Pures et Appliquées*, 92(5):499–527.

[Bellettini et al., 2002] Bellettini, G., Caselles, V., and Novaga, M. (2002). The total variation flow in $\mathbb{R}^N$. *J. Differential Equations*, 184(2):475–525.

[Bellettini et al., 2005a] Bellettini, G., Caselles, V., and Novaga, M. (2005a). Explicit solutions of the eigenvalue problem $-\mathrm{div}\,(Du/|Du|) = u$. *SIAM J. Math. Anal.*, 36:1095–1129.

[Bellettini et al., 2005b] Bellettini, G., Caselles, V., and Novaga, M. (2005b). Explicit solutions of the eigenvalue problem $div\left(\frac{Du}{|Du|}\right) = u$ in $\mathbb{R}^2$. *SIAM J. Math. Anal.*, 36:1095–1129.

[Bellettini et al., 2001] Bellettini, G., Novaga, M., and Paolini, M. (2001). Characterization of facet–breaking for nonsmooth mean curvature flow in the convex case. *Interfaces and Free Boundaries*, 3:415–446.

[Bergeaud and Mallat, 1995] Bergeaud, F. and Mallat, S. (1995). Matching pursuit of images. In *Image Processing, 1995. Proceedings., International Conference on*, volume 1, pages 53 –56 vol.1.

[Besdok and Yüksel, 2005] Besdok, E. and Yüksel, M. E. (2005). Impulsive noise suppression from images with jarque-bera test based median filter. *AEU - International Journal of Electronics and Communications*, 59(2):105 – 110.

[Beucher and Lantuéjoul, 1979] Beucher, S. and Lantuéjoul, C. (1979). Use of watersheds in contour detection. In *International workshop on image processing, real-time edge and motion detection.*

[Bilcu and Vehvilainen, 2008] Bilcu, R. C. and Vehvilainen, M. (2008). Combined non-local averaging and intersection of confiden intervals for image de-noising. In *ICIP*, pages 1736–1739.

[Blomgren and Chan, 1998] Blomgren, P. and Chan, T. (1998). Color TV: Total variation methods for restoration of vector valued images. *IEEE Transactions on Image Processing*, 7(3):304–309.

[Blu and Luisier, 2007] Blu, T. and Luisier, F. (2007). The SURE-LET approach to image denoising. *IEEE_J_IP*, 16(11):2778–2786.

[Bouraoui et al., 2008] Bouraoui, B., Ronse, C., Baruthio, J., Passat, N., and Germain, P. (2008). Fully automatic 3d segmentation of coronary arteries based on mathematical morphology. In *ISBI 2008, 5th International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1059–1062. IEEE.

[Boykov and Kolmogorov, 2004] Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(9):1124–1137.

[Braides, 2002] Braides, A. (2002). *Gamma-convergence for Beginners*. Oxford University Press, USA.

[Brakke, 1991] Brakke, K. (1991). Minimal cones on hypercubes. *J. Geom. Anal.*, 1:329–338.

[Bresson and Chan, 2008] Bresson, X. and Chan, T. (2008). Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problems and Imaging*, 2(4):455–484.

[Bresson et al., 2007] Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.-P., and Osher, S. (2007). Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167.

[Brox and Cremers, 2007] Brox, T. and Cremers, D. (2007). Iterated nonlocal means for texture restoration. In Sgallari, F., Murli, A., and Paragios, N., editors, *Proc. International Conference on Scale Space and Variational Methods in Computer Vision*, volume 4485 of *LNCS*, pages 13–24, Ischia, Italy. Springer.

[Brox et al., 2008] Brox, T., Kleinschmidt, O., and Cremers, D. (2008). Efficient nonlocal means for denoising of textural patterns. *IEEE Transactions on Image Processing*, 17(7):1083–1092.

[Buades et al., 2005] Buades, A., Coll, B., and Morel, J.-M. (2005). A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.*, 4(2):490–530.

[Buades et al., 2006] Buades, A., Coll, B., and Morel, J.-M. (2006). Image enhancement by non-local reverse heat equation. Technical Report 2006-22, Préprint CMLA, ENS Cachan.

[Buades et al., 2009] Buades, A., Coll, B., and Morel, J.-M. (2009). Non-local means denoising. *Image Processing on Line*.

[Buades et al., 2010] Buades, A., Le, T., Morel, J.-M., and Vese, L. (2010). Fast cartoon + texture image filters. *IEEE Transactions on Image Processing*, 19(8):1978–1986.

[Buttazzo et al., 2007] Buttazzo, G., Carlier, G., and Comte, M. (2007). On the selection of maximal cheeger sets. *Differential and Integral Equations*, 20(9):991–1004.

[Cai et al., 2008] Cai, J.-F., Chan, R., and Nikolova, M. (2008). Two phase methods for deblurring images corrupted by impulse plus gaussian noise. *AIMS Journal on Inverse Problems and Imaging*, 2(2):187–204.

[Cai et al., 2010] Cai, J.-F., Chan, R., and Nikolova, M. (2010). Fast two-phase image deblurring under impulse noise. *Journal of Mathematical Imaging and Vision*, 36(1):46–53.

[Carlier and Comte, 2007] Carlier, G. and Comte, M. (2007). On a weighted total variation minimization problem. *J. Funct. Anal.*, 250:214–226.

[Carlier et al., 2009] Carlier, G., Comte, M., and Peyré, G. (2009). Approximation of maximal cheeger sets by projection. *ESAIM: Mathematical Modelling and Numerical Analysis*, 43(1):131–150.

[Caselles et al., 2008] Caselles, V., Chambolle, A., Moll, S., and Novaga, M. (2008). A characterization of convex calibrable sets in with respect to anisotropic norms. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, 25(4):803 – 832.

[Caselles et al., 2007a] Caselles, V., Chambolle, A., and Novaga, M. (2007a). The discontinuity set of solutions of the tv denoising problem and some extensions. *Multiscale Modeling and Simulation*, 6(3):879–894.

[Caselles et al., 2007b] Caselles, V., Chambolle, A., and Novaga, M. (2007b). Uniqueness of the cheeger set of a convex body. *Pacific J. Math.*, 232(1):77–90.

[Caselles et al., 2009] Caselles, V., Facciolo, G., and Meinhardt, E. (2009). Anisotropic cheeger sets and applications. *SIAM J. Imaging Sciences*, 2(4):1211–1254.

[Caselles et al., 1997] Caselles, V., Kimmel, R., and Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision*, 22:61–79. 10.1023/A:1007979827043.

[Caselles et al., 2010] Caselles, V., Miranda, M. J., and Novaga, M. (2010). Total variation and cheeger sets in gauss space. *Journal of Functional Analysis*, 259(6):1491 – 1516.

[Chambolle, 2000] Chambolle, A. (2000). Inverse problems in image processing and image segmentation: some mathematical and numerical aspects. In *ICTP Lecture Notes Series Volume II (ISBN 92-95003-04-7)*.

[Chambolle, 2004a] Chambolle, A. (2004a). An algorithm for mean curvature motion. *Interfaces Free Bound.*, 6(2):195–218.

[Chambolle, 2004b] Chambolle, A. (2004b). An algorithm for total variation minimization and its applications. *Journal of Mathematical Imaging and Vision*, 20:89–97.

[Chambolle, 2005] Chambolle, A. (2005). Total variation minimization and a class of binary mrf models. In Rangarajan, A., Vemuri, B., and Yuille, A., editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 3757 of *Lecture Notes in Computer Science*, pages 136–152. Springer Berlin / Heidelberg.

[Chambolle et al., 2009] Chambolle, A., Caselles, V., Novaga, M., Cremers, D., and Pock, T. (2009). An introduction to Total Variation for Image Analysis.

[Chambolle et al., 2008] Chambolle, A., Cremers, D., and Pock, T. (2008). A convex approach for computing minimal partitions. Technical Report 649, CMAP, Ecole Polytechnique, France.

[Chambolle and Darbon, 2009] Chambolle, A. and Darbon, J. (2009). On total variation minimization and surface evolution using parametric maximum flows. *Int. J. Comput. Vision*, 84:288–307.

[Chambolle et al., 2010] Chambolle, A., Giacomini, A., and Lussardi, L. (2010). Continuous limits of discrete perimeters. *ESAIM Math. Model. Numer. Anal.*, 44(2):207–230.

[Chambolle and Pock, 2011] Chambolle, A. and Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145.

[Chan et al., 2005] Chan, R., Ho, C.-W., and Nikolova, M. (2005). Salt-and-pepper noise removal by median-type noise detector and detail-preserving regularization. *IEEE Trans. on Image Processing*, 14(10):1479–1485.

[Chan et al., 2004] Chan, R., Hu, C., and Nikolova, M. (2004). An iterative procedure for removing random-valued impulse noise. *IEEE Signal Processing Letters*, 11:921–924.

[Chan and Esedoglu, 2005] Chan, T. and Esedoglu, S. (2005). Aspects of total variation regularized L^1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837.

[Chan et al., 2006] Chan, T. F., Esedoglu, S., and Nikolova, M. (2006). Algorithms for finding global minimizers of denoising and segmentation models. *SIAM J. Appl. Math.*, 66:1632–1648.

[Chan et al., 2000] Chan, T. F., Kang, S. H., and Shen, J. (2000). Total variation denoising and enhancement of color images based on the cb and hsv color models. *J. Visual Comm. Image Rep*, 12:2001.

[Cheeger, 1970] Cheeger, J. (1970). A lower bound for the smallest eigenvalue of the laplacian. *Problems in Analysis*, pages 195–199.

[Chen et al., 2005] Chen, T., Huang, T., Yin, W., and Zhou, X. (2005). A new coarse to fine framework for 3d brain mr image registration. *Computer Vision for Biomedical Image*, 3765:114–124.

[Chen and Wu, 2001a] Chen, T. and Wu, H. (2001a). Adaptive impulse detection using center-weighted median filters. *IEEE Trans. Circuits Syst.-II*, 48:785–789.

[Chen and Wu, 2001b] Chen, T. and Wu, H. (2001b). Space variant median filters for the restoration of impulse noise corrupted images. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, 48(8):784 –789.

[Chen et al., 1991] Chen, Y., Giga, Y., and Goto, S. (1991). Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *J. Differential Geom*, 33(3):749–786.

[Cheng and Venetsanopoulos, 1992] Cheng, F. and Venetsanopoulos, A. N. (1992). An adaptive morphological filter for image processing. *IEEE Trans. Image Process.*, 1:533–539.

[Coifman and Donoho, 1995] Coifman, R. and Donoho, D. (1995). Translation invariant denoising. Technical Report 475, Dept. of Statistics, Stanford University.

[Combettes and Pesquet, 2004] Combettes, P. L. and Pesquet, J.-C. (2004). Wavelet-constrained image restoration. *Int. J. Wavelets, Multires. Inf. Process.*, 2:371?389.

[Condat, 2010] Condat, L. (2010). A Simple Trick to Speed Up and Improve the Non-Local Means. *submited*.

[Criminisi et al., 2004] Criminisi, A., Pérez, P., and Toyama, K. (2004). Region filling and object removal by exemplar-based inpainting. *IEEE Trans. Image Processing*, 13(9):1200–1212.

[Crnojevic et al., 2004] Crnojevic, V., Senk, V., and Trpovski, Z. (2004). Advanced impulse detection based on pixel-wise mad. *Signal Processing Letters, IEEE*, 11(7):589 – 592.

[Crow, 1984] Crow, F. C. (1984). Summed-area tables for texture mapping. In *SIGGRAPH*, pages 207–212.

[Cunha et al., 2007] Cunha, A., Darbon, J., Chan, T., and Toga, A. (2007). Fast and accurate feature detection and triangulation using total variation filtering of biological images. In *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pages 680 –683.

[Dabov et al., 2007] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. O. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095.

[Dalalyan and Tsybakov, 2007] Dalalyan, A. S. and Tsybakov, A. B. (2007). Aggregation by exponential weighting, sharp oracle inequalities and sparsity. In *COLT*, pages 97–111.

[Dalalyan and Tsybakov, 2008] Dalalyan, A. S. and Tsybakov, A. B. (2008). Aggregation by exponential weighting, sharp pac-bayesian bounds and sparsity. *Mach. Learn.*, 72(1-2):39–61.

[Darbon, 2005] Darbon, J. (2005). Total variation minimization with L^1 data fidelity as a contrast invariant filter. *4th International Symposium on Image and Signal Processing and Analysis (ISPA 2005)*, pages 221–226.

[Darbon et al., 2008] Darbon, J., Cunha, A., Chan, T. F., Osher, S., and Jensen, G. J. (2008). Fast nonlocal filtering applied to electron cryomicroscopy. In *ISBI*, pages 1331–1334.

[Darbon and Sigelle, 2006] Darbon, J. and Sigelle, M. (2006). Image restoration with discrete constrained total variation part I: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 26(3):261–276.

[Daubechies and Teschke, 2005] Daubechies, I. and Teschke, G. (2005). Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising. *Applied and Computational Harmonic Analysis*, 19:1–16.

[De Haan and Lodder, 2002] De Haan, G. and Lodder, R. (2002). De-interlacing of video data using motion vectors and edge information. In *International Conference on Consumer Electronics*, pages 70–71.

[Deledalle et al., 2011] Deledalle, C., Duval, V., and Salmon, J. (2011). Non-local methods with shape adaptive patches (nlm-sap). *Journal of Mathematical Imaging and Vision*. To appear.

[Deledalle et al., 2009] Deledalle, C.-A., Denis, L., and Tupin, F. (2009). Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Trans. on Image Processing*, 18.

[Deledalle et al., 2010a] Deledalle, C.-A., Denis, L., and Tupin, F. (2010a). Nl-insar: Non-local interferogram estimation. *IEEE Trans. on Geoscience and Remote Sensing*.

[Deledalle et al., 2010b] Deledalle, C.-A., Denis, L., and Tupin, F. (2010b). Poisson nl-means: unsupervised non local means for poisson noise. In *IEEE International Conference on Computer Vision*, Hong-kong, China.

[Dibos and Koepfler, 2001] Dibos, F. and Koepfler, G. (2001). Total variation minimization by the fast level sets transform. *IEEE Workshop Variational and Level Set Methods in Computer Vision.*, pages 179–185.

[Do and Vetterli, 2005] Do, M. and Vetterli, M. (2005). The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions Image on Processing*, 14(12):2091–2106.

[Dong et al., 2007] Dong, Y., Chan, R., and Xu, S. (2007). A detection statistic for random-valued impulse noise. *IEEE Trans. Image Processing*, 16:1112–1120.

[Donoho and Johnstone, 1994] Donoho, D. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

[Donoho and Johnstone, 1995] Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90(432):1200–1224.

[Dore and Cheriet, 2009] Dore, V. and Cheriet, M. (2009). Robust NL-Means filter with optimal pixel-wise smoothing parameter for statistical image denoising. *IEEE Transactions on Signal Processing*, 57.

[Durand and Dorsey, 2002] Durand, F. and Dorsey, J. (2002). Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics (TOG)*, 21(3):257–266.

[Duval et al., 2009] Duval, V., Aujol, J.-F., and Gousseau, Y. (2009). The TVL1 Model: a Geometric Point of View. *Multiscale Model. Simul.*, 8(1):154–189.

[Duval et al., 2011] Duval, V., Aujol, J.-F., and Gousseau, Y. (2011). A bias-variance approach for the non-local means. *SIAM Journal on Imaging Sciences*. To appear.

[Duval et al., 2010] Duval, V., Aujol, J.-F., and Vese, L.-A. (2010). Mathematical modeling of textures: Application to color image decomposition with a projected gradient algorithm. *Journal of Mathematical Imaging and Vision*, 37:232–248.

[Efros and Leung, 1999] Efros, A. A. and Leung, T. K. (1999). Texture synthesis by non-parametric sampling. In *IEEE International Conference on Computer Vision*, pages 1033–1038, Corfu, Greece.

[Ekeland and Temam, 1999] Ekeland, I. and Temam, R. (1999). *Convex Analysis and Variational Problems*. Classics in Applied Mathematics. Siam, Philadelphia.

[Elad, 2002] Elad, M. (2002). On the origin of the bilateral filter and ways to improve it. *IEEE Transactions on Image Processing*, 11(10):1141–1151.

[Esedoglu and Osher, 2004] Esedoglu, S. and Osher, S. (2004). Decomposition of images by the anisotropic rudin - osher - fatemi model. *Comm. Pure Appl. Math.*, 57:1609–1626.

[Federer, H., 1969] Federer, H. (1969). *Geometric Measure Theory*, volume 153. Springer-Verlag New York Inc.

[Fonseca, 1991] Fonseca, I. (1991). The wulff theorem revisited. *Proceedings: Mathematical and Physical Sciences*, 432(1884):pp. 125–145.

[Ford and Fulkerson, 1956] Ford, L. R. and Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*, pages 399–404.

[Garnett et al., 2007] Garnett, J., Le, T., Meyer, Y., and Vese, L. (2007). Image decompositions using bounded variation and generalized homogeneous besov spaces. *Appl. Comput. Harmon. Anal.*, 23:25–56.

[Garnett et al., 2005] Garnett, R., Huegerich, T., Chui, C., and He, W. (2005). A universal noise removal algorithm with an impulse detector. *Image Processing, IEEE Transactions on*, 14(11):1747 –1754.

[Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6:721–741.

[Gilboa and Osher, 2007] Gilboa, G. and Osher, S. (2007). Nonlocal linear image regularization and supervised segmentation. *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, 6:595–630.

[Gilboa and Osher, 2008] Gilboa, G. and Osher, S. (2008). Nonlocal operators with applications to image processing. *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, 7:1005–1028.

[Gilboa et al., 2006] Gilboa, G., Sochen, N. A., and Zeevi, Y. Y. (2006). Variational denoising of partly textured images by spatially varying constraints. *IEEE Transactions on Image Processing*, pages 2281–2289.

[Gilles and Meyer, 2010] Gilles, G. and Meyer, Y. (2010). Properties of bv-g structures + textures decomposition models. application to road detection in satellite images. *IEEE Transactions on Image Processing*, 19(11):2793–2800.

[Gilles, 2007] Gilles, J. (2007). Noisy image decomposition: a new structure, texture and noise model based on local adaptivity. *Journal of Math. Imag. and Vision (JMIV)*, 28(3):285–295.

[Gilles, 2009] Gilles, J. (2009). Image decomposition: Theory, numerical schemes, and performance evaluation. *Advances in Imaging and Electron Physics*, 158:89–137.

[Giusti, 1978] Giusti, E. (1978). On the equation of surfaces of prescribed mean curvature. *Invent. Math.*, 46(2):111–137.

[Glaunès and Joshi, 2006] Glaunès, J. and Joshi, S. (2006). Template estimation from unlabeled point set data and surfaces for computational anatomy. In Pennec, X. and Joshi, S., editors, *Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2006)*, pages 29–39.

[Glaunès, 2005] Glaunès, J. (2005). *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique*. PhD thesis, Université Paris 13 en Mathématiques.

[Goldluecke and Cremers, 2010] Goldluecke, B. and Cremers, D. (2010). An approach to vectorial total variation based on geometric measure theory. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Goossens et al., 2008] Goossens, B., Luong, H. Q., Pizurica, A., and Philips, W. (2008). An improved non-local denoising algorithm. In *LNLA*, pages 143–156.

[Grasmair and Lenzen, 2010] Grasmair, M. and Lenzen, F. (2010). Anisotropic total variation filtering. *Applied Mathematics & Optimization*, 62:323–339. 10.1007/s00245-010-9105-x.

[Greig et al., 1989] Greig, D., Porteous, B., and Seheult, A. (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society Series B*, 51:271–279.

[Guichard et al., 2004] Guichard, F., Morel, J.-M., and Ryan, R. (2004). Contrast invariant image analysis and PDE's. preprint.

[Haas et al., 1967] Haas, A., Matheron, G., and Serra, J. (1967). Morphologie mathématique et granulométries en place. *Annales des Mines*, 11:736–753.

[Haddad, 2007] Haddad, A. (2007). Texture separation: $BV - G$ and $BV - L^1$ models. *Multiscale Modeling and Simulation*, 6(1):273–286.

[Haddad and Meyer, 2007] Haddad, A. and Meyer, Y. (2007). An improvement of rudin-osher-fatemi model. *Applied and Computational Harmonic Analysis*, 22(3):319 – 334.

[Hanbury and Serra, 2002] Hanbury, A. and Serra, J. (2002). Analysis of oriented textures using mathematical morphology. In *in OAGM*, pages 49–52.

[Harvey and Lawson, 1982] Harvey, R. and Lawson, H. (1982). Calibrated geometries. *Acta Mathematica*, 148(1):47–157.

[Hochbaum, 2001] Hochbaum, D. (2001). An efficient algorithm for image segmentation, markov random fields and related problems. *Journal of the ACM.*, 48(2):686–701.

[Hudson, 1978] Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.*, 6(3):473–484.

[Kawohl and Lachand-Robert, 2006] Kawohl, B. and Lachand-Robert, T. (2006). Characterization of cheeger sets for convex subsets of the plane. *Pacific J. Math.*, 225(1):103–118.

[Kawohl and Novaga, 2008] Kawohl, B. and Novaga, M. (2008). The p-Laplace eigenvalue problem as $p \to 1$ and Cheeger sets in a Finsler metric. *Journal of Convex Analysis*, 15(3):623.

[Kervrann and Boulanger, 2006] Kervrann, C. and Boulanger, J. (2006). Optimal spatial adaptation for patch-based image denoising. *IEEE Trans. Image Process.*, 15(10):2866–2878.

[Kervrann and Boulanger, 2008] Kervrann, C. and Boulanger, J. (2008). Local adaptivity to variable smoothness for exemplar-based image denoising and representation. *Int. J. Computer Vision*, 79:45–69.

[Kim and Vese, 2009] Kim, Y. and Vese, L. (2009). Image recovery using functions of bounded variation and sobolev spaces of negative differentiability. *Inverse Problems and Imaging*, 3:43–68.

[Kimia et al., 1995] Kimia, B. B., Tannenbaum, A. R., and Zucker, S. W. (1995). Shapes, shocks, and deformations i: The components of two-dimensional shape and the reaction-diffusion space. *International Journal of Computer Vision*, 15:189–224. 10.1007/BF01451741.

[Kindermann and Osher, 2005] Kindermann, S. and Osher, S. (2005). Saddle point formulation for a cartoon-texture decomposition. Technical Report 05-42, UCLA CAM report.

[Kindermann et al., 2005] Kindermann, S., Osher, S., and Jones, P. (2005). Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, 4:1091–1115.

[Kindermann et al., 2006] Kindermann, S., Osher, S., and Xu, J. (2006). Denoising by bv-duality. *J. Sci. Comput.*, 28:411–444.

[Koenderink, 1984] Koenderink, J. (1984). The structure of images. *Biological Cybernetics*, 50:363–370.

[Koko and Jehan-Besson, 2010] Koko, J. and Jehan-Besson, S. (2010). An augmented lagrangian method for $tv_g + l^1$ minimization. *Journal of Mathematical Imaging and Vision*, 38(3):182–196.

[Lawlor, 2010] Lawlor, G. (2010). Metacalibrations: A unifying framework for minimization proof. Technical report, Preprint.

[Lax, 1965] Lax, P. (1965). Numerical solution of partial differential equations. *Amer. Math. Monthly*, 72:74–85.

[Le and Vese, 2005] Le, T. and Vese, L. (2005). Image decomposition using total variation and div(bmo). *Multiscale Modeling and Simulation*, 4(2):390–423.

[Le Pennec and Mallat, 2005] Le Pennec, E. and Mallat, S. (2005). Sparse geometric image representations with bandelets. *IEEE Trans. Image Process.*, 14(4):423–438.

[Lee, 1983] Lee, J.-S. (1983). Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, 24(2):255–269.

[Lefton and Wei, 1997] Lefton, L. and Wei, D. (1997). Numerical approximation of the first eigenvalue of the p-laplacian using finite elements and the penalty method. *Numer. Funct. Anal. Optim.*, 18:389–399.

[Lepski, 1990] Lepski, O. V. (1990). On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and its Applications*, 35(3):454–466.

[Leung and Barron, 2006] Leung, G. and Barron, A. R. (2006). Information theory and mixing least-squares regressions. *IEEE Trans. Inf. Theory*, 52(8):3396–3410.

[Li, 1985] Li, K.-C. (1985). From Stein's unbiased risk estimates to the method of generalized cross validation. *Ann. Statist.*, 13(4):1352–1377.

[Lieu and Vese, 2009] Lieu, L. and Vese, L. (2009). Image restoration and decomposition via bounded total variation and negative Hilbert-Sobolev spaces. *Applied Mathematics & Optimization*, 58:167–193.

[Liu et al., 2010] Liu, J., Huang, H., Huan, Z., and Zhang, H. (2010). Adaptive variational method for restoring color images with high density impulse noise. *International Journal of Computer Vision*, 90:131–149. 10.1007/s11263-010-0351-9.

[Louchet, 2008] Louchet, C. (2008). *Modèles variationnels et bayésiens pour le débruitage d'images: de la variation totale vers les moyennes non-locales*. PhD thesis, Université Paris Descartes.

[Louchet and Moisan, 2010] Louchet, C. and Moisan, L. (2010). Total variation as a local filter. Technical report, Université Paris Descartes (MAP5).

[Louchet and Moisan, 2008] Louchet, M. and Moisan, L. (2008). Total variation denoising using posterior expectation. In *Proceedings EUSIPCO (best student paper award)*.

[Luisier, 2010] Luisier, F. (2010). *The SURE-LET approach to image denoising*. PhD thesis, EPFL.

[Mahmoudi and Sapiro, 2005] Mahmoudi, M. and Sapiro, G. (2005). Fast image and video denoising via nonlocal means of similar neighborhoods. *IEEE Signal Process. Lett.*, 12:839–842.

[Mairal et al., 2009] Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2009). Non-local sparse models for image restoration. *ICCV*.

[Mairal et al., 2008] Mairal, J., Sapiro, G., and Elad, M. (2008). Learning multiscale sparse representations for image and video restoration. *Multiscale Model. Simul.*, 7(1):214–241.

[Mallat and Peyré, 2007] Mallat, S. and Peyré, G. (2007). A review of bandlet methods for geometrical image representation. *Numerical Algorithms*, 44(3):205–234.

[Mallat and Zhang, 1993] Mallat, S. and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397 –3415.

[Maragos, 1989] Maragos, P. (1989). Pattern spectrum and multiscale shape representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7).

[Martin et al., 2001] Martin, D., Fowlkes, C., Tal, D., and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423.

[Maruani et al., 1995] Maruani, A., Pechersky, E., and Sigelle, M. (1995). On gibbs fields in image processing. *Markov Processes and Related Fields*, 1(3):419–442.

[Masnou and Morel, 1998] Masnou, S. and Morel, J.-M. (1998). Level lines based disocclusion. In *Proc. IEEE ICIP (Chicago 1998)*, volume 3, pages 259–263.

[Matheron, 1964] Matheron, G. (1964). Etude théorique des granulométries. Technical Report 57, Ecole des Mines de Paris.

[Matheron, 1975] Matheron, G. (1975). *Random Sets and Integral Geometry*. John Wiley & Sons, New York.

[Matheron and Serra, 2002] Matheron, G. and Serra, J. (2002). The birth of mathematical morphology. In Talbot, H. and Beare, R., editors, *Mathematical Morphology and its application to Image and Signal Processing (ISMM 2002)*, pages 1–16, Sidney.

[Meyer, 2001] Meyer, Y. (2001). *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

[Moll, 2005] Moll, J. S. (2005). The anisotropic total variation flow. *Mathematische Annalen*, 332(1):177–218–218.

[Monasse and Guichard, 2000] Monasse, P. and Guichard, F. (2000). Scale-space from a level lines tree. *Journal of Visual Communication and Image Representation*, 11(2):224 – 236.

[Morgan, 2009] Morgan, F. (2009). *Geometric Measure Theory: A Beginner's guide*. Academic Press, fourth edition.

[Morgan and Lawlor, 1994] Morgan, F. and Lawlor, G. (1994). Paired calibrations applied to soap films, immiscible fluids, and surfaces or networks minimizing other norms. *Pac. J Math.*, 166:55–83.

[Morgan and Vixie, 2007] Morgan, S. and Vixie, K. (2007). L^1TV computes the flat norm for boundaries. *Abstract and Applied Analysis*, 2007. Article ID 45153, doi:10.1155/2007/45153.

[Mumford and Shah, 1989] Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684.

[Nemirovski, 2000] Nemirovski, A. S. (2000). *Topics in non-parametric statistics*, volume 1738 of *Lecture Notes in Math.* Springer, Berlin.

[Nikolova, 2000] Nikolova, M. (2000). Local strong homogeneity of a regularized estimator. *SIAM Journal on Applied Mathematics*, 61(2):633–658.

[Nikolova, 2002] Nikolova, M. (2002). Minimizers of cost-functions involving non-smooth data-fidelity terms. application to the processing of outliers. *SIAM Journal on Numerical Analysis*, 40(3):965–994.

[Nikolova, 2004a] Nikolova, M. (2004a). A variational approach to remove outliers and impulse noise. *JMIV*, 20(1-2):99–120.

[Nikolova, 2004b] Nikolova, M. (2004b). Weakly constrained minimization. application to the estimation of images and signals involving constant regions. *Journal of Mathematical Imaging and Vision*, 21(2):155–175.

[Nikolova, 2005] Nikolova, M. (2005). Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *SIAM Journal on Multiscale Modeling and Simulation*, 4(3):960–991.

[Olver et al., 1994] Olver, P. J., Sapiro, G., and Tannenbaum, A. (1994). Differential invariant signatures and flows in computer vision: A symmetry group approach. In *Geometry Driven Diffusion in Computer Vision*. Kluwer.

[Osher et al., 2003] Osher, S., Sole, A., and Vese, L. (2003). Image decomposition and restoration using total variation minimization and the H^{-1} norm. *SIAM journal on Multiscale Modeling and Simulation*, 1(3):349–370.

[Pang et al., 2009] Pang, C., Au, O., Dai, J., Yang, W., and F.Zou (2009). A fast nl-means method in image denoising based on the similarity of spatially sampled pixels. In *MMSP*.

[Perona and Malik, 1990] Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639.

[Peyré, 2008] Peyré, G. (2008). Image processing with non-local spectral bases. *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, 7:703–730.

[Peyré, 2009] Peyré, G. (2009). Manifold models for signals and images. *Computer Vision and Image Understanding*, 113:249–260.

[Pock et al., 2010] Pock, T., Cremers, D., Bischof, H., and Chambolle, A. (2010). Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences*.

[Pock et al., 2008] Pock, T., Unger, M., Cremers, D., and Bischof, H. (2008). Fast and exact solution of total variation models on the gpu. In *CVPR Workshop on Visual Computer Vision on GPU's*.

[Pok et al., 2003] Pok, G., Liu, and Nair, A. (2003). Selective removal of impulse noise based on homogeneity level information. *Image Processing, IEEE Transactions on*, 12(1):85 – 92.

[Polzehl and Spokoiny, 2000] Polzehl, J. and Spokoiny, V. G. (2000). Adaptive weights smoothing with applications to image restoration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 62(2):335–354.

[Polzehl and Spokoiny, 2003] Polzehl, J. and Spokoiny, V. G. (2003). Image denoising: pointwise adaptive approach. *Ann. Statist.*, 31(1):30–57.

[Ramani et al., 2008] Ramani, S., Blu, T., and Unser, M. (2008). Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms. *IEEE Trans. Image Process.*, 17(9):1540–1554.

[Raphan and Simoncelli, 2007] Raphan, M. and Simoncelli, E. P. (2007). Learning to be Bayesian without supervision. In *NIPS*, volume 19, page 1145.

[Rudin et al., 1992] Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268.

[S. Setzer and Teuber, 2008] S. Setzer, G. S. and Teuber, T. (2008). Restoration of images with rotated shapes. *Numerical Algorithms*, 48(1):49–66.

[Salembier et al., 1998] Salembier, P., Oliveras, A., and Garrido, L. (1998). Anti-extensive connected operators for image and sequence processing. *IEEE Transactions on Image Processing*, 7(4):555–570.

[Salmon, 2010a] Salmon, J. (2010a). *Aggregation of estimators and patches methods to denoise numerical images*. PhD thesis, Université Paris Diderot - Paris 7.

[Salmon, 2010b] Salmon, J. (2010b). On two parameters for denoising with non-local means. *IEEE Signal Process. Lett.*

[Salmon and Le Pennec, 2009a] Salmon, J. and Le Pennec, E. (2009a). An aggregator point of view on NL-Means. In *Wavelets XIII*, volume 7446, page 74461E. SPIE.

[Salmon and Le Pennec, 2009b] Salmon, J. and Le Pennec, E. (2009b). NL-Means and aggregation procedures. In *ICIP*, pages 2977–2980.

[Salmon and Strozecki, 2010a] Salmon, J. and Strozecki, Y. (2010a). From patches to pixels in non-local methods: Weighted-Average reprojection. In *ICIP*.

[Salmon and Strozecki, 2010b] Salmon, J. and Strozecki, Y. (2010b). From patches to pixels in semi-local means: Weighted average reprojection. *Submitted*.

[Sapiro and Ringach, 1996] Sapiro, G. and Ringach, D. L. (1996). Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Transactions on Image Processing*, 5(11):1582–1586.

[Schoenemann et al., 2011] Schoenemann, T., Kahl, F., Masnou, S., and Cremers, D. (2011). A linear framework for region-based image segmentation and inpainting involving curvature penalization. *CoRR*, abs/1102.3830.

[Serra, 1982] Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, San Diego, CA.

[Serra, 1988] Serra, J. (1988). *Image Analysis and Mathematical Morphology. Volume 2: Theoretical Advances*. Academic Press.

[Singer et al., 2009] Singer, A., Shkolnisky, Y., and Nadler, B. (2009). Diffusion interpretation of nonlocal neighborhood filters for signal denoising. *SIAM Journal on Imaging Sciences*, 2:118–139.

[Smith and Brady, 1997] Smith, M. and Brady, J. M. (1997). Susan - a new approach to low level image processing. *International Journal of Computer Vision*, 23:45–78.

[Sochen et al., 1998] Sochen, N., Kimmel, R., and Malladi, R. (1998). A general framework for low level vision. *IEEE Transactions on Image Processing*, 7(3):310–318.

[Soille, 2003] Soille, P. (2003). *Morphological Image Analysis: Principles and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[Solo, 1996] Solo, V. (1996). A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems. In *ICIP*, volume 3, pages 89–92.

[Spohn, 1993] Spohn, H. (1993). Interface motion in models with stochastic dynamics. *Journal of statistical physics*, 71(5):1081–1132.

[Starck et al., 2005] Starck, J., ELad, M., and Donoho, D. (2005). Image decomposition via the combination of sparse representations and a variational approach. *IEEE Trans. On Image Processing*, 14(10):1570–1582.

[Starck et al., 2002] Starck, J.-L., Candès, E. J., and Donoho, D. L. (2002). The curvelet transform for image denoising. *IEEE Trans. Image Process.*, 11(6):670–684.

[Steidl and Teuber, 2009] Steidl, G. and Teuber, T. (2009). Diffusion tensors for processing sheared and rotated rectangles. *IEEE Transactions on Image Processing*, 18(12):2640–2648.

[Steidl et al., 2004] Steidl, G., Weickert, J., Brox, T., Mrazek, P., and Welk, M. (2004). On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides. *SIAM J. Numer. Anal.*, 42(2):686 – 713.

[Stein, 1981] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9:1135–1151.

[Strang, 1983] Strang, G. (1983). Maximal flow through a domain. *Mathematical Programming*, 26:123–143. 10.1007/BF02592050.

[Strang, 2009] Strang, G. (2009). Maximum flows and minimum cuts in the plane. *Advances in Applied Mathematics and Global Optimization*, pages 1–11.

[Strong et al., 1996] Strong, D. M., Chan, T. F., and R (1996). Exact solutions to total variation regularization problems. Technical report, UCLA CAM Report.

[Sun and Neuvo, 1994] Sun, T. and Neuvo, Y. (1994). Detail-preserving median based filters in image processing. *Pattern Recogn. Lett.*, 15:341–347.

[Szlam, 2006] Szlam, A. (2006). *Non-Stationary Analysis on Datasets and Applications*. PhD thesis, Yale University.

[Szlam, 2008] Szlam, A. (2008). Non-local means for audio denoising. Technical Report 08-56, UCLA CAM report.

[Tasdizen, 2009] Tasdizen, T. (2009). Principal neighborhood dictionaries for nonlocal means image denoising. *IEEE Transactions on Signal Processing*, 18:2649–2660.

[Taylor, 1978] Taylor, J. (1978). Crystalline variational problems. *AMERICAN MATHEMATICAL SOCIETY*, 84(4):568.

[Taylor et al., 1992] Taylor, J., Cahn, J., and Handwerker, C. (1992). Geometric models of crystal growth. *Acta metallurgica et materialia*, 40:1443–1474.

[Tikhonov and Arsenin, 1977] Tikhonov, A. and Arsenin, V. (1977). *Solutions of Ill-Posed Problems*. Winston, Washington, DC.

[Tomasi and Manduchi, 1998] Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Proceedings of the 1998 IEEE International Conference on Computer Vision*, Bombay, India.

[Tschumperlé and Brun, 2008] Tschumperlé, D. and Brun, L. (2008). Image denoising and registration by PDE's on the space of patches. In *International Workshop on Local and Non-Local Approximation in Image Processing (LNLA'08), Lausanne/Switzerland*.

[Tschumperlé and Brun, 2009] Tschumperlé, D. and Brun, L. (2009). Non-local image smoothing by applying anisotropic diffusion PDE's in the space of patches. In *IEEE International Conference on Image Processing (ICIP'09), Cairo/Egypt*.

[Tschumperlé and Deriche, 2005] Tschumperlé, D. and Deriche, R. (2005). Vector-valued image regularization with PDEs: A common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4).

[Tsybakov, 2003] Tsybakov, A. B. (2003). Optimal rates of aggregation. In *COLT*, pages 303–313.

[Vaillant and Glaunes, 2005] Vaillant, M. and Glaunes, J. (2005). Surface matching via currents. In *IPMI*, pages 381–392.

[Van De Ville and Kocher, 2009] Van De Ville, D. and Kocher, M. (2009). SURE-based Non-Local Means. *IEEE Signal Process. Lett.*, 16:973–976.

[Van De Ville and Kocher, 2011] Van De Ville, D. and Kocher, M. (2011). Non-local means with dimensionality reduction and sure-based parameter selection. *IEEE Transactions on Image Processing*. In press.

[Van De Ville et al., 2003] Van De Ville, D., Nachtegael, M., Van der Weken, D., Kerre, E., Philips, W., and Lemahieu, I. (2003). Noise reduction by fuzzy image filtering. *Fuzzy Systems, IEEE Transactions on*, 11(4):429 – 436.

[Vanrell and Vitria, 1993] Vanrell, M. and Vitria, J. (1993). Mathematical morphology, texture perception and granulometries. In *SPIE Image Algebra and Morphological Image Proceessing IV*, pages 152–161.

[Vese and Osher, 2003] Vese, L. and Osher, S. (2003). Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19:553–572.

[Vincent, 1993] Vincent, L. (1993). Grayscale area openings and closings, their efficient implementation and applications. In Serra, J. and Salembier, P., editors, *Proceedings of the 1st Workshop on Mathematical Morphology and its Applications to Signal Processing*, pages 22–27, Barcelona, Spain.

[Viola and Jones, 2001] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *CVPR*, 1:511–518.

[Vixie et al., 2010] Vixie, K., Clawson, K., Asaki, T., Sandine, G., Morgan, S., and Price, B. (2010). Multiscale flat norm signature for shapes and images. *Applied Mathematical Sciences*, 4(14):667–680.

[Vixie and Esedoglu, 2007] Vixie, K. and Esedoglu, S. (2007). Some properties of minimizers for the Chan-Esedoglu L^1TV functional. Technical report, arXiv [http://arXiv.org/oai2] (United States).

[Wang et al., 2006] Wang, J., Guo, Y.-W., Ying, Y., Liu, Y.-L., and Peng, Q.-S. (2006). Fast non-local algorithm for image denoising. In *ICIP*, pages 1429–1432.

[Wang et al., 2004] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Signal Process.*, 13(4):600–612.

[Wedel et al., 2008a] Wedel, A., Pock, T., Braun, J., Franke, U., and Cremers, D. (2008a). Duality tv-l1 flow with fundamental matrix prior. In *Image Vision and Computing*, Auckland, New Zealand.

[Wedel et al., 2008b] Wedel, A., Pock, T., Zach, C., Cremers, D., and Bischof, H. (2008b). An improved algorithm for TV-L1 optical flow. In *Proc. of the Dagstuhl Motion Workshop*, LNCS. Springer.

[Weickert, 1996] Weickert, J. (1996). Theoretical foundations of anisotropic diffusion in image processing. *Computing Suppl.*, 11:221–236.

[Weickert, 1998] Weickert, J. (1998). *Anisotropic Diffusion in Image Processing*. ECMI Series, Teubner-Verlag, Stuttgart, Germany.

[Windyga, 2001] Windyga, P. (2001). Fast impulsive noise removal. *Image Processing, IEEE Transactions on*, 10(1):173 –179.

[Witkin, 1983] Witkin, A. P. (1983). Scale-space filtering. In *Proc. 8th Int. Joint Conf. Art. Intell.*, pages 1019–1022, Karlsruhe, Germany.

[Xu et al., 2004] Xu, H., Zhu, G., Peng, H., and Wang, D. (2004). Adaptive fuzzy switching filter for images corrupted by impulse noise. *Pattern Recognition Letters*, 25(15):1657 – 1663.

[Yaroslavsky, 1985] Yaroslavsky, L. P. (1985). *Digital Picture Processing - An Introduction*. Springer Verlag.

[Yin et al., 2005] Yin, W., Chen, T., Zhou, X., and Chakraborty, A. (2005). Background correction for cdna microarray image using the tv+l1 model. *Bioinformatics*, 21(10):2410–2416.

[Yin et al., 2007a] Yin, W., Goldfarb, D., and Osher, S. (2007a). A comparison of three total variation based texture extraction models. *Journal of Visual Communication and Image Representation*, 18(3):240–252.

[Yin et al., 2007b] Yin, W., Goldfarb, W., and Osher, S. (2007b). The total variation regularized L^1 model for multiscale decomposition. *SIAM Journal on Multiscale Modeling and Simulation*, 6(1):190–211.

[Yuan et al., 2010] Yuan, J., Bae, E., and Tai, X.-C. (2010). A study on continuous max-flow and min-cut approaches. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2217 –2224.

[Zhang and Luo, 1999] Zhang, X.-P. and Luo, Z.-Q. (1999). A new time-scale adaptive denoising method based on wavelet shrinkage. In *ICASSP '99: Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference*, pages 1629–1632, Washington, DC, USA. IEEE Computer Society.