



HAL
open science

Vision nocturne numérique : restauration automatique et recalage multimodal des images à bas niveau de lumière

Camille Sutour

► **To cite this version:**

Camille Sutour. Vision nocturne numérique : restauration automatique et recalage multimodal des images à bas niveau de lumière. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université de Bordeaux, 2015. Français. NNT : 2015BORD0099 . tel-01225612

HAL Id: tel-01225612

<https://theses.hal.science/tel-01225612>

Submitted on 6 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ DE BORDEAUX
ÉCOLE DOCTORALE DE MATHÉMATIQUES ET
D'INFORMATIQUE

THÈSE

présentée pour obtenir le grade de
Docteur en Mathématiques appliquées
de l'Université de Bordeaux
Spécialité : Mathématiques appliquées et calcul scientifique

Camille SUTOUR

Vision nocturne numérique : Restauration automatique et recalage multimodal des images à bas niveau de lumière

Soutenue le 10 juillet 2015 devant le jury composé de

Laure	BLANC-FÉRAUD	CNRS - Sophia Antipolis	Présidente de jury
Agnès	DESOLNEUX	ENS Cachan	Rapporteur
Charles	KERVANN	INRIA Rennes	Rapporteur
Jean-François	AUJOL	Université de Bordeaux	Directeur de thèse
Jean-Philippe	DOMENGER	Université de Bordeaux	Directeur de thèse
Charles-Alban	DELEDALLE	Université de Bordeaux	Co-encadrant de thèse

À mes neveux

Remerciements

Je souhaite tout d'abord faire part de ma gratitude envers tous ceux qui m'ont soutenue et accompagnée dans cette aventure.

Je tiens à remercier les membres du jury d'avoir accepté d'examiner cette thèse. Merci à Agnès Desolneux et Charles Kervrann de m'avoir fait l'honneur d'être les rapporteurs de cette thèse. Je remercie également Laure Blanc-Féraud d'avoir accepté d'être examinatrice. Et je remercie naturellement mes directeurs de thèse, Jean-François Aujol, Jean-Philippe Domenger et Charles-Alban Deledalle pour leur encadrement et leur soutien.

Lorsque j'ai envisagé de commencer cette thèse, pesant le pour et le contre entre plusieurs options, Jean-François m'a éclairée dans mon choix en me conseillant d'évaluer le potentiel de la thèse selon trois critères : le sujet, le lieu, et les encadrants.

Si je n'ai pas exactement signé pour les encadrants dont j'ai finalement hérité, je n'aurais pas pu espérer mieux, aussi bien sur le plan scientifique qu'humain. Merci Jean-Philippe d'avoir supervisé et accompagné ma thèse. Merci Jean-François pour ton encadrement, tes conseils, ton soutien. Merci d'avoir cru en moi et de n'avoir pas baissé les bras. Merci pour tes idées, ta disponibilité et ta bonne humeur. Et merci Charles d'avoir été l'encadrant bonus de ma thèse. Au delà des heures passées au tableau, moments que j'affectionnais particulièrement, tu as su endosser les multiples casquettes de stimulateur d'idées, de service informatique (merci, entre autres, à MooseTEX!), de relecteur (bon... pas pour l'orthographe), et d'ami.

Concernant le choix du sujet, j'ai eu la chance de bénéficier d'un sujet polyvalent qui m'a permis d'explorer plusieurs aspects de la recherche. Merci à la DGA et la région Aquitaine d'avoir financé ce projet. Je remercie également Thales Avionics d'avoir donné un sens à mon travail, et en particulier Matthieu Grossetête qui a su me consacrer une précieuse part de son temps. Merci également à Anne-Sophie Auger d'avoir effectué son stage à Thales et de m'avoir permis de découvrir l'encadrement de stages, à Emilie Darrin et Max Clogenson d'avoir égayé mes journées à Thales, ainsi qu'à Jean-Noël Perbet et le GIS Albatros d'avoir répondu à mes besoins matériels.

Ma thèse se caractérise aussi par l'environnement dans lequel j'ai eu la chance d'évoluer pendant trois ans, en particulier l'IMB. Je remercie tout d'abord l'ensemble de l'équipe Image pour avoir su créer une atmosphère agréable, d'un point de vue professionnel et humain, ainsi que les collègues du LaBRI. Nicolas, merci pour les conseils et l'aide que tu as su me donner parfois. Bravo et merci pour l'organisation de SSVM, c'est incontestablement la conférence qui aura le plus marqué ma thèse ! Baudouin, merci d'avoir participé à mes travaux sur le recalage, et d'avoir égayé ma deuxième année de thèse. Charles Do., merci d'avoir été mon pilier, mon inspiration.

Je remercie également l'ensemble du personnel de l'IMB, et en particulier ceux qui m'ont aidée dans ma seconde aventure de l'année, mon agrégation. Je ne serais pas parvenue à gérer cette agrégation sans les oraux blancs, les livres, les conseils et le soutien que m'ont fournis mes collègues ou anciens professeurs.

Et parce que cette thèse n'aurait pas la même saveur sans eux, je remercie

chaleureusement les doctorants qui ont évolué avec moi ces trois dernières années. Je commence naturellement par mes co-bureaux, Romain et Marc. Romain, merci d'avoir été là, merci pour ta gentillesse et ton humour, et merci d'avoir veillé à notre intégrité intellectuelle ! Tu m'as manqué cette année, ça m'a fait du bien de partager ces deux années avec toi. Marc, merci pour ta coolitude, pour les moments détente et le soutien. Merci également à Fabien d'avoir été un doctorant plus sérieux que nous, et d'avoir veillé à la régularité de nos pauses par tes visites quotidiennes. Je remercie ensuite l'ensemble des doctorants avec qui l'on a partagé des pauses, des soirées, des séminaires, etc. Zoé, merci d'accepter de fumer pour donner un sens à nos pauses, et merci de les agrémenter de ta bonne humeur.

Je ne suis pas poète, je n'ai pas ton talent,
Mais je ne peux manquer dans mes remerciements,
De louer ton humour, ton écoute et tes rimes,
Merci à toi Bruno, et tous tes pseudonymes.

Samuel, merci pour ton innocence. Merci également à Marie, Alice, Elsa, Alan, JB, Pierre, Sami, Thomas, Corentin et les autres. Vous formez une belle équipe.

J'ai donc eu la chance de bénéficier pour ma thèse d'un sujet intéressant, encadré par des personnes exceptionnelles dans un environnement privilégié. Mais une thèse c'est aussi une personne, qui fait évoluer le sujet durant trois ans avec ses idées, ses envies, ses doutes, ses caprices, son caractère. Et je ne serais pas là aujourd'hui sans ces personnes qui m'entourent et qui me sont si chères. Je remercie tout particulièrement mes piliers, mes conseillères, mes confidentes, mes meilleures amies, mes jumelles, Caroline et Laure. Je n'ai pas les mots pour vous dire combien je vous aime, mais vous savez que je ne serais rien sans vous. Merci bien évidemment à ma Maman, qui me soutient quelles que soient mes décisions et qui est toujours là pour nous. Merci à Touti, mon Papa, de m'aider, de me soutenir et de m'accompagner dans chacune de mes aventures. Merci en particulier pour l'assistance informatique ! Merci à mon cocon, ma famille. Mes Grand-mères, mes taties et mes oncles, mes cousins, David, Manon, Fanny, Théo, Matthis, Anaëlle et Julien, ainsi que Lolotte, Philippe, Anthony, et enfin Alex, fraîchement intégré mais déjà promu le meilleur beau-frère dont on puisse rêver. Et un merci tout particulier à mes neveux, Samuel et Sacha. Vous êtes mon rayon de soleil, merci d'illuminer ma vie.

Et merci à ma deuxième famille, mes amis. Merci Maïlys et Gersende de faire perdurer 20 ans d'amitié. Vous êtes exceptionnelles. Si je ne suis pas certaine que mes directeurs vous remercient de m'avoir si souvent divertie de mes tâches professionnelles, je tiens à remercier les danseurs de swing de Bordeaux (et d'ailleurs !) de partager cette passion pour la danse. Je remercie en particulier les Swing-Gum, Aurélie, Max, Julie, Benoit, Claire, Raph et Mike pour ces si bons moments vécus ensemble. Vous avez été ma dose hebdomadaire de rire et de bonne humeur, mon énergie. Merci également aux potes du rock pour tous ces moments partagés.

Je tiens également à remercier ces personnes qui ont eu une influence particulière sur ma vie professionnelle, et sans qui je ne serais pas là aujourd'hui. J'ai eu notamment deux professeurs de mathématiques qui ont contribué à m'amener jusqu'ici. Merci à M. Roussel, mon professeur en première année de classe prépa à Montaigne,

d'avoir allumé en moi la flamme des mathématiques et de l'enseignement. Je pense que je n'aurais pas eu l'agrégation cette année si je n'avais pas eu la chance de commencer mon parcours mathématique avec un professeur aussi exceptionnel. Et merci également à Charles Do. de m'avoir fait découvrir le traitement d'images. Tu as accompagné mes premiers pas dans le monde du traitement d'images et de la recherche, tu as suivi l'évolution de ma thèse, et tu as endossé le rôle de coach sportif et mental ces derniers mois. Je ne serais tout simplement pas là aujourd'hui si je n'avais pas déserté un TD de licence il y a cinq ans pour assister au tien à la place. Merci de m'avoir permis de découvrir ce milieu. Et merci à Simon Masnou de ne pas m'avoir laissée le quitter.

Et le meilleur pour la fin, merci Marcel, merci d'être là, merci d'être toi.

Résumé des travaux

L'objectif de cette thèse est d'améliorer la vision de nuit des pilotes d'hélicoptère à l'aide d'un dispositif de vision nocturne numérique. Ce dernier est constitué d'un intensificateur de lumière (IL) couplé à une caméra CCD (*Charge Coupled Device*) d'une part et d'une caméra infrarouge (IR) d'autre part. Nous nous consacrons en particulier à l'amélioration des images optiques issues de l'intensificateur de lumière. Cela repose dans un premier temps sur une étude du dispositif d'acquisition, afin de caractériser les dégradations dont souffrent les images IL, et sur l'estimation de la nature du bruit présent en fonction des conditions et de la luminosité. Une telle étude permet ensuite d'effectuer le débruitage des images IL, à l'aide d'un algorithme adaptatif de débruitage vidéo. Enfin, les images optiques et infrarouge peuvent être fusionnées pour combiner les informations issues des deux modalités. Cela nécessite d'effectuer un recalage multimodal des deux capteurs, basé sur l'alignement des contours.

Estimation automatique du bruit

La méthode d'estimation automatique du bruit que nous avons développée repose sur l'analyse des statistiques du bruit sur des zones homogènes de l'image. En effet, sur de petites régions de l'image dans lesquelles le signal peut être considéré constant, seules les statistiques du bruit interviennent. Il s'agit donc dans un premier temps de mettre en place une méthode automatique de détection des zones homogènes, afin de pouvoir à partir de ces statistiques en déduire un modèle pour le bruit.

Détection non paramétrique des régions homogènes

Les images naturelles ont la propriété de contenir des zones homogènes, c'est-à-dire des régions dans lesquelles le rapport signal à bruit est très faible. Ces dernières fournissent une estimation des statistiques du bruit. La détection des régions homogènes que nous proposons est non paramétrique et requiert simplement que le bruit soit non corrélé spatialement. Elle repose sur la corrélation de rang de Kendall [Kendall 1938], un test non paramétrique qui évalue la corrélation entre deux séquences $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^n$ de réalisations des variables aléatoires X et Y de la façon suivante :

$$\tau(x, y) = \frac{1}{n(n-1)} \sum_{1 \leq i, j \leq n} \text{sign}(x_i - x_j) \text{sign}(y_i - y_j). \quad (1)$$

Cette formulation fait l'hypothèse qu'il n'y a pas d'égalités, c'est-à-dire de couples tels que $i \neq j$ et $x_i = x_j$ ou $y_i = y_j$. Dans un souci de clarté, on néglige dans ce résumé le cas des égalités, mais une extension du coefficient $\tau(x, y)$ qui en tient compte est proposée dans [Kendall 1945], et est prise en compte dans ce manuscrit.

Le coefficient de Kendall est lié à l'ordre relatif des valeurs, et non aux valeurs directement. $\tau(x, y)$ est compris entre -1 et 1, et une valeur nulle indique l'absence de corrélation significative entre x et y .

Sous l'hypothèse d'indépendance de X et Y , notée H_0 , le z -score défini par

$$z(x, y) = \frac{\tau(x, y)}{\sqrt{2(2n+5)/9n(n+1)}} \quad (2)$$

est asymptotiquement distribué selon une loi normale centrée réduite lorsque la taille de la séquence n augmente [Kendall 1948] :

$$z(X, Y) \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1). \quad (3)$$

Cette propriété nous permet de contrôler la probabilité de fausse alarme du test d'hypothèse, c'est-à-dire qu'il est possible pour une probabilité de fausse alarme P_{FA} fixée de déterminer un seuil α tel que

$$\lim_{n \rightarrow \infty} \mathbb{P}(|z(X, Y)| > \alpha \mid H_0) = P_{FA}, \quad (4)$$

dès lors que X et Y sont décorrélés, et ce quelle que soit leur distribution. En d'autres termes, cela quantifie la probabilité de détecter du signal dans une zone homogène, c'est-à-dire de manquer la détection d'une zone homogène (considérer qu'elle est inhomogène alors qu'elle est homogène).

On cherche également à contrôler la probabilité de fausse détection, c'est-à-dire la probabilité sous une hypothèse alternative H_1 que $|z(X, Y)|$ soit inférieur au seuil α . On considère l'hypothèse alternative d'un signal saut corrompu par un bruit blanc additif gaussien défini par

$$\forall 1 \leq k \leq n, X_k = \begin{cases} \varepsilon_k & \text{si } k \leq n/2 \\ a + \varepsilon_k & \text{si } k > n/2 \end{cases} \quad \text{et} \quad Y_k = \begin{cases} \eta_k & \text{si } k \leq n/2 \\ a + \eta_k & \text{si } k > n/2 \end{cases}, \quad (5)$$

où ε et η sont des variables aléatoires indépendantes distribuées selon $\mathcal{N}(0, \sigma^2)$. Nous avons démontré dans cette étude que sous cette hypothèse, l'espérance et la variance du z -score sont déterminés asymptotiquement par

$$\mathbb{E}[z(X, Y) \mid H_1] = O(\sqrt{n}) \quad \text{et} \quad \text{Var}[z(X, Y) \mid H_1] = O(1). \quad (6)$$

Cela permet ainsi de contrôler la probabilité de fausse détection associée au seuil α :

$$\mathbb{P}(|z(X, Y)| < \alpha \mid H_1) = O\left(\frac{1}{n}\right). \quad (7)$$

Cette probabilité de fausse détection correspond à la probabilité de considérer qu'une zone est homogène alors qu'elle contient du signal (erreur la plus gênante en pratique). Les équations (4) et (7) garantissent ainsi que le test est asymptotiquement un détecteur à taux de fausse alarme constant.

Nous appliquons ce test d'hypothèses à la détection des zones homogènes en considérant que si un bloc b de l'image est homogène, cela signifie qu'il n'y a pas de

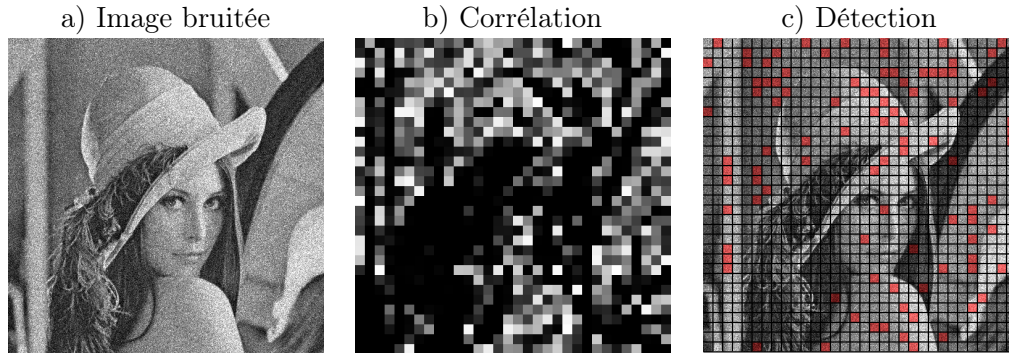


FIGURE 1 – Détection des zones homogènes. a) Image bruitée (sur $[0, 255]$), PSNR initial 17.93dB, b) Évaluation de la corrélation associée au τ de Kendall calculé sur des blocs de taille $N_b = 16 \times 16$, et c) Sélection des blocs homogènes (en rouge) associée à la probabilité de détection $P_D = 1 - P_{FA} = 0.7$.

structure liée au signal mais seulement du bruit, et donc pas de corrélation au sein du bloc (pourvu que le bruit soit décorrélé). On évalue donc la corrélation entre les séquences voisines $x = (b_{2k})$ et $y = (b_{2k+1})$, où $2k$ et $2k + 1$ représentent les indices des pixels voisins pour un parcours donné d'un bloc b de l'image. Quatre tests dans des directions différentes sont effectués et les résultats sont agrégés afin d'obtenir un estimateur plus sélectif.

La figure 1 illustre le processus de sélection de ces zones homogènes. La corrélation est calculée sur des blocs de taille 16×16 pixels, pour des voisinages horizontaux, verticaux et diagonaux. La significativité du test (selon la direction horizontale seulement) est représentée sur la figure 1-b, où le blanc indique l'absence de corrélation significative. Les zones détectées comme homogènes sont représentées en rouge sur la figure 1-c.

À partir de ces zones homogènes sélectionnées sont extraites les statistiques du bruit, c'est-à-dire les couples $(\hat{\mu}, \hat{\sigma}^2)$ des intensités moyennes et des variances évaluées sur chacun de ces blocs. L'étape suivante consiste alors à associer un modèle à ces statistiques.

Estimation robuste de la fonction de niveau de bruit

Afin de pouvoir estimer un modèle de bruit à partir des estimations $(\hat{\mu}, \hat{\sigma}^2)$ des moyennes et des variances issues des zones homogènes, on suppose à ce stade que la fonction de niveau de bruit (NLF, pour *Noise Level Function*), c'est-à-dire la fonction qui relie la variance σ^2 du bruit à l'intensité sous-jacente f de l'image, est une fonction polynomiale de degré deux donnée par

$$\sigma^2 = \text{NLF}_{(a,b,c)}(f) = af^2 + bf + c. \quad (8)$$

On estime alors les coefficients (a, b, c) de la NLF à l'aide de l'estimateur LAD

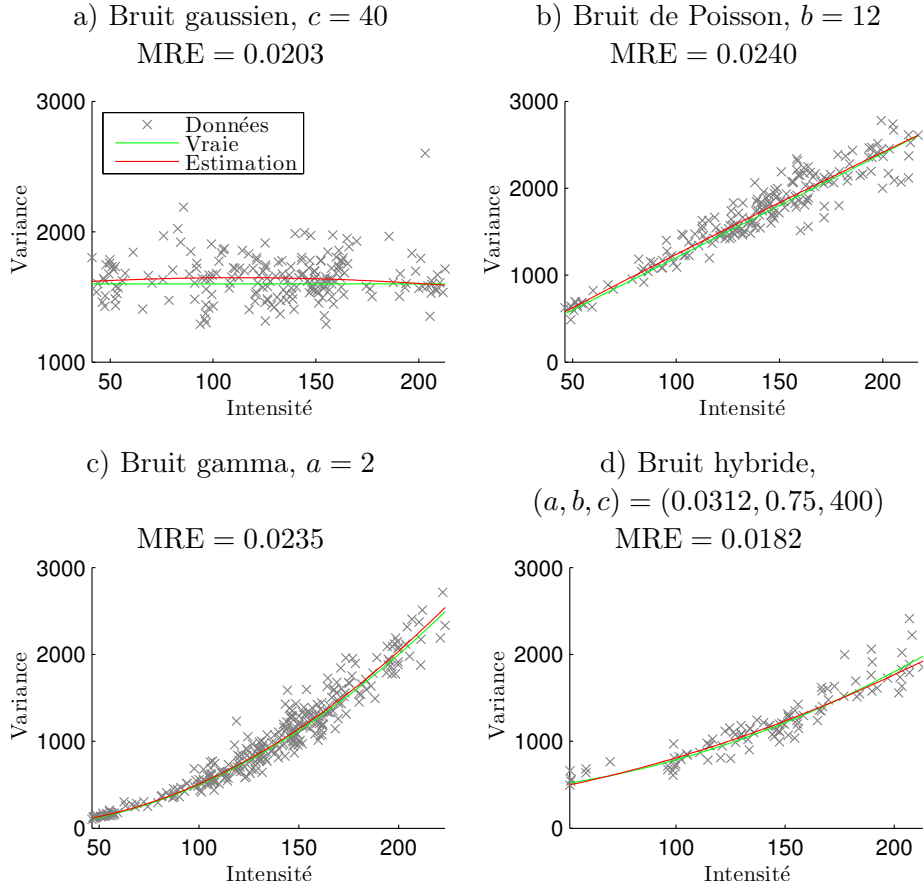


FIGURE 2 – Estimation des paramètres pour différents types de bruit avec l'estimateur LAD.

(*Least Absolute Deviation*) qui minimise l'erreur ℓ^1 :

$$(\widehat{a}, \widehat{b}, \widehat{c}) = \arg \min_{(a,b,c) \in (\mathbb{R}^+)^3} \|\text{NLF}_{(a,b,c)}(\hat{\mu}) - \hat{\sigma}^2\|_1 = \arg \min_{(a,b,c) \in (\mathbb{R}^+)^3} \|a\hat{\mu}^2 + b\hat{\mu} + c - \hat{\sigma}^2\|_1. \quad (9)$$

Cet estimateur est plus robuste que son analogue des moindres carrés qui minimise l'erreur quadratique. Il assure une meilleure gestion des outliers, susceptibles de se produire en cas de fausse détection.

La résolution de ce problème de minimisation s'effectue alors à l'aide de l'algorithme primal-dual préconditionné de Chambolle et Pock [Chambolle 2011].

La figure 2 illustre les résultats d'estimation des paramètres de la NLF pour plusieurs types de bruit. Sur la figure 2-a le bruit est additif gaussien, donc la vraie NLF est une fonction constante, représentée par la droite horizontale verte. Sur la figure 2-b, le bruit est de nature poissonnienne, et sa NLF est une fonction linéaire, tandis que sur la figure 2-c, il s'agit d'un bruit gamma dont la NLF est parabolique. Enfin sur la figure 2-d, le bruit est hybride et correspond à la somme de bruit additif gaussien, d'un bruit de Poisson et d'un bruit gamma.

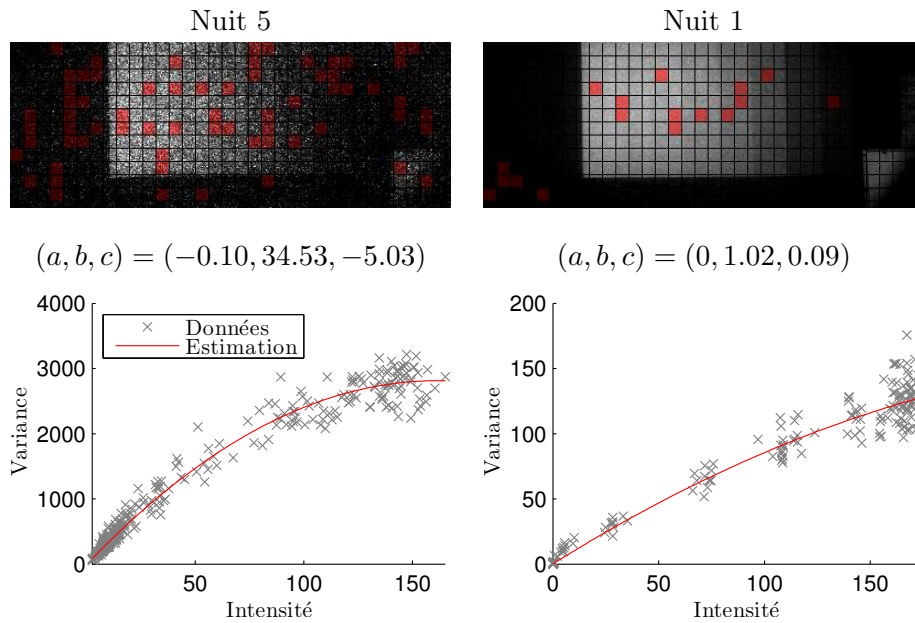


FIGURE 3 – Détection des zones homogènes et estimation de la NLF pour des images d'une mire de contraste issues d'un banc de test en salle noire, à différents niveaux de nuit.

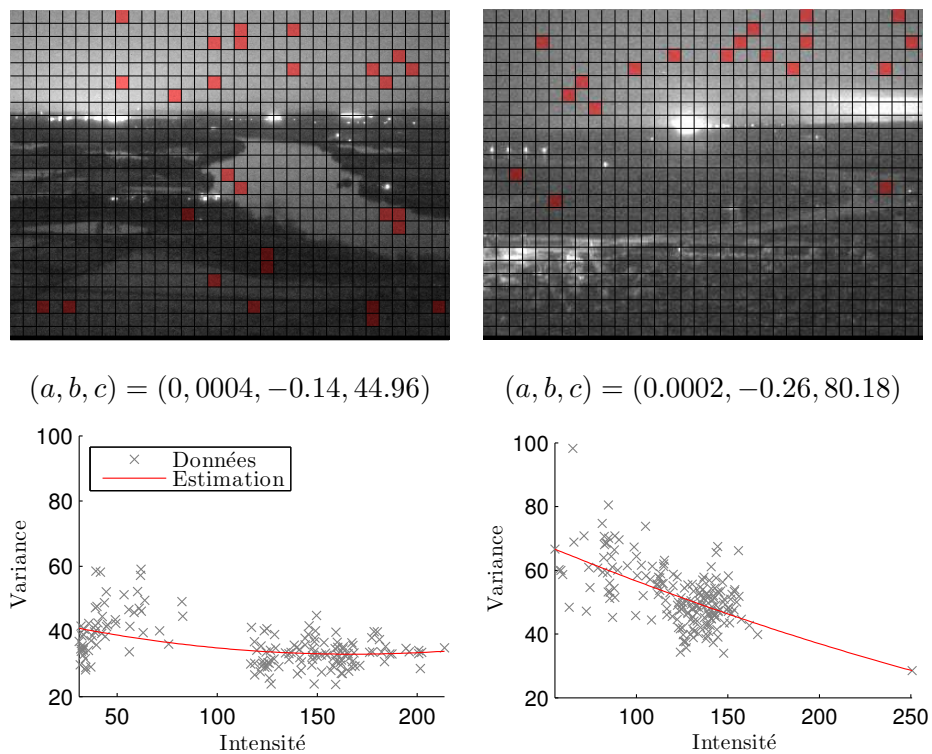


FIGURE 4 – Détection des zones homogènes et estimation de la NLF sur images réelles.

La méthode d'estimation du bruit peut alors être utilisée sur les images BNL. Ces dernières sont très fortement corrélées, donc on effectue un sous-échantillonnage des images pour réduire la corrélation spatiale et rendre le test de Kendall applicable. Il est possible d'effectuer l'estimation du bruit a priori sur des images issues de bancs de test, acquises en salle noire calibrée. Cela peut permettre de fournir, pour un dispositif donné et dans des conditions de luminosité prédéfinies, une loi pour les paramètres de la NLF. C'est illustré sur la figure 3, où les mesures sont effectuées sur une mire de contraste, à deux niveaux de luminosité calibrés différents.

Grâce à la rapidité d'implémentation et au caractère automatique de la méthode, l'estimation peut également s'envisager directement sur des données de vol, comme illustré sur la figure 4.

Ces estimations permettent ainsi d'associer un modèle de distribution au bruit présent sur les images BNL. Munis d'une meilleure appréhension du bruit, nous pouvons alors nous consacrer à la suppression de celui-ci en développant une méthode de débruitage adaptée.

La coopération entre les méthodes variationnelles et les moyennes non locales

Le débruitage des séquences d'images BNL se doit de satisfaire deux contraintes. Conformément aux résultats de l'analyse effectuée en première partie, la nature du bruit présent sur les images IL est fortement non gaussienne, et varie en fonction des conditions d'éclairage. Il est donc nécessaire de développer un algorithme de débruitage qui soit adaptable en fonction du modèle de bruit. Les méthodes variationnelles [Rudin 1992] et les moyennes non locales [Buades 2005b] répondent à ce besoin. En effet, l'adaptation de l'attache aux données du modèle de ROF à la log-vraisemblance permet de tenir compte de la statistique du bruit, et des algorithmes de résolution adaptés ont été proposés par exemple dans [Aubert 2008, Anthoine 2012], tandis que les travaux de [Kervrann 2007, Deledalle 2010, Delon 2012, Deledalle 2014] étendent les NL-means au bruit multiplicatif, impulsif ou au bruit de Poisson. De plus, le débruitage doit s'appliquer à des séquences d'images, donc tirer partie de la redondance d'information apportée par le flux vidéo, tout en limitant les coûts calculatoires en vue d'une implémentation embarquée. Nous avons donc choisi de nous affranchir de la compensation de mouvement entre les trames, coûteuse et souvent imparfaite, grâce à l'implémentation vidéo des moyennes non locales qui étend la recherche des candidats à un voisinage spatio-temporel [Buades 2005a]. Après avoir réalisé une étude approfondie des NL-means et des méthodes variationnelles et identifié leurs défauts respectifs, la coopération des deux méthodes permet de tirer le meilleur de chacune.

Dejittering des NL-means

Les NL-means classiques souffrent d'un effet de gigue, ou *jittering*, qui se manifeste par un sur-lissage lié à la sélection de candidats non pertinents dans la moyenne

pondérée. Cet effet introduit un biais lié à une réduction de variance trop importante [Kervrann 2007]. L'approche utilisée consiste à rééquilibrer le compromis biais variance en réintroduisant de la donnée bruitée lorsque le débruitage est jugé non pertinent, c'est-à-dire lorsque la réduction de variance est trop importante. Pour cela, on réalise une combinaison convexe adaptative de la solution des NL-means d'origine u^{NL} et de l'image bruitée g donnée en chaque pixel i par :

$$u_i^{\text{NLDJ}} = (1 - \alpha_i)u_i^{\text{NL}} + \alpha_i g_i = \sum_{j \in \Omega} w_{i,j}^{\text{NLDJ}} g_j, \quad (10)$$

où $w_{i,j}^{\text{NLDJ}} = (1 - \alpha_i)w_{i,j}^{\text{NL}} + \alpha_i \delta_{i,j}$ ($\delta_{i,j}$ désignant le symbole de Kronecker), et α_i est un indice de confiance, ou indice de *jittering* donné par :

$$\alpha_i = \frac{|(\hat{\sigma}_i^{\text{NL}})^2 - (\sigma_i^{\text{noise}})^2|}{|(\hat{\sigma}_i^{\text{NL}})^2 - (\sigma_i^{\text{noise}})^2| + (\sigma_i^{\text{noise}})^2}. \quad (11)$$

$(\sigma_i^{\text{noise}})^2$ désigne la variance du bruit, supposée à ce stade connue, et $(\hat{\sigma}_i^{\text{NL}})^2$ la variance non locale du signal évaluée de la façon suivante :

$$(\hat{\sigma}_i^{\text{NL}})^2 = \sum_{j \in \Omega} w_{i,j}^{\text{NL}} g_j^2 - \left(\sum_{j \in \Omega} w_{i,j}^{\text{NL}} g_j \right)^2. \quad (12)$$

Si au sein d'un voisinage non local les candidats sélectionnés sont issus de la même population, alors la variance $(\hat{\sigma}_i^{\text{NL}})^2$ doit être proche de celle du bruit. Cela entraîne un indice de *jittering* α_i proche de zéro, donc la solution d'origine u_i^{NL} est conservée. Si au contraire les candidats au sein d'un voisinage sont issus de différentes populations, ce qui est responsable de l'effet de *jittering*, alors la variance $(\hat{\sigma}_i^{\text{NL}})^2$ sera significativement plus importante que celle du bruit, et donc l'indice de *jittering* α_i est plus proche de 1. Cela conduit donc à réinjecter de la donnée bruitée g_i . De plus, la variance résiduelle en un pixel i de la solution u^{NLDJ} après débruitage et *dejittering* peut être estimée par

$$(\hat{\sigma}_i^{\text{résiduel}})^2 = \left[\sum_{j \in \Omega} (w_{i,j}^{\text{NLDJ}})^2 \right] (\sigma_i^{\text{noise}})^2. \quad (13)$$

La quantité $\sum_{j \in \Omega} (w_{i,j}^{\text{NLDJ}})^2$ reflète la quantité de bruit qui a été ôtée au pixel i , et traduit la qualité du débruitage obtenu, vis à vis du compromis biais-variance.

La figure 5 illustre l'effet du *dejittering*. On note en particulier sur la solution NLDJ de la figure 5-b que du bruit a été réintroduit au niveau du chapeau, restaurant ainsi de l'information perdue sur la solution des NL-means de la figure 5-a. L'information réintroduite est visible sur la carte des différences figure 5-c.

Si la correction du *jittering* développée permet d'affiner la réduction de bruit en terme de compromis biais variance, elle réintroduit néanmoins du bruit. À celui-ci vient s'ajouter un bruit résiduel déjà présent, lié à la difficulté de trouver assez de candidats sur les structures singulières pour effectuer un débruitage suffisant. Cela résulte donc en un effet de patch rare accru, que nous corrigeons à l'aide d'une régularisation adaptative.



FIGURE 5 – Illustration du *de-jittering*. a) Solution des NL-means classiques, b) Solution des NL-means *de-jittered*, et c) Différence entre les deux solutions.

Régularisation adaptative

La réduction de l'effet de patch rare, qui se manifeste sur les images par la présence de bruit résiduel autour des contours et des structures singulières, doit être effectuée de façon adaptative, afin de ne pas sur-lisser les zones correctement débruitées ou détruire le travail réalisé par l'étape de *de-jittering*. Nous nous basons pour cela sur l'indice de qualité du débruitage $(\hat{\sigma}_i^{\text{résiduel}})^2$ défini dans l'équation 13 afin de développer une régularisation localement adaptative. Celle-ci repose sur une régularisation TV associée à une attache aux données non locale, pondérée par l'indice de confiance de la façon suivante :

$$\begin{aligned} u^{\text{R-NL}} &= \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \lambda_i \sum_{j \in \Omega} w_{i,j} (g_j - u_i)^2 + \text{TV}(u) \\ &= \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \lambda_i (u_i - u_i^{\text{NL}})^2 + \text{TV}(u), \end{aligned} \quad (14)$$

où λ_i est un paramètre de régularisation adaptatif donné par

$$\lambda_i = \gamma \left(\frac{\hat{\sigma}_i^{\text{résiduel}}}{\sigma_i^{\text{noise}}} \right)^{-1} = \gamma \left(\sum_{j \in \Omega} w_{i,j}^2 \right)^{-1/2}. \quad (15)$$

Intuitivement, ce modèle s'interprète de la façon suivante : lorsque le débruitage est satisfaisant, la quantité de bruit résiduel $(\hat{\sigma}_i^{\text{résiduel}})^2$ est faible, donc le paramètre de régularisation λ_i est élevé. Dans l'équation 14, cela se traduit par une attache aux données prépondérante, donc la solution $u_i^{\text{R-NL}}$ est maintenue proche de u_i^{NL} . Si au contraire du bruit résiduel subsiste, cela signifie que la quantité $(\hat{\sigma}_i^{\text{résiduel}})^2$ est trop importante, et donc le paramètre de régularisation λ_i est faible. Dans l'équation 14, cela se traduit par un terme de régularisation prépondérant sur l'attache aux données, ce qui pousse à régulariser d'avantage la solution $u_i^{\text{R-NL}}$.

La figure 6 illustre l'effet de la régularisation adaptative. La solution de R-NL présentée sur la figure 6-b est plus lisse que la solution de NLDJ présentée

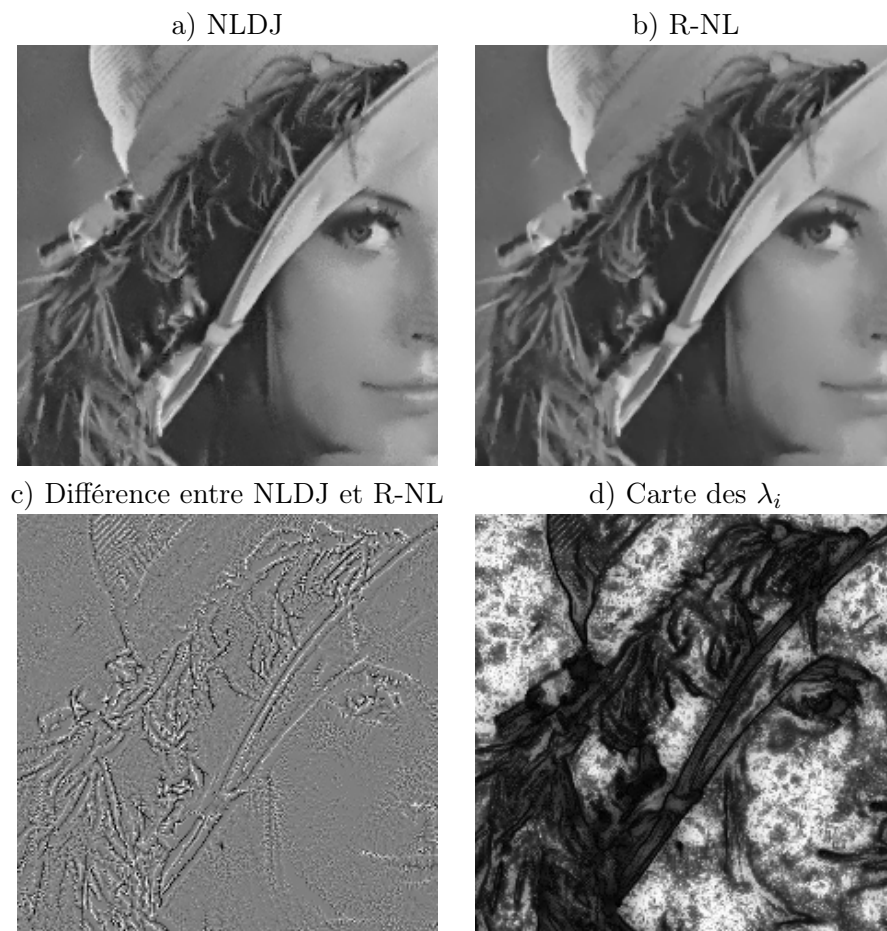


FIGURE 6 – Illustration de la régularisation adaptative. a) Solution des NL-means (*dejittered*), b) Solution régularisée R-NL, c) Différence entre les deux solutions, et d) Carte du paramètre de régularisation adaptatif λ_i qui reflète l'indice de confiance en les NL-means.

sur la figure 6-a. On l'observe notamment autour de l'œil et du chapeau, et c'est confirmé par la carte des différences de la figure 6-c. La figure 6-d représente la carte du paramètre de régularisation adaptatif λ_i ; conformément à l'intuition, ce dernier est fort sur les zones plutôt homogènes et redondantes où les NL-means sont performants, et faible au niveau de contours et des structures singulières. On note en particulier le caractère très local de celui-ci, ce qui permet à la régularisation de s'appliquer de façon sélective.

De plus, grâce aux bonnes propriétés des NL-means et des méthodes variationnelles, cet algorithme peut s'adapter naturellement à de nombreux types de bruit, et les implémentations rapides des NL-means [Darbon 2008] ainsi que les méthodes d'optimisation [Chambolle 2011, Combettes 2005, Anthoine 2012] fournissent un algorithme rapide et modulable. Nous avons en particulier pu adapter cet algorithme à des bruits à variance polynomiale comme ceux utilisés pour l'estimation de bruit

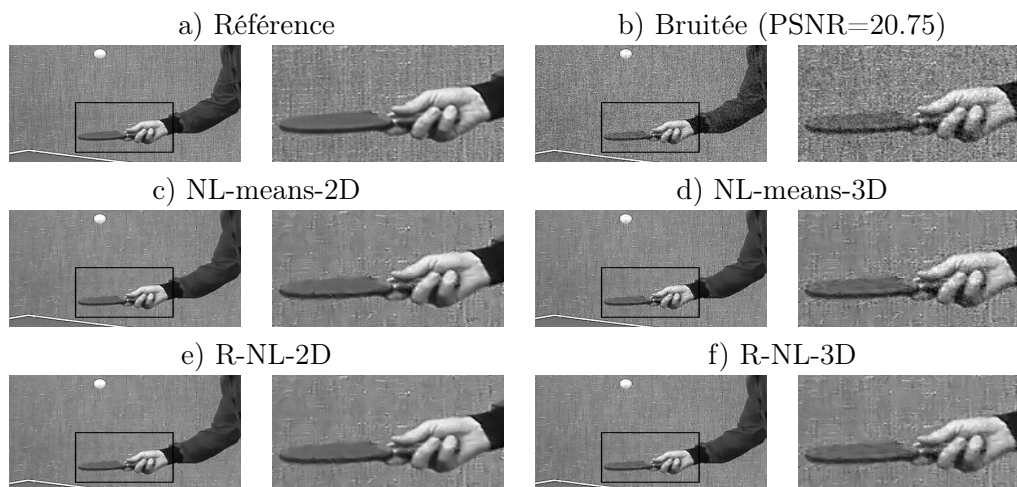


FIGURE 7 – Différence entre patches 2D et patches 3D. Extraits de séquences d’images débruitées par des méthodes à patches 2D, à savoir NL-means-2D, R-NL-2D et V-BM3D, et par des méthodes à patches 3D, NL-means-3D, R-NL-3D et BM4D. La texture du fond de l’image est perdue avec NL-means-2D et R-NL-2D. NL-means-3D préserve mieux le fond mais souffre de bruit résiduel autour du bras en mouvement, effet corrigé par la régularisation adaptative de R-NL-3D.

dans la première partie, développant ainsi l’algorithme des NLF-means et sa version régularisée R-NLF.

Adaptation à la restauration de vidéos

Le choix d’étudier les NL-means en vue de développer une méthode de débruitage des images BNL reposait d’une part sur leur capacité à s’adapter à d’autres modèles de bruit, mais aussi sur leur extension au débruitage vidéo. L’algorithme R-NL s’applique donc naturellement au débruitage de séquences d’images. D’après [Buades 2008], les NL-means permettent un débruitage efficace des vidéos sans compensation préalable de mouvement entre les trames en utilisant une fenêtre de recherche spatio-temporelle. En effet, puisque les moyennes non locales utilisent la redondance des structures au sein d’une image, il semble naturel et pertinent d’exploiter également la redondance temporelle. Une recherche de candidats spatio-temporelle non locale permet ainsi de garantir qu’un maximum de candidats sont susceptibles d’être trouvés pour assurer un débruitage efficace.

Cependant, cette méthode souffre d’un manque de régularité temporelle. En effet, la compensation de mouvement permet en principe de respecter l’hypothèse lambertienne, qui stipule qu’un pixel doit au cours de sa trajectoire temporelle garder la même intensité. Cette hypothèse n’est plus garantie par les NL-means spatio-temporels, ce qui crée une absence de régularité qui se manifeste par un scintillement dont l’effet est proche de celui du bruit. Afin de contrer cet effet indésirable, nous suivons l’approche adoptée par [Wexler 2007] dans un contexte d’inpainting vidéo,

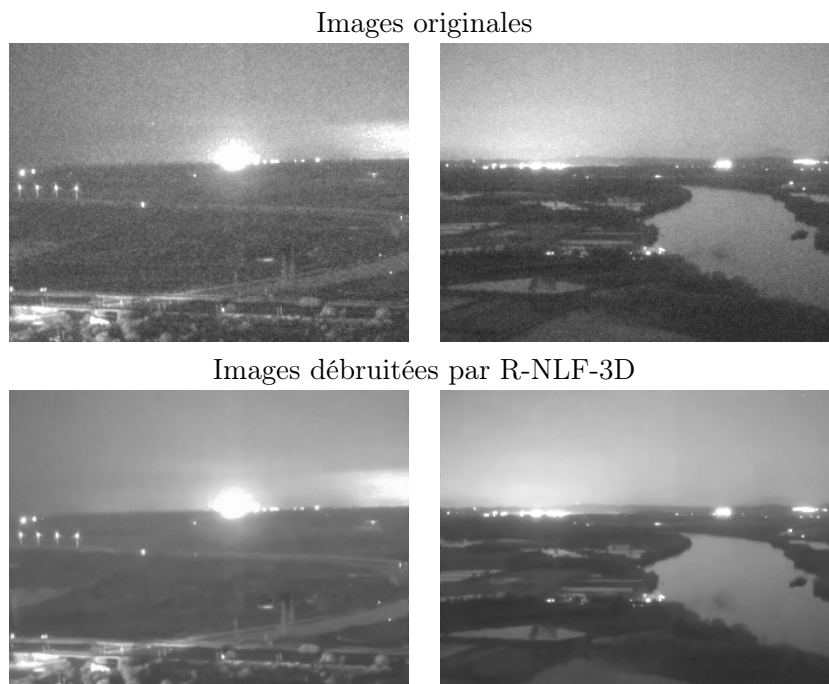


FIGURE 8 – Images acquises en vol et débruitage automatique réalisé par l’algorithme R-NLF-3D.

et nous utilisons des patchs spatio-temporels. Cela impose de la cohérence temporelle dans la recherche de candidats, ce qui assure plus de régularité temporelle. Cependant, les patchs tridimensionnels sont de taille plus importante, et rendent la recherche de candidats plus sélective, ce qui accroît le bruit résiduel. Celui-ci se manifeste particulièrement autour des structures en mouvement. De façon analogue au débruitage d’images, l’effet de patch rare est alors corrigé par la régularisation TV adaptative. Sauf que dans ce cas, toujours afin de maintenir la cohérence temporelle, la variation totale est spatio-temporelle également. Si l’emploi de la variation totale temporelle peut sembler peu appropriée lorsqu’elle est appliquée de façon globale, ici elle n’est appliquée que très localement et de façon adaptative.

Les apports des patchs spatio-temporels et de la régularisation adaptative sont illustrés qualitativement sur de séquences d’images, comme sur la figure 7, où l’on peut observer (à défaut de l’instabilité temporelle, qui ne s’observe que sur les vidéos) la meilleure préservation des textures garantie par l’utilisation des patchs 3D, ainsi que la réduction du bruit résiduel assurée par la régularisation adaptative.

Ces méthodes de débruitage adaptatif peuvent ensuite s’appliquer au débruitage des séquences d’images BNL, grâce à l’estimation de bruit réalisée en première partie et à l’algorithme NLF-means, ou la version régularisée R-NLF, qui permettent de débruiter des bruits à variance polynomiale dépendant du signal. Sur la figure 8 sont représentées deux images issues du dispositif IL, et les résultats débruités à l’aide de l’algorithme R-NLF-3D, qui estime automatiquement les paramètres du bruit

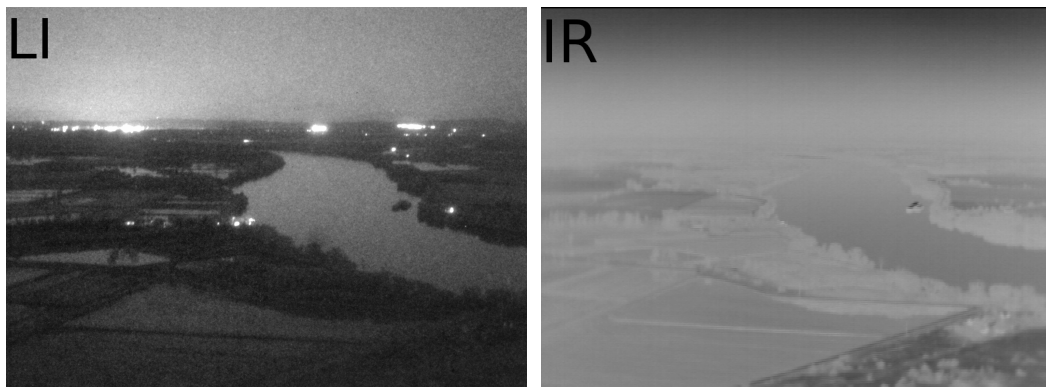


FIGURE 9 – Exemple d’une image optique (IL, à gauche) et d’une image infrarouge (IR, à droite) acquises simultanément durant le vol d’un hélicoptère. La résolution et le type d’information révélé, ainsi que les distributions des intensités, sont très différentes d’une modalité à l’autre.

et les réinjecte dans l’algorithme de débruitage spatio-temporel à patchs 3D. Des améliorations sont également étudiées, notamment l’idée suggérée dans [Liu 2010] d’effectuer malgré tout un recalage des trames afin de maximiser la redondance et optimiser la sélection de candidats. Afin de réduire les coûts calculatoires, nous proposons pour cela d’utiliser les capteurs de mouvement embarqués pour estimer le déplacement de la caméra et optimiser la recherche temporelle de candidats à moindres coûts.

Les algorithmes de débruitage, combinés à l’estimation préalable du bruit, permettent d’obtenir des images IL de qualité rehaussée. D’autres traitements seraient envisageables, notamment effectuer un rehaussement de contraste, un rééquilibrage de la dynamique, une correction des artefacts comme la saturation des sources lumineuses et les reflets, et un rehaussement des contours. Ce dernier peut en fait être contourné en utilisant les informations apportées par l’autre modalité, à savoir la caméra infrarouge. En effet celle-ci laisse bien apparaître les contours comme les routes ou les bâtiments. Réaliser la fusion des deux modalités permettrait ainsi de renforcer l’information des contours au sein de l’image IL. Cependant, en raison de l’emplacement différent des caméras, une telle fusion nécessite d’effectuer un recalage préalable des deux modalités dans le même référentiel.

Recalage multimodal

La troisième partie de cette thèse est donc consacrée au recalage des données issues de la caméra IL, située sur le casque du pilote, avec celles issues de la caméra infrarouge, située sous l’hélicoptère. Ces deux caméras observent la scène sous différents angles de vue, se déplacent indépendamment l’une de l’autre et présentent des caractéristiques différentes : les deux capteurs n’ont pas la même résolution et

fournissent une réponse différente en terme d'intensités et du type d'information représenté. La figure 9 illustre les caractéristiques de chacune des modalités : on note en particulier les différences d'intensité des deux images, mais aussi les différences de texture, de détail et de contraste.

Réaliser le recalage des deux modalités consiste à transformer l'image courante notée u , qui correspond ici à l'image IL, dans le référentiel de l'image de référence, notée v , ici l'image infrarouge. Pour cela, on recherche la transformation spatiale T telle que l'image transformée $u(T)$ soit alignée avec v . Les techniques de recalage reposent souvent sur l'appariement de points d'intérêts [Lowe 2004], sur la maximisation de critères statistiques comme la corrélation croisée [Pratt 1978, Roshni 2008] ou l'information mutuelle [Viola 1997, Maes 1997] ou sur l'utilisation de modèles de représentation des images [Irani 1998, Heinrich 2012]. Dans le contexte des images BNL, les caractéristiques de chacune étant très différentes, il faut s'appuyer sur le contenu commun à ces deux modalités ; c'est pourquoi nous utilisons l'information des contours.

Alignement des contours

Suivant les idées proposées dans [Sun 2004, Haber 2006], nous définissons une métrique qui mesure l'alignement des contours des deux images de la façon suivante :

$$C(T) = \int_{\Omega_c} |\nabla u(T(\mathbf{x})) \cdot \nabla v(\mathbf{x})| \, d\mathbf{x}, \quad (16)$$

qui peut également s'écrire sous la forme suivante :

$$C(T) = \int_{\Omega_c} w_T(\mathbf{x}) |\cos(\Delta\theta_T(\mathbf{x}))| \, d\mathbf{x}, \quad (17)$$

où Ω_c désigne le domaine (continu) de l'image, et $w_T(\mathbf{x})$ et $\Delta\theta_T(\mathbf{x})$ sont définis à partir de l'amplitude M et de l'orientation θ des gradients des images localisés en \mathbf{x} par :

$$\begin{aligned} w_T(\mathbf{x}) &= M_u(T(\mathbf{x}))M_v(\mathbf{x}), \\ \Delta\theta_T(\mathbf{x}) &= \theta_u(T(\mathbf{x})) - \theta_v(\mathbf{x}). \end{aligned} \quad (18)$$

Maximiser ce critère correspond à maximiser le produit scalaire entre les gradient des images. Il favorise les contours forts, grâce à la pondération par les amplitudes, ce qui permet de ne prendre en compte que les contours présents dans les deux images. De plus, la valeur absolue du cosinus permet de rendre le critère insensible aux inversions de contraste en ne prenant en compte que la direction du gradient et non son orientation. Ces deux aspects en font un critère d'alignement particulièrement adapté à la multimodalité du problème.

Dans un premier temps, la recherche de la transformation optimale peut s'effectuer sous forme d'une recherche exhaustive sur l'espace des paramètres. On considère pour cela un modèle de transformation constitué d'un zoom uniforme de paramètre z

et d'une translation dans les deux directions, horizontale et verticale, de paramètres respectifs t_1 et t_2 .

À partir du critère proposé, une analyse théorique sur un signal en une dimension soumis à un zoom et à une translation permet de montrer que malgré la présence de maxima locaux, la maximisation du critère conduit effectivement à un unique maximum global qui correspond aux paramètres de transformation recherchés. Des expériences pratiques sur des images synthétiques permettent également de valider la capacité du critère à estimer correctement la transformation optimale, ainsi que sa résistance au bruit et son comportement autour du maximum global en terme de résistance aux maxima locaux. Ce dernier point est crucial pour pouvoir envisager une alternative à la recherche exhaustive par une méthode d'optimisation.

En effet, au delà de la transformation constituée du changement d'échelle et des translations, un modèle plus général qui permet de tenir compte des différences de perspective entre les deux caméras serait à envisager, notamment lorsque l'hélicoptère vole à basse altitude. Pour cela, on utilise le modèle des transformations projectives [Hartley 2003] à huit degrés de liberté. Ce modèle est une extensions des transformations affines ; il généralise le modèle précédent et permet de définir notamment des mouvements de rotation et des changements de perspective. Cependant, si la recherche exhaustive pour trois paramètres était viable en pratique, elle n'est plus compatible avec une implémentation temps réel sur un espace à huit dimensions. C'est pourquoi on utilise une méthode d'optimisation qui réalise une montée de gradient du critère proposé.

Schéma d'optimisation

Grâce à son expression basée sur un produit scalaire, le gradient du critère peut se calculer explicitement. Dans le cas de la transformation simple, celui-ci s'exprime de la façon suivante :

$$\begin{cases} t_1^{n+1} &= t_1^n + \lambda_1 \partial_{t_1} C(T_{t_1^n, t_2^n, z^n}) \\ t_2^{n+1} &= t_2^n + \lambda_2 \partial_{t_2} C(T_{t_1^n, t_2^n, z^n}) \\ z^{n+1} &= z^n + \lambda_3 \partial_z C(T_{t_1^n, t_2^n, z^n}), \end{cases} \quad (19)$$

où les dérivées de la fonction $C(T_{t_1, t_2, z})$ sont données à chaque itération par :

$$\begin{aligned} \partial_{t_1} C(T_{(t_1, t_2, z)}) &= \int_{\Omega_c} \text{sign}(\nabla u(T_{t_1, t_2, z}(\mathbf{x}))) \cdot \nabla v(\mathbf{x}) D^2 u(T_{t_1, t_2, z}(\mathbf{x})) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \nabla v(\mathbf{x}) d\mathbf{x}, \\ \partial_{t_2} C(T_{(t_1, t_2, z)}) &= \int_{\Omega_c} \text{sign}(\nabla u(T_{t_1, t_2, z}(\mathbf{x}))) \cdot \nabla v(\mathbf{x}) D^2 u(T_{t_1, t_2, z}(\mathbf{x})) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \nabla v(\mathbf{x}) d\mathbf{x}, \\ \partial_z C(T_{(t_1, t_2, z)}) &= \int_{\Omega_c} \text{sign}(\nabla u(T_{t_1, t_2, z}(\mathbf{x}))) \cdot \nabla v(\mathbf{x}) D^2 u(T_{t_1, t_2, z}(\mathbf{x})) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \nabla v(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (20)$$

Un résultat analogue peut être obtenu dans le cas des transformations projectives, où cette fois le gradient est à huit dimensions.

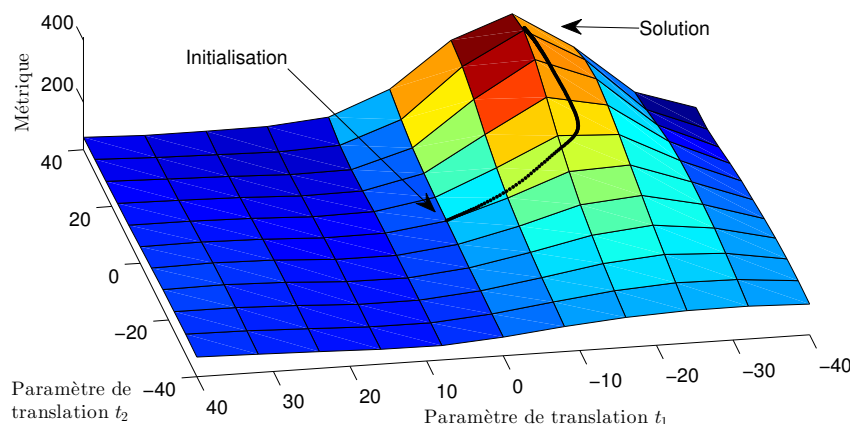


FIGURE 10 – Carte du critère de similarité pour une grille de paramètres de translations, à zoom fixé, et évolution des paramètres estimés à chaque itération de la montée de gradient. L’algorithme d’optimisation permet effectivement de maximiser la métrique, et converge vers le maximum.

Cependant dans les deux cas, le critère n’est pas convexe, donc la montée de gradient peut converger vers un maximum local au lieu du maximum global. Il convient donc de s’assurer qu’autour du maximum global la fonctionnelle est suffisamment lisse, et que l’on peut fournir une bonne initialisation raisonnablement proche du maximum global.

Sur la figure 10 est représentée l’évolution du critère en fonction des deux paramètres de translation, tandis que le zoom est fixé au paramètre optimal. Le chemin parcouru à chaque étape de la montée de gradient est également représenté. On observe qu’autour du maximum global, la variation du critère est lisse et ne souffre pas de maxima locaux. Cela confirme également la capacité de la montée de gradient à atteindre le maximum global, pourvu que l’initialisation soit pertinente.

Une telle initialisation est garantie par le flux temporel fourni par la vidéo. En effet, chaque couple d’images peut bénéficier des informations de recalage obtenues aux trames précédentes. Mieux encore, le mouvement indépendant de chacune des caméras peut être estimé rapidement à l’aide du détecteur de posture présent sur le casque et des systèmes inertiels de l’hélicoptère. Ainsi, la connaissance des paramètres de recalage multimodal à l’instant $t - 1$ et du mouvement de chacune des caméras permet de fournir une bonne estimation des paramètres de recalage à l’instant t , comme illustré sur le schéma 11. Une telle implémentation fournit non seulement une bonne initialisation plus susceptible de converger vers le maximum global, mais aussi un système de contrôle qui permet de comparer la solution obtenue à l’estimation issue de la propagation du résultats des trames précédentes. Une estimation du déplacement de chaque caméra permet également de gérer les mouvements plus importants, comme un déplacement brusque de la tête du pilote, qui déplacerait les paramètres optimaux plus loin de l’initialisation non propagée, au risque de ne pas converger vers le maximum global. La prise en compte de tels

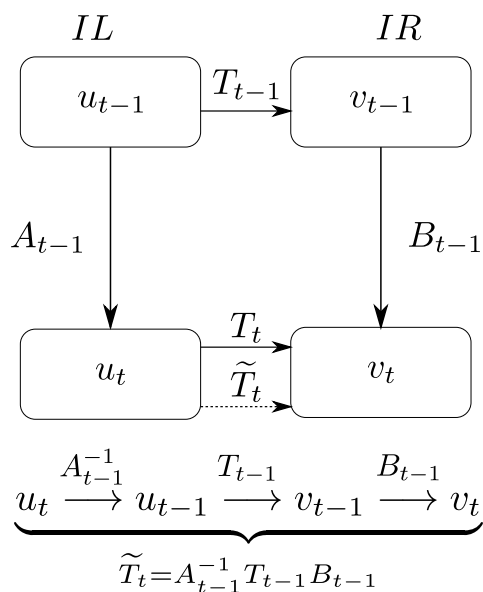


FIGURE 11 – Schéma temporel garantissant la convergence rapide de la montée de gradient, ainsi qu'un contrôle de l'erreur.

mouvements par le schéma temporel permet notamment de gérer ces cas de figure.

Ainsi, la mise en place de la montée de gradient couplée au schéma temporel permet d'effectuer un recalage rapide et robuste des séquences d'images BNL. La figure 12 illustre le recalage de trois couples d'images réelles acquises en vol. La précision du recalage est soulignée par la superposition des contours de l'image de référence ainsi que par la création des images mosaïques qui alternent par bandes l'image de référence et l'image courante, afin de constater l'alignement des contours.

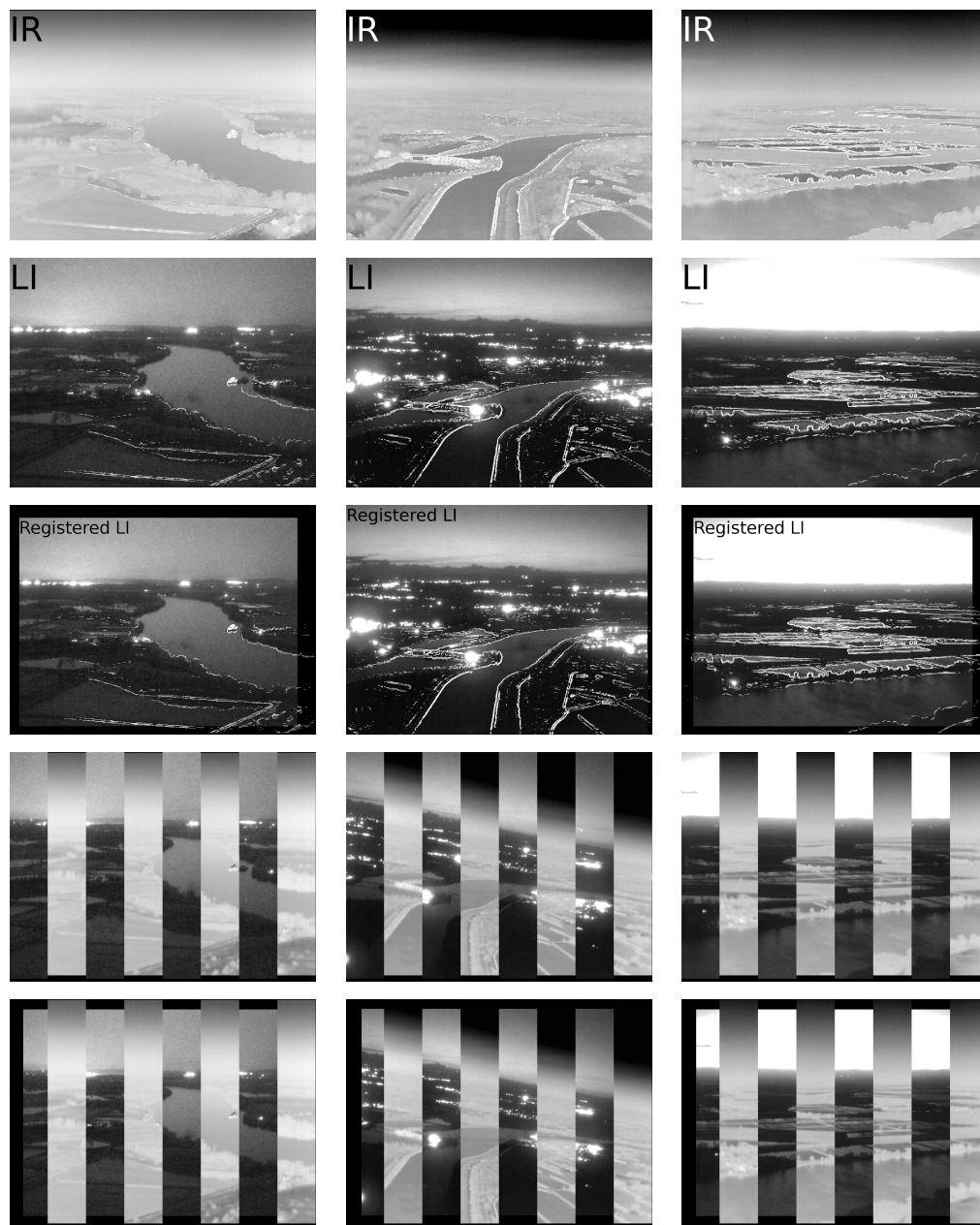


FIGURE 12 – Résultats de recalage obtenus sur données réelles à l'aide de la métrique proposée. L'image de référence IR est représentée sur la première ligne, l'image courante IL (à recaler) sur la deuxième, et l'image recalée sur la troisième. Les contours extraits de l'image IR sont superposés sur chaque image pour appréhender visuellement la qualité du recalage. Sur les quatrième et cinquième lignes sont représentées des images mosaïques avant et après recalage qui alternent des bandes de l'image IR et de l'image IL à recaler, pour mieux révéler les différences d'alignement. Les séquences complètes peuvent être téléchargées sur <http://image.math.u-bordeaux1.fr/Registration>.

Table des matières

1	Introduction	35
1.1	La vision de nuit	35
1.1.1	Les images IL	36
1.1.2	Combinaison aux images infrarouge	38
1.1.3	État de l'art des traitements développés à Thales	39
1.2	Synthèse des travaux et contributions	40
I	Analyse du bruit	45
2	Modélisation du bruit pour les systèmes BNL	49
2.1	Introduction	49
2.2	Le bruit en imagerie	50
2.2.1	Notations	50
2.2.2	Le bruit blanc additif gaussien	50
2.2.3	Le bruit de Poisson	52
2.2.4	Le bruit de speckle	52
2.2.5	Le bruit impulsif	53
2.2.6	Le bruit poivre et sel	53
2.2.7	Le bruit Poisson-gaussien	53
2.2.8	Retour au bruit gaussien : stabilisation de variance	54
2.3	Acquisition des images BNL	55
2.3.1	Le tube intensificateur de lumière	55
2.3.2	La caméra CCD	57
2.3.3	Modélisation du bruit	58
2.3.4	Estimation empirique du bruit dans les images IL	59
2.4	Conclusion	60
3	Estimation non paramétrique du bruit	63
3.1	Introduction	63
3.1.1	Contexte et notations (rappel)	65
3.2	Détection des zones homogènes	66
3.2.1	Homogénéité et décorrélation	66
3.2.2	Test d'indépendance non paramétrique	68
3.2.3	Illustration de la détection	75
3.3	Estimation du modèle	76
3.3.1	Estimation ponctuelle issue d'une seule zone homogène	76
3.3.2	Estimation de la fonction de niveau de bruit à partir de l'ensemble des zones homogènes	79
3.4	Résultats sur images synthétiques	84

3.4.1	Application au débruitage d'images	84
3.4.2	Comparaison à l'état de l'art	85
3.5	Application aux images BNL	87
3.5.1	Sur un banc d'images test	87
3.5.2	Sur des données de vol	88
3.6	Conclusion	91
3.7	Annexes du chapitre : preuves	92
3.7.1	Preuve du théorème 3.2.5	92
3.7.2	Preuve de la proposition 3.2.6	92
3.7.3	Preuve de la proposition 3.2.7	97
3.7.4	Preuve du corollaire 3.2.8	103
3.7.5	Preuve du théorème 3.2.9	103
3.7.6	Preuve de la proposition 3.3.2	104
II Débruitage		105
4	État de l'art du débruitage d'images et motivations	109
4.1	Introduction	109
4.2	Le débruitage	109
4.2.1	Évaluation du débruitage	110
4.2.2	Filtres linéaires	111
4.2.3	Filtrage non linéaire	113
4.3	Les méthodes variationnelles	116
4.3.1	Le modèle de ROF	116
4.3.2	Résolution du problème de minimisation (4.17)	117
4.3.3	Adaptation à d'autres types de bruit	121
4.3.4	Limites et discussion	124
4.4	Les méthodes non locales	126
4.4.1	L'algorithme des NL-means	126
4.4.2	Implémentation	130
4.4.3	Adaptation à d'autres types de bruit	132
4.4.4	Extensions et discussion	133
4.5	Conclusion	135
4.6	Annexes du chapitre : Éléments d'analyse convexe	137
4.6.1	Généralités	137
4.6.2	Méthode de Chambolle	139
4.6.3	Algorithme forward-backward	140
4.6.4	Generalized forward-backward	141
4.6.5	Algorithme primal-dual	143

5	Régularisation adaptative des moyennes non locales	145
5.1	Introduction	145
5.2	<i>Dejittering</i> des NL-means (NLDJ)	146
5.3	NL-means régularisés (R-NL)	148
5.4	D'autres coopérations des modèles variationnels et non locaux	152
5.5	R-NL pour la famille exponentielle	155
5.5.1	Cas du bruit gaussien	156
5.5.2	Cas du bruit de Poisson	158
5.5.3	Cas du bruit gamma	159
5.5.4	Cas d'un bruit hybride	160
5.6	Résultats et discussion	162
5.6.1	Sélection des paramètres	162
5.6.2	Débruitage d'images	162
5.6.3	Discussion	163
5.7	Conclusion	168
6	R-NL pour le débruitage vidéo	171
6.1	Introduction	171
6.2	Adaptation 3D	171
6.3	Restauration de vidéos	173
6.4	Application au débruitage de vidéos BNL	176
6.4.1	Choix algorithmiques et sélection des paramètres	178
6.4.2	Compensation de mouvement	178
6.4.3	Illustrations sur données réelles	179
6.5	Conclusion	179
III	Recalage multimodal	183
7	Critère de recalage basé sur l'alignement des contours	187
7.1	Introduction	187
7.1.1	Contexte opérationnel et motivations	187
7.1.2	État de l'art	188
7.2	Définition du critère	193
7.2.1	Le modèle de transformation	194
7.3	Analyse du modèle proposé	194
7.3.1	Analyse théorique	194
7.3.2	Étude de la performance du critère proposé	202
7.4	Extension aux transformations projectives	206
7.4.1	Un peu de géométrie projective	206
7.4.2	Influence des paramètres	206
7.5	Conclusion	208

8	Schéma d'optimisation du critère de recalage	211
8.1	Introduction	211
8.2	Montée de gradient	211
8.2.1	Cas d'une transformation linéaire : zoom et translation	211
8.2.2	Cas projectif	214
8.3	Implémentation temporelle	215
8.4	Analyse du schéma proposé	217
8.4.1	Étude du schéma d'optimisation	217
8.4.2	Étude du schéma temporel	218
8.5	Résultats	220
8.5.1	Implémentation et temps de calcul	220
8.5.2	Validation expérimentale sur données réelles	221
8.5.3	Recalage sur données réelles	224
8.5.4	Résultats sur simulations projectives	224
8.6	Conclusion	227
9	Conclusion et perspectives	229
	Bibliographie	237

Notations et abréviations

Notations et abréviations associées à la vision de nuit

Le dispositif qui fait l'objet de cette thèse est un dispositif de vision bas niveau de lumière, noté BNL. Il est constitué d'un tube intensificateur de lumière, noté IL, couplé à un dispositif à transfert de charge appelé caméra CCD, pour *Charge Coupled Device*. Par abus de langage, on désigne par IL le dispositif numérique composé du tube et de la caméra, ainsi que les images issues de ce dispositif.

L'autre appareil de vision bas niveau de lumière qui constitue le dispositif est une caméra infrarouge. On désigne par IR le système d'acquisition des images infrarouge, ainsi que les images elles-mêmes.

Notations générales

On note $g : \Omega \rightarrow \mathbb{R}$ une image définie sur le domaine 2D discret $\Omega \subset \mathbb{Z}^2$, de sorte que g soit identifiée à un élément de \mathbb{R}^N avec $N = \#\Omega$ le nombre de pixels de l'image. Ainsi, pour une image $g \in \mathbb{R}^N$, $i \in [1, \dots, N]$ est l'indice du i^e pixel et $g_i \in \mathbb{R}$ désigne la valeur de g au pixel i .

g est supposée être l'observation d'une image « idéale » inconnue $f \in \mathbb{R}^N$ corrompue par un bruit dépendant du signal, stationnaire et décorréolé spatialement. Plus précisément, g peut être modélisée comme une réalisation d'un vecteur aléatoire G qui vérifie

$$\mathbb{E}[G] = f \quad \text{et} \quad \text{cov}[G] = \begin{pmatrix} \text{NLF}(f_1) & & & 0 \\ & \text{NLF}(f_2) & & \\ & & \ddots & \\ 0 & & & \text{NLF}(f_N) \end{pmatrix}. \quad (21)$$

où $\text{NLF} : \mathbb{R} \rightarrow \mathbb{R}^+$ est appelée la fonction de niveau de bruit, notée NLF pour *Noise Level Function*. Puisque $\text{cov}[G]$ est supposée ici diagonale, ce modèle est restreint à du bruit décorréolé spatialement. Mais comme le i^e élément diagonal dépend de f_i , le bruit est dit dépendant du signal. On remarque de plus que la même fonction de niveau de bruit est supposée s'appliquer à chaque élément de la diagonale, donc le bruit est supposé stationnaire. Les modèles de bruit à fonction de niveau de bruit variant spatialement ne sont pas considérés dans cette étude.

Lorsque le besoin s'y prête, par exemple lorsque l'on effectue des transformations de domaine, on assimile l'image g à une fonction continue $g : \Omega_c \rightarrow \mathbb{R}$, où Ω_c désigne le domaine continu de l'image. On désigne alors par $\mathbf{x} \in \Omega_c$ les coordonnées continues et $g(\mathbf{x})$ la valeur de la fonction g au point \mathbf{x} .

Dans le cadre des problèmes de débruitage, on note de façon unifiée g l'image observée, bruitée, et f l'image de référence sous-jacente. On désigne par u la solution du problème de débruitage.

Notations statistiques

Comme pour les images, une lettre minuscule g désigne une réalisation de la variable aléatoire G notée en majuscule. La notation $G|G' = g'$ désigne la variable aléatoire G conditionnellement à l'événement $G' = g'$, et $G|f$ correspond à la paramétrisation de la variable aléatoire G par f . On note p la probabilité discrète ou la densité de probabilité, \mathbb{E} l'espérance et Var la variance. On se consacre en particulier à certaines distributions.

Loi normale. On parle de loi normale ou loi gaussienne, associée en imagerie au bruit gaussien. On désigne par $\mathcal{N}(f, \sigma^2)$ la loi normale de moyenne f et de variance σ^2 . En général, σ^2 désigne la variance d'une distribution (quand elle existe).

Loi de Poisson. On note $\mathcal{P}(f)$ la distribution de Poisson de moyenne et de variance égales à f .

Loi gamma. La loi gamma de moyenne f et de paramètre de forme L est notée $\mathcal{G}(f, L)$. L est également appelé nombre de vues, et la loi gamma est parfois paramétrée selon $\gamma = \frac{1}{L}$.

Notations associées aux patches et aux blocs

On note ρ_i un patch d'indice i , c'est-à-dire un voisinage du pixel d'indice i , en général une fenêtre rectangulaire à deux ou trois dimensions. Sa taille, c'est-à-dire le nombre de pixels qu'il contient est désignée par $|\rho|$. Le patch extrait de l'image g en position i est noté g^{ρ_i} .

W_i désigne une fenêtre de recherche centrée autour du pixel i , en général une fenêtre rectangulaire à deux ou trois dimensions. Sa taille, c'est-à-dire le nombre de pixels qu'elle contient est désignée par $|W|$.

Lorsque l'on considère des blocs disjoints extraits d'une image, on note $(\omega_k)_{k \in [1, \dots, K]} \subset \Omega$ les K blocs de taille N_b de sorte que $KN_b \leq N$. On note ω un tel bloc quelconque ω_k et $g^\omega \in \mathbb{R}^N$ (resp. G^ω et f^ω) la restriction de g sur ω (resp. la restriction de G et f sur ω).

Notations associées aux méthodes variationnelles

∇ désigne l'opérateur de gradient. Pour une image $g \in \mathbb{R}^N$, $\nabla g = (\nabla g^{(1)}, \nabla g^{(2)})$ désigne le gradient discret de l'image g , défini selon [Chambolle 2004]. Il s'agit d'une matrice de taille $\mathbb{R}^{N,2}$, où la seconde dimension correspond aux gradients verticaux et horizontaux. L'opérateur de divergence est noté div .

La variation totale de l'image g est donnée par $\text{TV}(g) = \|\nabla g\|_{2,1} = \sum_{i \in \Omega} \|(\nabla g)_i\|$, où l'opérateur $\|\cdot\|$ désigne ici la norme ℓ^2 calculée par $\|(\nabla g)_i\| = \sqrt{\left((\nabla g)_i^{(1)}\right)^2 + \left((\nabla g)_i^{(2)}\right)^2}$.

De façon générale, $\|\cdot\|_p$ désigne la norme ℓ^p , et lorsque l'on omet l'indice p elle désigne par défaut la norme ℓ^2 (ou norme euclidienne).

Notations associées à la vidéo

Une vidéo est une suite d'images, appelées trames ou *frames*, se succédant dans le temps. Une séquence d'images peut être considérée comme une image en trois dimensions $g : \Omega \times \Omega_t \rightarrow \mathbb{R}$ définie sur un domaine 3D $\Omega \times \Omega_t \subset \mathbb{Z}^2 \times \mathbb{Z}$. On note N le nombre de pixels du domaine spatial $N = \#\Omega$ et T le nombre de trames du bloc $T = \#\Omega_t$. Les patches et blocs peuvent alors être définis comme des objets en trois dimensions.

On peut de même définir un gradient et une variation totale spatio-temporels. Le gradient $\nabla g = (\nabla g^{(1)}, \nabla g^{(2)}, \nabla g^{(3)})$ est une matrice de taille $\mathbb{R}^{N,3}$, où la seconde dimension correspond au gradient selon la direction verticale, horizontale et temporelle. La variation totale spatio-temporelle se définit alors comme $\text{TV}(g) = \|\nabla g\|_{2,1} = \sum_{i \in \Omega} \|(\nabla g)_i\|$, où l'opérateur $\|\cdot\|$ désigne ici la norme ℓ^2 calculée par $\|(\nabla g)_i\| = \sqrt{\left((\nabla g)_i^{(1)}\right)^2 + \left((\nabla g)_i^{(2)}\right)^2 + \left((\nabla g)_i^{(3)}\right)^2}$.

Introduction

1.1 La vision de nuit

La vision est liée à la perception par l'œil de rayonnements lumineux, qui sont ensuite interprétés par le système cognitif. L'homme possède une capacité naturelle à voir dans le noir, qui repose sur les cellules en bâtonnets présentes dans la rétine. Ces dernières sont plus sensibles à la lumière, et sont responsables également de la vision périphérique et de la détection de mouvement. En revanche, elles ne perçoivent pas la couleur et sont sensibles à l'éblouissement.

L'éclairement peut être quantifié par la quantité de flux lumineux reçu par unité de surface. Il est exprimé en lux, qui correspond à l'éclairement d'une surface recevant un flux uniforme d'un lumen par mètre carré, et peut être mesuré à l'aide d'un luxmètre. Une classification des niveaux de nuit en fonction de l'éclairement ainsi que les conditions associées est fournie dans le tableau 1.1.

L'œil humain peut s'accommoder à des niveaux d'éclairement très variables, de 130 000 lux (une journée ensoleillée d'été) à 1 lux (une nuit de pleine Lune). Certains animaux nocturnes ou semi-nocturnes comme la chouette ou le lynx ont une sensibilité bien plus grande à la lumière grâce à un nombre plus important de bâtonnets. D'autres animaux comme le serpent sont dotés d'une vision infrarouge permettant la détection de corps chauds la nuit. La capacité humaine à voir de nuit est plus limitée, c'est pourquoi la vision de nuit est artificiellement assistée.

On distingue pour cela trois principaux types de dispositifs. Les caméras dites thermiques sont sensibles au rayonnement thermique (dans le domaine infrarouge) et reflètent la température de la scène. Les caméras proche infrarouge, elles, ne détectent pas les rayons infrarouge mais utilisent des rayons proche infrarouge non visibles à l'œil nu pour éclairer la scène. Enfin, les intensificateurs de lumière fonctionnent dans le domaine visible et augmentent artificiellement la luminosité en multipliant le nombre de photons.

Dans le cadre de cette étude, nous considérons un dispositif numérique de vision bas niveau de lumière (BNL) constitué d'un intensificateur de lumière (IL) d'une part et d'une caméra infrarouge d'autre part, voués à améliorer la vision de nuit pour les pilotes d'hélicoptère. En collaboration avec Thales Avionics, l'objectif est de développer un système entièrement numérique qui fournirait des images de qualité afin de permettre leur exploitation pour des applications ultérieures de haut niveau comme de la détection. Cette thèse se consacre en particulier à l'amélioration des images issues de l'intensificateur de lumière, puis au recalage des images issues des deux modalités optique et infrarouge en vue d'une fusion ultérieure.

TABLE 1.1 – Classification des niveaux de nuit.

Niveaux de nuit	Éclairement (mlux)	Conditions de ciel associées
1	40 à 1000	Pleine lune ou gibbeuse
2	10 à 40	Pleine lune et nuages
3	2 à 10	Quartier ou croissant de lune
4	0.7 à 2	Illumination stellaire sans lune, ciel dégagé
5	< 0.7	Nuit nuageuse sans lune

1.1.1 Les images IL

Les images optiques sont obtenues grâce à un intensificateur de lumière, dont le rôle est de multiplier les photons afin d’augmenter artificiellement la luminosité. Le dispositif est fixé sur le casque du pilote, et les images sont projetées directement sur la visière du casque. Jusqu’à lors le système était optique, et nécessitait la mise en place d’un système de projection coûteux et encombrant. Afin d’alléger le dispositif, mais aussi de permettre l’enregistrement des missions et de nombreuses applications de traitement d’images, le système est désormais couplé à une caméra CCD (*Charge Coupled Device*) pour obtenir un dispositif entièrement numérique. Ce couplage s’accompagne malheureusement de contraintes et de dégradations. En effet, le dispositif de raccord entre les deux systèmes introduit des perturbations qui viennent s’ajouter aux dégradations associées au tube IL d’une part et à la caméra CCD d’autre part.

L’objectif de cette thèse est d’étudier la faisabilité de ce système de vision nocturne numérique. Cela repose dans un premier temps sur l’étude et l’amélioration des images IL issues du couplage IL-CCD.

Souvent représentées en vert ou en niveau de gris, les images IL sont faciles d’interprétation car elles reflètent les propriétés du domaine visible de la scène, naturelles pour l’homme. Cependant, elles souffrent de nombreuses dégradations qui peuvent rendre leur exploitation difficile. Caractéristique inhérente à la vision nocturne, les images IL présentent une très grande dynamique, associée à un contraste très faible. Cela se traduit par une différence d’éclairement entre les zones sombres et les zones claires, observée sur la figure 1.1. Sur l’image 1.1-a, le gain de l’intensificateur de lumière, c’est-à-dire le taux de multiplication des photons et donc l’augmentation de la luminosité, est faible, donc la plupart de la scène est noyée dans l’obscurité. Le réglage du gain de l’intensificateur permet d’adapter l’augmentation de la luminosité en fonction de l’éclairement de la scène. Lorsque ce dernier est uniforme, par exemple en l’absence de source lumineuse, cela permet de rehausser globalement l’information. En revanche les différences de dynamique au sein d’une même scène sont difficiles à équilibrer. Sur l’image 1.1-b par exemple, le gain est plus fort pour éclairer les zones sombres, mais cela sature les sources lumineuses et masque une partie de l’information. On observe ce phénomène de saturation des sources lumineuses sur

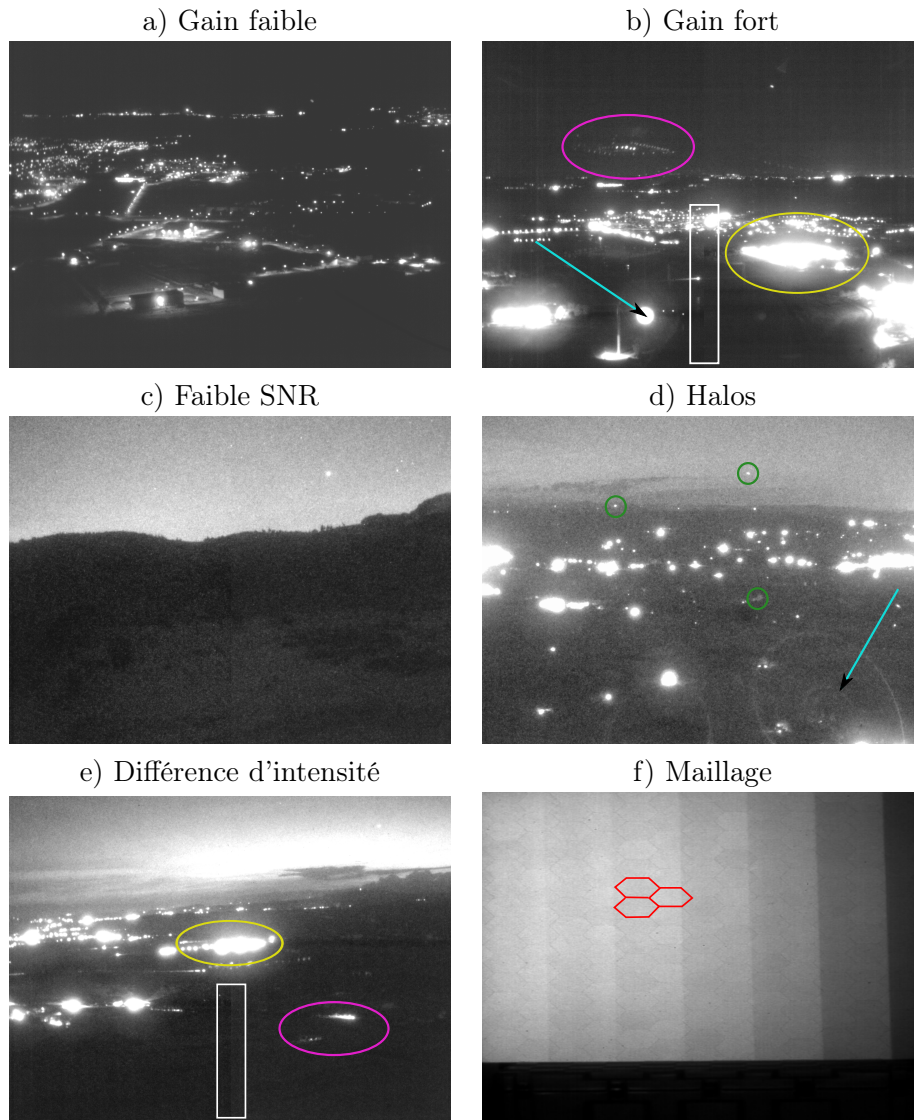


FIGURE 1.1 – Illustration des défauts des images BNL. Les images a) et b) présentent une très forte dynamique en raison des différences d'illuminations entre les sources lumineuses et le reste de la scène. Le faible rapport signal à bruit lié à la présence de bruit est illustré sur les figures c) et d). Cette dernière présente également des taches (cercles verts). La présence de reflets et la saturation des sources lumineuses sont soulignées sur les figures b) et e) par des ellipses respectivement magenta et jaune, tandis que des halos (flèches bleues) sont observés sur les figures b) et d). On observe également sur les figures b) et e) la différence d'intensité entre les parties gauche et droite de l'image, manifestée par les rectangles blancs. Enfin, l'image d'une mire après suppression du bruit sur la figure f) laisse apparaître la présence d'un maillage, surligné en rouge, lié à la constitution du système d'acquisition. L'illustration des défauts ci-dessus est optimisée pour un affichage numérique en couleur, éventuellement zoomé.

les figures 1.1-d et 1.1-e également. Les images 1.1-c et 1.1-d souffrent d'un faible contraste qui cumulé à la présence de bruit résulte en un rapport signal à bruit très faible. Ce type d'images nécessite donc un rééquilibrage de la dynamique de la scène, qui peut passer par des corrections globales ou locales d'histogrammes ou des traitements fréquentiels. Il faut envisager également une correction spécifique des défauts comme les halos ou les taches, observables sur la figure 1.1-d, ou la présence de reflets, visibles sur les figures 1.1-b et 1.1-e. Ces défauts sont plus particulièrement observables sur les séquences vidéos, car leur déplacement ne correspond pas au mouvement global de la scène. La correction de ces artefacts nécessite une étape de détection puis de suppression. D'autres défauts systématiques liés au dispositif d'acquisition caractérisent ces images IL. On peut observer sur les figures 1.1-b et 1.1-e une différence de luminosité entre les parties gauche et droite de l'image, et sur la figure 1.1-f la présence d'un maillage sur toute l'image.

Par ailleurs, le système d'acquisition introduit toujours une part de bruit sur les images. Dans le cas des images IL, en fonction de la luminosité et du gain du tube intensificateur, le bruit peut s'avérer élevé, et dépendant du signal. La figure 1.2 présente deux exemples d'images fortement bruitées. Ce phénomène est gênant pour l'analyse des images en général, et sa correction suscite beaucoup d'intérêt en traitement d'images. Sur les séquences d'images, il crée de plus un effet de scintillement qui peut gêner l'interprétation de la scène. Combiné à un faible contraste, cela engendre un rapport signal à bruit très faible, où les informations importantes de la scène sont masquées par le bruit et l'obscurité, comme c'est illustré sur la figure 1.1-c. De plus, la présence de bruit dans les images peut nuire à l'exploitation de celles-ci pour des applications de haut niveau comme de la segmentation automatique, de la détection ou encore de la reconnaissance d'objets. Il est donc primordial de restaurer ces images pour fournir une image plus facile à la fois à interpréter et à exploiter.

Cependant, pour pouvoir traiter le bruit de façon adaptée, il est nécessaire de connaître ses caractéristiques. En traitement d'images, on suppose souvent que la statistique du bruit est connue et on développe une méthode de débruitage en fonction du modèle. Sur les images BNL en revanche, la nature du bruit est a priori inconnue, et une analyse du système d'acquisition révèle la complexité du modèle, ce qui rend inexploitable son application pratique. Il est donc nécessaire d'estimer directement les caractéristiques du bruit à partir des images à traiter, pour pouvoir ensuite développer des traitements adaptés.

1.1.2 Combinaison aux images infrarouge

Second élément du dispositif de vision nocturne, les images infrarouge reflètent la température de la scène. Elles sont plus difficiles à interpréter car elles ne reflètent pas les caractéristiques associées au domaine visible, mais elles fournissent des informations précieuses comme la présence de bâtiments qui sont des sources chaudes par rapport au reste du paysage. Cependant, elles souffrent d'une plus faible résolution et d'un contraste assez faible, ce qui empêche leur seule utilisation. La caméra



FIGURE 1.2 – Illustration d’images IL fortement bruitées.

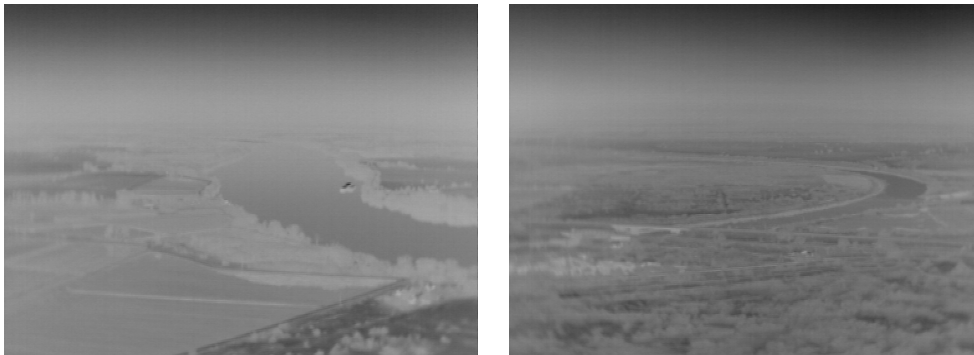


FIGURE 1.3 – Illustration d’images IR correspondant aux scènes de la figure ci-dessus.

infrarouge est située sous l’hélicoptère, et peut être dirigée par le pilote. Des images infrarouge, correspondant aux scènes observées sur la figure 1.2, sont représentées sur la figure 1.3.

La fusion des deux modalités optique et infrarouge permettrait de bénéficier conjointement des informations complémentaires fournies par chacune des deux modalités. L’image infrarouge permet par exemple de bien distinguer les routes, donc l’ajout de cette information sur les images optiques permettrait de les renforcer en s’affranchissant d’un rehaussement de contours. Cependant, ces deux caméras ne sont pas situées au même emplacement, donc elles observent la scène sous des angles différents, et elles se déplacent indépendamment l’une de l’autre. Une étape de recalage des deux modalités, qui consiste à aligner les images dans le même référentiel, est donc nécessaire avant de pouvoir envisager la fusion des capteurs.

1.1.3 État de l’art des traitements développés à Thales

Le système numérique de vision nocturne a fait l’objet de plusieurs études au sein de Thales Avionics, qui ont permis de mettre en avant le potentiel des traitements offerts par le dispositif mais aussi ses principales limitations.

Une caractérisation empirique du système en salle noire calibrée a permis d’ap-

préhender le comportement du tube IL et de la caméra CCD. La fonction de transfert en modulation caractérise le contraste et la résolution d'un système optique et permet d'évaluer la qualité du dispositif. Le bruit de référence des capteurs CCD et IL peut être évalué par des mesures en salle noire avec un cache sur l'objectif. Cela fournit une image moyenne d'obscurité qui, soustraite au moment de l'acquisition, permet de ramener le niveau de gris du noir à 0 et de supprimer les pixels défauts du capteur. Ce bruit de référence est caractérisé en fonction du temps d'intégration et du gain du tube IL. Néanmoins, mise à part la caractérisation du bruit systématique du capteur, aucune étude des composantes du bruit dégradant les images et permettant d'émettre des hypothèses sur la distribution statistique du bruit en fonction des conditions d'acquisition n'a été portée à notre connaissance.

L'absence d'une telle étude est justifiée entre autres par le fait que les méthodes de débruitage des images IL employées jusqu'à lors à Thales se sont appuyées sur un débruitage temporel. Si celui-ci est optimal sur des images statiques, lors de l'observation de scènes dynamiques une compensation de mouvement doit être effectuée. Des études au sein de Thales ont mis en place un algorithme de filtrage temporel associé au recalage des trames par «block matching» [Lu 1997]. Cette méthode peut s'avérer imparfaite, en raison notamment du faible contraste dont souffrent les images IL, et des mouvements non uniformes sur l'image (présence d'une cible en mouvement par exemple). Pour remédier à cela, un système de détection des mouvements du pilote a été mis en place. Ce dernier repose sur le mécanisme électromagnétique de «détection de posture» présent dans le cockpit qui permet d'évaluer le déplacement de la tête du pilote et d'en déduire la trajectoire des pixels d'une trame à la suivante. Le recalage précis des trames par block matching reste néanmoins une étape coûteuse du processus de traitement. L'affranchissement de la compensation de mouvement pourrait améliorer le débruitage, en termes de complexité mais aussi de performance.

D'autres traitements correctifs ont été développés. Des méthodes de rehaussement de dynamique et de contraste ont été étudiées [FitzHenry 1991, Dippel 2002], ainsi que la suppression des défauts locaux comme les halos ou les reflets [Fainstain 2010].

Une méthode de rehaussement envisagée également repose sur la fusion des capteurs IL et IR [Hall 1997]. Cela permettrait en effet de combiner les informations de nature différente issues des deux modalités. Des algorithmes de fusion ont été développés au sein de Thales, pourvu que les images à fusionner soient correctement alignées. Le recalage des deux modalités est donc à envisager.

1.2 Synthèse des travaux et contributions

L'objectif de cette thèse est d'étudier en partenariat avec Thales Avionics la faisabilité du dispositif de vision nocturne entièrement numérique destiné à améliorer la vision de nuit des pilotes d'hélicoptère, et de proposer des solutions pour le perfectionner. Le dispositif bas niveau de lumière est constitué d'un intensificateur de

lumière couplé à une caméra CCD d'une part et d'une caméra infrarouge d'autre part. Nous nous consacrons en particulier à l'amélioration des images optiques issues de l'intensificateur de lumière. Cela repose sur la réduction automatique du bruit présent dans les images IL et sur le recalage des images issues des capteurs optiques et infrarouge en vue d'une fusion ultérieure, destinée à combiner les informations issues des deux modalités.

Estimation automatique du bruit. L'étude repose dans un premier temps sur une étude du dispositif d'acquisition, afin de caractériser les dégradations dont souffrent les images IL, et notamment la nature du bruit présent en fonction des conditions et de la luminosité. Une modélisation des différents type de bruit rencontrés en imagerie ainsi que le descriptif et l'analyse du système d'acquisition des images IL sont présentés dans le chapitre 2. Cependant, la complexité du processus d'acquisition ne permet pas d'établir au préalable la nature précise du bruit, ce qui rend le modèle peu exploitable pour une application au débruitage par exemple. Il est donc souhaitable de pouvoir effectuer une estimation automatique du bruit présent dans les images. C'est l'objet du chapitre 3. L'estimation automatique du bruit à partir de l'image elle-même repose sur l'analyse des statistiques de ce dernier sur des zones homogènes de l'image. En effet, sur des zones dans lesquelles le rapport signal à bruit est faible, les fluctuations liées au signal sont négligeables par rapport à celles engendrées par le bruit, donc on peut considérer que seules les statistiques liées au bruit interviennent. La détection des régions homogènes constitue donc la première étape du processus d'estimation. Cependant, la caractérisation de ces zones à faible rapport signal à bruit ne peut s'appuyer sur des hypothèses sur le bruit, que l'on cherche à estimer. C'est pourquoi nous développons un test non paramétrique dont les performances de détection sont indépendantes de la distribution du bruit, et nous montrons que le taux d'erreur est lié au rapport signal à bruit ainsi qu'au nombre d'échantillons disponibles. Ce test est basé sur le coefficient de corrélation de rang de Kendall [Kendall 1938], et il mesure la corrélation entre deux séquences de variables aléatoires, en se basant sur l'ordre relatif des valeurs et non sur les valeurs elles-mêmes. Une fois les zones homogènes détectées, la fonction de niveau de bruit, c'est-à-dire la relation qui lie l'intensité de l'image à la variance du bruit, est estimée sous forme d'un polynôme d'ordre deux à l'aide de l'estimateur robuste des moindres déviations (LAD, pour *Least Absolute Deviation*) basé sur la minimisation de l'erreur d'approximation ℓ^1 . La méthode ainsi développée permet d'estimer de façon robuste des bruits stationnaires non corrélés dépendants du signal, que nous pouvons ensuite appliquer aux images BNL.

Ces travaux ont fait l'objet de la soumission d'un article au *Siam Journal on Imaging Sciences* [Sutour 2015b], ainsi qu'à la conférence internationale *ICIP'2015* et au colloque *Gretsi'2015*.

Régularisation adaptative des moyennes non locales. Une fois la nature du bruit estimée, la deuxième partie de cette thèse repose sur le débruitage des

images IL. Pour cela, nous nous focalisons sur deux méthodes de débruitage classiques, le modèle de ROF [Rudin 1992] basé sur la minimisation de la variation totale ainsi que les moyennes non locales (NL-means) [Buades 2005b]. Le choix s'est basé notamment sur la simplicité des méthodes, leurs extensions possibles à des types de bruit non gaussiens [Aubert 2008, Anthoine 2012, Deledalle 2010, Delon 2012, Deledalle 2014], ainsi que l'adaptation des NL-means à la restauration de séquences d'images qui ne nécessite pas d'effectuer de compensation de mouvement entre les trames [Buades 2005a]. Le chapitre 4 fournit un état de l'art des méthodes de débruitage et une description détaillée des deux techniques impliquées. Nous analysons en particulier les forces et les faiblesses des deux méthodes, et nous identifions les défauts majeurs associés à chacune : la minimisation TV souffre de l'effet de crénelage lié à l'approximation des zones lisses par des surfaces constantes par morceaux. Elle ne préserve pas bien les structures fines et les textures et souffre d'une perte de contraste globale. Les moyennes non locales souffrent de deux défauts qui s'opposent : l'effet de patch rare et l'effet de gigue, ou *jittering*. L'effet de patch rare est lié à la difficulté de trouver suffisamment de redondance sur les structures singulières pour effectuer un débruitage efficace, et résulte en la présence de bruit résiduel. Au contraire, l'effet de *jittering* résulte d'un débruitage non pertinent qui se manifeste par un effet de flou et des artefacts. Il se produit lorsque que des pixels issus de structures différentes sont confondus, en raison de la présence de bruit ou d'un faible contraste par exemple. Le chapitre 5 propose donc une méthode de débruitage adaptative qui combine la minimisation TV et les moyennes non locales afin de corriger leurs défauts respectifs. La première étape consiste à réduire l'effet de *jittering* produit par les NL-means, en évaluant la qualité de la sélection de candidats, afin de réinjecter de l'information bruitée quand le débruitage est jugé non pertinent. La seconde étape réduit l'effet de patch rare en réalisant une régularisation TV adaptative. Celle-ci se base sur une attache aux données non locale qui s'appuie sur l'évaluation de la réduction de bruit effectuée en amont par les moyennes non locales. Grâce aux bonnes propriétés de la minimisation TV et des NL-means, l'algorithme s'étend à de nombreux types de bruit, et on propose en particulier un algorithme de débruitage aveugle qui repose sur l'estimation du bruit réalisée dans la première partie. Enfin, la méthode est étendue au débruitage de séquences d'images dans le chapitre 6. Pour cela, les moyennes non locales utilisent une fenêtre de recherche spatio-temporelle qui permet de bénéficier de la redondance d'information fournie par les trames environnantes. Une telle implémentation permet de s'affranchir de la compensation de mouvement entre les trames, mais elle entraîne une instabilité temporelle qui se manifeste par un effet de scintillement. Pour remédier à cela, nous suggérons d'utiliser également des patchs spatio-temporels, afin d'assurer une meilleure stabilité temporelle et donc plus de confort visuel sur la vidéo restaurée. L'utilisation de patchs tridimensionnels s'accompagne d'un effet de patch rare accru, qui est ensuite corrigé par une régularisation TV adaptative spatio-temporelle. La méthode permet ainsi d'effectuer un débruitage adaptatif pertinent, qui peut être appliqué au débruitage des séquences d'images IL.

Ces travaux ont été publiés dans le journal *IEEE Transactions on Image Pro-*

cessing [Sutour 2014b], et ont été présentés au colloque *Gretsi'2013* [Sutour 2013] ainsi qu'à la conférence *ICIP'2014* [Sutour 2014a].

Recalage multimodal. Enfin, la troisième partie de cette thèse réalise le recalage multimodal des images optiques sur les images infrarouge, en vue d'une ultérieure fusion de capteurs. Les deux modalités impliquées présentent des caractéristiques particulières qui ne permettent pas l'appariement de descripteurs ni l'utilisation de critères statistiques, en raison des différences de distribution des intensités et des textures. Nous proposons donc dans le chapitre 7 un critère de recalage adapté à la multimodalité du problème, en se basant sur une information commune aux deux modalités : les contours [Sun 2004]. Nous définissons pour cela une métrique qui réalise l'alignement des contours présents dans les deux modalités. Ensuite, le chapitre 8 propose une optimisation basée sur la mise en place d'une montée de gradient qui maximise la métrique d'alignement, couplée à un schéma de validation temporel qui assure que la méthode est rapide et robuste. Cette optimisation et les résultats sur données réelles permettent d'envisager l'implémentation d'un tel dispositif sur des systèmes embarqués.

Les résultats de ces travaux ont été publiés dans le *Journal of Mathematical Imaging and Vision* [Sutour 2015a].

Première partie
Analyse du bruit

Table des matières

2	Modélisation du bruit pour les systèmes BNL	49
2.1	Introduction	49
2.2	Le bruit en imagerie	50
2.3	Acquisition des images BNL	55
2.4	Conclusion	60
3	Estimation non paramétrique du bruit	63
3.1	Introduction	63
3.2	Détection des zones homogènes	66
3.3	Estimation du modèle	76
3.4	Résultats sur images synthétiques	84
3.5	Application aux images BNL	87
3.6	Conclusion	91
3.7	Annexes du chapitre : preuves	92

Modélisation du bruit pour les systèmes BNL

2.1 Introduction

L'amélioration de la vision de nuit des pilotes d'hélicoptère repose sur la restauration des images BNL, en particulier des images issues du dispositif intensificateur de lumière. Cela nécessite donc de caractériser les dégradations dont souffrent les images IL, pour pouvoir ensuite les corriger de façon adaptée. On distingue des défauts locaux comme les halos autour des sources lumineuses et les reflets, et des défauts globaux qui perturbent l'interprétation de l'image dans sa globalité, par exemple le faible contraste et la présence de bruit. Des algorithmes performants de rehaussement de contraste basés par exemple sur l'équilibrage d'histogramme existent [Zimmerman 1988, Stark 2000], en revanche une telle opération modifie la distribution statistique des intensités des images, et donc du bruit. C'est pourquoi l'étape de suppression ou réduction du bruit, appelée débruitage, doit être effectuée au préalable. Dans [Genin 2012], l'estimation de bruit est réalisée conjointement à l'étape de soustraction de fond en vue d'effectuer de la détection de cibles. De façon générale en revanche, le débruitage constitue souvent la première étape de traitement d'une image, car la présence de bruit vient perturber les caractéristiques de l'image et donc empêcher la réalisation d'autres tâches comme de la segmentation ou de l'inpainting. Pour pouvoir traiter le bruit de façon adéquate, il est nécessaire de lui associer un modèle ou une distribution statistique. De nombreux modèles mathématiques ont été appliqués en imagerie pour décrire les perturbations fréquemment rencontrées. Certains types de bruit comme le bruit gaussien, le bruit de Poisson, ou le bruit de *speckle* font écho à des phénomènes physiques issus des systèmes d'acquisition, tandis que d'autres modèles comme le bruit impulsif font plutôt figure de cas d'école. La section 2.2 fournit un descriptif des modèles de bruit rencontrés dans les problèmes d'imagerie, académiques ou appliqués.

Cependant, si certains types de bruit sont associés à un phénomène physique ou à un dispositif d'acquisition, souvent en pratique la chaîne d'acquisition des images est un processus complexe qui fait intervenir de nombreuses étapes et donc possiblement de nombreux types de dégradations. Il est donc nécessaire d'analyser le dispositif d'acquisition pour tenter de dégager les principales sources de dégradations et en particulier les principales composantes de bruit. Dans la section 2.3, les différentes composantes du couplage du tube intensificateur de lumière et de la caméra CCD sont analysées, afin de dégager les principaux éléments et proposer un modèle pour

le bruit, en fonction de la luminosité et des conditions d'acquisition.

2.2 Le bruit en imagerie

Le processus d'acquisition d'une image consiste en de nombreuses étapes qui introduisent pour chacune d'entre elles des dégradations de nature différente. On note par exemple les distorsions géométriques liées à la projection d'une scène en trois dimensions sur un plan, un effet de flou dû au mouvement des objets de la scène et lié au temps d'exposition, des artefacts d'échantillonnage et de compression, etc.

On s'intéresse essentiellement ici à la corruption associée au bruit dans une image, lié lui-même à plusieurs phénomènes au cours de l'acquisition. Les différentes composantes de bruit dans les images BNL seront analysées à l'aide de l'étude de la chaîne d'acquisition dans la section suivante. Nous nous intéressons dans un premier temps à la modélisation mathématique des bruits fréquemment rencontrés en imagerie, ou utilisés dans des problématiques liées au traitement d'images.

2.2.1 Notations

Soit $g : \Omega \rightarrow \mathbb{R}$ une image définie sur le domaine discret Ω , et $N = \#\Omega$ le nombre de pixels de l'image. g est supposée être l'observation d'une image inconnue $f \in \mathbb{R}^N$ corrompue par un bruit stationnaire et décorréolé spatialement. Plus précisément, g peut être modélisée comme une réalisation d'un vecteur aléatoire G , et si le bruit admet une densité ou une fonction de masse, on notera $p_G(g|f)$ ou plus simplement $p(g|f)$ la densité de l'observation g sachant l'image de référence sous-jacente f , appelée également la vraisemblance. La variable aléatoire G vérifie

$$\mathbb{E}[G] = f \quad \text{et} \quad \text{cov}[G] = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_N^2 \end{pmatrix}. \quad (2.1)$$

Ce modèle est de plus restreint à du bruit spatialement indépendant, ce qui implique que $\text{cov}[G]$ est supposée diagonale, et la densité globale de l'image g est égale au produit des densités en chaque pixel :

$$p(g|f) = \prod_{i=1}^N p(g_i|f_i). \quad (2.2)$$

2.2.2 Le bruit blanc additif gaussien

Le modèle le plus utilisé en imagerie est celui de bruit blanc additif gaussien, souvent appelé plus simplement bruit gaussien. g est alors une réalisation de la variable aléatoire

$$G = f + \sigma\varepsilon, \quad (2.3)$$

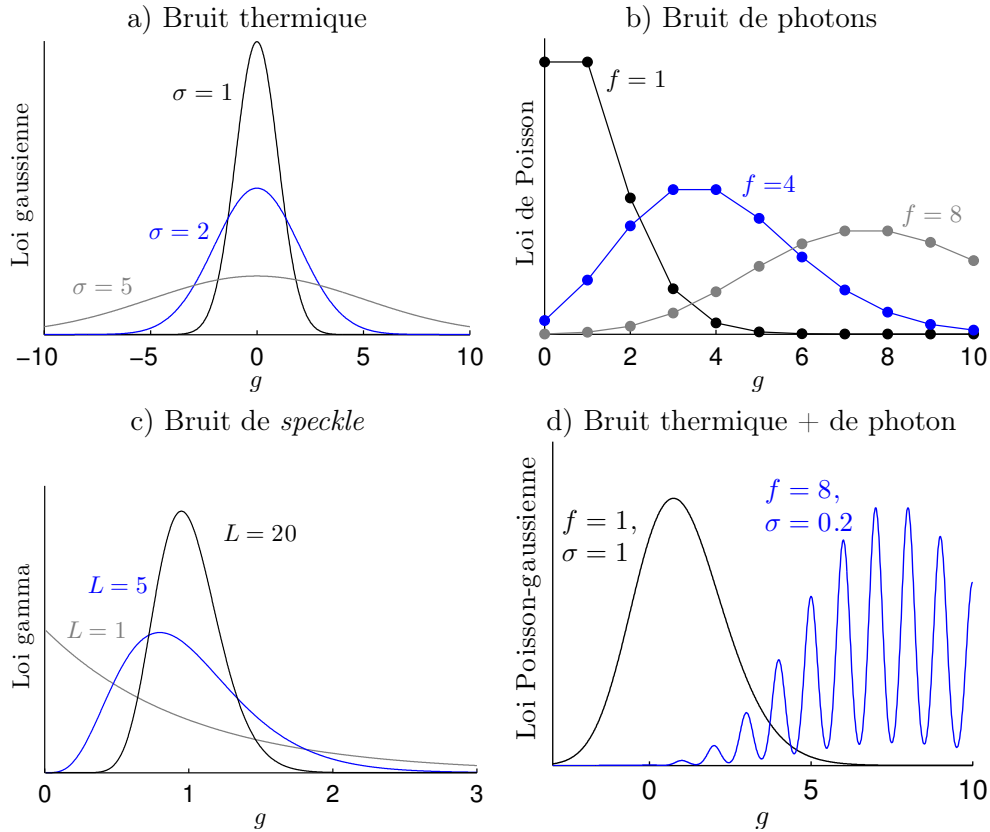


FIGURE 2.1 – Distributions de bruits fréquemment rencontrés en imagerie. a) Modèle de bruit thermique représenté par la loi normale de moyenne nulle, b) Modèle de bruit de photon représenté par la loi de Poisson, c) Modèle de bruit de *speckle* représenté par la loi gamma et d) Modèle de bruit thermique et de photon représenté par la loi Poisson-gaussienne.

où $\sigma \in \mathbb{R}$ et ε suit une loi normale centrée réduite notée $\mathcal{N}(0, 1)$. Dans ce cas, la variable aléatoire G suit une loi normale de moyenne f et de variance σ^2 notée $\mathcal{N}(f, \sigma^2)$, et sa densité est donnée par

$$p(g|f) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(g-f)^2}{2\sigma^2} \right]. \quad (2.4)$$

G vérifie alors l'équation (2.1) avec $\sigma_i^2 = \sigma^2$ pour tout i , ce qui montre que le bruit est indépendant du signal, ou homoscédastique. Comme illustré sur la figure 2.1-a, la loi normale est symétrique, centrée autour de sa moyenne et le paramètre d'écart type σ caractérise l'étalement de la gaussienne autour de sa moyenne. Le paramètre σ^2 de la distribution correspond donc à la variance du bruit ; plus σ^2 est élevé plus le bruit est important. En imagerie, le bruit gaussien est associé au bruit dit thermique.

2.2.3 Le bruit de Poisson

La loi de Poisson est un processus aléatoire discret beaucoup utilisé dans le comptage d'événements rares, ou sur un laps de temps donné. En imagerie, elle est souvent associée au bruit de photons (*shot noise* ou *photon count noise*) et se manifeste lorsque le nombre de photons (donc la luminosité) est faible, en astronomie, en imagerie médicale ou encore en vision de nuit par exemple.

Pour $f \in \mathbb{R}^+$, la variable aléatoire G à valeurs dans \mathbb{N} suit une loi de Poisson, $G \sim \mathcal{P}(f)$, dont la fonction de masse est :

$$p(g|f) = \frac{f^g e^{-f}}{g!}, \quad (2.5)$$

et on a $\mathbb{E}[G] = f$ et $\text{Var}[G] = \mathbb{E}[G] = f$. Dans ce cas, G vérifie l'équation (2.1) avec $\sigma_i = f_i$, ce qui montre que la variance du bruit varie localement en fonction de la moyenne ; on dit que la loi de Poisson est hétéroscédastique, et on parle en imagerie de bruit dépendant du signal. Des exemples de distributions de Poisson en fonction de la moyenne f sont représentés sur la figure 2.1-b.

En imagerie, on modélise la force du bruit de Poisson par un entier Q , de sorte que la variable aléatoire à valeurs entières $\frac{G}{Q} \sim \mathcal{P}\left(\frac{f}{Q}\right)$. On a alors $\mathbb{E}[G] = Q \times \frac{f}{Q} = f$ et $\text{Var}[G] = Q^2 \times \frac{f}{Q} = Qf$, auquel cas G vérifie l'équation (2.1) avec $\sigma_i = Qf_i$. Plus Q est grand, plus la dégradation est importante.

2.2.4 Le bruit de speckle

En imagerie radar, sonar ou ultrason par exemple, le signal est acquis après rétrodiffusion de celui-ci sur la scène observée. Des interférences entre plusieurs échos se produisent au sein d'un même pixel, ce qui génère après sommation le bruit dit de *speckle*. Ce dernier se traduit par des fluctuations du signal dont l'amplitude dépend de l'intensité du signal sous-jacent, donc il est qualifié de bruit dépendant du signal. On modélise alors le *speckle* par la variable aléatoire G telle que $G_i = f_i \times S_i$, avec $\mathbb{E}[S] = 1$ et $\mathbb{E}[SS^\top] = \frac{1}{L} \text{Id}$, $L > 0$, et de sorte que sa densité de probabilité soit donnée par

$$p(g | f) = \frac{L^L g^{L-1}}{\Gamma(L) f^L} \exp\left(-\frac{Lg}{f}\right). \quad (2.6)$$

On a $\mathbb{E}[G] = f$ et $\text{Var}[G] = \frac{f^2}{L}$. Dans ce cas, G vérifie l'équation (2.1) avec $\sigma_i^2 = \frac{1}{L}(f_i)^2$. On remarque en outre que la loi gamma est hétéroscédastique. Le paramètre $L \in \mathbb{N}^*$ règle la force du bruit ; ce dernier est maximal lorsque $L = 1$. La loi gamma est parfois donnée en fonction du paramètre $\gamma = \frac{1}{L}$. La forme de la loi gamma est représentée sur la figure 2.1-c, où l'on observe notamment la queue lourde responsable des très fortes intensités susceptibles d'apparaître sur des images corrompues par un tel bruit.

2.2.5 Le bruit impulsif

Le bruit impulsif est un bruit de nature discrète dont la fonction de masse est donnée par

$$p(g|f) = \begin{cases} P/L + 1 - P & \text{si } g = f, \\ P/L & \text{sinon.} \end{cases} \quad (2.7)$$

Il permet de modéliser la présence de pixels aberrants; avec une telle distribution en effet, les pixels de l'image g ont soit la même valeur que celle de référence f , soit une valeur uniformément distribuée sur l'intervalle $[1..L]$ des intensités de l'image. Le paramètre $P \in [0, 1]$ caractérise le taux de dégradation de l'image. Cela peut modéliser la corruption des pixels durant la transmission d'une image.

2.2.6 Le bruit poivre et sel

De façon assez analogue au bruit impulsif, le bruit poivre et sel est de nature discrète et sa fonction de masse est caractérisée par l'intervalle des intensités de l'image L et le taux de dégradation $P \in [0, 1]$ de la façon suivante :

$$p(g|f) = \begin{cases} P/2 + 1 - P & \text{si } g = f = 1 \text{ ou } g = f = L, \\ P/2 & \text{si } g = 1 \text{ et } f \neq 1 \text{ ou } g = L \text{ et } f \neq L, \\ (1 - P) & \text{si } g = f \text{ et } u \notin \{1, L\}, \\ 0 & \text{sinon.} \end{cases} \quad (2.8)$$

Il traduit la saturation des pixels à 0 ou à L , avec une probabilité P , et peut modéliser des taches sur des images par exemple.

2.2.7 Le bruit Poisson-gaussien

Enfin, un type de bruit souvent rencontré en imagerie est le bruit Poisson-gaussien, issu de la combinaison du bruit thermique et du bruit de photons. Il se traduit par la somme de deux variables aléatoires distribuées respectivement selon une loi normale et une loi de Poisson, et sa densité est donnée par la convolution d'une distribution de Poisson par une gaussienne :

$$p(g|f) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=0}^{+\infty} \exp\left[-\frac{(g-k)^2}{2\sigma^2}\right] \frac{f^k e^{-f}}{k!}, \quad (2.9)$$

où le paramètre $\sigma > 0$ correspond à la puissance du bruit thermique. Lorsque que la variable aléatoire G est distribuée selon un modèle Poisson-gaussien, on a $\mathbb{E}[G] = f$ et $\text{Var}[G] = f + \sigma^2$, ce qui caractérise un bruit dépendant du signal à variance affine. La figure 2.1-d représente deux distributions Poisson-gaussiennes pour des valeurs différentes de moyenne et de puissance du bruit thermique. On observe notamment que pour des niveaux de bruit thermique faible, la distribution est très instable.

De nombreux autres modèles de bruit existent et peuvent être rencontrés en imagerie; nous nous focaliserons sur ceux décrits ci-dessus, et en particulier sur

les trois grandes familles de bruits gaussien, gamma et poissonien qui sont à la fois étudiés d'un point de vue académique et rencontrés dans la modélisation de beaucoup de problèmes appliqués.

2.2.8 Retour au bruit gaussien : stabilisation de variance

La plupart des méthodes de traitement d'images, et en particulier le débruitage, sont avant tout adaptées au bruit gaussien, prenant en compte la distribution gaussienne du bruit et le caractère indépendant du signal. Si le bruit présente une distribution statistique différente, par exemple avec une queue de distribution plus lourde, ou devient dépendant du signal, une prise en compte non adaptée peut introduire un biais.

La méthode la plus simple pour adapter les méthodes de traitement d'images prévues pour du bruit gaussien à d'autres types de bruit consiste à transformer non pas la technique de traitement mais les données directement, pour que le bruit dans l'image transformée puisse être assimilé à du bruit gaussien à variance constante. Le principe est donc d'appliquer une transformation s de stabilisation de variance aux données g , effectuer le traitement conçu pour du bruit gaussien, puis revenir dans l'espace de départ en appliquant la transformation inverse s^{-1} . La transformation de stabilisation de variance doit naturellement être adaptée au type de bruit concerné.

Bruit gamma

Le bruit gamma étant de nature multiplicative, il est naturel d'appliquer une transformation logarithmique aux données pour faire apparaître un caractère additif. En effet si on définit

$$s(G) = \log(G), \text{ alors } \text{Var}[s(G)] = \text{Var}[\log(G)] = \Psi_1(L), \quad (2.10)$$

où $\Psi_1(L)$ est la fonction polygamma du premier ordre de degré L [Xie 2002b]. La transformation inverse est naturellement obtenue en passant à l'exponentielle. Cependant, l'approximation par du bruit gaussien n'est valable qu'asymptotiquement pour de grandes valeurs de L , tandis que pour de plus faibles valeurs la distribution reste asymétrique et à queue plus lourde.

Bruit de Poisson

La transformation d'Anscombe [Anscombe 1948] donnée par

$$s(G) = 2\sqrt{G + \frac{3}{8}} \quad (2.11)$$

permet dans le cas du bruit de Poisson d'assimiler le bruit à un bruit additif à variance égale à 1. Cette approximation n'est valable que pour des intensités supérieures à 4 environ, ce qui peut poser problème à faible nombre de photons.

Cependant lorsque s n'est pas linéaire, comme c'est le cas dans les deux exemples ci-dessus, l'étape d'inversion de cette transformation introduit un biais, c'est-à-dire que $\mathbb{E}[G] \neq s^{-1}(\mathbb{E}[s(G)])$. Une correction de ce biais a été proposée dans le cas du bruit gamma par [Xie 2002a], et pour du bruit de Poisson dans [Mäkitalo 2011].

Une stabilisation de variance a également été proposée dans le cas de bruit Poisson-gaussien dans [Murtagh 1995, Boulanger 2010]. Cette dernière nécessite cependant d'être ajustée en fonction des paramètres du bruit.

La stabilisation de variance est une approche simple qui permet d'adapter la multitude de traitements conçus pour du bruit gaussien à d'autres modèles de bruit. Cependant, l'approximation par du bruit gaussien reste imparfaite. De plus, la transformation s dépend du modèle de bruit impliqué, et elle n'est pas définie pour toutes les distributions non gaussiennes susceptibles d'être rencontrées.

Que l'on revienne dans le monde gaussien ou non, la connaissance du bruit reste primordiale pour l'application de traitements adaptés. Dans la problématique de vision nocturne, l'étude du processus d'acquisition des images IL permet de mettre en avant plusieurs types de dégradations.

2.3 Acquisition des images BNL

Le but de cette étude est de proposer un modèle de bruit pour les images issues des caméras IL [Williams Jr 1992, Johnson 1995]. Les étapes successives du dispositif d'acquisition sont présentées afin de déterminer les différentes composantes de bruit intervenant à chacun des niveaux. La liste obtenue est non exhaustive, elle se limite aux phénomènes jugés prépondérants et susceptibles d'influencer la qualité des images et d'être corrigés.

La caméra sur laquelle porte cette étude est un caméra IL-CCD composée d'un tube intensificateur de lumière (IL) couplé à une caméra CCD (*Charge Coupled Device*). Ce modèle ne fait pas partie des dernières générations de capteurs, mais il est la source des données réelles mises à disposition par Thales Avionics. Les données fournies sont pessimistes en terme de qualité, mais elles ne seront ainsi qu'améliorées par la suite.

2.3.1 Le tube intensificateur de lumière

Le principe du tube IL repose sur la multiplication des photons afin d'augmenter artificiellement la luminosité des images. Un tube IL est constitué d'une photocathode qui émet des électrons quand elle reçoit de la lumière (c'est-à-dire des photons). En présence d'un champ électrique, ces électrons sont accélérés voire multipliés, puis retransformés en photons (cette fois plus nombreux) pour former une image plus lumineuse.

On distingue plusieurs types (ou générations) de tubes IL. La première génération ne fait appel qu'à l'accélération des électrons, tandis que les tubes de deuxième génération et plus utilisent une galette de micro-canaux (MCP, pour *MicroChannel*

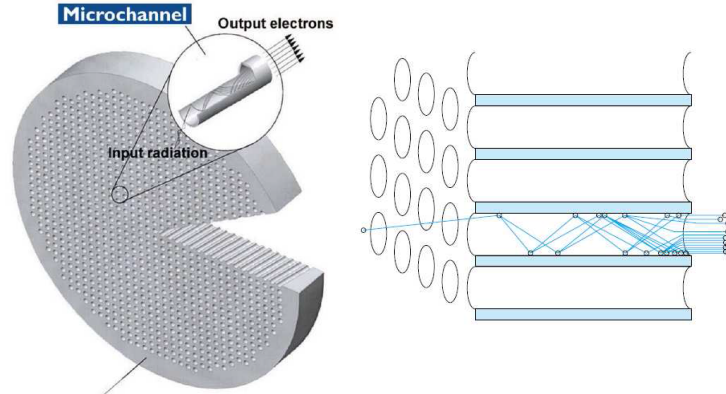


FIGURE 2.2 – Représentation d'une galette MCP et du principe de multiplication des électrons

Plate) afin de multiplier les électrons, comme représenté sur la figure 2.2. Les tubes utilisés dans cette étude sont de génération 2.

Le principe de fonctionnement d'un IL est le suivant. Une quantité de lumière (donc de photons) est captée par la photocathode. Le nombre de photons incidents est noté n_p^i . À faible luminosité, celui-ci est considéré suivre une loi de Poisson $\mathcal{P}(n_p^0)$, où n_p^0 est le nombre de photons moyen sous-jacent ; il s'agit du bruit de photons (*photon count noise*) dont il a été fait référence dans la section précédente. À plus forte luminosité on peut considérer en utilisant la loi des grands nombres que le nombre de photons suit une loi normale à moyenne et variance égales à l'intensité sous-jacente, mais ce n'est en général pas le cas lorsque que l'on fait appel à un tube intensificateur de lumière. La photocathode émet des électrons lorsqu'elle reçoit ces photons, avec un certain rendement η_p qui caractérise le capteur mais dépend de la longueur d'onde. Il y a donc en sortie de la photocathode un nombre d'électrons $n_e = \eta_p \cdot n_p^i$.

Une tension élevée (environ 10 à 15 kV) est appliquée en entrée et en sortie du tube, ce qui accélère les électrons émis par la photocathode. Ceux-ci sont ensuite multipliés dans la galette MCP, avec un certain gain K_{MCP} . Ce gain est dépendant du signal : il peut être assimilé à une variable aléatoire qui dépend de la puissance du signal, et dont la variance dépend elle aussi de la puissance. Lorsque beaucoup d'électrons sont présents, le gain de la galette est plus faible, et sa variance également. On assimile le gain $K_{MCP} = \varphi(n_e)$ à une fonction décroissante de n_e . Les électrons accélérés et multipliés viennent ensuite frapper une galette de phosphore qui reconvertit ces électrons en photons, avec un certain gain K_{ph} dit gain d'accélération (plusieurs photons sont issus pour un électron), considéré constant.

L'ensemble de ce dispositif doit fonctionner sous vide. En pratique, le vide à l'intérieur du tube étant imparfait, il subsiste toujours des particules parasites qui vont se trouver accélérés en même temps que les électrons. Cela résulte en un bruit résiduel observé même en l'absence totale de lumière, qui se caractérise par un bruit de scintillement noté b (certains pixels sont illuminés pendant un temps assez bref)

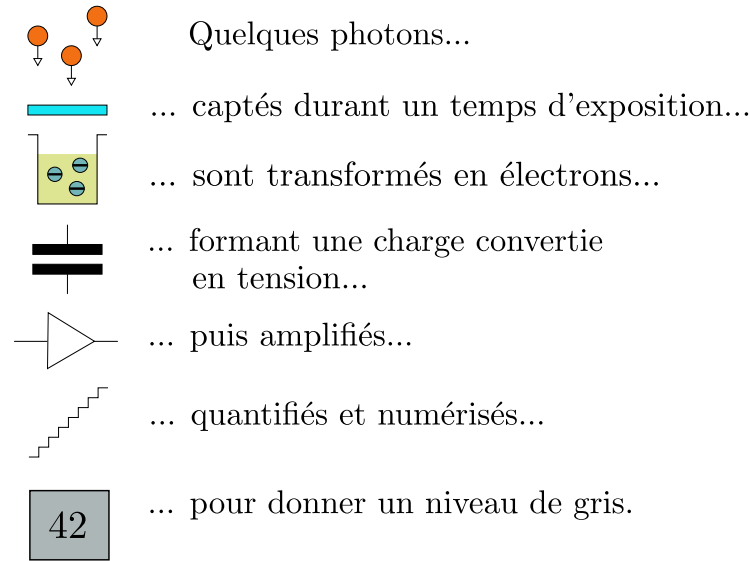


FIGURE 2.3 – Fonctionnement d'une caméra numérique.

assimilable à des outliers. On peut associer à ce dernier une distribution de type poivre et sel.

À ce stade, le nombre de photons est donc, en fonction du nombre de photons incidents :

$$n_p = \eta_p \cdot K_{ph} \cdot K_{MCP} \cdot n_p^i + b = \eta_p \cdot K_{ph} \cdot \varphi(\eta_p \cdot n_p^i) \cdot n_p^i + b, \text{ où } n_p^i \sim \mathcal{P}(n_p^0). \quad (2.12)$$

Les photons se propagent ensuite dans une galette de fibres avant d'arriver dans la caméra CCD. L'assemblage entre l'IL et la caméra CCD se fait à l'aide d'une colle qui est donc elle aussi traversée par les photons. Au cours de cette propagation, les photons se diffusent spatialement, ce qui résulte en une perte de FTM (fonction de transfert en modulation), c'est-à-dire en une perte de résolution : les photons sont "étalés". Cela peut être assimilable à une convolution par un noyau gaussien, ou à de la corrélation spatiale.

2.3.2 La caméra CCD

La caméra CCD (pour *Charge Coupled Device*, appelée parfois en français "dispositif à transfert de charge") assure la conversion du signal lumineux en signal électrique, numérique. Le fonctionnement d'une caméra CCD est schématisé sur la figure 2.3.

Les photons issus de l'intensificateur de lumière, dont le nombre est représenté par la variable aléatoire n_p décrite ci-dessus, sont convertis en électrons grâce à une photocathode, dont le rendement η dépend de la longueur d'onde. Comme précédemment, on a $n_e = \eta \cdot n_p$.

À cela viennent s'ajouter un certain nombre de bruits additifs. Le bruit de lecture n_0 est assimilable à un bruit blanc additif gaussien et correspond aux différents bruits

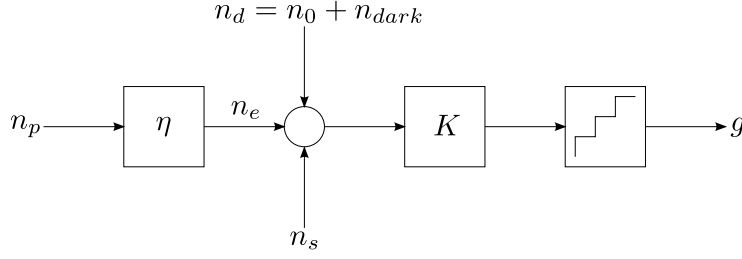


FIGURE 2.4 – Chaîne des bruits dans une caméra CCD.

de mesure dans la chaîne de traitement. Le courant d'obscurité ou *dark current* n_{dark} est lié à l'énergie thermique nécessaire à la libération des électrons. Ce bruit est proportionnel au temps d'exposition T_{exp} et dépend de la température à travers un coefficient N_T . Sa distribution suit une loi de Poisson. On note $n_d = n_0 + n_{dark}$ le bruit temporel. Enfin, la réponse n'est pas la même en chaque pixel en raison de l'hétérogénéité du capteur. Cela se traduit par un bruit spatial n_s .

Les électrons sont en suite multipliés par un gain K puis échantillonnés et convertis en une valeur de niveau de gris y . La chaîne est donc résumée par la formule suivante, et schématisée sur la figure 2.4.

$$y \leftarrow K(\eta.n_p + n_d + n_s) = K(\eta.n_p + n_0 + n_{dark} + n_s) \quad (2.13)$$

2.3.3 Modélisation du bruit

La complexité du dispositif d'acquisition ne permet pas de modéliser finement l'ensemble des perturbations dans le contexte du débruitage. Le comportement des différentes composantes décrites ci-dessus peut néanmoins être analysé afin de simplifier le modèle et déterminer les perturbations prépondérantes à prendre en compte dans la problématique de débruitage. Deux cas de figures sont distingués : à forte luminosité et à faible luminosité.

À forte luminosité, le nombre n_p^i de photons reçus par la photocathode est élevé, et le gain de la galette MCP $K_{MCP} = \varphi(\eta_p.n_p^i)$ est assez faible. Nous pouvons également supposer que le bruit de scintillement sera moins visible car noyé dans les photons présents. On peut donc considérer qu'à la sortie du tube IL le nombre de photons n_p peut être assimilé à un bruit gaussien à moyenne et variance égales. De plus, le temps d'intégration de la caméra CCD peut lui aussi être diminué. Cela diminue donc le courant d'obscurité n_{dark} proportionnel au temps d'intégration. Ainsi, nous pouvons considérer que le bruit résiduel de la caméra CCD sera essentiellement lié aux bruits de lecture n_0 ainsi qu'à la corrélation spatiale. L'hypothèse du bruit additif gaussien à variance dépendant du signal et d'une éventuelle convolution par un noyau gaussien semble donc être représentative de la configuration à fort éclairage.

À faible luminosité, le flux de photons en entrée est plus faible. Le gain de la galette MCP et le temps d'intégration sont augmentés, ce qui augmente la composante multiplicative et le courant d'obscurité. Le bruit de lecture peut alors être

considéré comme négligeable devant ces derniers. Dans cette configuration, il faut donc prendre en compte la statistique de Poisson liée au faible nombre de photons et au courant d'obscurité ainsi que le bruit de scintillement et la corrélation spatiale. L'influence du gain de la galette MCP K_{MCP} est plus délicate à prendre en compte. En effet, celui-ci se comporte comme une variable aléatoire dont la valeur moyenne serait inversement proportionnelle à la puissance du signal (c'est-à-dire au nombre d'électrons présents), puisque le gain est d'autant plus fort que le nombre d'électrons est faible, et dont la variance serait proportionnelle à sa moyenne (plus le gain est grand, plus l'incertitude sur la valeur de ce gain est élevée). L'influence de ce paramètre, ajouté à la statistique de Poisson qui décrit le nombre d'électrons présents en entrée de la galette, est donc difficile à quantifier. Nous considérons donc dans un premier temps que ce dernier joue sur le paramètre de la loi de Poisson : lorsque la luminosité est faible, nous observons une loi de Poisson de paramètre n_p/Q où n_p est le nombre de photons et Q serait proportionnel au gain de la galette K_{MCP} . Ainsi, plus l'éclairement est faible, plus le gain est élevé, et plus le bruit de Poisson est important. Viennent également s'ajouter à cela le bruit de scintillement qui cette fois sera observable en raison du faible nombre de particules provenant de la scène, ainsi que la corrélation spatiale.

2.3.4 Estimation empirique du bruit dans les images IL

La modélisation du bruit d'une caméra IL-CCD pour de l'imagerie microscopique a été réalisée dans [Roudot 2013]. Les auteurs montrent la nécessité de corriger le modèle théorique établi afin de représenter plus finement les dégradations observées sur les images.

De façon analogue ici, afin de valider les hypothèses établies par le modèle théorique, il faudrait pouvoir à partir des données estimer la nature du bruit. Sa distribution d'une part, et sa puissance d'autre part, en fonction du niveau de luminosité. Il s'agit donc de déterminer la relation entre l'intensité de l'image et la variance du bruit, qui peut être obtenue à partir des statistiques extraites des zones homogènes de l'image. Pour cela, deux approches sont possibles : la plus simple est de réaliser une étude a priori « à la main » : à partir d'une base de données d'images classée par niveaux de luminosité, on peut effectuer une analyse manuelle de chacune des images et associer une loi de distribution du bruit (distribution et paramètres) en fonction par exemple du niveau de nuit.

Nous avons mis en place une telle étude à l'aide de mires de contraste qui présentent divers niveaux d'intensité. À partir d'acquisitions réalisées en salle noire calibrée avec le dispositif IL-CCD fourni par Thales, nous avons pu effectuer une sélection manuelle des zones homogènes de l'image, représentée sur la première ligne de la figure 2.5. Il est possible à partir de cette sélection de récupérer sur ces zones les statistiques (moyenne, variance) du bruit. Celles-ci sont représentées sur la seconde ligne de la figure 2.5. L'ajustement d'une courbe par des outils Matlab révèle qu'une relation polynomiale d'ordre deux permet de caractériser l'évolution de la variance du bruit, à des niveaux de nuit plus ou moins forts.

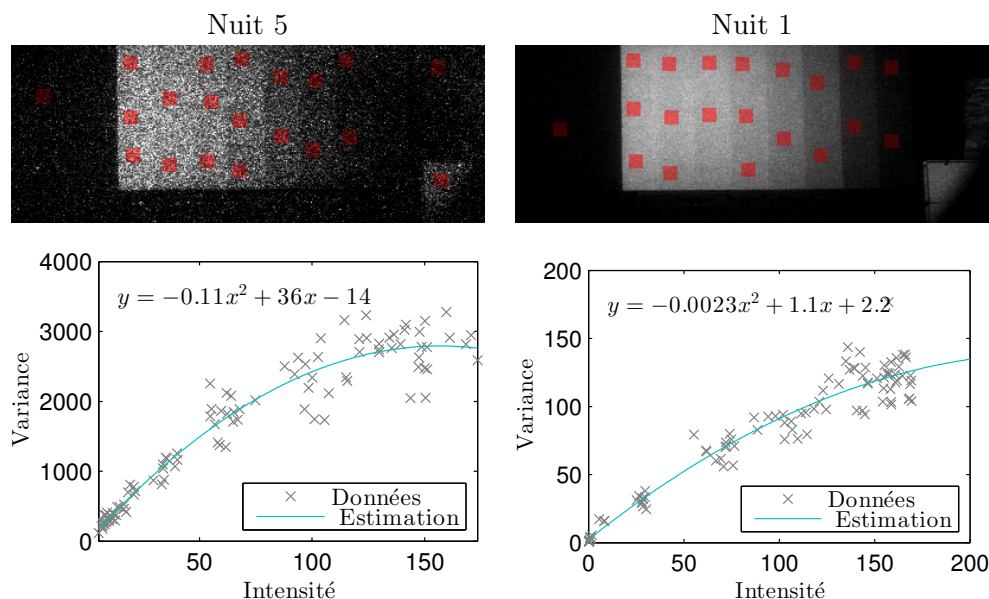


FIGURE 2.5 – Sélection manuelle de zones homogènes sur des images de mires de contraste et estimation de l'évolution de la variance du bruit, pour différents niveaux de nuit.

L'inconvénient majeur est que cette méthode est non automatique, donc difficile à utiliser en pratique. De plus, elle dépend naturellement du capteur et est peu flexible, donc elle ne s'adapterait pas à des conditions qui sortiraient du cadre des essais réalisés en amont.

La deuxième approche consiste donc à utiliser une méthode **automatique** d'estimation de la nature du bruit. Cette dernière doit être robuste aux différentes configurations de luminosité susceptibles d'être rencontrées, et utilisable en temps « raisonnable » : la contrainte temps réel n'est pas nécessaire car il ne s'agit pas d'estimer pour chaque image les paramètres précis du bruit, mais la caractérisation du bruit serait plutôt une étape d'initialisation effectuée en vol avant de lancer les algorithmes de débruitage et de traitement en général, puis mise à jour à intervalles réguliers. Le chapitre suivant est ainsi dédié à l'estimation automatique et robuste du bruit dans une image.

2.4 Conclusion

Ce chapitre s'est consacré à la modélisation du bruit en imagerie. Après avoir étudié les modèles mathématiques souvent utilisés pour décrire les dégradations associées aux systèmes d'imagerie, une analyse du dispositif d'acquisition des images IL a été effectuée. Celle-ci a révélé plusieurs composantes de bruit dépendant du signal, dont l'influence en fonction des conditions de luminosité est complexe et difficile à quantifier. À l'aide de cette analyse, nous avons proposé un modèle de comportement pour le bruit, à forte ou faible luminosité.

Pour valider ces hypothèses et surtout affiner le modèle en lui associant des paramètres de distribution et de puissance, il est nécessaire d'avoir recours à une méthode d'estimation automatique du bruit, afin de pouvoir estimer le modèle directement à partir des données acquises en vol. L'estimation automatique du bruit fait l'objet du chapitre suivant.

Estimation non paramétrique du bruit

3.1 Introduction

Beaucoup d'applications en traitement d'images, par exemple le débruitage [Buades 2005b, Sutour 2014b], la super-résolution [Freeman 2002], ou encore l'extraction de points d'intérêts [Lowe 2004], requièrent la connaissance préalable du niveau de bruit présent dans l'image. L'estimation du bruit à partir de plusieurs images est un problème sur-contraint résolu notamment dans [Healey 1994]. Dans le cas d'une seule image en revanche, l'estimation du bruit reste un problème délicat. Différentes approches dédiées à l'estimation de bruit non corrélé spatialement sont détaillées ci-dessous.

Une première approche pour l'estimation du niveau de bruit consiste à appliquer une transformation linéaire à l'image pour l'envoyer dans un domaine adapté dans lequel le signal et le bruit sont relativement bien séparés. Le niveau de bruit est alors estimé à l'aide de l'évaluation de statistiques dans le domaine associé. Lorsque le bruit est additif gaussien par exemple, un estimateur populaire de l'écart type du bruit est donné par l'écart type absolu moyen (MAD, pour *Mean Absolute Deviation*) des coefficients d'ondelettes les plus fins [Donoho 1995]. Cette approche a néanmoins tendance à surestimer le niveau de bruit car elle nécessite que le bruit et le signal soient suffisamment séparés, ce qui en pratique n'est pas toujours le cas. En effet, les singularités comme les contours sont souvent responsables de réponses fortes dans toutes les bandes du domaine de transformation. Pour contrer ce problème, certains auteurs [Olsen 1993, Tai 2008] ont proposé de préfiltrer l'image afin de supprimer les composantes haute-fréquence de l'image, à l'aide par exemple d'un filtre basé sur le Laplacien combiné à un détecteur de contour de Sobel [Rank 1999]. Cependant, ces approches conduisent toujours à une surestimation non négligeable du niveau de bruit. Plutôt que d'essayer de supprimer les hautes-fréquences, on peut chercher à les prendre directement en compte en supposant un modèle a priori sur la distribution des coefficients du signal. Dans [De Stefano 2004], les auteurs utilisent une méthode "d'apprentissage" des moments à partir d'une hypothèse de distribution laplacienne des coefficients du signal. Les auteurs de [Zoran 2009] exploitent eux l'invariance du kurtosis en faisant l'hypothèse d'une distribution gaussienne généralisée. Des techniques basées sur les patches ont également été proposées dans ce but, par exemple dans [Shin 2005, Liu 2013, Pyatykh 2013]. Si ces techniques sont puissantes, elles sont néanmoins limitées à des niveaux de bruit constants sur

toute l'image. Dans des applications pratiques, le bruit est en général dépendant du signal, et son niveau varie spatialement en fonction de l'intensité sous-jacente du signal. L'application de stabilisation de variance [Anscombe 1948] peut permettre de réduire la dépendance au signal. Les auteurs de [Pyatykh 2014] réalisent l'estimation de bruit Poisson-gaussien de façon itérative en appliquant une transformation de stabilisation de variance, puis ils estiment la variance du bruit stabilisé par une analyse en composantes principales. Dans [Boulanger 2010], le bruit Poisson-gaussien est stabilisé à l'aide de la transformation d'Anscombe généralisée. Cette dernière nécessite cependant l'estimation des paramètres du bruit, obtenus par régression linéaire sur des zones à variance homogène [Gasser 1986].

Les techniques de séparation ont été étendues à des modèles spécifiques de bruit dépendant du signal, par exemple à l'aide d'une transformée en ondelettes dans le cas d'un modèle Poisson-gaussien [Foi 2008, Azzari 2014a] ou à l'aide d'une modélisation de patches par des mélanges de gaussiennes pour du bruit additif à variance affine [Azzari 2014b]. Souvent malheureusement, il peut être difficile de développer une transformation linéaire qui sépare le bruit du signal de façon générale. Une alternative consiste à préfiltrer l'image pour ne garder que le résidu, supposé correspondre à la composante de bruit. [Hensel 2007, Chen 2007] ont par exemple proposé différentes techniques de préfiltrage associées à des types spécifiques de bruits dépendant du signal. Cependant dans le cas général, l'utilisation d'un filtre adapté nécessite d'avoir une certaine connaissance préalable de la nature du bruit, ce qui conduit à un problème de l'œuf et de la poule.

Une autre approche populaire rencontrée notamment dans [Chehdi 1992, Beaurepaire 1997] et reprise ici repose sur le fait que les images naturelles contiennent des zones homogènes, dans lesquelles le rapport signal à bruit est très faible. On peut alors considérer que seules les statistiques du bruit interviennent dans ces zones, sans perturbation liée au signal. Les statistiques du bruit peuvent donc être estimées en détectant dans un premier temps les zones homogènes, puis en estimant les paramètres du bruit dans ces régions. Cependant beaucoup de détecteurs classiques de zones uniformes nécessitent de faire des hypothèses sur la nature du bruit (par exemple gaussien), ce qui ne permet pas de les utiliser de façon fiable lorsque la nature du bruit est totalement inconnue.

D'autres techniques ont été consacrées au bruit corrélé spatialement. Ces méthodes estiment la matrice de covariance du bruit dans un domaine de transformation (DCT ou ondelettes), en séparant le bruit du signal par la représentation en mélange de gaussiennes [Portilla 2004], par analyse en composante principale [Colom 2014a] ou en ne gardant qu'une part représentative des variances estimées sur des blocs DCT [Colom 2014b, Lebrun 2015]. Nous nous focalisons dans le cadre de cette étude sur l'analyse du bruit non corrélé spatialement.

L'approche décrite dans cette partie consiste dans un premier temps à détecter les zones homogènes à l'aide d'une méthode non paramétrique qui nécessite seulement que le bruit soit décorréolé spatialement. Cela signifie que la performance du détecteur est indépendante de la distribution statistique du bruit, inconnue. Le détecteur de zones homogènes, basé sur le coefficient de corrélation de rang de Kendall

[Kendall 1938], est introduit dans la section 3.2. La performance de celui-ci est reliée au rapport signal à bruit, et le taux d'erreur décroît quand le nombre d'échantillons disponibles augmente. Une fois les zones homogènes détectées, on peut alors estimer la fonction de niveau de bruit (NLF, pour *Noise Level Function*) [Liu 2006], définie comme la relation entre la variance du bruit et l'intensité du signal. L'estimation de la NLF est présentée dans la section 3.3. À ce stade on suppose alors que le bruit est dépendant du signal, avec des moments d'ordre un et deux finis et reliés par un polynôme d'ordre deux. Cela permet d'englober une grande variété de distributions de bruit rencontrés en imagerie, en particulier le bruit additif gaussien, le bruit multiplicatif, le bruit de Poisson, ou encore des modèles hybrides. Une approche robuste permet d'estimer la NLF polynomiale à l'aide d'une approximation ℓ^1 de la relation entre les moyennes et les variances calculées sur les régions uniformes. Afin d'illustrer l'intérêt d'une telle estimation, une application au débruitage est ensuite proposée : l'algorithme des NL-means [Buades 2005b] est adapté pour traiter des bruits hybrides, à l'aide de l'estimation de la NLF.

3.1.1 Contexte et notations (rappel)

En se basant sur le modèle de bruit défini dans (2.1), on suppose que la variance est donnée en chaque pixel i par

$$\sigma_i^2 = \text{NLF}(f_i), \quad (3.1)$$

où $\text{NLF} : \mathbb{R} \rightarrow \mathbb{R}^+$ est appelée la fonction de niveau de bruit, notée NLF pour *Noise Level Function*. Comme la variance σ_i^2 dépend de f_i , le bruit est dit dépendant du signal. On remarque de plus que la même fonction de niveau de bruit est supposée s'appliquer à chaque élément de la diagonale, donc le bruit est supposé stationnaire. Les modèles de bruit à fonction de niveau de bruit variant spatialement ne sont pas considérés dans cette étude.

Le but de cette étude consiste à estimer la fonction de niveau de bruit NLF. Pour cela, on considère K blocs disjoints $(\omega_k)_{k \in [1, \dots, K]} \subset \Omega$ de taille N_b de sorte que $KN_b \leq N$. On note ω un tel bloc quelconque ω_k et $g^\omega \in \mathbb{R}^N$ (resp. G^ω et f^ω) la restriction de g sur ω (resp. la restriction de G et f sur ω). La motivation principale réside dans le fait que les images naturelles, pourvu que leur résolution soit assez bonne, contiennent des blocs ω pour lesquels le signal f^ω est constant. On dit dans ce cas que G^ω est homogène, et dans le cas contraire quand f^ω varie, G^ω est dit inhomogène. Ainsi, d'après le modèle donné par les équations (2.1) et (3.1), la moyenne et la variance empiriques au sein d'un tel bloc g^ω fournissent une estimation ponctuelle de la fonction de niveau de bruit.

En pratique f^ω est inconnu, de même que les blocs homogènes. C'est pourquoi on étudie dans la section 3.2 un test statistique qui permet de décider si un bloc est homogène ou non, sans faire d'hypothèses supplémentaires sur la statistique du bruit que celles introduites dans (2.1). Une fois les zones homogènes détectées, plusieurs estimations ponctuelles de la fonction de niveau de bruit sont disponibles. La section

3.3 décrit comment à partir de ces dernières estimer de façon robuste la fonction de niveau de bruit, avec peu d'hypothèses supplémentaires.

3.2 Détection des zones homogènes

Les statistiques du bruit sont estimées à partir de plusieurs régions homogènes, dans lesquelles le rapport signal à bruit est faible. En effet, dans ces zones les fluctuations liées au signal sont supposées négligeables par rapport aux fluctuations liées au bruit, donc seules les statistiques du bruit interviennent. Ainsi, pourvu qu'un nombre suffisant de régions homogènes soit détecté, les statistiques issues de chaque zones permettent de déduire la fonction de niveau de bruit sous-jacente. Le but de cette partie est donc de développer une méthode automatique de sélection des zones homogènes dans une image, en se basant sur le principe que les images naturelles contiennent effectivement de telles zones.

3.2.1 Homogénéité et décorrélation

Suivant le principe issu de la théorie de la détection, le but est de développer un détecteur de blocs homogènes qui à partir de la seule observation du bloc g^ω répond au test d'hypothèse statistique suivant :

$$\begin{aligned} H_0 : G^\omega \text{ est homogène, i.e., } f^\omega \text{ constant} & \quad (\text{hypothèse nulle}), \\ H_1 : G^\omega \text{ est inhomogène, i.e., } f^\omega \text{ varie} & \quad (\text{hypothèse alternative}). \end{aligned} \quad (3.2)$$

Comme f^ω n'est pas connu, on ne peut pas répondre directement à ce test. Pour contourner cette difficulté, on utilise la proposition suivante.

Proposition 3.2.1 *Soit $G^\omega \in \mathbb{R}^{N_b}$ un bloc homogène. On note G_I^ω le sous-vecteur aléatoire de G^ω dont les entrées sont indexées par $I \subset \{1, \dots, N_b\}$. Alors, pour toutes séquences disjointes I et J de n éléments de $\{1, \dots, N_b\}$, $2n \leq N_b$, $X = G_I^\omega$ et $Y = G_J^\omega$ sont décorrélés, c'est-à-dire :*

$$\text{Corr}(X, Y) = 0 .$$

Cette proposition assure que si deux séquences disjointes X et Y d'un bloc sont corrélées, alors le bloc est inhomogène. En pratique, les deux variables aléatoires X et Y ne sont pas connues, et seule une réalisation de chacune d'entre elles, les signaux bruités $x = g_I^\omega$ et $y = g_J^\omega$, sont disponibles. La corrélation entre X et Y est inconnue, et la question à résoudre est donc "Les réalisations x et y sont-elles corrélées de façon significative? ".

Le but dès lors est de construire un score $s : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ qui permette de répondre correctement au test d'hypothèses. Plus précisément, le score s devrait garantir que, pour toute constante $P_{\text{FA}} \in [0, 1[$, il existe un seuil $\alpha > 0$ tel que, quelle que soit la distribution de X et Y , les conditions suivantes sont vérifiées :

$$\lim_{n \rightarrow \infty} \underbrace{\mathbb{P}(|s(X, Y)| \geq \alpha \mid H_0)}_{\text{fausse alarme}} = P_{\text{FA}}, \quad (C1)$$

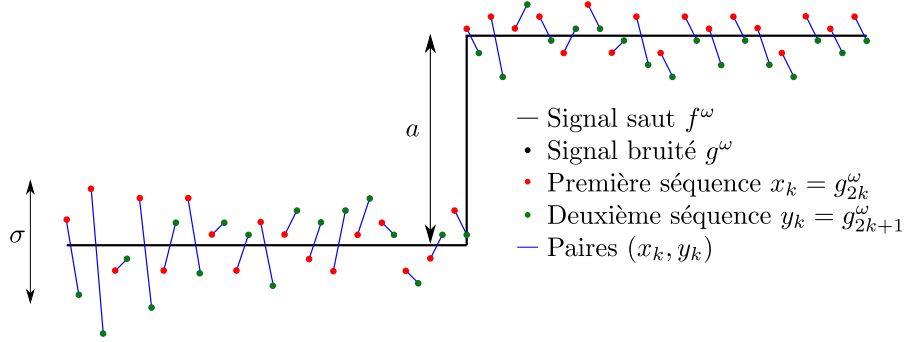


FIGURE 3.1 – Signal constitué d'un saut associé à l'hypothèse alternative H_1 (voir définition 4) et sélection des séquences voisines.

$$\lim_{n \rightarrow \infty} \underbrace{\mathbb{P}(|s(X, Y)| < \alpha \mid H_1)}_{\text{fausse détection}} = 0. \quad (\text{C2})$$

La première condition est primordiale et assure que lorsque le nombre d'échantillons n devient suffisamment important, un même seuil α permette de maintenir la probabilité de fausse alarme P_{FA} , dès lors que X et Y sont décorrélés, et ce quelle que soit leur distribution. On dit que le test est asymptotiquement un détecteur à taux de fausse alarme constant (CFAR, pour *Constant False Alarm Rate*). La seconde condition est naturelle et garantit que pour un même seuil α , la probabilité de fausse détection s'annule quand la taille n des séquences grandit.

Par la suite, on propose donc de construire un tel test statistique entre des paires de séquences disjointes d'un bloc. Comme on ne peut pas en pratique tester toutes les séquences disjointes, on étudie la corrélation entre les séquences voisines notées $x = (g_{2k}^\omega)$ et $y = (g_{2k+1}^\omega)$, où $2k$ et $2k + 1$ représentent les indices des pixels voisins pour un parcours donné du bloc ω . Un exemple de construction de deux séquences voisines $x = (g_{2k}^\omega)$ et $y = (g_{2k+1}^\omega)$ est représenté sur la figure 3.1. Si ces deux variables s'avèrent être corrélées de façon significative, cela traduit des dépendances entre les pixels d'un bloc et leurs voisins, ce qui permet de considérer qu'une certaine structure est présente au sein du bloc et que les variations ne sont pas seulement imputables au bruit.

L'approche naïve pour répondre au test d'hypothèse ci-dessus serait d'utiliser la corrélation empirique entre les deux séquences x et y , c'est-à-dire $s(x, y) = \text{Corr}(x, y)$. Malheureusement, la significativité de ce test dépend des paramètres du processus aléatoire qui génère les observations. On ne peut donc pas construire de détecteur CFAR à partir de ce dernier sans hypothèse supplémentaire sur la distribution de X et Y . Le but étant d'estimer la nature du bruit, il est primordial que le test d'homogénéité ne dépende pas de ces paramètres. On considère pour cela une approche non paramétrique dont la réponse statistique est indépendante du modèle du bruit.

3.2.2 Test d'indépendance non paramétrique

Comme il a été mentionné précédemment, la corrélation empirique ne permet pas de construire un détecteur CFAR car le seuil α dépendrait des paramètres inconnus du processus aléatoire à l'origine des observations. L'idée phare derrière le test non paramétrique est de se baser sur le rang (c'est-à-dire sur l'ordre relatif) des observations plutôt que sur leurs valeurs. Un bloc est alors considéré homogène si le rang des valeurs des pixels est distribué uniformément, quelle que soit l'organisation spatiale de ces valeurs. Une telle corrélation se basant sur le rang des séquences x et y peut être évaluée à l'aide du coefficient de Kendall.

3.2.2.1 Le coefficient de Kendall

Le coefficient de Kendall est une mesure de corrélation de rang [Kendall 1938] qui fournit un test non paramétrique de dépendance statistique. Afin de définir τ , le coefficient de Kendall, il nous faut introduire les notions de paires concordantes ou discordantes, et d'égalité.

Définition 1 Deux paires d'observations (x_i, y_i) et (x_j, y_j) sont dites **concordantes** si l'ordre des rangs des deux éléments concorde, c'est-à-dire si $(x_i < x_j$ et $y_i < y_j)$ ou $(x_i > x_j$ et $y_i > y_j)$. Elle sont qualifiées de **discordantes** si $(x_i < x_j$ et $y_i > y_j)$ ou si $(x_i > x_j$ et $y_i < y_j)$. Si $x_i = x_j$ ou $y_i = y_j$, la paire n'est ni concordante ni discordante ; on dit qu'il y a **égalité**.

Le coefficient de Kendall τ introduit dans [Kendall 1938] peut alors être défini comme suit.

Définition 2 Soient $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^n$ deux séquences sans cas d'égalités. Le coefficient de Kendall $\tau : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [-1, 1]$ est donné par

$$\tau(x, y) = \frac{n_c - n_d}{\frac{n(n-1)}{2}} = \frac{1}{n(n-1)} \sum_{1 \leq i, j \leq n} \text{sign}(x_i - x_j) \text{sign}(y_i - y_j), \quad (3.3)$$

où n_c est le nombre de paires concordantes et n_d le nombre de paires discordantes, et sign désigne le signe.

Le coefficient de Kendall $\tau(x, y)$ dépend seulement de l'ordre relatif entre les valeurs de x et y . Il est compris entre les valeurs -1 et 1 : $-1 \leq \tau(x, y) \leq 1$. Si la concordance entre les deux classements est parfaite, toutes les paires sont concordantes, et donc $\tau(x, y) = 1$. Si la concordance entre les deux classements est nulle, toutes les paires sont discordantes, et donc $\tau(x, y) = -1$. La valeur $\tau(x, y) = 0$ indique l'absence de corrélation significative entre x et y .

La définition 2 est en revanche limitée au cas de séquences qui ne contiennent pas d'égalités. Cela restreint la mesure de corrélation au cas où le processus aléatoire de génération du bruit est continu, par exemple dans le cas gaussien, où l'apparition d'égalités est alors un ensemble de mesure nulle. Dans le contexte d'imagerie et dans le cas par exemple des systèmes optiques, le processus de formation des images fait

intervenir une étape de comptage des photons, discret par nature. Des cas d'égalité peuvent alors apparaître avec une probabilité non nulle. Il est donc important de les prendre en compte dans le calcul de la corrélation.

Lorsque que les données sont susceptibles de contenir des égalités, la définition du coefficient de Kendall proposée dans [Kendall 1945] permet de les prendre en compte comme suit :

Définition 3 Soient $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^n$ deux séquences arbitraires. On définit à partir de x et y les quantités

$$\begin{aligned} n_0 &= n(n-1)/2 && \text{le nombre total de paires,} \\ n_1 &= \sum_i t_i(t_i-1)/2 && \text{le nombre total d'égalités pour } x, \\ n_2 &= \sum_j u_j(u_j-1)/2 && \text{le nombre total d'égalités pour } y, \\ t_i &&& \text{le nombre d'égalités pour le } i^{\text{e}} \text{ groupe pour } x, \\ u_j &&& \text{le nombre d'égalités pour le } j^{\text{e}} \text{ groupe pour } y. \end{aligned}$$

Le coefficient de Kendall $\tau : \mathbb{R}^n \times \mathbb{R}^n \in [-1, 1]$ est donné par

$$\tau(x, y) = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} . \quad (3.4)$$

On remarque qu'en l'absence d'égalités, la définition 3 est équivalente à la première définition 2. Munis d'une telle mesure de corrélation, on peut désormais étudier sa capacité à répondre au test d'hypothèses en satisfaisant les contraintes (C1) et (C2).

3.2.2.2 Distribution de τ dans les blocs homogènes (sous l'hypothèse nulle H_0)

On étudie dans cette section la distribution du coefficient de Kendall sous l'hypothèse nulle H_0 et on montre qu'il permet de satisfaire la condition (C1). En particulier, dans la mesure où il n'est basé que sur l'ordre relatif de x et y , il fournit une mesure non paramétrique de corrélation qui conduit à un détecteur CFAR.

Dans un premier temps, en l'absence d'égalités, la distribution de τ est régie comme suit.

Proposition 3.2.2 Soient X et Y deux séquences arbitraires sans égalités, sous l'hypothèse H_0 . La variable aléatoire $\tau(X, Y)$ a pour espérance 0 et une variance égale à $\frac{2(2n+5)}{9n(n-1)}$. De plus, si le nombre d'échantillons n est important, sa distribution tend approximativement vers une loi normale. Autrement dit, pour toutes séquences x et y , soit $z : \mathbb{R}^n \times \mathbb{R}^n \in \mathbb{R}$ le z -score défini par

$$z(x, y) = \frac{3(n_c - n_d)}{\sqrt{n(n-1)(2n+5)/2}} . \quad (3.5)$$

Le z -score est asymptotiquement distribué selon une loi normale centrée réduite

$$z(X, Y) \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1) .$$

La preuve est donnée dans [Kendall 1948]. La proposition suivante traite l'occurrence d'égalités.

Proposition 3.2.3 *Pour toutes séquences $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^n$, soit $z : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ le z -score défini par*

$$z(x, y) = \frac{n_c - n_d}{\sqrt{v}} \quad (3.6)$$

avec

$$\begin{aligned} v &= (v_0 - v_t - v_u)/18 + v_1 + v_2 \\ v_0 &= n(n-1)(2n+5) \\ v_t &= \sum_i t_i(t_i-1)(2t_i+5) \\ v_u &= \sum_j u_j(u_j-1)(2u_j+5) \\ v_1 &= \sum_i t_i(t_i-1) \sum_j u_j(u_j-1) / (2n(n-1)) \\ v_2 &= \sum_i t_i(t_i-1)(t_i-2) \sum_j \frac{u_j(u_j-1)(u_j-2)}{9n(n-1)(n-2)}. \end{aligned}$$

Alors, pour toutes séquences X et Y sous H_0 , le z -score est asymptotiquement distribué selon une loi normale centrée réduite

$$z(X, Y) \underset{n \rightarrow \infty}{\sim} \mathcal{N}(0, 1) .$$

La preuve est donnée dans [Kendall 1945]. Une conséquence directe de la proposition 3.2.3 est la suivante.

Corollaire 3.2.4 *Soient X et Y deux séquences arbitraires sous H_0 , alors*

$$\lim_{n \rightarrow \infty} \mathbb{P}(z(X, Y) > \alpha \mid H_0) = \int_{\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = 1 - \phi(\alpha), \quad (3.7)$$

où ϕ est la fonction de répartition (cdf, pour Cumulative Distribution Function) de la loi normale

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt. \quad (3.8)$$

La preuve est immédiate et repose sur la définition de la probabilité de fausse alarme donnée par (C1), combinée au fait que $z(X, Y)$ suit asymptotiquement une loi normale centrée réduite sous H_0 .

Le théorème suivant montre alors que le z -score satisfait la condition (C1), dans la mesure où sa distribution ne repose sur aucune hypothèse sur les distributions de X et Y .

Théorème 3.2.5 *Soient X et Y deux séquences arbitraires sous H_0 . On fixe $P_{FA} \in [0, 1[$ et on choisit*

$$\alpha = \phi^{-1}(1 - P_{FA}/2) = \sqrt{2} \operatorname{erf}^{-1}(1 - P_{FA}) , \quad (3.9)$$

où erf est la fonction d'erreur de la gaussienne, $\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt$. Alors,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|z(X, Y)| > \alpha \mid H_0) = P_{FA} . \quad (3.10)$$

La preuve est donnée dans la section 3.7.1. En pratique, la détection est effectuée en calculant la p -valeur

$$p(x, y) = 2 - 2\phi(|z(x, y)|), \quad (3.11)$$

c'est-à-dire la probabilité sous l'hypothèse nulle H_0 que $|z(X, Y)|$ soit plus grand que $|z(x, y)|$. On rejette alors l'hypothèse nulle si la p -valeur est inférieure à un seuil de significativité prédéterminé, donné par la probabilité de fausse alarme P_{FA} souhaitée. Plus la p -valeur est faible, plus le rejet de l'hypothèse H_0 est significatif.

3.2.2.3 Distribution de τ dans le cas d'un saut (une hypothèse alternative H_1)

L'évaluation de la distribution de τ sous l'hypothèse nulle reflète le caractère non paramétrique du test de Kendall, dans la mesure où aucune hypothèse sur la distribution de X et Y (à l'exception de l'absence de corrélation spatiale) n'est nécessaire pour garantir le caractère CFAR du test.

Lorsque X et Y sont corrélés, un modèle de dépendance doit être adopté afin de déduire des propriétés statistiques pour τ . On étudie alors la tendance du coefficient de Kendall à accepter ou rejeter l'hypothèse alternative H_1 lorsque que X et Y correspondent à des séquences extraites d'un signal unidimensionnel représentant un saut, et corrompu par du bruit blanc additif gaussien. La figure 3.1 illustre un tel signal.

Définition 4 Soit n un entier pair représentant la taille de la séquence et H_1 l'hypothèse de dépendance entre X et Y paramétrée par $a > 0$ et σ^2 de la façon suivante

$$\forall 1 \leq k \leq n, X_k = \begin{cases} \varepsilon_k & \text{si } k \leq n/2 \\ a + \varepsilon_k & \text{si } k > n/2 \end{cases} \quad \text{et} \quad Y_k = \begin{cases} \eta_k & \text{si } k \leq n/2 \\ a + \eta_k & \text{si } k > n/2 \end{cases},$$

où ε et η sont des variables aléatoires indépendantes distribuées selon $\mathcal{N}(0, \sigma^2)$.

Proposition 3.2.6 Soient X et Y deux séquences arbitraires satisfaisant H_1 . Alors,

$$\mathbb{E}[\tau(X, Y) | H_1] = \frac{n}{2(n-1)} \left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right) \right)^2, \quad (3.12)$$

et en particulier

$$\lim_{\frac{a}{\sigma} \rightarrow +\infty, n \rightarrow +\infty} \mathbb{E}[\tau(X, Y) | H_1] = \frac{1}{2}. \quad (3.13)$$

La preuve est détaillée dans la section 3.7.2. Cette proposition affirme qu'asymptotiquement par rapport à la taille des séquences n , l'espérance de τ dans le cas d'un saut tend vers $1/2$ lorsque le rapport signal à bruit (a versus σ) est suffisamment grand. En d'autres termes, puisque l'espérance de τ sous H_0 est nulle, les deux distributions peuvent être considérées bien distinctes dès lors que les séquences sont assez grandes et le rapport signal à bruit suffisant. Reste à montrer que la variance de la distribution empirique de τ décroît quand la taille des séquences n augmente.

Proposition 3.2.7 *Soient X et Y deux séquences arbitraires satisfaisant H_1 . Alors,*

$$\text{Var}[\tau(X, Y) \mid H_1] = O\left(\frac{1}{n}\right).$$

La preuve est détaillée dans la section 3.7.3. Cette proposition assure que lorsque le nombre d'échantillons n croît, la variance de la distribution empirique de τ sous l'hypothèse H_1 décroît vers zéro. Comme l'espérance de τ sous H_1 est strictement positive pour $a > 0$, cela permet de garantir que pour un nombre d'échantillons suffisamment élevé, les distributions de τ sous les deux hypothèses peuvent être distinguées.

Le corollaire suivant étend ce résultat au cas du z -score défini dans (3.5).

Corollaire 3.2.8 *Soient X et Y deux séquences arbitraires satisfaisant H_1 . Alors,*

$$\mathbb{E}[z(X, Y) \mid H_1] = O(\sqrt{n}) \quad \text{et} \quad \text{Var}[z(X, Y) \mid H_1] = O(1). \quad (3.14)$$

La preuve est détaillée dans la section 3.7.4.

Puisque la distribution de z sous H_0 a pour espérance zéro et pour variance 1, ce corollaire garantit que lorsque le nombre d'échantillons n croît, le z -score permet de discriminer les deux hypothèses H_0 et H_1 . En effet, en appliquant l'inégalité de Chebyshev, le théorème suivant montre que le z -score satisfait la condition (C2).

Théorème 3.2.9 *Soient X et Y deux séquences arbitraires satisfaisant H_1 . Pour tout $\alpha > 0$, on a*

$$\mathbb{P}(z(X, Y) < \alpha \mid H_1) \leq \frac{\text{Var}(z(X, Y) \mid H_1)}{(\mathbb{E}[z(X, Y) \mid H_1] - \alpha)^2} = O\left(\frac{1}{n}\right). \quad (3.15)$$

En particulier, comme $\mathbb{P}(|z(X, Y)| < \alpha \mid H_1) \leq \mathbb{P}(z(X, Y) < \alpha \mid H_1)$, alors

$$\mathbb{P}(|z(X, Y)| < \alpha \mid H_1) = O\left(\frac{1}{n}\right). \quad (3.16)$$

La preuve est donnée dans la section 3.7.5.

Illustrations

La figure 3.2 représente les distributions empiriques du z -score sous l'hypothèse nulle H_0 (en bleu) et sous l'hypothèse alternative du saut H_1 (en rouge), pour un nombre croissant d'échantillons ($n = 32, 128, 512$). Le seuil α associé à la probabilité de fausse alarme $P_{\text{FA}} = 0.15$ est représenté par la ligne verte. Sur la ligne supérieure, la puissance du bruit ($\sigma = 40$) est faible par rapport à la taille du saut ($a = 128$), ce qui conduit à un rapport signal à bruit $a/\sigma = 3.2$ assez important. Sur la ligne inférieure, la puissance du bruit ($\sigma = 60$) est plus importante par rapport à la taille du saut ($a = 64$), ce qui conduit à un rapport signal à bruit $a/\sigma = 1.07$ plus faible, ce qui rend la distinction des deux distributions plus délicate. Ces courbes montrent que le taux d'erreur est lié au rapport signal à bruit, c'est-à-dire au rapport entre

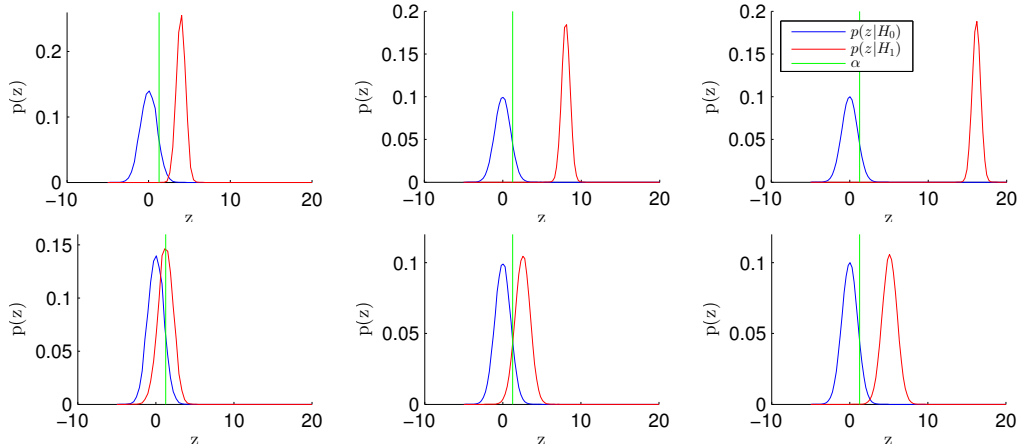


FIGURE 3.2 – Distributions empiriques du z -score sous l’hypothèse nulle H_0 (bleu) et sous l’hypothèse alternative H_1 (rouge) pour un nombre croissant d’échantillons ($n = 32, 128, 512$). Le seuil α associé à la probabilité de fausse alarme $P_{FA} = 0.15$ est représenté par la ligne verte. Sur la ligne supérieure, la puissance du bruit ($\sigma = 40$) est faible par rapport à la taille du saut ($a = 128$), ce qui conduit à un rapport signal à bruit $a/\sigma = 3.2$ assez important. Sur la ligne inférieure, la puissance du bruit ($\sigma = 60$) est plus importante par rapport à la taille du saut ($a = 64$), ce qui conduit à un rapport signal à bruit $a/\sigma = 1.07$ plus faible, rendant la distinction des deux distributions plus délicate.

la puissance du bruit σ et la taille du saut a , en raison de la proximité des deux distributions. En effet sur la ligne inférieure où a et σ sont proches, la proposition 3.2.6 assure que l’espérance de τ et donc de z sous H_1 est proche de zéro, et donc de l’espérance de τ (ou z) sous H_0 . La distinction entre les deux distributions est donc plus délicate. En revanche quand le nombre d’échantillons n croît, d’après le corollaire 3.2.8 l’espérance de z tend vers l’infini, ce qui permet de les distinguer avec certitude.

La figure 3.3 illustre la condition (C2). La probabilité empirique de fausse détection en fonction du nombre d’échantillons est représentée en échelle logarithmique. La borne théorique établie décroît bien asymptotiquement en $1/n$, tandis que la probabilité empirique de fausse détection offre un taux de décroissance encore meilleur.

3.2.2.4 Implémentation

L’étude conduite ci-dessus a montré la capacité du coefficient de Kendall à distinguer les deux hypothèses, c’est-à-dire à distinguer les régions uniformes de celles qui présentent un contour. En utilisant l’approximation des images naturelles par un modèle constant par morceaux, cela permet donc de garantir la détection de régions homogènes, et le taux d’erreur est régi par le nombre d’échantillons et le rapport signal à bruit.

En pratique, on effectue $K = 4$ tests de Kendall pour quatre paires de sé-

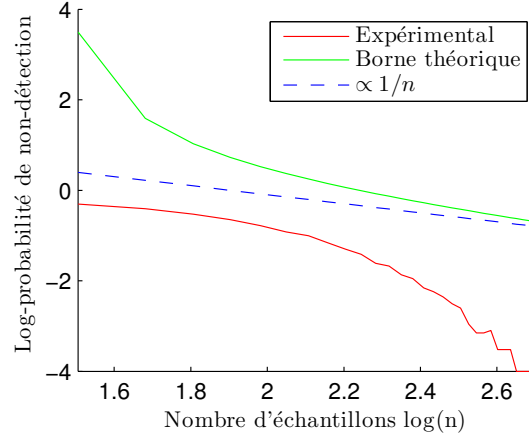


FIGURE 3.3 – Évolution de la probabilité empirique de fausse détection en fonction du nombre d'échantillons n , pour une probabilité de fausse alarme $P_{FA} = 0.15$. Représentée en échelle logarithmique pour plus de lisibilité, la borne de décroissance en $1/n$ est vérifiée.

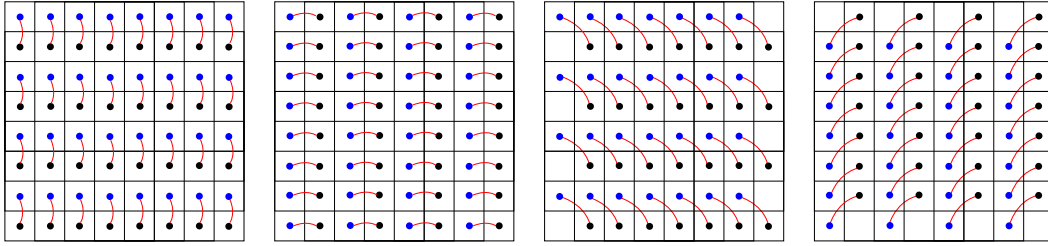


FIGURE 3.4 – Sélection des voisins pour les tests d'indépendance.

quences $(x^{(1)}, y^{(1)})$, $(x^{(2)}, y^{(2)})$, $(x^{(3)}, y^{(3)})$, $(x^{(4)}, y^{(4)})$ correspondant respectivement aux voisinages horizontaux, verticaux, et diagonaux (supérieur droit et inférieur droit), comme illustré sur la figure 3.4. D'autres relations pourraient être considérées également, par exemple des voisinages non pas directs mais situés à une distance de k pixels, pour tout k . La corrélation dans d'autres domaines est envisageable également, par exemple dans le domaine de Fourier, pour mettre en avant des dépendances fréquentielles. Une zone est considérée homogène si chacune des K p -valeurs $p(x^{(k)}, y^{(k)})$ associées au coefficient de Kendall pour chaque direction $\tau(x^{(k)}, y^{(k)})$ atteint le seuil de significativité α . Ainsi, le niveau de détection global α_{eq} après agrégation ne correspond plus à α ; il est plus sélectif et est donné par

$$\alpha_{eq} = \mathbb{P} \left(\bigcap_{k=1}^K \{p(X^{(k)}, Y^{(k)}) > \alpha\} \right). \quad (3.17)$$

Afin de contrôler le niveau global de détection α_{eq} , la relation entre α_{eq} et α est estimée empiriquement au préalable.

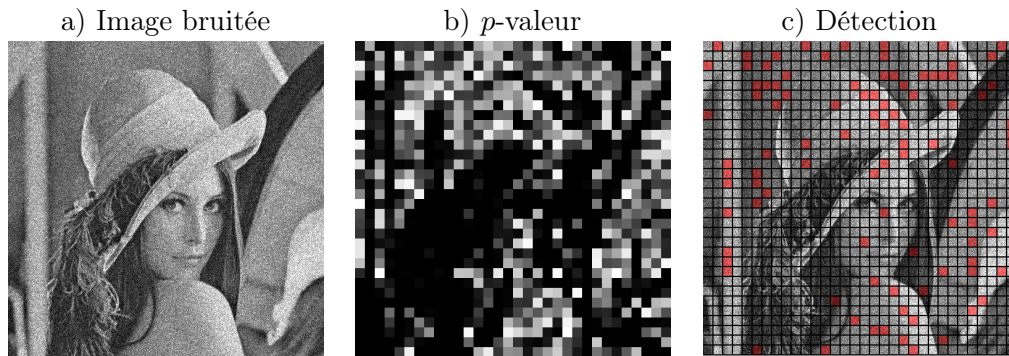


FIGURE 3.5 – Détection des zones homogènes pour une image corrompue par un bruit hybride défini comme la somme de bruit gaussien, poissonien et multiplicatif et associé à une NLF de paramètres $(a, b, c) = (0.0312, 0.75, 400)$. a) Image bruitée (sur $[0, 255]$), PSNR initial 17.93dB, b) p -valeur (dans [noir = 0, blanc = 1]) associée au τ de Kendall calculé sur des blocs de taille $N_b = 16 \times 16$, et c) Sélection des blocs homogènes (en rouge) obtenue en seuillant les p -valeurs pour atteindre la probabilité de détection $P_D = 1 - P_{FA} = 0.7$.

De plus, une implémentation rapide [Knight 1966] qui calcule le coefficient de Kendall en $O(N_b \log N_b)$ pour un bloc de taille N_b permet d'atteindre une complexité globale en $O(N \log N_b)$, pour une image de taille $N = 512 \times 512$ pixels et pour des blocs de taille $N_b = 16 \times 16 = 256$ pixels, l'implémentation rapide permet de calculer le coefficient de Kendall selon un direction en 0.12 secondes, et la détection des zones homogènes sur toute l'image ainsi que le calcul des statistiques sur chacune de ces régions s'effectue en environ 0.6 secondes.

3.2.3 Illustration de la détection

La figure 3.5 illustre le processus de sélection des zones homogènes. Un bruit synthétique hybride correspondant à la somme de bruit gaussien, poissonien et multiplicatif et associé à une NLF de paramètres $(a, b, c) = (0.0312, 0.75, 400)$ a été synthétisé. L'image bruitée obtenue, dont le PSNR est 17.93dB, est représentée sur la figure 3.5-a. Le coefficient de Kendall est calculé sur tous les blocs disjoints de taille $N_b = 16 \times 16$, pour les voisins horizontaux, verticaux et diagonaux. La figure 3.5-b représente la p -valeur associée à la corrélation horizontale. La sélection est obtenue en seuillant les p -valeurs, afin d'obtenir un niveau de détection global donné par la probabilité de détection souhaitée $P_D = 1 - P_{FA} = 0.7$. Les blocs considérés comme homogènes sont représentés en rouge sur la figure 3.5-d.

3.3 Estimation du modèle

Une fois que les blocs homogènes G^ω sont détectés, les statistiques du bruit peuvent être estimés sur ces derniers. Plus précisément, pour un bloc homogène G^ω de taille N_b , l'équation (2.1) entraîne le fait qu'il existe $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ tels que $\mathbb{E}[G_i^\omega] = \mathbb{E}[G_j^\omega] = \mu$ et $\text{Var}[G_i^\omega] = \text{Var}[G_j^\omega] = \sigma^2$ pour tous $i, j \in [1, \dots, N_b]$. Comme cette fois encore G^ω est inconnu, il faut se baser sur les statistiques empiriques issues de l'observation g^ω qui estiment μ et σ^2 comme suit :

$$\hat{\mu}(g^\omega) = \frac{1}{N_b} \sum_{i \in 1}^{N_b} g_i^\omega \quad \text{et} \quad \hat{\sigma}^2(g^\omega) = \frac{1}{N_b - 1} \sum_{i \in 1}^{N_b} (g_i^\omega - \hat{\mu})^2. \quad (3.18)$$

Cette estimation est non biaisée, dans le sens où $\mathbb{E}[\hat{\mu}(G^\omega)] = \mu$ et $\mathbb{E}[\hat{\sigma}^2(G^\omega)] = \sigma^2$. Ainsi, chaque bloc homogène fournit une estimation ponctuelle de la fonction de niveau de bruit définie dans (2.1) par $\hat{\sigma}^2(g^\omega) \approx \text{NLF}(\hat{\mu}(g^\omega))$. À partir des P blocs homogènes détectés g^{ω_p} qui fournissent P paires de statistiques $(\hat{\mu}(g^{\omega_p}), \hat{\sigma}^2(g^{\omega_p}))$, $p \in [1, \dots, P]$ la fonction de niveau de bruit peut alors être estimée de façon robuste.

Afin de se consacrer à l'estimation de cette fonction de niveau de bruit, il est important néanmoins de comprendre le comportement de l'estimateur de variance au sein d'une seule région homogène.

3.3.1 Estimation ponctuelle issue d'une seule zone homogène

L'estimateur non biaisé de la variance σ^2 calculé au sein d'un bloc G^ω de N_b pixels peut souffrir d'erreurs statistiques. On étudie ici la performance de l'estimateur de variance dans le meilleur cas, c'est-à-dire lorsque toutes les fluctuations au sein du bloc g^ω sont imputables au bruit (pas de fausse alarme). Plus précisément, on cherche à quantifier l'erreur relative $\frac{|\hat{\sigma}^2(G^\omega) - \sigma^2|}{\sigma^2}$ sous l'hypothèse H_0 . On s'appuie pour cela sur la proposition suivante.

Proposition 3.3.1 *Soit G^ω un bloc homogène satisfaisant H_0 . On a*

$$\text{Var}(\hat{\sigma}^2(G^\omega) \mid H_0) \underset{N_b \rightarrow \infty}{\sim} \frac{\sigma^4}{N_b} \left(\beta_2 - \frac{N_b - 3}{N_b - 1} \right), \quad (3.19)$$

où β_2 est le kurtosis donné, pour tout $i \in [1, \dots, N_b]$, par la formule suivante :

$$\beta_2 \underset{N_b \rightarrow \infty}{\sim} \mathbb{E} \left[\left(\frac{G_i^\omega - \mu}{\sigma} \right)^4 \mid H_0 \right]. \quad (3.20)$$

La démonstration peut se trouver dans [Cho 2008]. De façon non surprenante, cette proposition montre que lorsque la taille du bloc N croît, la variance de l'estimateur tend vers zéro, et puisque $\hat{\sigma}^2$ est un estimateur non biaisé, $\hat{\sigma}^2(B)$ tend vers σ^2 en probabilité. Cependant, cela montre également que la variance de l'estimateur σ^2 , et donc son efficacité, dépendent de la vraie variance sous-jacente σ^2 et du kurtosis β_2 , tous les deux associés au processus aléatoire inconnu du bruit qui génère g^ω .

L'erreur relative est alors donnée par le théorème suivant :

Théorème 3.3.2 Soit G^ω un bloc homogène satisfaisant H_0 . L'erreur relative d'estimation de la variance σ^2 a pour espérance et pour variance les quantités données par

$$\mathbb{E} \left[\left. \frac{|\hat{\sigma}^2(G^\omega) - \sigma^2|}{\sigma^2} \right| H_0 \right] \underset{N_b \rightarrow \infty}{\sim} \sqrt{\frac{1}{N_b} \left(\beta_2 - \frac{N_b - 3}{N_b - 1} \right) \times \frac{2}{\pi}}, \quad (3.21)$$

$$\text{Var} \left(\left. \frac{|\hat{\sigma}^2(G^\omega) - \sigma^2|}{\sigma^2} \right| H_0 \right) \underset{N_b \rightarrow \infty}{\sim} \frac{1}{N_b} \left(\beta_2 - \frac{N_b - 3}{N_b - 1} \right) \times \left(1 - \frac{2}{\pi} \right). \quad (3.22)$$

La preuve est détaillée dans la section 3.7.6. Ces formules fournissent l'espérance et la variance de l'erreur d'estimation de la variance sur une seule zone homogène. Il est intéressant de noter que, comme pour la variance de $\hat{\sigma}^2(G^\omega)$, ce théorème garantit que l'erreur relative disparaît lorsque la taille des blocs N croît. Mais contrairement à la variance de $\hat{\sigma}^2(G^\omega)$, l'erreur relative sous l'hypothèse H_0 dépend seulement du kurtosis et de la taille du bloc, mais pas de la variance sous-jacente σ^2 que l'on cherche à estimer. Ainsi, pour une taille de bloc N_b donnée, l'erreur d'estimation peut être prédite en fonction du kurtosis, comme le montre la figure 3.6-a dans le cas de blocs de taille $N_b = 16 \times 16$. Ainsi, sans hypothèse sur la variance sous-jacente du bruit, l'erreur relative moyenne peut être bornée sous l'hypothèse H_0 à l'aide d'hypothèses raisonnables sur le kurtosis. Par exemple, en supposant que le kurtosis est inférieur à 40 (ce qui correspond à un aplatissement du processus aléatoire qui génère le bruit très marqué, et donc peu susceptible d'être rencontré dans des problèmes d'imagerie), cela garantit que l'erreur relative moyenne est inférieure à 30%.

Les exemples suivants illustrent cette propriété dans le cas de distributions de bruit rencontrées fréquemment en imagerie.

Bruit blanc additif gaussien

Dans le cas d'un bruit blanc additif gaussien de paramètre σ^2 , le kurtosis β_2 est constant quelle que soit la valeur de σ^2 , égal à 3. Pour un bloc de taille $N_b = 16 \times 16$, le théorème 3.3.2 montre que l'erreur relative moyenne d'estimation de σ^2 est d'environ 7%.

La figure 3.6-b illustre le comportement de l'erreur d'estimation de la variance dans le cas du bruit gaussien. La ligne rouge correspond à l'erreur relative moyenne, qui décroît en $1/\sqrt{N_b}$ d'après le théorème 3.3.2, les courbes bleues correspondent à l'écart type autour de la valeur moyenne et tendent également vers zéro à la même vitesse, et les données en vert montrent l'erreur d'estimation pour chaque taille de bloc considérée, et pour des réalisations de bruit différentes.

Bruit gamma multiplicatif

Dans le cas d'un bruit gamma multiplicatif de paramètre L , le kurtosis est donné par la formule $\beta_2 = \frac{6}{L} + 3$. Contrairement au cas gaussien, l'erreur d'estimation

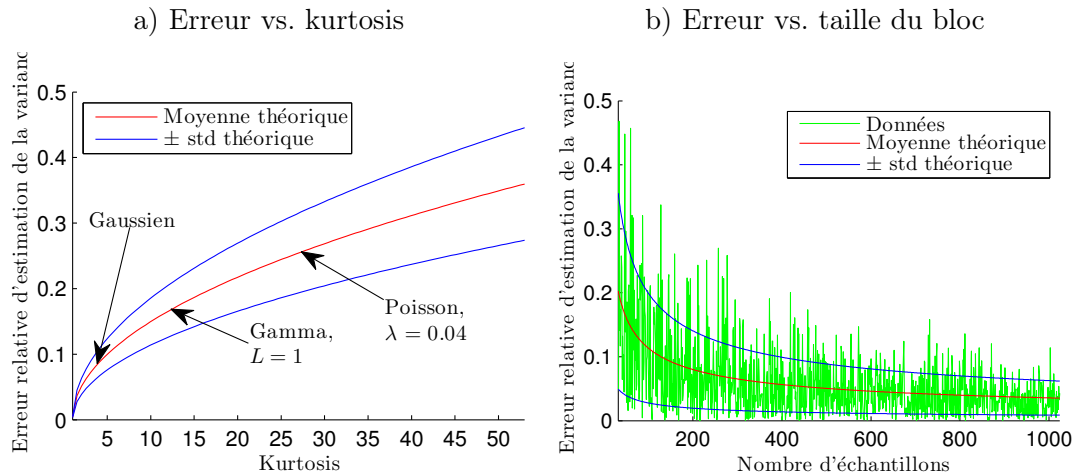


FIGURE 3.6 – a) Évolution de la valeur moyenne et de l'écart type de l'erreur relative de l'estimateur de variance en fonction du kurtosis β_2 pour des blocs de taille $N_b = 16 \times 16$, et b) Évolution de la valeur moyenne et de l'écart type de l'erreur relative de l'estimateur de variance en fonction de la taille du bloc dans le cas d'un bruit blanc additif gaussien.

dépend du paramètre L de la distribution. Dans le pire cas où $L = 1$, le théorème 3.3.2 garantit que pour une taille de bloc $N_b = 16 \times 16$ l'erreur relative moyenne est d'environ 14%.

Bruit de Poisson

Dans le cas d'un bruit de Poisson, le kurtosis varie spatialement en fonction de l'intensité de l'image, et il est donné par $\beta_2 = \frac{1}{f_i} + 3$, où f_i est l'intensité sous-jacente au pixel d'indice i de l'image non bruitée. Par exemple dans un cas pessimiste où $\min_{i \in \omega} f_i = 0.04$, ce qui correspond en imagerie à un pixel avec un très faible nombre de photons, et pour un bloc de taille $N_b = 16 \times 16$, le théorème 3.3.2 montre que l'erreur relative d'estimation de la variance est d'environ 28%.

L'erreur d'estimation calculée ci-dessus fournit une borne sur l'erreur qui affecte l'estimation de la variance sur un seul bloc, c'est-à-dire en un point de la fonction de niveau de bruit. L'étape suivante consiste à faire correspondre une fonction de niveau de bruit à l'ensemble des points estimés, donc l'erreur d'estimation globale devrait être réduite en fonction du nombre de blocs pris en compte, qui dépend lui-même de la taille de l'image et du niveau de détection souhaité.

Cependant, si de fausses détections se produisent, c'est-à-dire si des blocs inhomogènes sont sélectionnés par erreur, la variance est alors surestimée, ce qui résulte en une erreur d'estimation plus importante.

Les bornes sur l'erreur obtenues ci-dessus fournissent donc un moyen de contrôler l'erreur : si l'erreur globale est supérieure à la borne prédite, cela peut signifier ou

bien que trop d'outliers ont été pris en compte, ou bien que la fonction de niveau de bruit n'a pas été correctement estimée à partir des statistiques ponctuelles. D'où l'intérêt d'une estimation robuste.

3.3.2 Estimation de la fonction de niveau de bruit à partir de l'ensemble des zones homogènes

Le but désormais est d'estimer de façon robuste la fonction de niveau de bruit (NLF, pour *Noise Level Function*) définie dans l'équation 2.1 à partir de P paires de statistiques $(\hat{\mu}_p, \hat{\sigma}_p^2) = (\hat{\mu}(g^{\omega_p}), \hat{\sigma}^2(g^{\omega_p}))$, $p \in [1, \dots, P]$, issues des P blocs homogènes g^{ω_p} détectés. Au sein du p^e bloc homogène, les variations d'intensité devraient n'être liées qu'aux fluctuations de bruit, donc $\hat{\sigma}_p^2 \approx \text{NLF}(\hat{\mu}_p)$. Comme chaque bloc ne fournit qu'une estimation ponctuelle de la NLF, il s'agit d'un problème d'approximation sous-contraint qui doit associer à P points de $\mathbb{R} \times \mathbb{R}^+$ une fonction de $\mathbb{R} \rightarrow \mathbb{R}^+$. Il est donc nécessaire d'injecter des hypothèses de régularité supplémentaires. On suppose alors à ce stade que la NLF est une fonction polynomiale du second degré positive et croissante, paramétrée par $\theta = (a, b, c)^\top \in (\mathbb{R}^+)^3$ et donnée par

$$\sigma_p^2 = \text{NLF}_\theta(\mu_p) = a\mu_p^2 + b\mu_p + c. \quad (3.23)$$

Ce modèle englobe le bruit gaussien, avec $(a, b) = (0, 0)$, auquel cas la NLF est constante, le bruit de Poisson, avec $(a, c) = (0, 0)$, auquel cas la NLF est linéaire, le bruit multiplicatif, avec $(b, c) = (0, 0)$, auquel cas la NLF est une parabole, ainsi que les modèles hybrides comme le bruit Poisson-gaussien. Les contraintes de positivité sur les coefficients (a, b, c) pourraient de plus être soulevées pour inclure des modèles encore plus généraux et permettre d'envisager des NLF non croissantes.

Le but dès lors est de trouver les coefficients polynomiaux (a, b, c) de sorte que les variances estimées $\hat{\sigma}_p^2$ soient approchées de façon fine par $\text{NLF}_\theta(\hat{\mu}_p) = a\hat{\mu}_p^2 + b\hat{\mu}_p + c$. En notant $\hat{\mu} \in \mathbb{R}^P$, $\hat{\sigma}^2 \in (\mathbb{R}^+)^P$ et $\text{NLF}_\theta(\hat{\mu})$ les vecteurs obtenus en concaténant chacune des estimations, le problème d'approximation se traduit sous forme vectorielle par

$$\hat{\sigma}^2 \approx \text{NLF}_\theta(\hat{\mu}) = A\theta \quad \text{avec} \quad A = \begin{pmatrix} \hat{\mu}_1^2 & \hat{\mu}_1 & 1 \\ \hat{\mu}_2^2 & \hat{\mu}_2 & 1 \\ \vdots & \vdots & \vdots \\ \hat{\mu}_P^2 & \hat{\mu}_P & 1 \end{pmatrix} \quad (3.24)$$

Bien entendu, si la nature du bruit est supposée connue, l'opérateur A peut être adapté. Dans les cas respectivement de bruit gaussien, poissonien ou multiplicatif, la matrice A serait alors un vecteur colonne ne contenant que des uns, que les moyennes ou bien que les carrés des moyennes. Dans le cas du bruit Poisson-gaussien, seules les deux colonnes contenant les uns et les moyennes ne seraient conservées.

Deux méthodes robustes d'estimation de θ sont étudiées ici.

3.3.2.1 Estimateur des moindres carrés

La méthode des moindres carrés (LS, pour *Least Squares*) consiste à minimiser la norme ℓ^2 du résidu $\text{NLF}_\theta(\hat{\mu}) - \hat{\sigma}^2$. L'estimation est formalisée par la solution du problème d'optimisation contraint suivant :

$$\hat{\theta}^{\text{LS}} = \arg \min_{\theta \in (\mathbb{R}^+)^3} \|\text{NLF}_\theta(\hat{\mu}) - \hat{\sigma}^2\|_2 = \arg \min_{(a,b,c) \in (\mathbb{R}^+)^3} \|a\hat{\mu}^2 + b\hat{\mu} + c - \hat{\sigma}^2\|_2 \quad (3.25)$$

Les solutions sont obtenues à l'aide d'outils de programmation quadratique.

On remarque que sans les contraintes de positivité sur (a, b, c) la solution des moindres carrés est explicite et donnée par

$$\hat{\theta}^{\text{LS}} = (A^t A)^{-1} A^t \hat{\sigma}^2 \quad \text{et} \quad \text{NLF}_\theta^{\text{LS}}(\hat{\mu}) = \Pi_A \hat{\sigma}^2 \quad (3.26)$$

où $\Pi_A = A(A^t A)^{-1} A^t$ est la projection sur l'espace des polynômes du second ordre.

En supposant que $\hat{\sigma}^2$ possède approximativement une distribution normale et une covariance proportionnelle à l'identité, l'estimateur des moindres carrés décrit ci-dessus correspondrait à l'estimateur du maximum de vraisemblance qui converge en probabilité vers la vraie NLF de degré deux sous-jacente. Cependant, bien que $\hat{\sigma}^2$ possède une matrice de covariance diagonale (puisque les blocs sont indépendants), celle-ci n'est pas proportionnelle à l'identité. Pire encore, la distribution de $\hat{\sigma}^2$ s'écarte de la loi normale en raison de la présence d'outliers. En effet, de fortes erreurs sont susceptibles de se produire si des statistiques issues de zones non homogènes sont prises en compte. Cela motive l'étude de méthodes d'estimation plus robustes aux outliers.

3.3.2.2 Estimateur des moindres déviations

L'occurrence de fausses détections entraîne la sélection de régions non homogènes, et la présence de structure dans les échantillons conduit alors à une surestimation de la variance. Une alternative pour réduire l'impact de ces outliers est la méthode des moindres déviations (LAD, pour *Least Absolute Deviation*). Il a été établi [Pollard 1991] que cette dernière est plus robuste aux outliers que l'estimation des moindres carrés.

L'estimation des moindres déviations est basée sur la minimisation d'une norme ℓ^1 du résidu (au lieu d'une norme ℓ^2 dans le cas des LS), sous la forme du problème de minimisation sous contrainte suivant

$$\hat{\theta}^{\text{LAD}} = \arg \min_{\theta \in (\mathbb{R}^+)^3} \|\text{NLF}_\theta(\hat{\mu}) - \hat{\sigma}^2\|_1 = \arg \min_{(a,b,c) \in (\mathbb{R}^+)^3} \|a\hat{\mu}^2 + b\hat{\mu} + c - \hat{\sigma}^2\|_1. \quad (3.27)$$

Il n'existe pas de forme explicite pour la solution, même dans le cas non contraint. On peut cependant estimer la solution via un algorithme itératif rapide, par exemple à l'aide de l'algorithme primal-dual préconditionné de Chambolle-Pock [Chambolle 2011], décrit dans l'algorithme 1.

Tandis que l'estimateur LS correspond à l'estimateur du maximum de vraisemblance dans le cas où les variances estimées suivent une loi normale spécifique,

Algorithme 1 Algorithme primal-dual préconditionné [Chambolle 2011] pour LAD

Prendre $\gamma \in \mathbb{R}^{P,1}$ et $\tau \in \mathbb{R}^{1,P}$ tels que $\gamma_i = 1/\sum_{j=1}^3 |A_{ij}|$, et $\tau_j = 1/\sum_{i=1}^P |A_{ij}|$.

$\Pi_{|z| \leq 1}$ est la projection de \mathbb{R}^P telle que $|z_i| \leq 1 \forall 1 \leq i \leq p$,

$\Pi_{\theta \geq 0}$ est la projection de \mathbb{R}^3 telle que $\theta_j \geq 0 \forall 1 \leq j \leq 3$,

Initialiser $z_0 = 0 \in \mathbb{R}^P$, $\bar{\theta}_0 = 0 \in \mathbb{R}^3$ et $\theta_0 = 0 \in \mathbb{R}^3$,

Itérer pour $k \geq 0$

$$\begin{cases} z_{k+1} &= \Pi_{|z| \leq 1} (z_k + \gamma(A\bar{\theta}_k - \hat{\sigma})), \\ \theta_{k+1} &= \Pi_{\theta \geq 0} (\theta_k - \tau A^* z_{k+1}), \\ \bar{\theta}_{k+1} &= 2\theta_{k+1} - \theta_k. \end{cases}$$

l'estimateur LAD est l'estimateur du maximum de vraisemblance dans le cas d'une distribution laplacienne. Contrairement à la loi normale, la distribution laplacienne a une queue lourde, ce qui lui permet de décrire des données sujettes à des valeurs extrêmes ou des outliers. D'autres distributions à queue lourde comme la loi de Cauchy conduisent à des problèmes d'optimisation non convexe, et donc plus difficiles à minimiser. C'est pourquoi la distribution laplacienne offre un bon compromis entre robustesse et efficacité d'implémentation.

D'autres estimateurs robustes existent, comme l'estimateur de Theil-Sen [Theil 1992] (également connu sous le nom de *Kendall robust line-fit method*), qui estime des fonctions affines en se basant sur la médiane des pentes, sa généralisation à la régression quantile [Koenker 2005] pour estimer des courbes correspondant à la médiane conditionnelle ou autres quantiles associés aux données, ou les M-estimateurs de Hubert [Huber 1964], robustes lorsque la distribution des données n'est connue qu'approximativement.

3.3.2.3 Comparaisons des estimateurs

Pourvu que suffisamment de zones uniformes aient été détectées, les paramètres optimaux $\hat{\theta} = (\hat{a}, \hat{b}, \hat{c})$ de la NLF sont déduits des paires (moyenne, variance) extraites de ces zones à l'aide des estimateurs décrits dans la section 3.3.

Afin de comparer ces estimateurs, des images bruitées ont été générées à partir de 150 images naturelles, à différents types et différents niveaux de bruit. Grâce à la connaissance des vrais paramètres $\theta = (a, b, c)$ de la NLF, on peut calculer l'erreur relative moyenne (MRE, pour *Mean Relative Error*) donnée par

$$\text{MRE}(\hat{\theta}) = \frac{1}{|I|} \sum_{f_i \in I} \frac{|\text{NLF}_{\theta}(f_i) - \text{NLF}_{\hat{\theta}}(f_i)|}{\text{NLF}_{\theta}(f_i)}, \quad (3.28)$$

où I est une discrétisation de l'intervalle des intensités de l'image.

La figure 3.7 montre l'évolution de l'erreur relative moyenne sur la banque d'images en fonction de la probabilité de détection souhaitée $P_D = 1 - P_{FA}$. Sur la figure 3.7-a, les images ont été corrompues par du bruit de Poisson, à différents

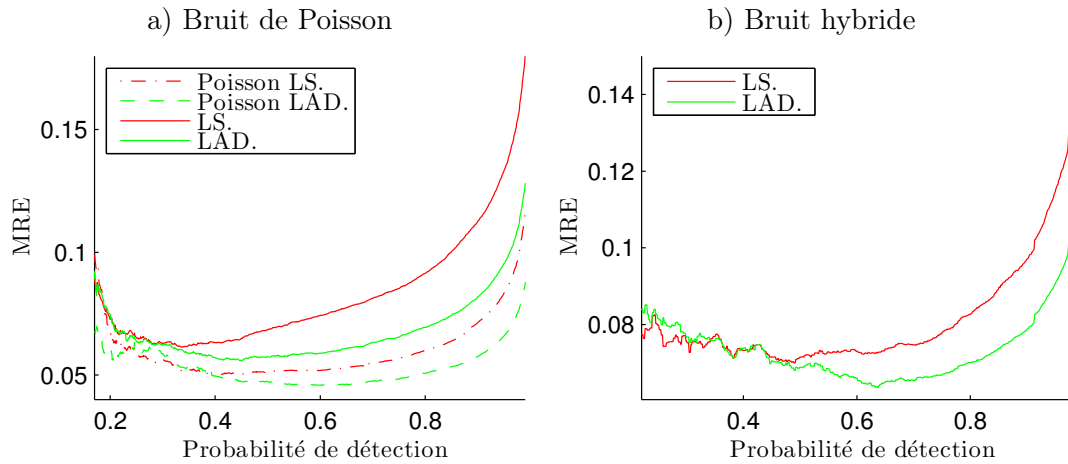


FIGURE 3.7 – Erreur relative moyenne en fonction de la probabilité de détection souhaitée. À gauche, le bruit est poissonien, avec différents niveaux de bruit. Il peut être estimé à l’aide de la connaissance préalable du modèle de bruit, en utilisant les moindres carrés ou les moindres déviations adaptés (*Poisson LS* ou *Poisson LAD*), ou directement à l’aide des estimateurs hybrides du modèle général du second ordre, LS ou LAD. À droite, le bruit est hybride, avec différents paramètres de NLF utilisés. La NLF est estimée à l’aide des estimateurs LS et LAD sous la forme générale d’un polynôme d’ordre 2. Que le modèle soit supposé connu ou non, l’estimateur LAD offre un taux d’erreur plus faible que LS.

niveaux de bruit. On peut d’une part supposer la nature poissonnienne du modèle connue et estimer uniquement le paramètre b de la NLF à l’aide des estimateurs des moindres carrés ou des moindres déviations, respectivement *Poisson LS* ou *Poisson LAD*. Ces estimations sont comparées au cas général (*LS* et *LAD*) où l’on estime directement un modèle général d’ordre deux, sans connaissance préalable du modèle de bruit. Sur la figure 3.7-b, les images ont été corrompues par des bruits hybrides à paramètres (a, b, c) variables. Dans ce cas, seul le modèle général du second ordre est estimé.

La forme convexe des courbes révèle deux influences opposées. Pour des taux de détection faibles, le nombre de zones homogènes sélectionnées est faible. Les fausses détections sont peu susceptibles de se produire, en revanche la NLF n’est pas estimée finement car le nombre de données est insuffisant. Quand le niveau de détection augmente, le nombre d’échantillons augmente, ce qui devrait réduire l’erreur d’estimation, mais le nombre de fausses détections augmente également, ce qui entraîne la présence d’outliers. Ces derniers expliquent la forte augmentation de l’erreur à haut taux de détection. Ils expliquent également la supériorité de l’estimateur des moindres déviations par rapport aux moindres carrés, plus marquée à haut niveau de détection. Ces courbes mettent en avant la nécessité d’effectuer un compromis entre détection et outliers. Elles indiquent que pour un taux de détection autour de 60%, une erreur relative inférieure à 10% est garantie sans connaissance préalable

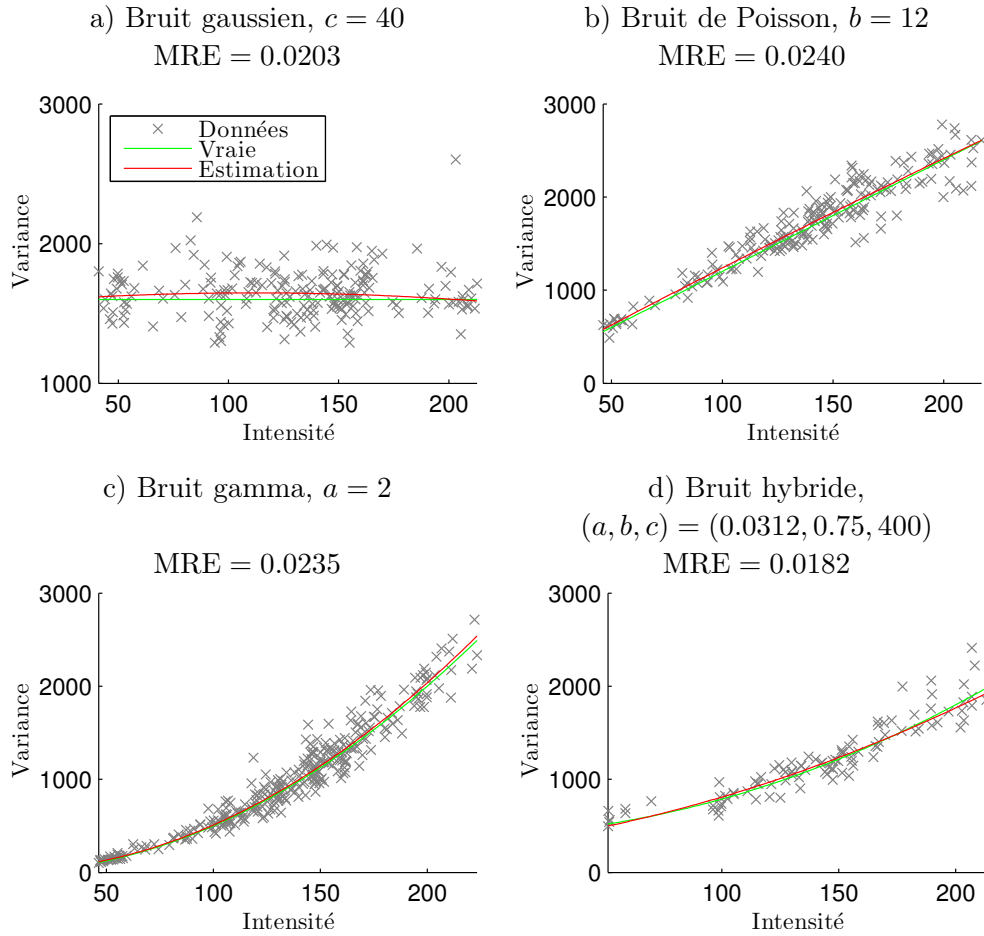


FIGURE 3.8 – Estimation des paramètres pour différents types de bruit avec l’estimateur LAD. Le modèle de bruit n’est pas supposé connu donc la NLF est estimée directement sous forme d’un polynôme du second ordre.

du modèle de bruit. Une procédure similaire a été utilisée pour régler la taille de bloc N_b . Ainsi, dans toutes les expériences, les paramètres $P_D = 0.6$ et $N_b = 16 \times 16$ ont été utilisés.

3.3.2.4 Estimation des paramètres

La figure 3.8 illustre l’estimation de la NLF pour différents modèles de bruit (gaussien, poissonien, multiplicatif ou hybride). Les paires (moyenne, variance) calculées sur les zones homogènes sont représentées sous forme de croix grises, la NLF vraie est représentée en vert et la NLF estimée à l’aide de l’estimateur des moindres déviations en rouge. Aucune hypothèse supplémentaire n’a été faite sur la nature du bruit, une NLF polynomiale d’ordre deux a été estimée dans chacun des cas. Les courbes offrent une belle approximation, même sans connaissance préalable du modèle. L’injection d’une telle hypothèse supplémentaire offrirait une approximation

TABLE 3.1 – PSNR moyens obtenus après débruitage d’un ensemble d’images corrompues par du bruit hybride. Les méthodes de débruitage utilisées sont l’algorithme des NL-means classique qui suppose que le bruit est blanc gaussien, les NLF-means se basant sur la NLF estimée et les NLF-means utilisant les vrais paramètres de la NLF.

Initial PSNR	NL-means (gaussien)	NLF-means (NLF estimée)	NLF-means (vraie NLF)
19.52	26.26	28.85	28.96
18.87	25.88	28.48	28.58
18.04	26.24	27.93	28.09
17.51	25.86	27.73	27.87

encore meilleure.

3.4 Résultats sur images synthétiques

3.4.1 Application au débruitage d’images

L’estimation automatique du bruit dans une image peut se montrer utile pour beaucoup d’applications qui nécessitent de connaître le niveau de bruit, notamment pour le débruitage d’images, qui fait l’objet de la seconde partie de cette thèse. Afin d’illustrer l’intérêt de la méthode d’estimation proposée, la NLF estimée est ici réinjectée pour permettre un débruitage «à l’aveugle» de l’image.

Pour cela, l’algorithme des moyennes non locales (NL-means) [Buades 2005b] a été adapté afin de ne prendre en compte que la connaissance des paramètres $(\widehat{a}, \widehat{b}, \widehat{c})$ fournis par la NLF estimée, donnant naissance à l’algorithme baptisé NLF-means. Une étude plus approfondie des NL-means ainsi que la description de l’algorithme des NLF-means pour traiter du bruit à variance polynomiale sont fournies dans la partie 4.4.

L’application de l’estimation de bruit au débruitage aveugle des images permet d’illustrer à la fois la performance de l’estimation ainsi que la nécessité de connaître les statistiques du bruit pour réaliser un débruitage adapté, ou d’autres tâches de traitement d’images. Le tableau 3.1 fournit des PSNR obtenus après débruitage à l’aveugle d’un jeu d’images corrompues par un bruit hybride, à quatre niveaux différents, dont le PSNR initial moyen est donné en première colonne. La figure 3.9 montre des résultats de débruitage de *Lena*, corrompue par du bruit hybride de paramètres $(a, b, c) = (0.0312, 1.875, 100)$, avec un PSNR initial de 18.70dB. Dans un premier temps, le bruit est (classiquement) supposé additif, blanc, gaussien. On estime donc seulement la variance σ^2 du bruit à l’aide de la méthode d’estimation décrite ci-dessus, puis on applique l’algorithme des NL-means classique, conçu pour traiter du bruit gaussien. Les résultats de PSNR sont représentés dans la deuxième colonne du tableau, et une illustration d’un tel débruitage est représenté sur la figure 3.9-a. Le caractère inapproprié du bruit conduit à des artefacts importants

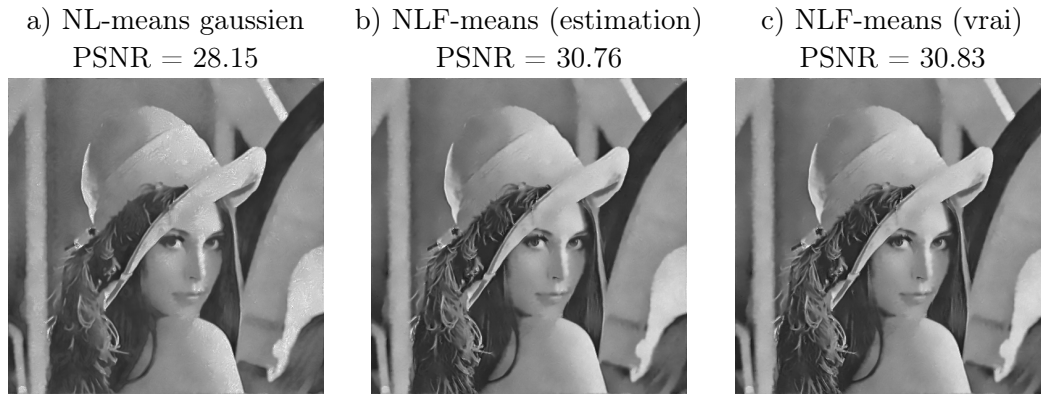


FIGURE 3.9 – Débruitage d’un bruit hybride de paramètres $(a, b, c) = (0.0312, 1.875, 100)$, avec un PSNR initial PSNR = 18.70dB. L’image bruitée est représentée sur la figure 3.5-a. a) NL-means classique qui présument et estiment un bruit blanc gaussien, b) NLF-means basé sur la NLF estimée et c) NLF-means basé sur la vraie NLF.

sur l’image débruitée : on note des artefacts sur les zones uniformes comme la peau, et une perte de détails sur les plumes notamment. On utilise ensuite l’algorithme des NLF-means adapté au bruit dépendant du signal. Dans la troisième colonne du tableau et sur la figure 3.9-b, les paramètres de bruit de la NLF sont estimés à l’aide de la méthode développée ici puis réinjectés dans l’algorithme des NLF-means. Les résultats sont comparés à l’algorithme des NLF-means pourvu des vrais paramètres (a, b, c) de la NLF, quatrième colonne du tableau et figure 3.9-c. Les deux débruitages issus des NLF-means offrent un résultat satisfaisant à la fois visuellement et en terme de PSNR, ce qui permet de considérer que le modèle de débruitage proposé par les NLF-means est adapté, et que l’estimation de la NLF est suffisamment précise pour offrir un débruitage adapté.

3.4.2 Comparaison à l’état de l’art

À notre connaissance, peu de méthodes permettent de gérer un cadre aussi général de modèle de bruit que celui étudié ici. Dans cette section, l’algorithme d’estimation est comparé aux estimateurs de bruit décorrélé spatialement : l’estimateur MAD [Donoho 1995] qui estime dans le domaine des ondelettes la variance d’un bruit blanc additif gaussien, et l’algorithme de Foi *et al.* [Foi 2008] qui permet d’estimer du bruit Poisson-gaussien. L’estimation est également comparée à la méthode PCA [Colom 2014a] qui s’appuie sur les travaux de [Pyatykh 2013], à la méthode des *Percentiles* [Colom 2013] qui étend l’estimation de [Ponomarenko 2010] au cas de bruit dépendant du signal, et à l’algorithme *Noise Clinic* [Lebrun 2014, Lebrun 2015] qui associe l’estimation de bruit dépendant du signal et corrélé spatialement [Colom 2014b] à un débruitage aveugle basé sur les NL-Bayes [Lebrun 2013].

TABLE 3.2 – Erreur moyenne relative (MRE) pour l’estimation de bruit additif gaussien à l’aide de l’estimateur MAD [Donoho 1995], du modèle Poisson-gaussien [Foi 2008], des extensions des méthodes PCA [Colom 2014a] et des percentiles [Colom 2013], de Noise Clinic [Colom 2014b, Lebrun 2014], ou de la méthode proposée qui peut estimer soit un modèle gaussien soit un modèle général du second ordre, et PSNR moyen obtenu en réinjectant les paramètres estimés dans l’algorithme des NLF-means.

Estimateur	$ \hat{a} $	$ \hat{b} $	$\frac{ c-\hat{c} }{c}$	MRE	PSNR
MAD [Donoho 1995]	\	\	0.08	0.081	26.33
Foi <i>et al.</i> [Foi 2008]	\	0.55	0.05	0.040	26.41
PCA [Colom 2014a]	\	\	0.51	0.513	24.12
Percentile [Colom 2013]	\	\	0.11	0.105	25.69
Noise Clinic [Colom 2014b]	\	\	0.39	0.389	23.69
[Lebrun 2014]					(24.29)
Proposé, gaussien	\	\	0.03	0.030	26.72
Proposé, hybride	0.00	1.56	0.12	0.056	26.64

TABLE 3.3 – Erreur moyenne relative (MRE) pour l’estimation de bruit Poisson-gaussien à l’aide du modèle Poisson-gaussien [Foi 2008], des extensions des méthodes PCA [Colom 2014a] et des percentiles [Colom 2013], de Noise Clinic [Colom 2014b, Lebrun 2014], ou de la méthode proposée qui peut estimer soit un modèle affine soit un modèle général du second ordre, et PSNR moyen obtenu en réinjectant les paramètres estimés dans l’algorithme des NLF-means.

Estimateur	$ \hat{a} $	$\frac{ b-\hat{b} }{b}$	$\frac{ c-\hat{c} }{c}$	MRE	PSNR
Foi <i>et al.</i> [Foi 2008]	\	0.07	0.37	0.053	27.00
PCA [Colom 2014a]	\	0.46	6.99	1.027	24.97
Percentile [Colom 2013]	\	0.25	1.56	0.202	26.80
Noise Clinic [Colom 2014b]	\	0.41	0.69	0.353	25.30
[Lebrun 2015]					(25.85)
Proposé, affine	\	0.09	0.56	0.063	27.04
Proposé, hybride	0.01	0.13	0.65	0.064	26.87

À partir d’une banque de 150 images naturelles¹ sont générés des jeux d’images bruitées, corrompues par du bruit gaussien puis par du bruit Poisson-gaussien, et les différents estimateurs adaptés sont comparés. Grâce à la connaissance des vrais paramètres de la NLF, l’erreur relative MRE définie dans l’équation (3.28) est calculée pour chaque estimateur. Les résultats sont synthétisés dans les tableaux 3.2 et 3.3, pour du bruit gaussien puis Poisson-gaussien. Sont représentées également les erreurs relatives d’estimation de chaque paramètre individuellement. Conformément à l’idée de la section précédente, les paramètres estimés de la NLF sont ensuite ré-

1. <http://www.gipsa-lab.grenoble-inp.fr/~laurent.condat/imagebase.html>

TABLE 3.4 – Calibration du dispositif de la salle noire et estimation des paramètres du bruit pour chaque niveau de luminosité.

Diamètre (mm)	Éclairement (mlux)	Niveau de nuit	Paramètres estimés (a, b, c)
1	0.3	5	(-0.10, 34.53, -5.03)
2	1.3	4	(-0.06, 24.69, 6.64)
2.6	2.3	3	(-0.04, 17.88, 5.75)
3.2	3	3	(-0.03, 12.52, 8.14)
5.3	9.74	3	(-0.01, 5.32, 7.30)
8.9	28	2	(-0.01, 2.49, 3.06)
15	79	1	(0, 1.02, 0.09)
19.3	125	1	(0, 0.82, 0.26)

injectés dans l’algorithme des NLF-means pour évaluer la capacité de l’estimation à fournir un débruitage adapté. Les PSNR obtenus pour chaque méthode sont fournis en dernière colonne.

L’estimateur MAD [Donoho 1995] n’est adapté que pour estimer du bruit gaussien. La méthode de Foi *et al.* [Foi 2008] permet d’estimer un bruit Poisson-gaussien, donc elle peut s’appliquer dans les deux cas envisagés ici, en revanche elle n’est pas adaptée à l’estimation plus générale de bruit à variance polynomiale. Les trois méthodes de Colom *et al.* permettent d’estimer un bruit dépendant du signal. Les méthodes PCA [Colom 2014a] et *Percentile* [Colom 2013] estiment des points de la fonction de niveau de bruit ; cette dernière est ensuite interpolée par l’estimateur des *Least absolute deviations*, puis injectée dans l’algorithme des NLF-means. L’algorithme *Noise Clinic* [Lebrun 2015] permet de même d’estimer des points de la fonction de niveau de bruit, exploités par LAD et les NLF-means comme précédemment. L’estimation peut également s’accompagner d’un débruitage par l’algorithme des NL-Bayes adapté au bruit dépendant du signal. Le PSNR obtenu est indiqué sur la seconde ligne associée à *Noise Clinic* des tableaux 3.2 et 3.3.

La méthode proposée s’avère compétitive ; elle égale ou surpasse l’état de l’art des méthodes d’estimation, dans le cas particulier des bruits gaussien ou Poisson-gaussien.

3.5 Application aux images BNL

3.5.1 Sur un banc d’images test

Afin de calibrer le bruit des caméras utilisées par Thales en fonction des niveaux de nuit, il est possible d’effectuer des acquisitions d’images en salle noire, dans laquelle le niveau de luminosité est calibré. Nous disposons pour cela d’un dispositif qui émet de la lumière, sur lequel on applique un cache avec un diaphragme afin de ne laisser passer qu’une quantité de lumière donnée. Ainsi, pour chaque diamètre de

diaphragme est associée une luminosité, et un niveau de nuit. Le tableau 3.4 fournit les données associées à ce dispositif.

Les mesures sont ensuite effectuées à l'aide d'une mire dite de contraste, qui présente des bandes d'intensité différentes afin d'évaluer la puissance du bruit pour toutes les intensités. Des séquences d'images de cette mire sont acquises avec le dispositif IL à tester pour chaque luminosité. Les zones homogènes sont alors détectées non pas sur une image, mais sur une séquence d'images, afin de bénéficier d'un maximum de détections. Il est important de noter que les images issues de ce dispositif sont très fortement corrélées spatialement, ce qui ne nous permet pas de les traiter directement avec le modèle de détection présenté ci-dessus. Les images ont donc été sous-échantillonnées d'un facteur 4, afin de réduire les dépendances entre pixels voisins. Les zones détectées à différents niveaux de nuit sont représentées sur la figure 3.10. Puis l'estimation des paramètres du bruit est obtenue à l'aide de la minimisation ℓ_1 , sans contrainte de positivité sur les paramètres (a, b, c) . Les courbes obtenues ainsi que les paramètres associés pour différents niveaux de nuit sont représentées sur la figure 3.11. L'ensemble des estimations est fourni dans le tableau 3.4. On observe tout d'abord qu'à très faible niveau de luminosité (nuit 5 et 4), le bruit est fortement non gaussien, ni même poissonien. Il possède une composante d'ordre deux caractérisée par un coefficient a négatif, que l'on peut associer au gain de la galette MCP (voir la description du dispositif dans la section 2.3). Lorsque la luminosité augmente, pour les niveaux de nuit 3 et 1 par exemple, le bruit semble tendre vers un modèle Poisson-gaussien, qui correspondrait au bruit de lecture et au *shot noise*, prépondérants cette fois devant les autres composantes en raison d'un gain plus faible.

Ces mesures permettent de confirmer les hypothèses émises à partir de la description du dispositif d'acquisition (voir section 2.3). En revanche, elles sont associées à un dispositif donné (un modèle de caméra IL, une caméra CCD et un système de couplage spécifiques) et dans les conditions particulières de la salle noire, ce qui ne permet pas d'exploiter le modèle de bruit tel quel pour des données de vol. Il serait donc intéressant d'évaluer la nature du bruit directement sur des données réelles.

3.5.2 Sur des données de vol

Les mêmes expériences peuvent donc être réalisées sur des séquences d'images IL issues de données de vol. Ces dernières présentent des caractéristiques de la vision nocturne telles que le très faible contraste, la forte dynamique et la saturation des sources lumineuses. Sur la première colonne de la figure 3.12 sont représentées de telles images ; elles présentent en fonction des conditions d'acquisition différents niveaux de bruit et différents artefacts. Ces défauts ne sont pas reflétés sur les images issues du banc de test, pourtant ce sont des dégradations typiques des systèmes BNL que l'on ne peut pas négliger en pratique. D'où l'importance de réaliser l'étude sur données réelles.

Sur la deuxième colonne de la figure 3.12 figure l'estimation des paramètres du bruit. On observe des comportements différents en fonctions des images et des

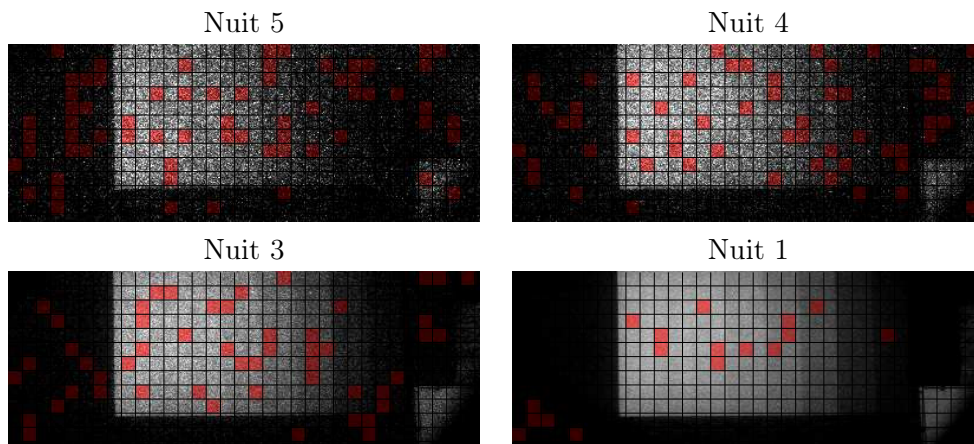


FIGURE 3.10 – Détection des zones homogènes pour des images d’une mire de contraste issues d’un banc de test en salle noire. Les mesures sont effectuées à différents niveaux de nuit, correspondant aux valeurs en gras dans le tableau 3.4.

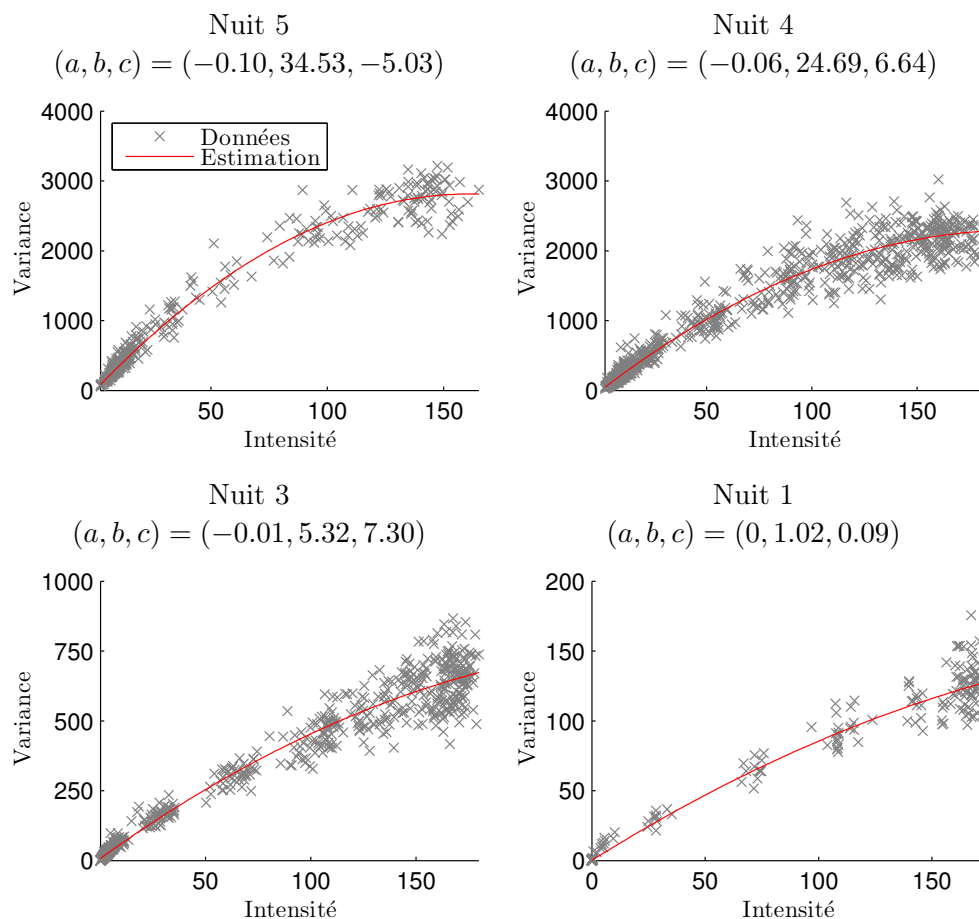


FIGURE 3.11 – Estimation de la NLF pour des images d’une mire de contraste issues d’un banc de test en salle noire, à différents niveaux de nuit.

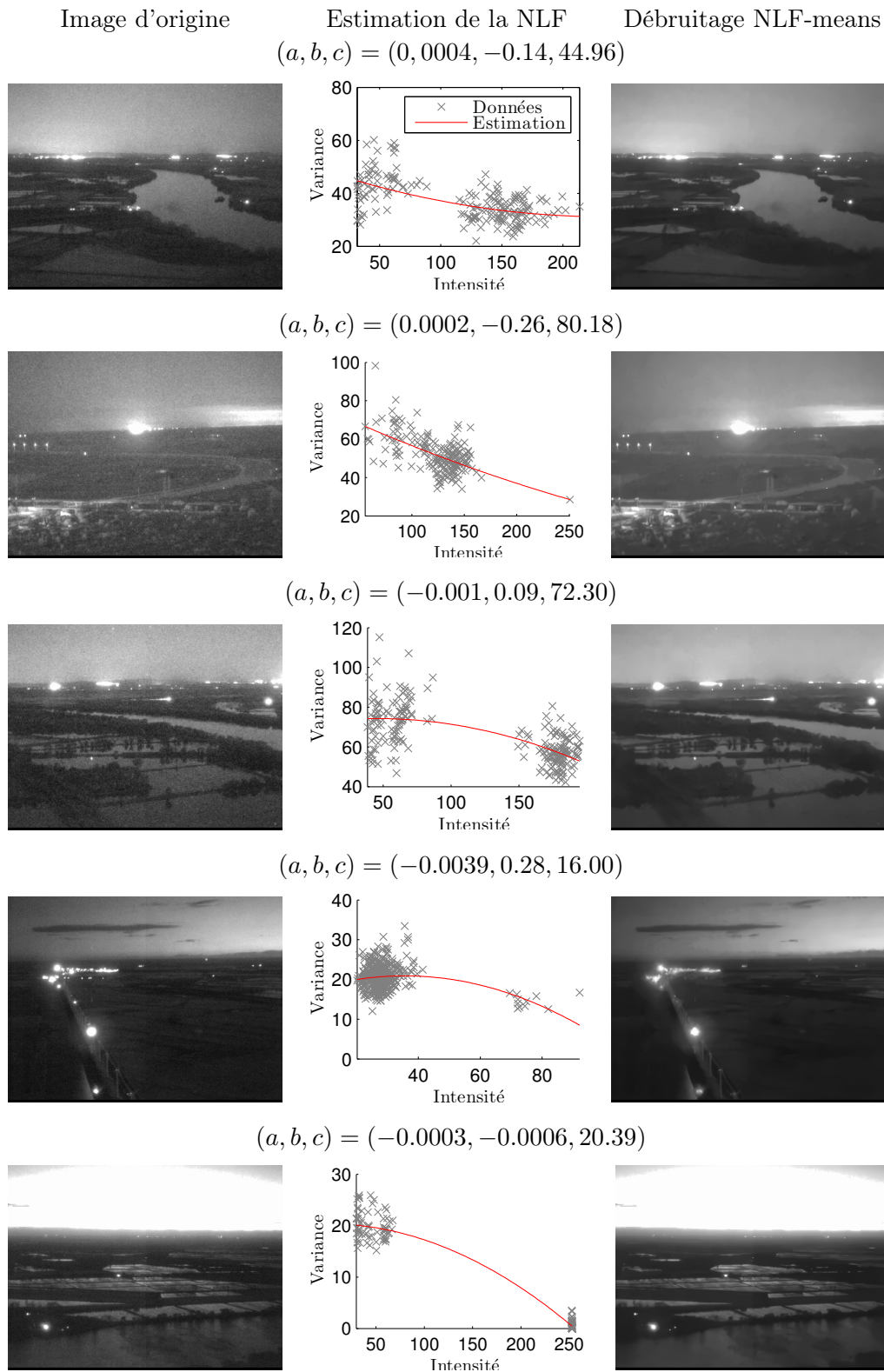


FIGURE 3.12 – Estimation de la NLF sur images réelles et débruitage NLF.

perturbations observées, en particulier la saturation des zones claires introduit une décroissance de la variance du bruit avec l'intensité peu habituelle. Faute de vérité terrain pour valider les estimations, les paramètres de bruit ont été ensuite réinjectés dans le modèle de débruitage NLF-means, afin de tester la pertinence de l'estimation. La colonne de droite de la figure 3.12 présente les résultats de débruitage obtenus. On observe une réduction notable du bruit, et on note la capacité du modèle à traiter différemment les zones claires des zones sombres. Cela permet en particulier de respecter le contraste et les structures fines, notamment les routes et les poteaux (sur la deuxième ligne par exemple).

L'estimation du modèle de bruit s'est appuyée comme pour les images du banc de test sur les séquences d'images afin de bénéficier d'un maximum de zones homogènes. Il serait judicieux d'utiliser la redondance temporelle de l'information pour le débruitage également, c'est-à-dire de débruiter non pas les images une par une mais le flux vidéo. Ce sera l'objet de la partie suivante, dans laquelle sera développée une méthode de débruitage adaptative basée sur les moyennes non locales, ainsi qu'une adaptation à la restauration de séquences d'images.

3.6 Conclusion

Une méthode automatique d'estimation du bruit dans une image a été développée à partir de la détection des régions homogènes. Ces zones homogènes sont sélectionnées à l'aide du coefficient de corrélation de rang de Kendall, évalué sur des séquences de voisinages de pixels. Ce test est non paramétrique et requiert peu d'hypothèses sur la distribution du bruit. En particulier, les taux d'erreur de détection (fausse alarme et fausse détection) sont contrôlés par le rapport signal à bruit et le nombre d'échantillons disponibles. Dans un second temps, la fonction de niveau de bruit est assimilée à un polynôme de degré deux, puis elle est estimée à l'aide de la méthode robuste des moindres déviations. Des résultats à partir d'images synthétiques permettent de fournir une estimation robuste avec peu d'hypothèses préalables sur le modèle de bruit. L'estimation de la NLF peut ensuite être utilisée pour d'autres applications de traitement d'image qui requièrent la connaissance du niveau de bruit. C'est illustré ici pour le débruitage aveugle d'images à l'aide de l'adaptation de l'algorithme des NL-means au bruit dépendant du signal caractérisé seulement par les paramètres de la NLF.

Beaucoup d'extensions de ce modèle sont envisageables. Dans un premier temps, la taille des zones homogènes pourrait être optimisée. Dans les applications présentées dans ce chapitre, la taille des blocs a été choisie empiriquement; de façon analogue au choix de la probabilité de détection, elle établit un compromis entre la volonté de détecter assez de zones, et la nécessité d'avoir suffisamment d'échantillons dans une région pour diminuer l'erreur sur le calcul des statistiques (moyenne, variance). Il serait donc intéressant d'adapter localement la taille voire la forme des blocs au contenu de l'image. De plus, l'estimation du bruit dans les images BNL pourrait utiliser le flux vidéo. Considérer des blocs temporels permet-

trait d'exploiter la redondance temporelle, et fournirait plus de régions homogènes, ce qui améliorerait l'estimation encore davantage.

Un modèle plus général encore que le second ordre serait à étudier. Il serait intéressant tout d'abord de gérer le bruit non stationnaire, pour lequel la fonction de niveau de bruit varierait spatialement. En effet dans beaucoup de dispositifs d'imagerie le niveau de bruit varie en fonction de la distance au capteur par exemple, ce qui nécessite d'adapter spatialement la NLF. Le bruit corrélé spatialement est une autre extension à envisager. Le détecteur de Kendall ne gère pas la corrélation spatiale, qui pourtant se manifeste à travers beaucoup de corruptions naturelles d'images. En outre la prise en compte de la corrélation spatiale étendrait l'estimation de bruit au problème plus général de détection d'artefacts, notamment les artefacts de compression. Au delà de l'estimation de bruit, l'approche développée ici mériterait en fait de s'étendre à d'autres types de corruption, comme la détection de flou par exemple.

Grâce au caractère automatique de la méthode, nous avons pu appliquer l'estimation du bruit aux images BNL acquises en vol. Sous peu d'hypothèses sur la nature du bruit présent sur les images IL, nous avons pu estimer les paramètres de la fonction de niveau de bruit. Munis de ces paramètres, nous pouvons désormais envisager un débruitage adapté de ces images BNL.

3.7 Annexes du chapitre : preuves

3.7.1 Preuve du théorème 3.2.5

Preuve. Le théorème établit la relation entre le niveau de détection α et la probabilité de fausse alarme. En fait, le niveau de détection α est déterminé pour garantir la probabilité de fausse alarme P_{FA} souhaitée, associée à la condition (C1). En effet, en utilisant la définition de la probabilité de fausse alarme (C1) ainsi que le corollaire 3.2.4, on a :

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(|z(X, Y)| > \alpha \mid H_0) &= \lim_{n \rightarrow \infty} \mathbb{P}(z(X, Y) > \alpha \mid H_0) + \mathbb{P}(-z(X, Y) < -\alpha \mid H_0) \\ &= (1 - \phi(\alpha)) + \phi(-\alpha) = 2(1 - \phi(\alpha)). \end{aligned} \quad (3.29)$$

En utilisant désormais le choix de α donné par $\alpha = \phi^{-1}(1 - P_{\text{FA}}/2)$, on obtient :

$$2(1 - \phi(\alpha)) = 2(1 - \phi(\phi^{-1}(1 - P_{\text{FA}}/2))) = P_{\text{FA}}. \quad (3.30)$$

Cela conclut la preuve. ■

3.7.2 Preuve de la proposition 3.2.6

Les démonstrations suivantes visent à évaluer l'espérance et la variance du coefficient de Kendall sous l'hypothèse alternative H_1 .

L'hypothèse H_1 suppose que le bruit est de nature gaussienne, donc les cas d'égalités ne sont pas à prendre en considération. On peut donc utiliser la définition

simplifiée 2 du coefficient de Kendall. De plus, afin de simplifier les notations, on néglige la dépendance en (X, Y) dans la notation de τ ou z . On néglige également le conditionnement à l'hypothèse H_1 dans les expressions $\mathbb{E}[\tau|H_1]$ et $\text{Var}[\tau|H_1]$, dans la mesure où tous les résultats démontrés ci-dessous sont développés dans le cas de l'hypothèse alternative H_1 .

Résultats préliminaires

Cette section fournit quelques résultats préliminaires nécessaires au calcul de l'espérance et de la variance de τ . Soit ε une variable aléatoire à densité φ . Sauf mentionné différemment, on ne suppose pas de distribution a priori pour ε . On note $\varepsilon_i, \varepsilon_j$ des réalisations de la variable aléatoire ε , et Φ la fonction de répartition (cdf) de ε :

$$\Phi(x) = \int_{-\infty}^x \varphi(\varepsilon) d\varepsilon. \quad (3.31)$$

Proposition 3.7.1 Soient ε_i et ε_j deux réalisations indépendantes de ε . Alors,

$$\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j)] = 0. \quad (3.32)$$

Preuve. En utilisant la définition de l'espérance ainsi que l'indépendance entre les deux variables aléatoires, on a :

$$\begin{aligned} & \mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j)] \\ &= \iint_{\mathbb{R}^2} \text{sign}(\varepsilon_i - \varepsilon_j) \varphi(\varepsilon_i) \varphi(\varepsilon_j) d\varepsilon_i d\varepsilon_j \\ &= \int_{\varepsilon_i \in \mathbb{R}} \left(\int_{\varepsilon_j = -\infty}^{\varepsilon_i} \varphi(\varepsilon_j) d\varepsilon_j - \int_{\varepsilon_j = \varepsilon_i}^{+\infty} \varphi(\varepsilon_j) d\varepsilon_j \right) \varphi(\varepsilon_i) d\varepsilon_i \\ &= \int_{\varepsilon_i \in \mathbb{R}} \left(\int_{\varepsilon_j = -\infty}^{\varepsilon_i} \varphi(\varepsilon_j) d\varepsilon_j - \int_{\varepsilon_j = \varepsilon_i}^{+\infty} \varphi(\varepsilon_j) d\varepsilon_j \right) \varphi(\varepsilon_i) d\varepsilon_i \\ &= \int_{\varepsilon_i \in \mathbb{R}} (\Phi(\varepsilon_i) - (1 - \Phi(\varepsilon_i))) \varphi(\varepsilon_i) d\varepsilon_i \\ &= \int_{\varepsilon_i \in \mathbb{R}} (2\Phi(\varepsilon_i) - 1) \varphi(\varepsilon_i) d\varepsilon_i = \left[\frac{1}{4} (2\Phi(\varepsilon_i) - 1)^2 \right]_{-\infty}^{+\infty} = \frac{1}{4} - \frac{1}{4} = 0. \end{aligned}$$

■

Proposition 3.7.2 Soient $\varepsilon_i, \varepsilon_j$ et ε_k des réalisations indépendantes de ε . Alors,

$$\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k)] = \frac{1}{3}. \quad (3.33)$$

Preuve. De façon analogue à la démonstration de la proposition 3.7.1 et en utilisant l'indépendance des trois variables aléatoires, on a :

$$\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k)]$$

$$\begin{aligned}
&= \iiint_{\mathbb{R}^3} \text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k) \varphi(\varepsilon_i) \varphi(\varepsilon_j) \varphi(\varepsilon_k) \, d\varepsilon_i \, d\varepsilon_j \, d\varepsilon_k \\
&= \int_{\varepsilon_i \in \mathbb{R}} \left(\int_{\varepsilon_j \in \mathbb{R}} \text{sign}(\varepsilon_i - \varepsilon_j) \varphi(\varepsilon_j) \, d\varepsilon_j \int_{\varepsilon_k \in \mathbb{R}} \text{sign}(\varepsilon_i - \varepsilon_k) \varphi(\varepsilon_k) \, d\varepsilon_k \right) \varphi(\varepsilon_i) \, d\varepsilon_i \\
&= \int_{\varepsilon_i \in \mathbb{R}} (2\Phi(\varepsilon_i) - 1)^2 \varphi(\varepsilon_i) \, d\varepsilon_i = \left[\frac{1}{6} (2\Phi(\varepsilon_i) - 1)^3 \right]_{-\infty}^{+\infty} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.
\end{aligned}$$

■

Proposition 3.7.3 Soient ε_i et ε_j deux réalisations indépendantes de ε et $a \in \mathbb{R}$.

$$\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j - a)] = 2 \int_{\mathbb{R}} \Phi(\varepsilon_i - a) \varphi(\varepsilon_i) \, d\varepsilon_i - 1. \quad (3.34)$$

de plus, si ε suit une loi normale $\mathcal{N}(0, \sigma^2)$, alors :

$$\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j - a)] = 2\phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right) - 1 = 1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right). \quad (3.35)$$

Preuve. Dans le cas général lorsqu'aucune distribution particulière n'est supposée pour ε , on utilise le même cheminement que pour les propositions précédentes.

$$\begin{aligned}
\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j - a)] &= \iint_{\mathbb{R}^2} \text{sign}(\varepsilon_i - \varepsilon_j - a) \varphi(\varepsilon_i) \varphi(\varepsilon_j) \, d\varepsilon_i \, d\varepsilon_j \\
&= \int_{\varepsilon_i \in \mathbb{R}} \left(\int_{\varepsilon_j = -\infty}^{\varepsilon_i - a} \varphi(\varepsilon_j) \, d\varepsilon_j - \int_{\varepsilon_j = \varepsilon_i - a}^{+\infty} \varphi(\varepsilon_j) \, d\varepsilon_j \right) \varphi(\varepsilon_i) \, d\varepsilon_i \\
&= \int_{\mathbb{R}} (2\Phi(\varepsilon_i - a) - 1) \varphi(\varepsilon_i) \, d\varepsilon_i = 2 \int_{\mathbb{R}} \Phi(\varepsilon_i - a) \varphi(\varepsilon_i) \, d\varepsilon_i - 1.
\end{aligned}$$

Si $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, on utilise le fait que $\varepsilon_i - \varepsilon_j = \mu_{ij}$ suit une loi normale $\mathcal{N}(0, 2\sigma^2)$, de densité $\varphi_{2\sigma^2}$, et

$$\begin{aligned}
\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j - a)] &= \mathbb{E}[\text{sign}(\mu_{ij} - a)] = \int_{\mathbb{R}} \text{sign}(\mu - a) \varphi_{2\sigma^2}(\mu) \, d\mu \\
&= \int_{\mathbb{R}} \text{sign}(\sqrt{2\sigma^2}t - a) \varphi(t) \, dt, \text{ avec } t = \frac{\mu}{\sqrt{2\sigma^2}} \\
&= - \int_{t=-\infty}^{a/\sqrt{2\sigma^2}} \varphi(t) \, dt + \int_{t=a/\sqrt{2\sigma^2}}^{+\infty} \varphi(t) \, dt \\
&= 1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) = 2\phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right) - 1.
\end{aligned}$$

■

Ce dernier résultat est confirmé par la proposition suivante :

Proposition 3.7.4 Soit $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ une variable aléatoire gaussienne. Pour tout $a \in \mathbb{R}$, on a :

$$\int_{\varepsilon} \phi(\varepsilon - a) \varphi(\varepsilon) \, d\varepsilon = \phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right). \quad (3.36)$$

Preuve. On utilise d'une part une relation fondamentale entre la fonction de répartition de la loi normale centrée réduite ϕ et sa fonction d'erreur.

$$\begin{aligned}\phi(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{\pi}} \int_{-x/\sqrt{2}}^{x/\sqrt{2}} e^{u^2} du \\ &\quad \text{en effectuant le changement de variable } u = t/\sqrt{2} \\ &= \phi(-x) + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right).\end{aligned}$$

Ainsi, en utilisant que $\phi(-x) = 1 - \phi_0(x)$, on a :

$$\phi(x) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right). \quad (3.37)$$

et dans le cas d'une loi normale de variance σ^2 , on obtient :

$$\phi_\sigma(x) = \phi\left(\frac{x}{\sigma}\right) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x}{\sigma\sqrt{2}}\right) \right). \quad (3.38)$$

Avec ces notations, la proposition 3.7.4 peut se redémontrer dans le cas général à partir du lemme suivant (qui sera démontré ultérieurement) :

Lemme 3.7.5 *Pour tout $a, b, c, d \in \mathbb{R}$,*

$$\operatorname{erf}\left(\frac{b-ac}{\sqrt{1+2a^2d^2}}\right) = \int_{\mathbb{R}} \frac{\operatorname{erf}(ax+b)}{\sqrt{2\pi d^2}} \exp\left(-\frac{(x+c)^2}{2d^2}\right) dx \quad (3.39)$$

Ainsi :

$$\begin{aligned}\int_{\mathbb{R}} \phi(x-a)\varphi(x)dx &= \int_{\mathbb{R}} \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x-a}{\sigma\sqrt{2}}\right) \right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &= \frac{1}{2} \left(1 + \int_{\mathbb{R}} \operatorname{erf}\left(\frac{x-a}{\sigma\sqrt{2}}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \right),\end{aligned}$$

puis en appliquant le lemme 3.7.5 avec $d = \sigma, c = 0, b = -\frac{a}{\sigma\sqrt{2}}$ et $a = \frac{1}{\sigma\sqrt{2}}$:

$$\begin{aligned}\int_{\mathbb{R}} \phi(x-a)\varphi(x)dx &= \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{-a/\sigma\sqrt{2}}{\sqrt{1+\frac{2\sigma^2}{2\sigma^2}}}\right) \right) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{-a}{2\sigma}\right) \right) \\ &= \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{-a}{\sqrt{2} \times \sqrt{2}\sigma}\right) \right) = \phi\left(\frac{-a}{\sqrt{2}\sigma}\right).\end{aligned}$$

Reste à montrer le lemme. Afin de simplifier les notations, nous allons démontrer le lemme dans le cas (qui nous concerne) où $c = 0$. On utilise pour cela la dérivée de la fonction erf donnée par

$$\frac{d}{dx} \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}. \quad (3.40)$$

Soit $I(a, b, d) = \int_{\mathbb{R}} \frac{\operatorname{erf}(ax+b)}{\sqrt{2\pi d^2}} \exp\left(-\frac{(x+c)^2}{2d^2}\right) dx$. On remarque tout d'abord que $I(a, 0, d) = 0$, car la fonction erf est impaire.

On peut calculer la dérivée de I par rapport à b :

$$\begin{aligned}
I'(b) &= \frac{d}{db} I(a, b, d) = \frac{2}{\pi d} \int_{\mathbb{R}} e^{-(ax+b)^2} e^{-x^2/2d^2} dx \\
&= \frac{2}{\pi d} \int_{\mathbb{R}} e^{-a^2x^2 - 2abx - b^2 - x^2/2d^2} dx = \frac{2}{\pi d} \int_{\mathbb{R}} e^{-\frac{(1+2a^2d^2)}{2d^2}x^2 - 2abx - b^2} dx \\
&= \frac{2}{\pi d} \int_{\mathbb{R}} e^{-\frac{(1+2a^2d^2)}{2d^2}\left(x + \frac{2abd^2}{1+2a^2d^2}\right)^2} e^{\frac{2abd^2}{1+2a^2d^2} - b^2} dx \\
&= e^{\frac{-b^2}{1+2a^2d^2}} \times \frac{d}{\sqrt{1+2a^2d^2}} \times \frac{2}{\sqrt{\pi}} \times \underbrace{\frac{1}{\frac{d}{\sqrt{1+2a^2d^2}}\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{\left(x + \frac{2abd^2}{1+2a^2d^2}\right)^2}{2\left(\frac{d}{\sqrt{1+2a^2d^2}}\right)^2}} dx}_{=1} \\
&= \frac{2}{\sqrt{\pi}} \frac{d}{\sqrt{1+2a^2d^2}} e^{\frac{-b^2}{1+2a^2d^2}}.
\end{aligned}$$

Ainsi, en réintégrant la dérivée de I par rapport à b , on a :

$$\begin{aligned}
I(a, b, d) &= \int_0^b I'(u) du = \int_0^b \frac{2}{\sqrt{\pi}} \frac{d}{\sqrt{1+2a^2d^2}} e^{\frac{-u^2}{1+2a^2d^2}} du \\
&= \operatorname{erf}\left(\frac{b}{\sqrt{1+2a^2d^2}}\right).
\end{aligned}$$

Ceci clôt la démonstration du lemme 3.7.5 et de la proposition 3.7.4. Les quatre propositions qui ont été démontrées ici seront utilisées dans les preuves suivantes.

■

Preuve. [Retour à la preuve de la proposition 3.2.6]

À partir de la définition de τ donnée dans Déf. 2, et en utilisant l'expression de X et Y associée à l'hypothèse H_1 donnée dans (4), il est possible de distinguer les cas en fonction de la position des indices i et j par rapport au saut dans le signal :

$$\begin{aligned}
\tau &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \operatorname{sign}(X_i - X_j) \operatorname{sign}(Y_i - Y_j) \\
&= \frac{1}{n(n-1)} \left[\sum_{\substack{i, j \leq n/2 \\ \text{ou } i, j > n/2 \\ i \neq j}} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j) \operatorname{sign}(\eta_i - \eta_j)}_{(*)} \right. \\
&\quad + \sum_{\substack{i \leq n/2, \\ j > n/2}} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j - a) \operatorname{sign}(\eta_i - \eta_j - a)}_{(**)} \\
&\quad \left. + \sum_{\substack{i > n/2, \\ j \leq n/2}} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j + a) \operatorname{sign}(\eta_i - \eta_j + a)}_{(***)} \right].
\end{aligned}$$

Chacune de ces trois sommes peut être évaluée séparément à l'aide des propositions démontrées dans la première partie de la preuve.

$\mathbb{E}[(*)] = 0$ grâce à la proposition 3.7.1 et à l'indépendance de ε_i et ε_j .

$\mathbb{E}[(**)] = \mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j - a)] \times \mathbb{E}[\text{sign}(\eta_i - \eta_j - a)] = \left(1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)(a)\right)^2$, en utilisant l'indépendance entre les deux variables ε, η ainsi que la proposition 3.7.3. De plus, la somme impliquant $(**)$ correspond aux cas où i et j sont dans les parties distinctes du signal, donc elle contient $\frac{n}{2} \times \frac{n}{2}$ termes.

$\mathbb{E}[(***)] = \mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j + a)] \times \mathbb{E}[\text{sign}(\eta_i - \eta_j + a)] = \left(2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)(a) - 1\right)^2$, grâce à la proposition 3.7.3 et à l'indépendance des variables. Comme pour $(**)$, $(***)$ intervient quand i et j sont dans les deux moitiés distinctes du signal, donc la somme contient $\frac{n}{2} \times \frac{n}{2}$ termes.

En combinant les trois résultats, on obtient alors que

$$\begin{aligned} \mathbb{E}[\tau|H_1] &= \frac{1}{n(n-1)} \left[2 \times \left(\frac{n}{2}\right)^2 \left(1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)\right)^2 \right] \\ &= \frac{n}{2(n-1)} \left(1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)\right)^2. \end{aligned}$$

■

3.7.3 Preuve de la proposition 3.2.7

Preuve. Afin d'évaluer la variance de τ sous l'hypothèse alternative H_1 , il faut dans un premier temps calculer $\mathbb{E}[\tau^2]$. À partir de la définition de τ donnée dans la définition 2, il faut donc évaluer τ^2 et distinguer les cas où les variables sont indépendantes ou non, c'est-à-dire considérer les cas où les indices coïncident ou pas.

$$\begin{aligned} \tau^2 &= \frac{1}{n^2(n-1)^2} \times \\ &\left[\sum_{\substack{i,j=1 \\ i \neq j}}^n \sum_{\substack{k,l=1 \\ k \neq l \neq i \neq j}}^n \underbrace{\text{sign}(X_i - X_j) \text{sign}(Y_i - Y_j) \text{sign}(X_k - X_l) \text{sign}(Y_k - Y_l)}_{(*)} \right. \\ &+ 4 \sum_{\substack{i,j=1 \\ i \neq j}}^n \sum_{\substack{k=1 \\ k \neq i,j}}^n \underbrace{\text{sign}(X_i - X_j) \text{sign}(Y_i - Y_j) \text{sign}(X_i - X_k) \text{sign}(Y_i - Y_k)}_{(**)} \\ &\left. + 2 \underbrace{\sum_{\substack{i,j=1 \\ i \neq j}}^n (\text{sign}(X_i - X_j) \text{sign}(Y_i - Y_j))^2}_{=n(n-1)} \right]. \end{aligned}$$

Chacune de ces sommes est alors évaluée indépendamment, en prenant en compte la localisation des indices par rapport au saut du signal.

(*) : $k \neq l \neq i \neq j$

$$\begin{aligned}
(*) &= \sum_{\substack{i,j=1 \\ i \neq j}}^n \sum_{\substack{k,l=1 \\ k \neq l \neq i \neq j}}^n \text{sign}(X_i - X_j) \text{sign}(Y_i - Y_j) \text{sign}(X_k - X_l) \text{sign}(Y_k - Y_l) \\
&= \sum_{\substack{i,j \leq n/2 \\ \text{ou } i,j > n/2}}^n \left[\sum_{\substack{k,l \leq n/2 \\ \text{ou } k,l > n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\eta_i - \eta_j) \text{sign}(\varepsilon_k - \varepsilon_l) \text{sign}(\eta_k - \eta_l)}_{\mathbb{E}=0} \right. \\
&\quad + \sum_{\substack{k \leq n/2 \\ l > n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\eta_i - \eta_j) \text{sign}(\varepsilon_k - \varepsilon_l - a) \text{sign}(\eta_k - \eta_l - a)}_{\mathbb{E}=0} \\
&\quad + \sum_{\substack{k > n/2 \\ l \leq n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\eta_i - \eta_j) \text{sign}(\varepsilon_k - \varepsilon_l + a) \text{sign}(\eta_k - \eta_l + a)}_{\mathbb{E}=0} \left. \right] \\
&\quad + \sum_{\substack{i \leq n/2 \\ j > n/2}}^n \left[\sum_{\substack{k,l \leq n/2 \\ \text{ou } k,l > n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j - a) \text{sign}(\eta_i - \eta_j - a) \text{sign}(\varepsilon_k - \varepsilon_l) \text{sign}(\eta_k - \eta_l)}_{\mathbb{E}=0} \right. \\
&\quad + \sum_{\substack{k \leq n/2 \\ l > n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j - a) \text{sign}(\eta_i - \eta_j - a) \text{sign}(\varepsilon_k - \varepsilon_l - a) \text{sign}(\eta_k - \eta_l - a)}_{(*)1} \\
&\quad + \sum_{\substack{k > n/2 \\ l \leq n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j - a) \text{sign}(\eta_i - \eta_j - a) \text{sign}(\varepsilon_k - \varepsilon_l + a) \text{sign}(\eta_k - \eta_l + a)}_{(*)2} \left. \right] \\
&\quad + \sum_{\substack{i > n/2 \\ j \leq n/2}}^n \left[\sum_{\substack{k,l \leq n/2 \\ \text{ou } k,l > n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j + a) \text{sign}(\eta_i - \eta_j + a) \text{sign}(\varepsilon_k - \varepsilon_l) \text{sign}(\eta_k - \eta_l)}_{\mathbb{E}=0} \right. \\
&\quad + \sum_{\substack{k \leq n/2 \\ l > n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j + a) \text{sign}(\eta_i - \eta_j + a) \text{sign}(\varepsilon_k - \varepsilon_l - a) \text{sign}(\eta_k - \eta_l - a)}_{(*)2} \\
&\quad + \sum_{\substack{k > n/2 \\ l \leq n/2}}^n \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j + a) \text{sign}(\eta_i - \eta_j + a) \text{sign}(\varepsilon_k - \varepsilon_l + a) \text{sign}(\eta_k - \eta_l + a)}_{(*)3} \left. \right].
\end{aligned}$$

Comme $k \neq l \neq i \neq j$ et en utilisant l'indépendance, certaines de ces espérances sont nulles. Les autres termes sont évalués à l'aide de la proposition 3.7.3 et en exploitant l'indépendance, et on a :

$$\mathbb{E}[(*)1] = \left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma^2}\right) \right)^4, \quad (3.41)$$

$$\mathbb{E}[(\ast 2)] = \left(1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)\right)^4, \quad (3.42)$$

$$\mathbb{E}[(\ast 3)] = \left(1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)\right)^4. \quad (3.43)$$

(\ast 1) et (\ast 3) correspondent aux cas où i, j et k, l sont dans les deux moitiés distinctes du signal, donc chaque terme apparaît $\left(\frac{n}{2}\left(\frac{n}{2} - 1\right)\right)^2$ fois. (\ast 2) apparaît deux fois et correspond au cas où i, j et k, l sont dans les deux moitiés distinctes du signal, donc le terme est comptabilisé en tout $2 \times \left(\frac{n}{2}\left(\frac{n}{2} - 1\right)\right)^2$ fois. En additionnant les espérances, on obtient alors que

$$\mathbb{E}[(\ast)] = \frac{n^2(n-2)^2}{4} \left(1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)\right)^4. \quad (3.44)$$

(\ast\ast) : $i \neq j \neq k$

De façon analogue, il faut différencier les cas en fonction de la localisation des indices par rapport au saut du signal, ce qui permet de distinguer plusieurs sommes, évaluées ensuite indépendamment.

$$\begin{aligned} (\ast\ast) &= \sum_{\substack{i, j = 1 \\ i \neq j}}^n \sum_{\substack{k = 1 \\ k \neq i, j}}^n \text{sign}(X_i - X_j) \text{sign}(Y_i - Y_j) \text{sign}(X_i - X_k) \text{sign}(Y_i - Y_k) \\ &= \sum_{i, j \leq n/2} \sum_{k \leq n/2} \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k) \text{sign}(\eta_i - \eta_j) \text{sign}(\eta_i - \eta_k)}_{(\ast 1)} \\ &+ \sum_{i, j \leq n/2} \sum_{k > n/2} \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k - a) \text{sign}(\eta_i - \eta_j) \text{sign}(\eta_i - \eta_k - a)}_{(\ast 2)} \\ &+ \sum_{i \leq n/2} \sum_{j > n/2} \sum_{k \leq n/2} \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j - a) \text{sign}(\varepsilon_i - \varepsilon_k) \text{sign}(\eta_i - \eta_j - a) \text{sign}(\eta_i - \eta_k)}_{(\ast 2)} \\ &+ \sum_{i \leq n/2} \sum_{j > n/2} \sum_{k > n/2} \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j - a) \text{sign}(\varepsilon_i - \varepsilon_k - a) \text{sign}(\eta_i - \eta_j - a) \text{sign}(\eta_i - \eta_k - a)}_{(\ast 4)} \\ &+ \sum_{i > n/2} \sum_{j \leq n/2} \sum_{k \leq n/2} \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j + a) \text{sign}(\varepsilon_i - \varepsilon_k + a) \text{sign}(\eta_i - \eta_j + a) \text{sign}(\eta_i - \eta_k + a)}_{(\ast 5)} \\ &+ \sum_{i > n/2} \sum_{j \leq n/2} \sum_{k > n/2} \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j + a) \text{sign}(\varepsilon_i - \varepsilon_k) \text{sign}(\eta_i - \eta_j + a) \text{sign}(\eta_i - \eta_k)}_{(\ast 3)} \\ &+ \sum_{i, j > n/2} \sum_{k \leq n/2} \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k + a) \text{sign}(\eta_i - \eta_j) \text{sign}(\eta_i - \eta_k + a)}_{(\ast 3)} \\ &+ \sum_{i, j > n/2} \sum_{k > n/2} \underbrace{\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k) \text{sign}(\eta_i - \eta_j) \text{sign}(\eta_i - \eta_k)}_{(\ast 1)}. \end{aligned}$$

$\mathbb{E}[(*)1] = \frac{1}{9}$ à partir de la proposition 3.7.3, et cette expression apparaît dans deux sommes pour lesquelles i, j, k sont dans le même sous-ensemble, ce qui comptabilise $2 \times \frac{n}{2} \left(\frac{n}{2} - 1\right) \left(\frac{n}{2} - 2\right) = \frac{n(n-2)(n-4)}{4}$ termes.

(*2) est le produit de termes indépendants, donc on évalue dans un premier temps

$$\begin{aligned}
\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j) \cdot \text{sign}(\varepsilon_i - \varepsilon_k - a)] &= \iiint_{\mathbb{R}^3} \text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k - a) \varphi(\varepsilon_i) \varphi(\varepsilon_j) \varphi(\varepsilon_k) d\varepsilon_i d\varepsilon_j d\varepsilon_k \\
&= \int_{\varepsilon_i \in \mathbb{R}} \left(\int_{\varepsilon_j \in \mathbb{R}} \text{sign}(\varepsilon_i - \varepsilon_j) \varphi(\varepsilon_j) d\varepsilon_j \int_{\varepsilon_k \in \mathbb{R}} \text{sign}(\varepsilon_i - \varepsilon_k - a) \varphi(\varepsilon_k) d\varepsilon_k \right) \varphi(\varepsilon_i) d\varepsilon_i \\
&= \int_{\mathbb{R}} (2\phi(\varepsilon_i) - 1) (2\phi(\varepsilon_i - a) - 1) \varphi(\varepsilon_i) d\varepsilon_i \text{ grâce à la prop. 3.7.3,} \\
&= \int_{\mathbb{R}} 2\phi(\varepsilon_i - a) (2\phi(\varepsilon_i) - 1) \varphi(\varepsilon_i) d\varepsilon_i \\
\text{car } \int_{\mathbb{R}} (2\phi(\varepsilon_i) - 1) \varphi(\varepsilon_i) d\varepsilon_i &= 0 \text{ grâce à la Prop. 3.7.1,} \\
&= 4 \int_{\mathbb{R}} \phi(\varepsilon_i - a) \phi(\varepsilon_i) \varphi(\varepsilon_i) d\varepsilon_i - 2 \underbrace{\int_{\mathbb{R}} \phi(\varepsilon_i - a) \varphi(\varepsilon_i) d\varepsilon_i}_{= \phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right) \text{ grâce à la Prop. 3.7.4}}
\end{aligned}$$

Afin d'évaluer l'intégrale de gauche, on utilise le fait que $a > 0$ et ϕ est une croissante (puisque'il s'agit d'une fonction de répartition), donc $\phi(\varepsilon_i - a) \leq \phi(\varepsilon_i)$. Cela permet d'écrire que

$$\int_{\varepsilon_i} \phi(\varepsilon_i - a) \phi(\varepsilon_i) \varphi(\varepsilon_i) d\varepsilon_i \leq \int_{\varepsilon_i} \phi(\varepsilon_i)^2 \varphi(\varepsilon_i) d\varepsilon_i = \frac{1}{3}, \text{ donc}$$

$$\begin{aligned}
\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k - a)] &\leq \frac{4}{3} - 2\phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right) \\
&\leq \frac{1}{3} + \left(1 - 2\phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right)\right) = \frac{1}{3} + \left(2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) - 1\right).
\end{aligned}$$

En multipliant les deux termes (positifs, car $a > 0$),

$$\mathbb{E}[(*)2] \leq \left(\frac{1}{3} + \left(2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) - 1\right)\right)^2, \quad (3.45)$$

et l'expression (*2) apparaît $2 \times \frac{n}{2} \frac{n}{2} \left(\frac{n}{2} - 1\right) = \frac{n^2(n-2)}{4}$ fois.

(*3) est également le produit de deux quantités indépendantes, donc de façon analogue on évalue d'abord

$$\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j) \cdot \text{sign}(\varepsilon_i - \varepsilon_k + a)]$$

$$\begin{aligned}
&= \iiint_{\mathbb{R}^3} \text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k + a) \varphi(\varepsilon_i) \varphi(\varepsilon_j) \varphi(\varepsilon_k) \, d\varepsilon_i \, d\varepsilon_j \, d\varepsilon_k \\
&= \int_{\varepsilon_i} \left(\int_{\varepsilon_j} \text{sign}(\varepsilon_i - \varepsilon_j) \varphi(\varepsilon_j) \, d\varepsilon_j \int_{\varepsilon_k} \text{sign}(\varepsilon_i - \varepsilon_k + a) \varphi(\varepsilon_k) \, d\varepsilon_k \right) \varphi(\varepsilon_i) \, d\varepsilon_i \\
&= \int_{\mathbb{R}} (2\phi(\varepsilon_i) - 1) (2\phi(\varepsilon_i + a) - 1) \varphi(\varepsilon_i) \, d\varepsilon_i \text{ avec la Prop. 3.7.3,} \\
&= \int_{\mathbb{R}} 2\phi(\varepsilon_i + a) (2\phi(\varepsilon_i) - 1) \varphi(\varepsilon_i) \, d\varepsilon_i \\
&\text{car } \int_{\mathbb{R}} (2\phi(\varepsilon_i) - 1) \varphi(\varepsilon_i) \, d\varepsilon_i = 0 \text{ avec la Prop. 3.7.1,} \\
&= 4 \int_{\mathbb{R}} \phi(\varepsilon_i + a) \phi(\varepsilon_i) \varphi(\varepsilon_i) \, d\varepsilon_i - 2 \int_{\mathbb{R}} \phi(\varepsilon_i + a) \varphi(\varepsilon_i) \, d\varepsilon_i.
\end{aligned}$$

Afin d'évaluer cette intégrale, on utilise le fait que $\phi(\varepsilon) \leq 1$ pour tout $\varepsilon \in \mathbb{R}$:

$$\begin{aligned}
\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j) \text{sign}(\varepsilon_i - \varepsilon_k + a)] &\leq 4 \int_{\mathbb{R}} \phi(\varepsilon_i + a) \varphi(\varepsilon_i) \, d\varepsilon_i - 2 \int_{\mathbb{R}} \phi(\varepsilon_i + a) \varphi(\varepsilon_i) \, d\varepsilon_i \\
&= 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right).
\end{aligned}$$

En multipliant les deux termes, cela conduit à

$$\mathbb{E}[(\ast 3)] \leq 4\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)^2, \quad (3.46)$$

et l'expression $(\ast 3)$ se produit $2 \times \frac{n}{2} \frac{n}{2} \left(\frac{n}{2} - 1\right) = \frac{n^2(n-2)}{4}$ fois.

$(\ast 4)$ est gérée de la même manière :

$$\begin{aligned}
&\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j - a) \cdot \text{sign}(\varepsilon_i - \varepsilon_k - a)] \\
&= \iiint_{\mathbb{R}^3} \text{sign}(\varepsilon_i - \varepsilon_j - a) \text{sign}(\varepsilon_i - \varepsilon_k - a) \varphi(\varepsilon_i) \varphi(\varepsilon_j) \varphi(\varepsilon_k) \, d\varepsilon_i \, d\varepsilon_j \, d\varepsilon_k \\
&= \int_{\varepsilon_i} \left(\int_{\varepsilon_j} \text{sign}(\varepsilon_i - \varepsilon_j - a) \varphi(\varepsilon_j) \, d\varepsilon_j \int_{\varepsilon_k} \text{sign}(\varepsilon_i - \varepsilon_k - a) \varphi(\varepsilon_k) \, d\varepsilon_k \right) \varphi(\varepsilon_i) \, d\varepsilon_i \\
&= \int_{\mathbb{R}} (2\phi(\varepsilon_i - a) - 1)^2 \varphi(\varepsilon_i) \, d\varepsilon_i \text{ avec la prop. 3.7.3,} \\
&= 4 \int_{\mathbb{R}} \phi(\varepsilon_i - a)^2 \varphi(\varepsilon_i) \, d\varepsilon_i - 4 \underbrace{\int_{\mathbb{R}} \phi(\varepsilon_i - a) \varphi(\varepsilon_i) \, d\varepsilon_i}_{=\phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right)} + \underbrace{\int_{\mathbb{R}} \varphi(\varepsilon_i) \, d\varepsilon_i}_{=1}.
\end{aligned}$$

En utilisant que $\phi(\varepsilon_i - a) \leq \phi(\varepsilon_i)$, $\int_{\mathbb{R}} \phi(\varepsilon_i - a)^2 \varphi(\varepsilon_i) \, d\varepsilon_i \leq \int_{\mathbb{R}} \phi(\varepsilon_i)^2 \varphi(\varepsilon_i) \, d\varepsilon_i = \frac{1}{3}$, donc

$$\mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j - a) \cdot \text{sign}(\varepsilon_i - \varepsilon_k - a)] \leq \frac{4}{3} - 4\phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right) + 1$$

$$\leq \frac{7}{3} - 4\phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right) = \frac{1}{3} + 2\left(1 - 2\phi\left(\frac{-a}{\sqrt{2\sigma^2}}\right)\right) = \frac{1}{3} + 2\left(2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) - 1\right).$$

Et en multipliant les deux quantités (positives),

$$\mathbb{E}[(*4)] \leq \left(\frac{1}{3} + 2\left(2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) - 1\right)\right)^2. \quad (3.47)$$

Pour les quantités impliquant (*4), j et k sont dans le même sous-ensemble, donc le nombre de termes de la somme est $\frac{n}{2} \frac{n}{2} \left(\frac{n}{2} - 1\right) = \frac{n^2(n-2)}{8}$.

Enfin, (*5) est évaluée comme suit :

$$\begin{aligned} \mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j + a) \cdot \text{sign}(\varepsilon_i - \varepsilon_k + a)] &= \int_{\mathbb{R}} (2\phi(\varepsilon_i + a) - 1)^2 \varphi(\varepsilon_i) d\varepsilon_i \\ &= 4 \int_{\mathbb{R}} \phi(\varepsilon_i + a)^2 \varphi(\varepsilon_i) d\varepsilon_i - 4 \int_{\mathbb{R}} \phi(\varepsilon_i + a) \varphi(\varepsilon_i) d\varepsilon_i + \int_{\mathbb{R}} \varphi(\varepsilon_i) d\varepsilon_i. \end{aligned}$$

Comme $\phi(\varepsilon) \leq 1$ pour tout $\varepsilon \in \mathbb{R}$, $\phi(\varepsilon)^2 \leq \phi(\varepsilon)$, et

$$\begin{aligned} \mathbb{E}[\text{sign}(\varepsilon_i - \varepsilon_j + a) \cdot \text{sign}(\varepsilon_i - \varepsilon_k + a)] \\ \leq 4 \int_{\mathbb{R}} \phi(\varepsilon_i + a) \varphi(\varepsilon_i) d\varepsilon_i - 4 \int_{\mathbb{R}} \phi(\varepsilon_i + a) \varphi(\varepsilon_i) d\varepsilon_i + \underbrace{\int_{\mathbb{R}} \varphi(\varepsilon_i) d\varepsilon_i}_{=1} \leq 1. \end{aligned}$$

Cela conduit (de façon peu informative!) à :

$$\mathbb{E}[(*5)] \leq 1. \quad (3.48)$$

Comme pour (*4), le nombre de termes est $\frac{n}{2} \frac{n}{2} \left(\frac{n}{2} - 1\right) = \frac{n^2(n-2)}{8}$.

En associant les résultats obtenus pour (*1) \rightarrow (*5), l'espérance de (**) est bornée par :

$$\begin{aligned} \mathbb{E}[(**)] &\leq \frac{n(n-2)(n-4)}{4 \times 9} + \frac{n^2(n-2)}{4} \left[\left(\frac{1}{3} + 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) - 1\right)^2 \right. \\ &\quad \left. + 4\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)^2 + \frac{1}{2} \left(\frac{1}{3} + 2\left(2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) - 1\right)\right)^2 + \frac{1}{2} \right] \\ &\leq \frac{n(n-2)(n-4)}{4 \times 9} + \frac{n^2(n-2)}{4} \left[\frac{2}{3} + \frac{4}{3} \left(2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) - 1\right) \right. \\ &\quad \left. + 3 \left(2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right) - 1\right)^2 + 4\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)^2 \right]. \end{aligned}$$

En reprenant le résultat obtenu pour (*) et en l'ajoutant à (**), on obtient :

$$\mathbb{E}[\tau^2 | H_1] \leq \frac{1}{n^2(n-1)^2} \left[\frac{n^2(n-2)^2}{4} \left(1 - 2\phi\left(\frac{a}{\sqrt{2\sigma^2}}\right)\right)^4 + \frac{n(n-2)(n-4)}{4 \times 9} \right]$$

$$\begin{aligned}
& + \frac{n^2(n-2)}{4} \left(\frac{2}{3} + \frac{4}{3} \left(2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) - 1 \right) + 3 \left(2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) - 1 \right)^2 + 4\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right)^2 \right) \Big] \\
& \leq \frac{1}{(n-1)^2} \left[\frac{(n-2)^2}{4} \left(1 - 2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) \right)^4 + \frac{(n-2)(n-4)}{4 \times 9n} \right. \\
& \left. + \frac{(n-2)}{4} \left(\frac{2}{3} + \frac{4}{3} \left(2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) - 1 \right) + 3 \left(2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) - 1 \right)^2 + 4\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right)^2 \right) \right].
\end{aligned}$$

Et finalement, une borne est obtenue pour la variance de τ sous l'hypothèse alternative H_1 :

$$\begin{aligned}
\text{Var}[\tau|H_1] &= \mathbb{E}[\tau^2|H_1] - \mathbb{E}[\tau|H_1]^2 \\
&\leq \frac{(n-2)^2}{4(n-1)^2} \left(1 - 2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) \right)^4 + \frac{1}{(n-1)^2} \left[\frac{(n-2)(n-4)}{4 \times 9n} \right. \\
&+ \frac{(n-2)}{4} \left(\frac{2}{3} + \frac{4}{3} \left(2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) - 1 \right) + 3 \left(2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) - 1 \right)^2 + 4\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right)^2 \right) \Big] \\
&- \frac{n^2}{4(n-1)^2} \left(1 - 2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) \right)^4 \\
&\leq \frac{1}{(1-n)} \left(1 - 2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) \right)^4 + \frac{1}{(n-1)^2} \left[\frac{(n-2)(n-4)}{4 \times 9n} \right. \\
&+ \frac{(n-2)}{4} \left(\frac{2}{3} + \frac{4}{3} \left(2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) - 1 \right) + 3 \left(2\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right) - 1 \right)^2 + 4\phi \left(\frac{a}{\sqrt{2\sigma^2}} \right)^2 \right) \Big].
\end{aligned}$$

Cette borne permet d'établir le taux de décroissance de τ lorsque le nombre d'échantillons n augmente, ce qui conclut la preuve. ■

3.7.4 Preuve du corollaire 3.2.8

Le corollaire 3.2.8 utilise les résultats obtenus pour l'espérance et la variance de τ sous l'hypothèse H_1 pour décrire le comportement du z -score associé.

Preuve. On rappelle que sous l'hypothèse H_1 le bruit est de nature gaussienne, donc les cas d'égalités ne sont pas à prendre en compte. z est donc défini par $z = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}}$. En utilisant les propositions 3.2.6 et 3.2.7 qui fournissent l'espérance et la variance de τ sous H_1 , le corollaire est alors immédiat. ■

3.7.5 Preuve du théorème 3.2.9

Le théorème 3.2.9 permet de contrôler la probabilité de fausse détection, telle qu'elle a été définie dans la condition (C2).

Preuve. L'inégalité de Chebyshev assure que si X est une variable aléatoire d'espérance finie μ et de variance non nulle σ^2 , alors

$$\forall k > 0, \Pr(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}. \quad (3.49)$$

En appliquant cette inégalité au z -score, on obtient

$$\Pr(z(X, Y) \leq \mathbb{E}[z(X, Y)|H_1] - k) \leq \Pr(|z(X, Y) - \mu| \geq k) \leq \frac{\text{Var}[z(X, Y)|H_1]}{k^2}. \quad (3.50)$$

Puis en prenant $\alpha = \mathbb{E}[z(X, Y)|H_1] - k$,

$$\mathbb{P}(z(X, Y) < \alpha \mid H_1) \leq \frac{\text{Var}(z(X, Y)|H_1)}{(\mathbb{E}[z(X, Y)|H_1] - \alpha)^2}. \quad (3.51)$$

La borne en $O\left(\frac{1}{n}\right)$ est alors garantie par le corollaire 3.2.8. ■

3.7.6 Preuve de la proposition 3.3.2

La proposition 3.3.2 évalue l'erreur d'estimation entre $\hat{\sigma}^2$ et la vraie valeur σ^2 .

Preuve. La proposition 3.3.1 nous dit que la variance de l'estimateur de variance $\hat{\sigma}^2$ est donnée par

$$\text{Var}(\hat{\sigma}^2) = \frac{\sigma^4}{N_b} \left(\beta_2 - \frac{N_b - 3}{N_b - 1} \right), \quad (3.52)$$

où $\beta_2 = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] = \frac{\mu_4}{\sigma^4}$ est le kurtosis associé à la distribution du bruit et N_b le nombre de pixels dans un bloc de l'image.

Si la taille N_b du bloc est suffisamment grande, la loi des grands nombres permet de considérer que $\hat{\sigma}^2$ est distribué selon une loi normale, c'est-à-dire :

$$\hat{\sigma}^2 \sim \mathcal{N}(\sigma^2, \text{Var}(\hat{\sigma}^2)) \Leftrightarrow \hat{\sigma}^2 - \sigma^2 \sim \mathcal{N}(0, \text{Var}(\hat{\sigma}^2)) \quad (3.53)$$

On utilise la propriété des moments absolus centrés de la loi normale $\mathcal{N}(0, \sigma^2)$ donnés par la formule suivante :

$$\mathbb{E}[|X|^p] = \sigma^p \frac{2^{\frac{p}{2}} \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}. \quad (3.54)$$

On en déduit alors l'espérance et la variance de l'erreur relative d'estimation :

$$\begin{aligned} \mathbb{E} \left[\frac{|\hat{\sigma}^2 - \sigma^2|}{\sigma^2} \right] &= \frac{\sqrt{\text{Var}(\hat{\sigma}^2)}}{\sigma^2} \times \frac{\sqrt{2}}{\sqrt{\pi}}, \quad \text{et} \quad (3.55) \\ \text{Var} \left(\frac{|\hat{\sigma}^2 - \sigma^2|}{\sigma^2} \right) &= \mathbb{E} \left[\frac{|\hat{\sigma}^2 - \sigma^2|^2}{\sigma^4} \right] - \mathbb{E} \left[\frac{|\hat{\sigma}^2 - \sigma^2|}{\sigma^2} \right]^2 \\ &= \frac{\text{Var}(\hat{\sigma}^2)}{\sigma^4} - \frac{\text{Var}(\hat{\sigma}^2)}{\sigma^4} \times \frac{2}{\pi} \\ &= \frac{\text{Var}(\hat{\sigma}^2)}{\sigma^4} \times \left(1 - \frac{2}{\pi} \right). \quad (3.56) \end{aligned}$$

En injectant la variance de l'estimateur $\hat{\sigma}^2$ donnée dans la proposition 3.3.1, cela conclut la preuve. ■

Deuxième partie

Débruitage

Table des matières

4	État de l’art du débruitage d’images et motivations	109
4.1	Introduction	109
4.2	Le débruitage	109
4.3	Les méthodes variationnelles	116
4.4	Les méthodes non locales	126
4.5	Conclusion	135
4.6	Annexes du chapitre : Éléments d’analyse convexe	137
5	Régularisation adaptative des moyennes non locales	145
5.1	Introduction	145
5.2	<i>Dejittering</i> des NL-means (NLDJ)	146
5.3	NL-means régularisés (R-NL)	148
5.4	D’autres coopérations des modèles variationnels et non locaux	152
5.5	R-NL pour la famille exponentielle	155
5.6	Résultats et discussion	162
5.7	Conclusion	168
6	R-NL pour le débruitage vidéo	171
6.1	Introduction	171
6.2	Adaptation 3D	171
6.3	Restauration de vidéos	173
6.4	Application au débruitage de vidéos BNL	176
6.5	Conclusion	179

État de l'art du débruitage d'images et motivations

4.1 Introduction

Le débruitage est un problème central en traitement d'images. C'est notamment une étape de pré-traitement nécessaire à beaucoup d'applications de haut niveau comme la segmentation, la reconstruction ou la super-résolution. Le débruitage est le procédé qui consiste à retrouver (ou estimer) l'image idéale f à partir de l'observation bruitée g . Toute la difficulté réside dans la nécessité de supprimer efficacement la composante de bruit tout en préservant le signal, ses contours et ses textures sans introduire d'artefacts.

De nombreuses méthodes ont été proposées dans la littérature pour tenter de résoudre ce problème, la plupart consacrées au cas du bruit gaussien. La section 4.2 réalise un état de l'art des grandes familles de méthodes de débruitage, ainsi que des critères d'évaluation du débruitage. Dans le contexte de la vision de nuit, la méthode de débruitage que nous envisageons doit satisfaire plusieurs contraintes : elle doit pouvoir s'implémenter dans un contexte embarqué et temps réel, ce qui rend par exemple les techniques d'apprentissage en ligne peu envisageables. Il faut également que l'algorithme soit facilement adaptable à différents types de bruit, en raison du comportement différent en fonction de la luminosité mis en avant dans la partie précédente. Enfin, nous souhaiterions pouvoir tirer partie de la redondance d'information apportée par le flux vidéo. L'ensemble de ces contraintes nous ont conduits à étudier plus spécifiquement dans les sections 4.3 et 4.4 deux méthodes de débruitage, les méthodes variationnelles et les moyennes non locales, susceptibles de remplir ces critères. Une étude de chacune de ces méthodes permet de dégager leurs points forts et leur points faibles, dans l'optique de les associer.

4.2 Le débruitage

On peut distinguer plusieurs familles de méthodes de débruitage : les techniques de filtrage peuvent être linéaires et non linéaires, et elles peuvent s'interpréter comme des filtres spatiaux, locaux ou non locaux, fréquentiels ou plus généralement dans un domaine de transformation, ou encore se baser sur des méthodes d'apprentissage. Ce chapitre réalise une description non exhaustive des techniques répandues de débruitage, et une taxonomie des plus grandes familles est schématisée sur la figure 4.1.

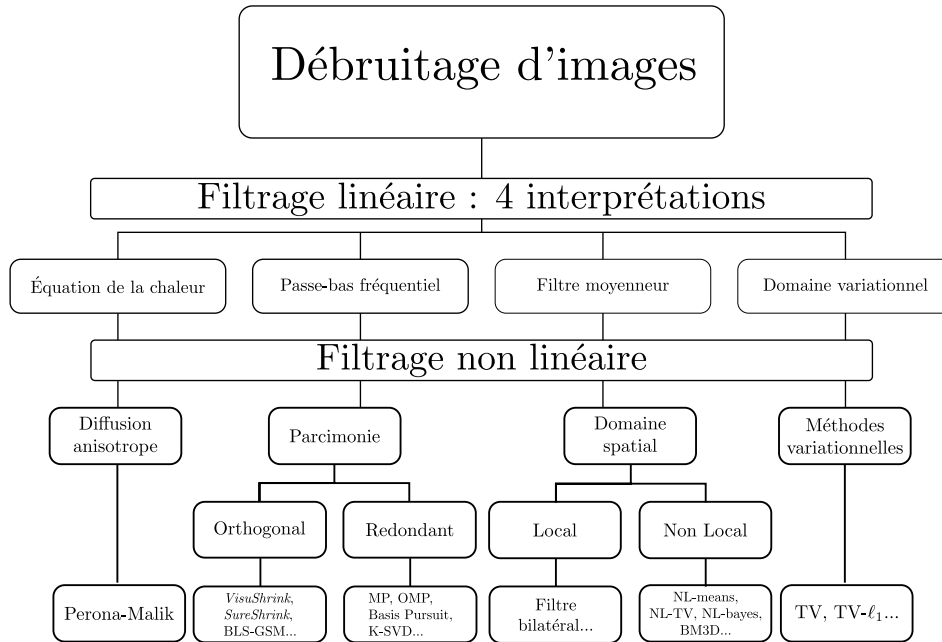


FIGURE 4.1 – Taxonomie des méthodes de débruitage.

4.2.1 Évaluation du débruitage

Si l’objet de ce chapitre n’est pas de comparer les performances des différentes méthodes de débruitage décrites par la suite, il est néanmoins important pour l’étude sur le débruitage de définir des critères d’évaluation de performance du débruitage. La plupart de ces méthodes nécessitent la connaissance de l’image de référence f , afin de comparer le résultat de débruitage obtenu à l’image idéale.

4.2.1.1 SNR/PSNR

Une mesure d’erreur intuitive repose sur l’erreur quadratique moyenne entre l’image de référence f et l’image débruitée \hat{u} : $\frac{1}{N} \|\hat{u} - f\|_2^2$. Cette erreur, quadratique peut varier sur une très large gamme de valeurs, aussi on utilise plutôt un critère logarithmique, par exemple le rapport signal sur bruit (SNR, pour *Signal to Noise Ratio*) donné par

$$\text{SNR}(\hat{u}, f) = 10 \log_{10} \frac{\text{Var}(f)}{\frac{1}{N} \|\hat{u} - f\|_2^2}, \quad (4.1)$$

où $\text{Var}(f)$ est la variance empirique des intensités de f , ou plus souvent lorsque l’image est codée sur 8 bits le PSNR (*Peak Signal to Noise Ratio*) :

$$\text{PSNR}(\hat{u}, f) = 10 \log_{10} \frac{255^2}{\frac{1}{N} \|\hat{u} - f\|_2^2}. \quad (4.2)$$

Une image restaurée est considérée de bonne qualité si son SNR est supérieur à 20dB ou son PSNR à 30dB.

Ces critères ne sont que des mesures de différences pixel à pixel et ne prennent pas en compte la préservation des structures autant que la perte de contraste par exemple.

4.2.1.2 SSIM

L'indice de qualité SSIM (pour *Structural SIMilarity*) [Wang 2004], a été introduit initialement pour mesurer la qualité visuelle des images compressées, en se basant sur des statistiques locales. Il est également utilisé pour comparer des méthodes de débruitage et discriminer la présence d'artefacts (comme les artefacts de bloc associés à la compression JPEG).

4.2.1.3 Method noise

Une méthode d'évaluation visuelle très simple proposée dans [Buades 2008] appelée *method noise* consiste à étudier le résidu entre l'image d'origine bruitée et le résultat débruité $\hat{u} - g$. Ce critère permet de contrôler la quantité d'information qui a été supprimée pendant le processus de débruitage. Un débruitage idéal ne devrait avoir supprimé que du bruit, donc le résidu devrait être (dans le cas gaussien) de nature gaussienne, avec une variance la plus proche possible de celle du bruit, et ne pas contenir de structure apparente. Néanmoins dans le cas de bruit dépendant du signal, la variance du bruit n'est pas la même sur toutes les zones de l'image donc on ne peut estimer la qualité du débruitage à l'aide du résidu.

Des techniques d'évaluation à l'aveugle [Moorthy 2011, Mittal 2012, Saad 2012] permettent d'estimer la qualité d'une image sans connaissance de l'image d'origine ni du type de dégradation perturbant l'image, mais elles nécessitent de définir un modèle d'image naturelle qui ne s'applique pas à tous les domaines de l'imagerie (télétection, imagerie médicale, etc). En particulier dans le contexte de vision nocturne, les caractéristiques des images bas niveau de lumière sont très variables en fonction du gain de l'IL et des conditions d'acquisition. L'évaluation de leur qualité reposerait davantage sur une évaluation qualitative réalisée auprès des pilotes, et basée sur des aspects facteurs humains.

4.2.2 Filtres linéaires

La méthode la plus élémentaire de filtrage exploite l'hypothèse d'ergodicité du signal : en supposant que les échantillons de signal bruité dans un voisinage autour de chaque pixel suivent la même distribution, alors en effectuant une moyenne locale des valeurs des pixels voisins la variance du bruit est réduite sans introduire de biais. Cela conduit au filtre *boxcar* qui consiste à remplacer chaque pixel bruité par une moyenne locale des pixels alentours. En moyennant K variables indépendantes de variance σ^2 , la variance résiduelle est alors de σ^2/K .

Un autre filtre classique est le filtre gaussien. Au lieu d'effectuer une moyenne des pixels sur une fenêtre uniforme, il consiste à convoluer l'image g par un noyau

gaussien G de bande passante h donné par $G_h(i) = \exp\left(-\frac{i^2}{2h^2}\right)$.

Ces deux filtres constituent des cas particuliers de filtres linéaires, dont la relation générale est donnée par

$$\hat{u} = Ag, \quad (4.3)$$

où A est une matrice (en générale circulante) de taille $N \times N$.

L'implémentation d'un filtre linéaire circulant est rapide et simple, mais les contours et les textures ont tendance à être trop fortement lissés, ce qui résulte en une perte de résolution. La bande passante h du filtre gaussien ou la largeur de la fenêtre du filtre *boxcar* jouent un rôle essentiel dans le compromis biais-variance entre débruitage et préservation des textures, qui résulte en une perte de résolution manifestée par un effet de flou, ou un débruitage insuffisant.

Le filtrage linéaire peut s'accompagner de nombreuses interprétations qui ont conduit par la suite à différentes familles de méthodes non linéaires de débruitage.

4.2.2.1 Équation de la chaleur

La convolution par le noyau gaussien est solution de l'équation de la chaleur donnée par

$$\begin{cases} \frac{\partial u_i(t)}{\partial t} = \operatorname{div}(\nabla u_i(t)), \\ u_i(0) = g_i. \end{cases} \quad (4.4)$$

La solution de cette équation est obtenue de façon itérative par un schéma aux différences finies [Aubert 2006]. La régularisation apportée par la convolution dans le cas du filtre gaussien est ici interprétée par une diffusion isotrope. Cette dernière opère dans toutes les directions sans distinction, ce qui permet un débruitage efficace des régions d'intensité homogène mais lisse les contours, résultant en une perte de résolution. Cette interprétation a conduit aux filtres de diffusion anisotrope.

4.2.2.2 Interprétation fréquentielle

Lorsque la matrice A est une matrice circulante, le filtre linéaire se traduit par une convolution, donc une multiplication dans le domaine fréquentiel. Le filtre *boxcar* revient par exemple à multiplier dans le domaine de Fourier par un sinus cardinal, et le noyau gaussien est identique dans le domaine fréquentiel.

L'interprétation fréquentielle du filtrage linéaire se traduit donc par

$$\hat{u} = F^{-1}\Psi F(g), \quad (4.5)$$

où Ψ est une fonction linéaire (par exemple la multiplication par un noyau gaussien) et F la transformée de Fourier. Ces filtres sont qualifiés de passe-bas, car ils conservent les basses fréquences en supprimant les hautes fréquences, assimilées au bruit. Cependant, un contour est représenté sur une harmonique de fréquences, donc l'application d'un filtre passe-bas supprime une partie de l'information portée par les contours, ce qui résulte encore une fois en une perte de résolution. La suppression des composantes associées au bruit dans un espace approprié est au cœur des approches parcimonieuses utilisant des dictionnaires.

4.2.2.3 Filtre moyennneur

Le filtre gaussien peut s'exprimer sous la forme d'un filtre moyennneur de la façon suivante :

$$\hat{u}_i = \frac{\sum_{j \in \Omega} w_{ij} g_j}{\sum_{j \in \Omega} w_{ij}}, \quad \text{avec} \quad w_{ij} = \exp\left(-\frac{(i-j)^2}{2h^2}\right). \quad (4.6)$$

On remarque que l'on peut également écrire \hat{u} comme la solution du problème des moindres carrés pondérés

$$\hat{u} = \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \sum_{j \in \Omega} w_{ij} (u_i - g_j)^2. \quad (4.7)$$

Cette formulation peut s'appliquer à d'autres filtres linéaires en adaptant le choix des poids w_{ij} .

4.2.2.4 Régularisation de Tikhonov

Le filtrage linéaire peut aussi s'interpréter comme un cas particulier de la régularisation de Tikhonov [Tikhonov 1977] associée au problème de minimisation

$$\hat{u} = \arg \min_{u \in \mathbb{R}^N} \lambda \|u - g\|^2 + \|Tu\|^2, \quad (4.8)$$

dont la solution est explicitement donnée par

$$\hat{u} = (Id + \lambda T^t T)^{-1} g. \quad (4.9)$$

Dans le cas de problèmes de débruitage, la matrice T peut correspondre par exemple au gradient, menant à un filtrage linéaire circulant, donc convolutif. Le gradient permet de privilégier des solutions qui ne présentent pas de variations trop rapides, assimilées à du bruit. Cependant, un contour présente également une forte variation, ce qui conduit à une solution dont les contours sont atténués.

De façon générale, le filtrage linéaire s'accompagne d'une perte de résolution liée au sur-lissage des contours de l'image. Selon l'interprétation réalisée, cette perte de résolution peut s'expliquer d'un point de vue spatial, par une diffusion isotrope ou un processus de moyenne locale, fréquentiel, par l'application d'un filtre passe-bas, ou variationnel, par le choix d'un critère de régularisation quadratique pénalisant les fortes variations. Ces différentes interprétations ont toutes conduit à la mise en place de techniques de débruitage non linéaires visant à mieux préserver les contours, en cherchant à remédier aux limitations exposées ci-dessus.

4.2.3 Filtrage non linéaire

4.2.3.1 Le modèle anisotrope de Perona-Malik

Afin de limiter la perte de résolution des contours liée à la diffusion isotrope, le filtre de Perona et Malik [Perona 1990] effectue un filtrage dit anisotrope en ne

lissant que dans la direction perpendiculaire au gradient de l'image (donc parallèle aux contours), afin de préserver ces derniers. Celui-ci se traduit de la façon suivante :

$$\begin{cases} \frac{\partial u_i(t)}{\partial t} = \operatorname{div}(\varphi(|\nabla u_i(t)|)\nabla u_i(t)) \\ u_i(t) = 0 \end{cases}, \quad (4.10)$$

où φ est une fonction décroissante, donnée par exemple par $\varphi(x) = \exp(-x^2/c^2)$ ou $\varphi(x) = \frac{1}{1+x^2/c^2}$. Elle permet d'adapter la diffusion au contenu de l'image (plus c est grand, plus la diffusion est isotropique). Dans la direction où le gradient est fort (perpendiculairement aux contours), la diffusion est nulle, tandis que si le gradient est faible (donc sur les régions homogènes), la diffusion est isotrope.

4.2.3.2 Dans un domaine de transformation

De façon générale, le procédé de débruitage associé au modèle (4.5) peut se réinterpréter sous la forme du problème de minimisation

$$\hat{u} = \arg \min_{\alpha} \|D\alpha - g\|^2 + R(\alpha), \quad (4.11)$$

où D est la transformation de domaine (dans (4.5), la transformée de Fourier) et R une contrainte de régularisation sur les coefficients dans le domaine de transformation (traduisant dans (4.5) l'influence de la fonction Ψ).

Ainsi, les techniques de filtrage ont envisagé d'adapter d'autre part le domaine de représentation, à travers le choix de la transformation D , et d'autre part la régularisation, en étudiant des régularisation non quadratiques plus adaptées aux caractéristiques des images.

Le débruitage dans un domaine de transformation consiste donc de façon générale à se placer dans un domaine dans lequel la représentation de l'image est intéressante, par exemple telle que le bruit et le signal soient relativement bien séparés, afin de pouvoir filtrer les coefficients de l'un sans altérer l'autre.

Pour cela, on décompose l'image sur un espace défini par une collection d'atomes, appelée dictionnaire. La décomposition sur un dictionnaire consiste à modéliser l'image par une combinaison linéaire de K atomes d_k de dimension N la taille de l'image. La matrice des atomes $(d_k)_{k \leq K}$ D de taille $N \times K$ est appelée le dictionnaire. On suppose alors que l'image est sparse pour D , c'est-à-dire qu'elle peut s'écrire sous la forme d'un produit $D\alpha$, avec α un vecteur de taille K sparse, c'est-à-dire possédant peu de composantes non nulles. On cherche la décomposition parcimonieuse de l'image g de la façon suivante :

$$\min_{D,\alpha} \|D\alpha - g\|_2^2 \text{ sous la contrainte } \|\alpha\|_0 < \varepsilon, \quad (4.12)$$

où ε contrôle le degré de parcimonie et $\|\cdot\|_0$ désigne la pseudo-norme ℓ_0 qui compte le nombre de composantes non nulles. Sans restriction sur le dictionnaire D , ce problème est NP-difficile, aussi on relaxe le problème à l'aide d'une norme ℓ_1 :

$$\min_{D,\alpha} \|D\alpha - g\|_2^2 \text{ sous la contrainte } \|\alpha\|_1 < \varepsilon, \quad (4.13)$$

Lorsque le dictionnaire est fixé à l'avance, ce problème, connu sous le nom de LASSO (pour *Least Absolute Shrinkage and Selection Operator*), permet de privilégier les solutions avec peu de coefficients non nuls [Tibshirani 1996].

Afin de résoudre directement le problème (4.12), il faut imposer des restrictions sur le dictionnaire D . Il peut être fixé d'avance, déterminé à partir des données ou adapté de façon itérative, et il peut être orthogonal ou redondant.

Décomposition orthogonale

Lorsque le dictionnaire D est une base orthonormale, avec $K = N$, chaque coefficient α est la projection de l'image g sur un atome d_k du dictionnaire. On utilise notamment la base orthogonale des fonctions sinusoïdales qui fournit la transformation en cosinus discret (DCT, pour *Discrete Cosine Transform*). Une telle transformation décompose le signal en combinaison de sinusoïdes, mais elle n'offre pas une bonne séparation du signal et du bruit, en raison notamment de la localisation des contours sur toutes les fréquences.

La transformée en ondelettes [Mallat 1999] permet de capturer les informations à la fois d'échelle et d'orientation. Le dictionnaire des ondelettes est composé des translations et dilatations de l'ondelette mère ψ :

$$d_{j,k}(t) = \sqrt{2^j} \psi(2^j t - k), \quad (4.14)$$

où les coefficients de dilatation j et de translation k sont des entiers. Le choix de l'ondelette mère permet d'une part de représenter finement les discontinuités ou les orientations, et d'autre part de garantir des conditions de reconstruction [Daubechies 1992]. Beaucoup de variantes ont été proposées, comme les wedgelets [Donoho 1999], curvelets [Candes 1999], bandlets [Le Pennec 2005] ou encore les contourlets [Do 2003].

Le principe des décompositions orthogonales permet alors de séparer les composantes de bruit du signal. Les coefficients les plus importants représentent les variations basse-fréquence tandis que le bruit et les coefficients de détail sont situés dans les bandes les plus fines. On peut donc débruiter en seuillant ou filtrant ces coefficients. La minimisation de (4.12) conduit au seuillage dur (HT, pour *Hard Thresholding*) [Mallat 2008] :

$$\hat{u} = D\hat{\alpha}^{HT}, \text{ avec } \hat{\alpha}_k^{HT} = \alpha_k \text{ si } |\alpha_k| > \lambda, 0 \text{ sinon,} \quad (4.15)$$

et la minimisation de (4.13) au seuillage doux (ST, pour *Soft Thresholding*) qui a donné lieu aux algorithmes *VisuShrink* et *SureShrink* [Donoho 1994, Donoho 1995] :

$$\hat{u} = D\hat{\alpha}^{ST}, \text{ avec } \hat{\alpha}_k^{ST} = \text{sign}(\alpha_k) \cdot \max(0, |\alpha_k| - \lambda). \quad (4.16)$$

Le paramètre de seuillage λ règle le degré de parcimonie de la solution. Un autre filtrage des coefficients repose sur le filtre de Wiener [Muresan 2003, Zhang 2010a].

L'une des méthodes les plus performantes de débruitage basé sur les ondelettes est l'approche BLS-GSM [Portilla 2003]. Elle consiste à modéliser les dépendances

entre les coefficients d'ondelettes voisins dans la pyramide multi-échelle à l'aide de mélanges de gaussiennes (GSM, pour *Gaussian Scale Mixture*) puis à mettre à jour les coefficients par une estimation bayésienne des moindres carrés (BLS, pour *Bayesian Least Square*).

Décomposition redondante

Lorsque le dictionnaire est redondant ou *over-complete*, le problème (4.12) est NP-complet. La solution est estimée par des algorithmes gloutons comme l'algorithme de *matching pursuit* [Mallat 1993], qui consiste de façon itérative à projeter l'image bruitée résiduelle sur l'atome qui minimise l'erreur quadratique moyenne, ou ses variantes comme *l'orthogonal matching pursuit* [Pati 1993].

Il est également possible d'apprendre le dictionnaire à partir des données bruitées. L'algorithme K-SVD [Elad 2006] est un algorithme en deux étapes qui adapte de façon itérative le dictionnaire :

1. Approximation parcimonieuse : mise à jour des coefficients α de la représentation parcimonieuse qui résout (4.12) pour le dictionnaire courant, à l'aide d'un algorithme de poursuite, et
2. Mise à jour du dictionnaire : chaque atome d_k du dictionnaire est mis à jour à l'aide d'une décomposition en K valeurs singulières (K-SVD, pour *K-Single Value Decomposition*) qui minimise l'erreur d'approximation entre la donnée bruitée et sa décomposition sur le dictionnaire.

Lorsque l'utilisation d'un dictionnaire redondant est associée à une minimisation ℓ_1 conformément au problème (4.13), le problème de *basis pursuit* [Chen 1998] est convexe et de nombreux algorithmes d'optimisation, dont certains seront fournis dans la suite du chapitre, permettent sa résolution.

4.3 Les méthodes variationnelles

L'interprétation variationnelle du filtre gaussien et l'étude de contraintes de régularisation non quadratiques ont donné naissance à la famille des méthodes variationnelles, qui visent à forcer la régularité des zones homogènes de l'image tout en préservant ses contours.

4.3.1 Le modèle de ROF

Les méthodes variationnelles consistent à chercher pour solution une image qui minimise une énergie donnée, visant à respecter les données tout en satisfaisant des contraintes de régularité. Parmi ces méthodes, le modèle de Rudin, Osher et Fatemi, appelé modèle de ROF [Rudin 1992], impose la régularité de l'image tout en préservant ses contours en minimisant la variation totale (TV) de l'image. L'image restaurée u^{TV} est obtenue à partir de l'image dégradée g en minimisant l'énergie

suivante :

$$u^{\text{TV}} = \arg \min_{u \in \mathbb{R}^N} \frac{\lambda}{2\sigma^2} \|u - g\|_2^2 + \text{TV}(u) \quad (4.17)$$

avec $\|u - g\|_2^2 = \sum_{i \in \Omega} (u_i - g_i)^2$ et $\text{TV}(u) = \sum_{i \in \Omega} \|(\nabla u)_i\|_2$,

où ∇u désigne le gradient discret de l'image u , défini selon [Chambolle 2004]. Le terme $\frac{\|u - g\|_2^2}{2\sigma^2}$ est un terme dit d'attache aux données, qui correspond dans le cas gaussien à la log-vraisemblance. $\text{TV}(u)$ est un terme de régularisation appelé variation totale. $\lambda > 0$ est le paramètre de régularisation qui règle le compromis entre attache aux données et régularisation. Quand λ est grand, le poids de l'attache aux données est fort, ce qui force la solution à rester proche de l'image bruitée afin de minimiser le terme de fidélité. Quand λ tend vers 0, le terme prépondérant de l'énergie à minimiser devient la variation totale, donc la solution u^{TV} tend vers une solution constante par morceaux. Contrairement à la régularisation de Tikhonov basée sur la norme ℓ_2 qui pénalise les forts contours, l'utilisation de la norme ℓ_1 du gradient, non lisse en zéro, permet de mieux respecter les singularités de l'image. Le modèle de ROF convient bien pour le débruitage d'images assimilables à des images constantes par morceaux, en revanche il ne permet pas une bonne prise en compte simultanée des textures et des régions homogènes.

Le modèle de ROF a pu par la suite être adapté, en modifiant par exemple l'attache aux données. Le modèle TV-L1 [Alliney 1992] utilise une norme ℓ_1 au lieu de la norme ℓ_2 utilisée dans le modèle de ROF :

$$\hat{u} = \arg \min_{u \in \mathbb{R}^N} \frac{\lambda}{\sigma} \|u - g\|_1 + \text{TV}(u), \quad \text{avec} \quad \|u - g\|_1 = \sum_{i \in \Omega} |u_i - g_i|. \quad (4.18)$$

Ce dernier permet de limiter la perte de contraste inhérente au modèle de ROF ; en revanche il a tendance à faire disparaître les structures de petites taille [Chan 2005a, Duval 2009].

D'autres termes de régularisation peuvent également être utilisés, notamment le modèle de Potts [Wu 1982] associé à la norme ℓ_0 .

De façon générale, ces modèles sont sensibles au paramètre de régularisation λ [Hansen 1993, Ramani 2008] et ne permettent pas de traiter de façon uniforme les zones homogènes et les textures. Il est alors possible d'ajuster localement le paramètre de régularisation au contenu de l'image, en fonction de la présence de contours [Strong 1997] ou de variance locale [Gilboa 2006].

4.3.2 Résolution du problème de minimisation (4.17)

L'analyse convexe fournit des algorithmes d'optimisation qui permettent la résolution du problème de minimisation (4.17). Des éléments d'analyse convexe sont donnés dans la section 4.6. Les principaux algorithmes de minimisation qui en découlent sont les suivants.

Algorithme 2 Algorithme de descente de gradient régularisée

Initialisation: choisir $u^0 = g$ et $\kappa > 0$

Itérations ($k \geq 0$):

$$u^{k+1} = u^k - \tau \nabla F_\varepsilon(u^k).$$

Algorithme 3 Algorithme de Chambolle [Chambolle 2004]

Initialisation: choisir $p^0 \in \mathbb{R}^{N,2}$ et $\kappa > 0$

Itérations ($k \geq 0$):

$$p^{k+1} = \frac{p^k + \tau \left(\nabla \left(\frac{\sigma^2}{\lambda} \operatorname{div} p^k - g \right) \right)}{1 + \tau \left| \nabla \left(\frac{\sigma^2}{\lambda} \operatorname{div} p^k - g \right) \right|}.$$

Renvoyer $u = g - \frac{\sigma^2}{\lambda} \operatorname{div} p$.

Par analogie à la minimisation d'une fonction dans \mathbb{R} où l'on cherche à annuler sa dérivée, la minimisation d'une fonctionnelle repose sur l'annulation de son gradient ou sous-gradient. Pour cela, et cette étape est fondamentale dans la résolution du problème, on cherche à calculer le gradient de la fonctionnelle (4.17) : $F(u) = \frac{\lambda}{2\sigma^2} \|u - g\|_2^2 + \operatorname{TV}(u)$.

4.3.2.1 Descente de gradient

La variation totale n'étant pas différentiable partout, une première approche pour résoudre le problème (4.17) consiste à régulariser celle-ci en posant $\operatorname{TV}_\varepsilon(u) = \sum_i \sqrt{\varepsilon^2 + |(\nabla u)_i|^2} = \sum_i \|(\nabla u)_i\|_\varepsilon$. La méthode la plus intuitive pour minimiser la fonctionnelle $F_\varepsilon(u) = \frac{1}{2\lambda} \|u - g\|_2^2 + \operatorname{TV}_\varepsilon(u)$ consiste alors à chercher à annuler le gradient de celle-ci en effectuant une descente de gradient. On obtient alors:

$$\nabla F_\varepsilon(u) = \frac{u - g}{\lambda} + \nabla \operatorname{TV}_\varepsilon(u) = \frac{u - g}{\lambda} - \operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|_\varepsilon} \right), \quad (4.19)$$

et la descente de gradient s'effectue alors à l'aide de l'algorithme 2. Cette méthode est très longue à converger pour de faibles valeurs de ε , même en utilisant des variantes à pas variable. L'utilisation d'une valeur de ε trop élevée mène en revanche à un sur-lissage des contours, limitant l'intérêt du modèle de ROF.

4.3.2.2 Méthode de Chambolle

D'après les calculs effectués dans la section 4.6.2, la variation totale peut s'écrire sous la forme

$$\operatorname{TV}(u) = \sum_{i=1}^N \|(\nabla u)_i\| = \|\nabla u\|_{2,1} = \sup_{p \in \mathbb{R}^{N,2}} \{ \langle p, \nabla u \rangle, |p_i| \leq 1 \}$$

Algorithme 4 Algorithme Forward-Backward [Beck 2010]

Initialisation: choisir $u^0 = g$ et $\kappa > 0$

Itérations ($k \geq 0$):

$$\begin{cases} y^k & = u^k - \tau \nabla G(u^k), \\ u^k & = \text{prox}_{\tau F} y^k, \\ t^{k+1} & = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}, \\ y^{k+1} & = u^k + \frac{t^k - 1}{t^{k+1}} (u^k - u^{k+1}). \end{cases}$$

$$= \sup_{p \in \mathbb{R}^{N,2}} \{ -\langle \text{div } p, u \rangle, |p_i| \leq 1 \}. \quad (4.20)$$

Ainsi, en écrivant la fonctionnelle (4.17) sous sa forme duale, le problème à résoudre devient

$$\min_{p \in \mathbb{R}^{N,2}} \left\{ \left\| \frac{\sigma^2}{\lambda} \text{div } p - g \right\|^2, |p_i|^2 \leq 1 \right\}. \quad (4.21)$$

L'algorithme se résout alors à l'aide de de l'algorithme 3.

4.3.2.3 Algorithme forward-backward

L'algorithme forward-backward [Combettes 2005] permet de résoudre des problèmes de minimisation de type

$$\min_{u \in \mathbb{R}^N} F(u) + G(u) \quad (4.22)$$

avec F et G deux fonctions convexes semi-continues inférieurement, avec G différentiable, et ∇G Lipschitz de constante $1/\beta$. Dans le cas du problème (4.17), on a donc $F(u) = \text{TV}(u)$ et $G(u) = \lambda \frac{\|u - g\|_2^2}{2\sigma^2}$. G est bien différentiable, avec $\nabla G(u) = \lambda \frac{u - g}{\sigma^2}$, et ∇G est Lipschitz de constante $1/\beta = \frac{\sigma^2}{\lambda}$.

D'après la section 4.6.1, on définit pour $x \in \mathbb{R}^N$ l'opérateur proximal associé à F , par :

$$\text{prox}_{\tau F} x = \arg \min_{u \in \mathbb{R}^N} F(u) + \frac{1}{2\tau} \|x - u\|^2. \quad (4.23)$$

En utilisant la sous-différentielle ∂F de F (voir section 4.6.1), on note souvent l'opérateur proximal sous la forme

$$\text{prox}_{\tau F} x = (I + \tau \partial F)^{-1}(x). \quad (4.24)$$

L'algorithme de forward-backward s'effectue alors d'après la section 4.6.3 de la façon suivante :

$$u^{k+1} = (I + \tau \partial F)^{-1} \left(u^k - \tau \nabla G(u^k) \right), \quad (4.25)$$

Algorithme 5 Algorithme de Chambolle-Pock [Chambolle 2011]

Initialisation: choisir $\tau, \gamma > 0$

$$u^0 = g, y^0 = \nabla u^0, \bar{u}^0 = u^0$$

Itérations ($k \geq 0$):

$$\begin{cases} y^{k+1} &= (I + \gamma \partial F^*)^{-1} (y^k + \gamma \nabla \bar{u}^k) \\ u^{k+1} &= (I + \tau \partial G)^{-1} (u^k - \tau \nabla^* y^{k+1}) \\ \bar{u}^{k+1} &= 2u^{k+1} - u^k \end{cases}$$

Il fait intervenir une étape explicite *forward* qui réalise la descente de gradient par rapport à l'attache aux données G , et une étape implicite *backward* qui calcule l'opérateur proximal $(I + \tau \partial F)^{-1}$ associé à la variation totale. Celui-ci n'est pas calculable explicitement, donc il est évalué de façon itérative dans une boucle interne, à l'aide de l'algorithme forward-backward accéléré FISTA [Beck 2009a], ou en utilisant la forme duale. Le paramètre τ doit être tel que $\tau \leq \beta^{-1}$ où β est la constante de Lipschitz associée à ∇G . Ici, il faut donc satisfaire $\tau \leq \frac{\sigma^2}{\lambda}$.

Des accélérations [Beck 2009b] et adaptations [Raguet 2013] ont été proposées suite à ce modèle. L'algorithme 4 décrit l'accélération proposée par Beck et Teboulle [Beck 2010].

4.3.2.4 Algorithme primal-dual

L'algorithme de Chambolle-Pock [Chambolle 2011] permet de résoudre un problème de la forme

$$\min_{u \in \mathbb{R}^N} F(Ku) + G(u), \quad (4.26)$$

où F et G sont dans le problème (4.17) données par $F = \|\cdot\|_1$ et $K = \nabla$, donc $F(\nabla u) = \text{TV}(u) = \|\nabla u\|_1$, et $G(u) = \lambda \frac{\|u - g\|_2^2}{2\sigma^2}$. On utilise alors la notion de dualité en considérant le conjugué de F , noté F^* , donné par

$$F^*(u) = \sup_{x \in \mathbb{R}^N} \langle u, x \rangle - F(x). \quad (4.27)$$

D'après la section 4.6.5, le problème (4.17) peut alors s'écrire sous la forme primale-duale suivante :

$$\min_{u \in \mathbb{R}^N} \max_{y \in \mathbb{R}^{N,2}} \langle u, -\text{div } y \rangle + \lambda \frac{\|u - g\|_2^2}{2\sigma^2} - \chi_{\mathcal{K}}(y), \quad (4.28)$$

où $\chi_{\mathcal{K}}$ désigne la fonction indicatrice du convexe $\mathcal{K} = \{y \in \mathbb{R}^{N,2}, |y_i| \leq 1\}$.

La résolution fait donc intervenir une étape (duale) de maximisation sur y et une étape (primale) de minimisation sur u . Elle est donnée dans l'algorithme 5, où $\nabla^* = -\text{div}$ désigne l'adjoint du gradient, et les termes $(I + \gamma \partial F^*)^{-1}$ et $(I + \tau \partial G)^{-1}$ sont les opérateurs proximaux associés aux fonctions F^* et G .

L'opérateur proximal associé à F^* est la projection sur l'ensemble $\mathcal{K} = \{y \in \mathbb{R}^{N,2}, |y_i| \leq 1\}$, exprimée par le seuillage doux [Chambolle 2011] :

$$y = (I + \gamma \partial F^*)^{-1}(\tilde{y}) \Leftrightarrow y_i = \frac{\tilde{y}_i}{\max(1, |\tilde{y}_i|)}, \quad (4.29)$$

et celui associé à l'attache aux données G est [Chambolle 2011] :

$$u = (I + \tau \partial G)^{-1}(\tilde{u}) \Leftrightarrow u_i = \frac{\tilde{u}_i + \tau/\sigma^2 \lambda g_i}{1 + \tau/\sigma^2 \lambda}. \quad (4.30)$$

4.3.3 Adaptation à d'autres types de bruit

Les méthodes de résolution proposées ci-dessus ont été mises en place et implémentées dans le cas (le plus simple !) d'un bruit additif gaussien. Nous nous intéressons désormais au cas où l'image est perturbée par un bruit non Gaussien, et nous considérons les adaptations possibles du modèle et des algorithmes de résolution présentés ci-dessus.

La première méthode, la plus naïve, consiste à transformer le bruit en bruit additif gaussien en stabilisant sa variance, conformément à la section 2.2.8. Comme il a été mentionné précédemment, cette technique présente l'inconvénient majeur de ne pas être applicable à toutes les distributions de bruit. De plus, elle peut se révéler inexacte de part la définition de la transformation inverse ou l'approximation par du bruit gaussien.

D'autre part, le modèle (4.17) est particulièrement adapté au modèle de bruit gaussien car il peut s'interpréter d'un point de vue bayésien comme un maximum a posteriori avec un terme d'attache aux données correspondant à la log-vraisemblance des observations, avec un modèle a priori TV sur l'image. Ce modèle peut donc être étendu à d'autres types de bruits en utilisant une énergie de la forme :

$$u^{\text{TV}} = \arg \min_{u \in \mathbb{R}^N} -\lambda \log p(g|u) + \text{TV}(u) \quad (4.31)$$

où $p(g_i|u_i)$ est la vraisemblance conditionnelle de la vraie valeur du pixel u_i sachant l'observation de la valeur bruitée g_i .

La prise en compte du modèle de bruit à travers une approche bayésienne intervient également dans les approches basées parcimonie, dans [Argenti 2002, Argenti 2006] pour la réduction de bruit de type *speckle*, ou dans [Harmany 2009, Raginsky 2010] pour l'adaptation à du bruit de Poisson.

4.3.3.1 Cas du bruit de Poisson

Le modèle de ROF peut donc s'adapter au cas de bruit de Poisson en modifiant directement le terme d'attache aux données afin de l'adapter au bruit poissonnien [Le 2007, Bardsley 2009, Figueiredo 2009, Willett 2010]. En effet, on rappelle (voir section 2.2.3) que la vraisemblance de l'image f étant donnés les observations g et

le paramètre $Q \in \mathbb{N}$ est donnée par :

$$p(g|f) = \sum_{i \in \Omega} \frac{f_i^{\frac{g_i}{Q}}}{(g_i/Q)!} \exp\left(-\frac{f_i}{Q}\right), \quad (4.32)$$

où Ω est le domaine de l'image et g/Q désigne un entier positif. Il vient aisément la log-vraisemblance négative :

$$\mathcal{L}(g|f) = \sum_{i \in \Omega} \frac{f_i}{Q} - \frac{g_i}{Q} \log \frac{f_i}{Q} + \log(g_i/Q)!. \quad (4.33)$$

Ainsi, en ne gardant que les termes dépendant de l'inconnue f , maximiser la log-vraisemblance revient à minimiser la fonctionnelle suivante :

$$\min_{u \in \mathbb{R}^N} G(u) = \sum_{i \in \Omega} -\frac{g_i}{Q} \log \frac{u_i}{Q} + \frac{u_i}{Q} + \chi_{\mathbb{R}^+}(u_i), \quad (4.34)$$

où $\chi_{\mathbb{R}^+}(u_i) = \begin{cases} 0 & \text{si } u_i \geq 0 \\ +\infty & \text{sinon} \end{cases}$ permet d'assurer la positivité (non stricte) de la solution ; la valeur zéro est possible et même probable en bruit de Poisson : cela correspond au cas où aucun photon n'est détecté. Le paramètre Q pourrait être mis en facteur dans la log-vraisemblance et négligé dans la minimisation, mais la prise en compte des constantes permet de réduire l'influence du bruit dans le choix du paramètre de régularisation λ , qui peut ainsi être maintenu constant quelle que soit la distribution et la force du bruit.

Étude de la fonctionnelle

Le problème de minimisation se réécrit donc de la façon suivante :

$$u^{\text{TV}} = \arg \min_{u \geq 0} \lambda \sum_{i \in \Omega} \left(\frac{u_i}{Q} - \frac{g_i}{Q} \log \frac{u_i}{Q} \right) + \text{TV}(u) \quad (4.35)$$

Nous notons tout d'abord que la fonction $u \mapsto \lambda \sum_{i \in \Omega} \left(\frac{u_i}{Q} - \frac{g_i}{Q} \log \frac{u_i}{Q} \right) + \text{TV}(u)$ est convexe, donc l'existence d'un unique minimiseur est garantie. Cependant, l'attache aux données $G(u) = \sum_{i \in \Omega} \frac{u_i}{Q} - \frac{g_i}{Q} \log \frac{u_i}{Q}$ n'est pas différentiable en zéro. Cela nécessite donc pour les algorithmes de descente de gradient ou forward-backward de régulariser celle-ci. C'est pourquoi nous privilégions l'algorithme primal-dual de Chambolle-Pock, qui ne nécessite pas forcément que la fonction G soit différentiable.

Algorithme primal-dual

À partir de la fonctionnelle définie ci-dessus, le problème (4.35) se réécrit sous la forme :

$$\min_{u \in \mathbb{R}^N} F(Ku) + G(u) \quad (4.36)$$

avec $G(u) = \lambda \sum_{i \in \Omega} \frac{u_i}{Q} - \frac{g_i}{Q} \log \frac{u_i}{Q} + \chi_{\mathbb{R}^+}(u_i)$ et $F(Ku) = \|\nabla u\|_1$, où $K = \nabla$ et $F = \|\cdot\|_1$.

Le problème peut alors se résoudre à l'aide de l'algorithme de Chambolle-Pock, en calculant les résolvantes $(I + \gamma \partial F^*)^{-1}$ et $(I + \tau \partial G)^{-1}$.

Le calcul de $(I + \gamma \partial F^*)^{-1}$ est identique à celui effectué dans le cas gaussien, et on peut déterminer le proximal associé à G . Le calcul est détaillé dans [Anthoine 2012], et on a la formule suivante :

$$y = (I + \tau \partial G)^{-1}(\tilde{y}) \Leftrightarrow y_i = \begin{cases} \frac{1}{2} \left(\tilde{y}_i - \tau \frac{\lambda}{Q} + \sqrt{\left(\tilde{y}_i - \tau \frac{\lambda}{Q} \right)^2 + 4\tau \frac{\lambda}{Q} g_i} \right) & \text{si } g_i > 0, \\ \max(\tilde{y}_i - \tau \frac{\lambda}{Q}, 0) & \text{si } g_i = 0. \end{cases} \quad (4.37)$$

4.3.3.2 Cas de la loi gamma

Nous nous intéressons également au cas d'un bruit multiplicatif distribué selon une loi gamma. Nous avons vu précédemment (voir section 2.2.4) que la vraisemblance de l'image f étant données les observations g est donnée par :

$$p(g|f) = \sum_{i \in \Omega} \frac{L^L g_i^{L-1} e^{-L \frac{g_i}{f_i}}}{\Gamma(L) f_i^L}, \quad (4.38)$$

donc la log-vraisemblance négative a la forme suivante :

$$\mathcal{L}(g|f) = \sum_{i \in \Omega} \frac{L g_i}{f_i} + L \log f_i + \log(\Gamma(L)) - L \log L - (L-1) \log g_i. \quad (4.39)$$

Ainsi, la fonctionnelle à minimiser associée au maximum de la log-vraisemblance est donnée par la formule :

$$\min_{u \in \mathbb{R}^N} G(u) = \sum_{i \in \Omega} L \log u_i + L \frac{g_i}{u_i} + \chi(u_i > 0). \quad (4.40)$$

Le paramètre L est une constante qui pourrait être mise en facteur, mais sa prise en compte permet de diminuer la sensibilité du paramètre de régularisation λ à la puissance du bruit.

Ainsi, le problème de minimisation dans le cas gamma s'écrit [Aubert 2008] :

$$u^{\text{TV}} = \arg \min_{u > 0} \lambda \sum_{i \in \Omega} \left(L \frac{g_i}{u_i} + L \log u_i \right) + \text{TV}(u). \quad (4.41)$$

La loi gamma est souvent observée dans le domaine radar. Elle décrit le comportement du bruit observé sur l'intensité du signal (c'est-à-dire sur le module au carré du signal radar complexe), tandis que sur l'amplitude du signal le bruit est décrit par une loi de Rayleigh. Afin d'être cohérents avec ces configurations, on considère souvent que la loi gamma est appliquée sur le module au carré de l'image, tandis que la variation totale elle s'applique à l'image. Le problème s'écrit alors sous la forme :

$$u^{\text{TV}} = \arg \min_{u > 0} \lambda \sum_{i \in \Omega} \left(L \frac{g_i}{u_i^2} + 2L \log u_i \right) + \text{TV}(u). \quad (4.42)$$

Étude de la fonctionnelle

Contrairement aux cas gaussien et poissonnien, la fonctionnelle ci-dessus n'est pas convexe. Il n'y a donc aucune garantie de convergence des algorithmes vers un unique minimiseur. Cependant, on peut montrer qu'un algorithme de minimisation converge vers un point stationnaire [Aubert 2008]. De plus, en initialisant l'algorithme de minimisation à l'image bruitée, ce dernier converge vers un minimum local qui reste fidèle à l'observation. Un algorithme de type Graph-cut peut également garantir de converger vers le minimum global, mais la complexité mémoire et les temps de calculs n'en font pas un choix privilégié.

On note de plus que l'attache aux données $G(u) = \sum_{i \in \Omega} L \frac{g_i}{u_i} + 2L \log u_i$ est différentiable pour tout $u > 0$. De plus, en considérant que $u > \alpha > 0$ alors G est à gradient Lipschitz (mais avec une constante malheureusement peu exploitable).

Faute de bénéficier d'un algorithme rapide et sûr, les algorithmes de minimisation étudiés dans le cas gaussien ont été adaptés afin de sélectionner le plus robuste.

Descente de gradient

L'attache aux données G étant différentiable, la méthode de descente de gradient avec la variation totale régularisée peut s'appliquer. Cependant, cette méthode est lente à converger et réalise une approximation de la variation totale.

Algorithme forward-backward

L'attache aux données G ainsi que la fonction de régularisation TV présentent les conditions requises pour l'application d'un algorithme de type forward-backward, qui converge alors vers un minimum local [Attouch 2013]. Cependant, on ne peut calculer explicitement de borne pour le pas γ qui dépend de la constante de Lipschitz associée à l'attache aux données via ∇G . Les paramètres de cet algorithme sont donc difficiles à ajuster pour garantir une convergence optimale. On observe les mêmes comportements avec le *Generalized Forward Backward* [Raguet 2013] décrit dans la section 4.6.4, avec cependant un temps de convergence (en nombre d'itérations) beaucoup plus long.

4.3.4 Limites et discussion

La minimisation de la variation totale a été largement utilisée pour la résolution de nombreux problèmes comme la restauration d'images [Chan 2000, Chambolle 2004], la déconvolution [Chan 1998], l'inpainting [Chan 2005b], l'interpolation [Guichard 1998], ou encore la décomposition d'images en composantes cartoon/textures [Osher 2003, Aujol 2005]. De nombreuses solutions d'implémentation ont également été proposées [Aujol 2009, Chambolle 2004, Darbon 2005, Weiss 2009].

Les auteurs de [Strong 2003] caractérisent les propriétés du modèle TV, en particulier sa capacité à préserver la localisation des contours et les structures

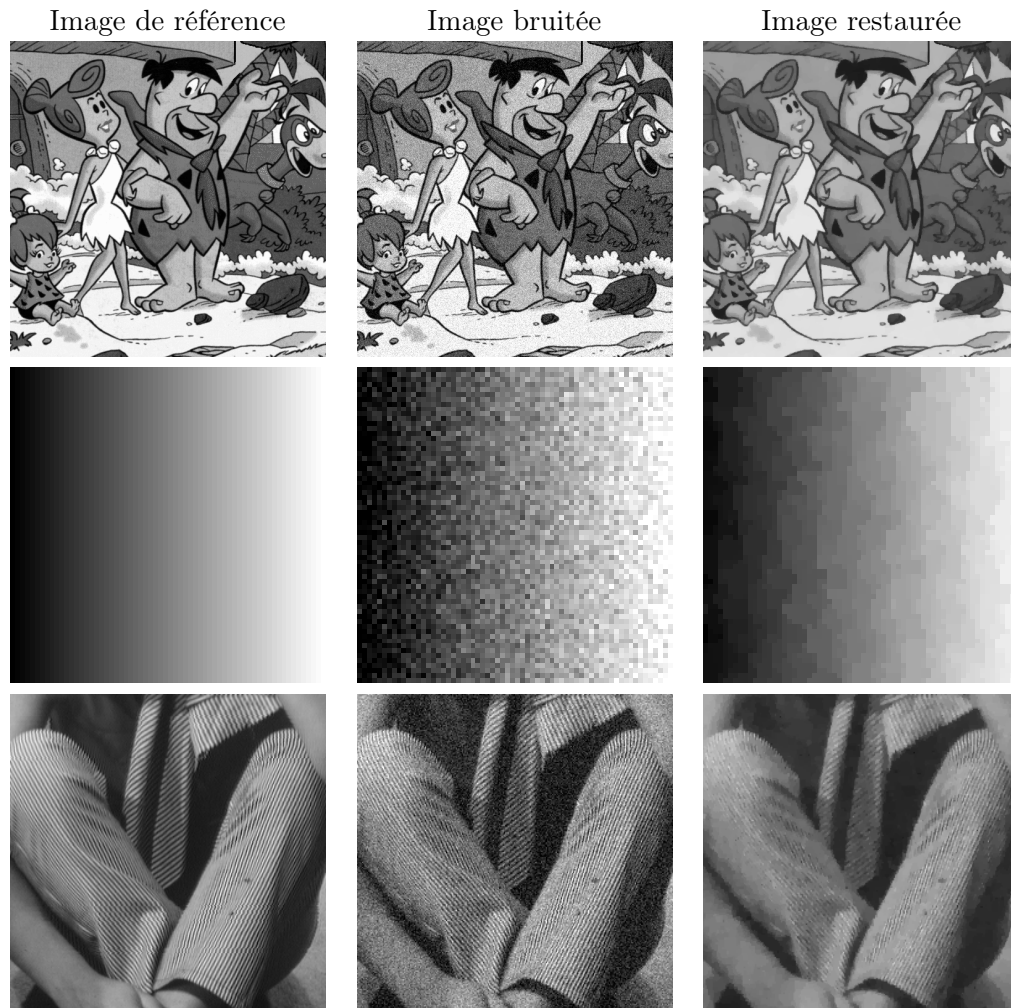


FIGURE 4.2 – Illustration du débruitage par minimisation TV d’une image de type cartoon, d’une image très lisse et d’une texture.

de grande taille, et le caractère plutôt local du filtre. Minimiser la variation totale force la solution à être régulière par morceaux, ce qui permet de débruiter tout en préservant les contours. Cependant, si la variation totale est adaptée aux images constantes par blocs, elle ne permet pas une bonne restauration des textures [Dobson 1996, Gousseau 2001]. De plus, un compromis doit être établi entre régularité sur les zones uniformes et préservation des textures. Celui-ci est basé sur le choix du paramètre de régularisation λ . Afin que les zones lisses soient proprement débruitées, λ doit être suffisamment faible, ce qui a tendance à sur-lisser les textures fines. En revanche, préserver les structures fines requiert un λ plus fort qui favorise l’attache aux données au détriment de la régularité des zones lisses, ce qui crée un effet de crénelage et une perte de contraste. Ce compromis rend le choix du paramètre de régularisation λ sensible et fortement dépendant du contenu de l’image [Hansen 1993, Ramani 2008]. Les auteurs de [Gilboa 2006] montrent qu’un

équilibre spatial peut être atteint en adaptant localement la valeur du paramètre de régularisation à la variance locale de l'image, afin de lisser davantage les surfaces lisses tout en préservant les textures.

On associe donc principalement trois défauts majeurs liés à la minimisation de la variation totale : l'approximation des surfaces par un modèle constant par morceaux entraîne sur les zones uniformes un effet de crénelage souvent désigné par *staircasing effect*, les textures sont souvent sur-lissées et l'image solution souffre globalement d'une baisse de contraste. La figure 4.2 illustre ces caractéristiques sur trois types d'images. Sur la première ligne, l'image *Flinstones* est de type cartoon ; elle peut être représentée finement par un modèle constant par morceaux, ce qui permet à la régularisation TV d'effectuer un débruitage adapté. Sur la deuxième ligne, l'image est très lisse ; on observe sur l'image restaurée l'effet de crénelage lié à l'approximation des zones uniformes par un modèle constant par morceaux. Enfin sur la dernière ligne, la texture du pantalon de *Barbara* a tendance à être sur-lissée, et on note par rapport à l'image de référence une baisse de contraste générale.

4.4 Les méthodes non locales

L'interprétation sous forme d'un filtre moyenneur (4.6) du filtre gaussien a conduit dans les années 90 aux filtres de type bilatéral qui combinent information spatiale et colorimétrique [Yaroslavsky 1985, Tomasi 1998, Smith 1997]. L'idée consiste à moyenner les pixels qui sont proches spatialement, et dont l'intensité est proche également, en espérant ainsi moyenner des structures similaires et donc limiter l'effet de flou au niveau des contours. La solution peut s'écrire de la façon suivante :

$$\hat{u}_i = \frac{\sum_{j \in \Omega} w_{ij} g_j}{\sum_{j \in \Omega} w_{ij}}, \quad \text{avec} \quad w_{ij} = \exp\left(-\frac{(i-j)^2}{2h^2}\right) \times \exp\left(-\frac{(g_i - g_j)^2}{2s^2}\right), \quad (4.43)$$

où h et s sont deux paramètres de filtrage. Ces filtres sont à l'origine des filtres non locaux comme les NL-means présentés dans la section suivante.

4.4.1 L'algorithme des NL-means

L'une des méthodes récentes de débruitage les plus populaires est l'algorithme des moyennes non locales (*NL-means*) proposé par Buades *et al.* dans [Buades 2005b]. Celui-ci se base sur la redondance naturelle des structures d'une image, localement mais aussi sur l'ensemble de l'image. Au lieu de moyenner seulement les pixels proches spatialement, l'algorithme des NL-means compare des patches, c'est-à-dire des petites fenêtres extraites autour de chaque pixel, afin de moyenner les pixels dont les voisinages sont similaires, comme illustré sur la figure 4.3. Pour chaque pixel $i \in \Omega$, la solution des NL-means est donnée par la moyenne pondérée suivante :

$$u_i^{\text{NL}} = \sum_{j \in \Omega} w_{i,j}^{\text{NL}} g_j, \quad (4.44)$$



FIGURE 4.3 – Illustration du principe de redondance et de la sélection de patches similaires.

équivalente à la solution du problème de minimisation suivant :

$$u^{\text{NL}} = \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j}^{\text{NL}} (g_j - u_i)^2. \quad (4.45)$$

Dans les deux cas, les poids $w_{i,j}^{\text{NL}} \in [0, 1]$ sont calculés afin de sélectionner les pixels j dont le voisinage ρ_j est similaire au patch ρ_i extrait autour du pixel d'intérêt i de la façon suivante :

$$w_{i,j}^{\text{NL}} = \frac{1}{Z_i} \varphi [d(g^{\rho_i}, g^{\rho_j})], \quad (4.46)$$

où $Z_i = \sum_{j \in \Omega} w_{i,j}^{\text{NL}}$ est une constante de normalisation qui veille à ce que les poids somment à 1, φ est une fonction décroissante et d une mesure de distance qui évalue la dissimilarité entre deux patches ρ_i et ρ_j de taille $|\rho|$ extraits respectivement autour des pixels d'indice i et j . De plus en pratique pour des raisons évidentes de temps de calcul, le calcul des poids est souvent restreint pour un pixel i à une fenêtre de recherche W_i centrée en i de sorte que $w_{i,j}^{\text{NL}} = 0$ pour $j \notin W_i$.

Comme illustré sur la figure 4.3, la sélection des pixels intervenant dans la moyenne pondérée est basée sur la similarité des voisinages, ce qui permet de moyenner des pixels issus de structures semblables. Cela permet d'éviter le phénomène de diffusion ou étalement des contours, donc de mieux les préserver.

4.4.1.1 Choix des paramètres

Les moyennes non locales font intervenir beaucoup de paramètres ; c'est à la fois une faiblesse car l'optimisation de ces derniers peut être délicate, mais aussi une force car cela laisse une place à l'adaptativité et à des améliorations possibles.

Le noyau φ

Le choix du noyau φ et de la mesure de dissimilarité d sont évidemment primordiaux. Dans l'algorithme d'origine proposé dans [Buades 2005b], φ est un noyau exponentiel décroissant donné par $\varphi(d) = \exp(-d/h^2)$, où h est un paramètre de filtrage. D'autres choix ont été également envisagés, par exemple l'utilisation de fonctions à support compact, qui peut fournir de meilleurs résultats [Goossens 2008] et réduire les temps de calcul.

La mesure de dissimilarité d

La mesure de distance d entre patchs est définie dans [Buades 2005b] comme la distance euclidienne convoluée à un noyau gaussien de largeur a , où a contrôle l'influence des pixels autour du pixel central ; quand $a \rightarrow 0$, seul le pixel central est pris en compte, tandis que si $a \rightarrow +\infty$, tous les pixels de l'image ont la même influence sur le calcul de la distance entre patchs. Le noyau gaussien est souvent abandonné dans les adaptations ultérieures des NL-means, afin de s'affranchir notamment d'un paramètre supplémentaire et d'un surcoût calculatoire.

Une amélioration notable des NL-means peut être obtenue en calculant la dissimilarité sur des patchs préfiltrés ou issus d'une image préfiltrée, ce qui permet de rendre le calcul des poids moins sensible au bruit, surtout si ce dernier est élevé. C'est le principe des NL-means itératifs [Brox 2007, Goossens 2008, Lebrun 2013].

De plus, de façon analogue au modèle de ROF le choix de la distance euclidienne fait écho à la nature gaussienne du bruit ; il peut donc être envisagé d'adapter cette mesure de dissimilarité lorsque le bruit n'est plus de nature additive gaussienne [Kervrann 2007, Delon 2012, Deledalle 2012].

Le paramètre de filtrage h

Outre le choix du noyau φ et de la mesure de dissimilarité d , les trois paramètres principaux qui influent sur les performances des NL-means sont le paramètre de filtrage h , qui règle la force du lissage, la taille des patchs $|\rho|$, qui influence sur la sélectivité des candidats, et la fenêtre de recherche W qui joue sur le temps de calcul mais également sur le caractère non local du filtre et la sélectivité des candidats. La figure 4.4 illustre l'influence de chacun de ces paramètres, lorsqu'ils sont choisis trop faibles ou trop élevés par rapport à la valeur qui garantit le débruitage optimal (visuellement).

Le paramètre de filtrage h joue un rôle sur la force du lissage appliqué à l'image. En effet lorsque h tend vers zéro, d'après la paramétrisation du noyau φ l'exponentielle tend vers zéro, ce qui conduit à des poids très faibles (en dehors du pixel



FIGURE 4.4 – Influence des paramètres des NL-means

central) et donc peu de débruitage. Quand au contraire h tend vers l'infini, les poids sont tous égaux et cela revient à moyenniser tous les pixels alentour. Ce comportement est illustré sur la première colonne de la figure 4.4. Le réglage de ce dernier est délicat et pourtant primordial. Il est souvent fixé pour toute l'image à l'aide d'une validation empirique sur un jeu de données synthétiques. Il est choisi automatiquement dans [Kervrann 2006] en fonction des quantiles de la distribution du χ_2 (qui provient de la distribution de la distance euclidienne entre deux patches gaussiens), ou dans [Van-De-Ville 2009] à l'aide d'un estimateur de risque. Le paramètre de filtrage est adapté localement en se basant sur l'estimation du risque également dans [Duval 2011].

La taille des patches $|\rho|$

La taille des patches $|\rho|$ joue sur la sélectivité des candidats et est liée à la redondance naturelle des structures au sein d’une image, en fonction de la résolution et de la taille des motifs de l’image. L’influence de ce paramètre peut être observé sur la deuxième colonne de la figure 4.4. Si la taille de patches est trop petite le filtre des NL-means tend vers le filtre bilatéral, puisqu’on ne compare plus des patches mais plutôt des pixels. Si elle est trop grande, le calcul des poids devient trop sélectif, ce qui produit un débruitage insuffisant autour des contours, appelé “effet de patch rare”.

La largeur des patches est en général fixée à 5,7 ou 9, et elle est fixée sur toute l’image. Dans [Deledalle 2011], la taille et la forme des patches est adaptée localement à la géométrie des structures de l’image, en se basant sur une estimation du risque. Cela permet de réduire efficacement l’effet de patch rare.

La fenêtre de recherche W

La recherche de patches similaires est restreinte à une fenêtre de recherche est justifiée avant tout par la réduction du temps de calcul, mais permet cependant d’obtenir de meilleurs résultats [Salmon 2010a]. Des études ont également montré qu’adapter automatiquement la taille de la fenêtre de recherche au contenu local [Kervrann 2006] peut être intéressant en terme de performances également.

4.4.2 Implémentation

Conformément à ce qui est proposé dans [Sutour 2014b], le noyau utilisé dans cette étude est le suivant :

$$\varphi[d] = \exp\left(-\frac{|d - m_d^\rho|}{s_d^\rho}\right) \quad (4.47)$$

où $m_d^\rho = \mathbb{E}[d(G^{\rho_i}, G^{\rho_i})]$ et $s_d^\rho = \sqrt{\text{Var}[d(G^{\rho_i}, G^{\rho_i})]}$ sont respectivement la valeur moyenne et l’écart type de la dissimilarité d calculée entre deux patches de taille $|\rho|$ identiquement distribués. Cela permet de s’affranchir du paramètre de filtrage h évoqué ci-dessus, en réduisant la sensibilité à la nature et au niveau du bruit, au choix de la mesure de dissimilarité d ainsi qu’à la taille des patches.

Un tel noyau ne fournit pas un poids maximal lorsque la distance d est nulle, mais lorsque celle-ci est égale à la valeur moyenne m_d^ρ , qui correspond au cas de patches identiquement distribués. Cela revient à ne pas privilégier les patches trop similaires au patch central. Ces derniers correspondent en effet à une réalisation de bruit trop proche de celle du patch central, donc leur participation dans la moyenne pondérée renforce la présence du bruit au lieu d’en diminuer la variance. Notons que conformément à [Salmon 2010a] le patch central est traité séparément ; le poids central est fixé à $w_{ii} = 1$, ce qui permet de garantir qu’aucun autre patch ne bénéficie d’un poids supérieur.

Dans le cas du bruit gaussien, et comme il a été décrit ci-dessus, la dissimilarité d est basée sur la distance euclidienne entre patchs et prend en compte la variance σ^2 du bruit (connue ou estimée) de la façon suivante :

$$d(g^{\rho_i}, g^{\rho_j}) = \frac{1}{|\rho|} \frac{\|g^{\rho_i} - g^{\rho_j}\|^2}{2\sigma^2} = \frac{1}{|\rho|} \sum_{k=1}^{|\rho|} \frac{(g_k^{\rho_i} - g_k^{\rho_j})^2}{2\sigma^2}. \quad (4.48)$$

Cela permet alors de donner une formule explicite pour les statistiques m_d^ρ et s_d^ρ . En effet, soient g^{ρ_i} et g^{ρ_j} deux patchs de g de taille $|\rho|$ identiquement distribués, dont l'intensité sous-jacente est donnée par f^ρ . g^{ρ_i} et g^{ρ_j} sont donc deux réalisations indépendantes de la variable aléatoire G^ρ , distribuée selon une loi normale de moyenne f^ρ et de variance σ^2 . Ainsi, $y^\rho = g^{\rho_i} - g^{\rho_j}$ est une réalisation de la variable Y^ρ , distribuée selon une loi normale de moyenne nulle et de variance $2\sigma^2$, et $Z^\rho = \frac{Y^\rho}{\sqrt{2}\sigma} \sim \mathcal{N}(0, 1)$. On a alors:

$$d(g^{\rho_i}, g^{\rho_j}) = \frac{1}{|\rho|} \sum_{k=1}^{|\rho|} \frac{(g_k^{\rho_i} - g_k^{\rho_j})^2}{2\sigma^2} = \frac{1}{|\rho|} \sum_{k=1}^{|\rho|} \frac{(y_k^\rho)^2}{2\sigma^2} = \frac{1}{|\rho|} \sum_{k=1}^{|\rho|} \left(\frac{y_k^\rho}{\sqrt{2}\sigma} \right)^2. \quad (4.49)$$

La quantité $\sum_{k=1}^{|\rho|} \left(\frac{Y_k^\rho}{\sqrt{2}\sigma} \right)^2 = \sum_{k=1}^{|\rho|} (Z^\rho)^2$ est distribuée selon une loi du χ_2 de paramètre $|\rho|$:

$$\sum_{k=1}^{|\rho|} (Z^\rho)^2 \sim \chi_2(|\rho|), \quad (4.50)$$

dont l'espérance et la variance sont respectivement égales à $|\rho|$ et à $2|\rho|$. Ainsi,

$$\mathbb{E}[d] = \frac{1}{|\rho|} \mathbb{E} \left[\sum_{k=1}^{|\rho|} (Z^\rho)^2 \right] = 1 \text{ et } \text{Var}[d] = \frac{1}{|\rho|^2} \text{Var} \left[\sum_{k=1}^{|\rho|} (Z^\rho)^2 \right] = \frac{2}{|\rho|},$$

et donc :

$$m_d^\rho = 1 \quad \text{et} \quad s_d^\rho = \sqrt{\frac{2}{|\rho|}}. \quad (4.51)$$

De plus, conformément à l'idée de préfiltrer les patchs afin d'affiner le calcul des poids [Dabov 2007b, Goossens 2008], la dissimilarité est calculée sur des patchs lissés. Ils sont donc convolués au préalable par un noyau gaussien de rayon 1. Cela permet d'assurer une meilleure sélection des candidats, en revanche les calculs de m_d^ρ et s_d^ρ détaillés ci-dessus ne sont plus valables. Dans ce cas, ils sont estimés empiriquement et au préalable en générant des patchs bruités, convolués, identiquement distribués, et en calculant la moyenne et l'écart type empiriques de la dissimilarité d mesurée entre ces patchs.

Enfin, plutôt que de débruiter indépendamment chaque pixel de l'image, on utilise la méthode d'agrégation des blocs. Cela consiste à effectuer la moyenne pondérée non pas seulement au niveau du pixel central g_i mais sur l'ensemble du patch g^{ρ_i} . Ainsi, en chaque pixel g_i on bénéficie de plusieurs estimations $(\hat{u}_i^{\text{NL}})^k$ correspondant

aux valeurs issues des k patchs qui contiennent le pixel i . On obtient ainsi une estimation $\widehat{u}_i^{\text{NL}}$ en effectuant la moyenne de ces k estimateurs, ce qui fournit un résultat plus lisse. Et dans ce cas, la solution u^{NL} se réécrit [Deledalle 2015] :

$$u^{\text{NL}} = \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j}^{\text{NL}} \|g^{\rho_j} - u^{\rho_i}\|^2. \quad (4.52)$$

4.4.3 Adaptation à d'autres types de bruit

4.4.3.1 Sélection de candidats

Les filtres bilatéraux ou basés sur les moyennes non locales font intervenir un critère de sélection de candidats qui mesure la similarité (locale ou non) entre pixels. Dans le cas gaussien, on utilise en général la distance euclidienne. Cette dernière fournit un estimateur non biaisé qui offre un taux de réduction de la variance constant. Cette propriété n'est plus vérifiée si le bruit n'est plus de nature gaussienne, d'où la nécessité d'adapter le critère de sélection.

Plusieurs extensions sont alors possibles, détaillées notamment dans [Delon 2012, Deledalle 2012]. On retiendra en particulier le rapport de vraisemblance généralisé [Deledalle 2012], qui définit la dissimilarité entre deux patchs g^{ρ_i} et g^{ρ_j} par :

$$d(g^{\rho_i}, g^{\rho_j}) = -\log \frac{\sup_u p(g^{\rho_i} | f^{\rho_i} = u) p(g^{\rho_j} | f^{\rho_j} = u)}{\sup_u p(g^{\rho_i} | f^{\rho_i} = u) \sup_u p(g^{\rho_j} | f^{\rho_j} = u)}. \quad (4.53)$$

où f^{ρ_i} et f^{ρ_j} sont les valeurs sous-jacentes des patchs non bruités.

De plus, conformément au principe bayésien du maximum a posteriori qui permet d'adapter l'attache aux données du modèle de ROF à d'autres statistiques de bruit, les moyennes non locales peuvent d'écrire dans le cas de bruit quelconque non corrélé comme une estimation pondérée du maximum de vraisemblance [Polzehl 2006, Deledalle 2009] :

$$u^{\text{NL}} = \arg \min_{u \in \mathbb{R}^N} - \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j}^{\text{NL}} \log p(g_j | u_i). \quad (4.54)$$

On note que dans le cas gaussien, cela équivaut bien à la définition de la solution des NL-means sous forme d'une moyenne pondérée donnée dans (4.44).

4.4.3.2 Bruit de Poisson

En prenant en compte la log-vraisemblance associée au bruit de Poisson calculée dans (4.33), l'équation (4.54) se réécrit

$$u^{\text{NL}} = \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j}^{\text{NL}} \left(\frac{u_i}{Q} - \frac{g_j}{Q} \log \frac{u_i}{Q} \right). \quad (4.55)$$

On remarque en particulier que la minimisation de ce problème aboutit (à la constante Q près) à la moyenne pondérée donnée par (4.44).

De plus, la distance entre deux patchs g^{ρ_i} et g^{ρ_j} est alors donnée par :

$$d(g^{\rho_i}, g^{\rho_j}) = g^{\rho_i} \log g^{\rho_i} + g^{\rho_j} \log g^{\rho_j} - (g^{\rho_i} + g^{\rho_j}) \log \left(\frac{g^{\rho_i} + g^{\rho_j}}{2} \right). \quad (4.56)$$

4.4.3.3 Loi gamma

En prenant en compte la log-vraisemblance associée à la loi gamma calculée dans (4.39), l'équation (4.54) se réécrit

$$u^{\text{NL}} = \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \sum_{j \in \Omega} w_{i,j}^{\text{NL}} \left(\frac{g_j}{u_i} + \log u_i \right). \quad (4.57)$$

On remarque également que la minimisation de ce problème aboutit à la moyenne pondérée donnée par (4.44).

La distance entre deux patchs g^{ρ_i} et g^{ρ_j} est dans ce cas donnée par :

$$d(g^{\rho_i}, g^{\rho_j}) = 2 \log(g^{\rho_i} + g^{\rho_j}) - \log g^{\rho_i} - \log g^{\rho_j} - 2 \log 2. \quad (4.58)$$

D'autres extensions ont été envisagées, par exemple celle proposée dans [Kervrann 2007] et appliquée au bruit dépendant du signal dans [Coupé 2008, Zhong 2011]. L'approche de [Kervrann 2007] diffère des NL-means classiques car elle s'appuie sur un schéma de filtrage en deux étapes, et son application à d'autres distributions de bruit spécifiques peut nécessiter des modifications supplémentaires, par exemple l'ajout d'une étape de "moyenne a priori" dans [Zhong 2011]. Conformément à [Deledalle 2009, Deledalle 2012, Deledalle 2014], l'extension utilisée ici s'applique directement à de nombreux modèles de bruit, notamment le bruit de Poisson et le bruit gamma. De plus, elle correspond exactement aux NL-means classiques dans le cas du bruit gaussien.

Un défaut de la mesure de dissimilarité proposée dans (4.53) est son comportement dans des régimes discrets de type impulsionnel, auquel cas le critère proposé dans [Delon 2012] peut se révéler plus adéquat.

4.4.4 Extensions et discussion

Les moyennes non locales offrent de belles performances générales mais souffrent de deux défauts qui s'opposent. D'une part, l'algorithme peut sélectionner des candidats non pertinents, ce qui crée l'effet de gigue, ou *jittering*. En associant des patchs issus de structures sous-jacentes différentes, ce qui peut se produire en raison de la présence de bruit ou d'un faible contraste, les structures sont sur-lissées, ce qui crée visuellement un effet de flou voire des artefacts. D'autre part, autour des structures singulières et des contours, il peut être difficile de trouver assez de patchs similaires pour effectuer un débruitage suffisant, ce qui laisse un bruit résiduel appelé l'effet de patch rare. Ces deux problèmes s'opposent. Ils sont contrôlés à la fois par le paramètre de filtrage h , la taille de la fenêtre de recherche et la taille des patchs. Ces paramètres assurent le compromis biais-variance, dont une interprétation est proposée dans [Duval 2011]. La Figure 4.5 illustre ces défauts : autour de la tête de *Cameraman* et de l'appareil, l'effet de patch rare se manifeste par la présence de bruit résiduel : la redondance est faible en raison de la singularité du détail, ce qui ne permet pas d'effectuer un débruitage suffisant. Sur les bâtiments à l'arrière plan en revanche, le sur-lissage lié à la confusion entre différentes structures entraîne

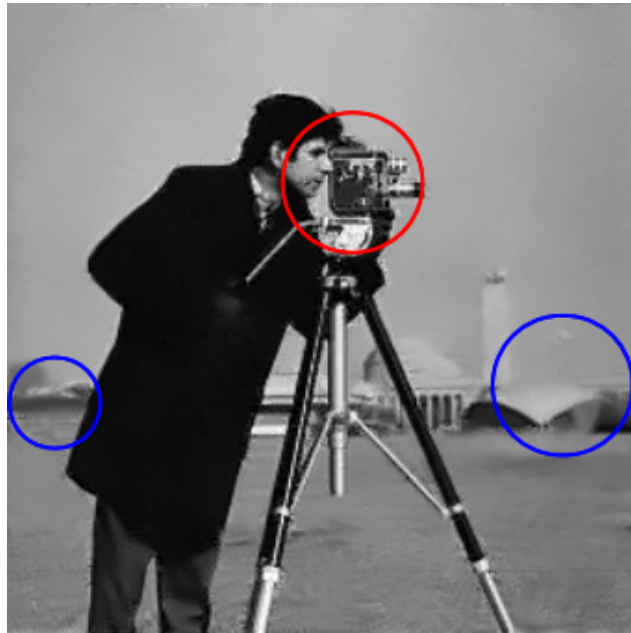


FIGURE 4.5 – Illustration des défauts des NL-means : l'effet de patch rare (cercle rouge), qui se manifeste par la présence de bruit résiduel, peut être observé autour de la tête et de l'appareil, tandis que l'effet de *jittering* (cercles bleus) est visible en arrière plan où les détails sont sur-lissés.

une perte de résolution et l'apparition d'artefacts. Ici aussi, un compromis doit être établi entre ces deux phénomènes, ce qui rend le choix des différents paramètres délicat et sensible au type d'image [Duval 2011, Van-de Ville 2011].

Le succès des moyennes non locales a conduit à de nombreuses adaptations et améliorations, par exemple l'interprétation bayésienne des NL-means de [Kervrann 2007]. Des méthodes d'implémentation ont également été proposées afin d'améliorer les temps de calcul [Mahmoudi 2005, Coupé 2006, Dauwe 2008, Darbon 2008]. Il est notamment possible de réduire la dimension de l'espace des patchs par apprentissage de dictionnaire [Azzabou 2007], SVD [Orchard 2008], ou à l'aide d'une analyse en composante principale (PCA) [Van-de Ville 2011] basée sur l'estimateur de risque SURE [Stein 1981]. L'algorithme *Patch match* [Barnes 2009] permet lui d'effectuer une recherche rapide des patchs candidats en effectuant une recherche aléatoire qui fournit une approximation des meilleurs candidats pour chaque pixel. Beaucoup d'études ont également porté sur l'adaptation des paramètres des NL-means [Duval 2011]. Dans [Kervrann 2008], la fenêtre de recherche est adaptée au contenu local de l'image afin de réaliser le meilleur compromis biais-variance. La taille des patchs est elle optimisée dans [Salmon 2010b], et la forme des patchs est adaptée localement aux structures de l'image à l'aide de l'estimateur de risque SURE dans [Deledalle 2011]. D'autre part, des versions itérative des NL-means permettent d'affiner le calcul de similarité entre patchs [Brox 2008, Deledalle 2009]. L'algorithme SAIF [Talebi 2013] propose également un algorithme itératif adaptatif localement

basé sur un estimateur de risque performant, qui fournit des performances proches de l'état de l'art.

D'autre part, les NL-Bayes [Lebrun 2013] proposent à l'aide d'une approche bayésienne une méthode simple et performante de débruitage basé sur les patches. Pour chaque patch ρ de l'image bruitée, on sélectionne les patches les plus similaires à ρ , puis une première estimation $\hat{\rho}^{\text{basic}}$ est obtenue à l'aide de l'estimation du maximum a posteriori déduit de la formule de Bayes de la façon suivante :

$$\hat{\rho}^{\text{basic}} = \bar{\rho} + [C_\rho - C_n] C_\rho^{-1} (\rho - \bar{\rho}), \quad (4.59)$$

où $\bar{\rho}$ est la moyenne des patches similaires à ρ , C_n est la matrice de covariance du bruit et C_ρ est la matrice de covariance des patches similaires à ρ . L'estimation est ensuite affinée en sélectionnant les patches similaires à partir des patches débruités $\hat{\rho}^{\text{basic}}$, ce qui fournit une estimation plus fine de la moyenne $\bar{\rho}^{\text{basic}}$ et de la matrice de covariance C_ρ^{basic} . La solution est alors la suivante :

$$\hat{\rho} = \bar{\rho}^{\text{basic}} + [C_\rho^{\text{basic}} - C_n] (C_\rho^{\text{basic}})^{-1} (\rho - \bar{\rho}^{\text{basic}}). \quad (4.60)$$

La valeur débruitée en chaque pixel i est alors obtenue en agrégeant les estimateurs obtenus pour chaque patch contenant le pixel i . Cette approche est présentée dans un cadre qui unifie l'ensemble des méthodes basées sur les patches et atteint l'état de l'art en restauration d'images.

Enfin, l'algorithme de débruitage BM3D (*Block Matching and 3D collaborative filter*) introduit dans [Dabov 2007b] associe filtrage collaboratif, représentation par patches et DCT. Le principe repose sur la construction de blocs 3D de patches similaires, qui sont ensuite décomposés sur la base des cosinus discrets et filtrés par le seuillage de Wiener en un schéma en deux étapes. Grâce à l'aspect collaboratif qui permet à chaque pixel de bénéficier de multiples estimations, il offre des performances peu égalées qui en font l'une des méthodes actuelles les plus puissantes de débruitage. Des méthodes récentes de débruitage réalisent également un filtrage collaboratif, en exploitant les similarités entre patches dans l'apprentissage de dictionnaire [Mairal 2009], ou en appliquant un filtre de Wiener sur des clusters de patches obtenus par une pré-segmentation [Chatterjee 2012].

De façon générale, les méthodes non locales dans leur ensemble sont susceptibles de souffrir des deux défauts inhérents à la comparaison de patches : l'effet de patch rare, lié à la difficulté de trouver suffisamment de redondance, et l'effet de *jittering*, lié à l'introduction de biais lorsque des populations différentes sont confondues. Le chapitre suivant se consacre à la réduction de ces deux défauts. Nous associons pour cela les moyennes non locales aux méthodes variationnelles, qui sont simples d'implémentation et facilement adaptables à des bruits non gaussiens, donc potentiellement intéressantes dans l'application au débruitage des images BNL.

4.5 Conclusion

Après avoir fourni un état de l'art des grandes familles de méthodes de débruitage, ce chapitre s'est consacré en particulier aux méthodes variationnelles et aux

méthodes non locales.

Les méthodes variationnelles cherchent à minimiser une énergie en forçant l'image solution à être régulière. L'une des méthodes variationnelles les plus connues est le modèle de Rudin, Osher et Fatemi (ROF) [Rudin 1992], qui minimise la variation totale (TV, pour *Total Variation*) de l'image, ce qui pousse l'image débruitée vers une solution constante par morceaux. Une telle méthode est adaptée pour préserver les contours lors du débruitage, cependant elle présente trois défauts majeurs : les textures fines ont tendance à être sur-lissées, les zones constantes sont approchées par des surfaces constantes par morceaux, ce qui crée un effet de crénelage (*staircasing effect*), et l'image souffre d'une baisse de contraste général. Une solution possible, proposée par exemple dans [Gilboa 2006], pour résoudre ces artefacts indésirables peut consister à adapter spatialement la régularisation.

D'autre part, l'algorithme des moyennes non locales (NL-means) [Buades 2005b] effectue un filtrage spatial en exploitant la redondance naturelle des structures au sein d'une image. Au lieu de moyenniser les pixels qui sont spatialement proches les uns des autres, les moyennes non locales comparent des patches, c'est-à-dire des petites fenêtres de pixels, extraits autour de chaque pixel, et effectuent une moyenne pondérée des pixels dont les voisinages sont similaires. La comparaison des voisinages permet de sélectionner des pixels susceptibles d'être issus de la même structure dans l'image. Ce principe a été largement repris et amélioré [Mahmoudi 2005, Kervrann 2006, Kervrann 2008, Van-de Ville 2011, Talebi 2013]. Ce genre de méthode offre de bonnes performances sur les zones homogènes et les textures répétitives où la redondance est élevée, en revanche sur les structures singulières trop peu de patches similaires risquent d'être sélectionnés, ce qui ne permet pas d'effectuer un débruitage suffisant. La présence de bruit résiduel dans ce cas est appelée l'effet de patch rare, ou *rare patch effect*, et a été notamment étudiée dans [Louchet 2011] et [Deledalle 2011]. De plus, la présence de bruit lors de la comparaison des patches peut conduire à de fausses détections, c'est-à-dire que des patches seraient considérés comme similaires alors qu'ils ne représentent pas la même structure sous-jacente. Cela conduit à moyenniser des pixels de différentes populations, créant un effet de sur-lissage [Kervrann 2008] appelé parfois *patch jittering blur effect* ou effet de *jittering* [Louchet 2011].

Les approches variationnelles et les méthodes non locales présentent des atouts communs requis pour le débruitage des séquences d'images BNL : elles sont simples d'implémentation et adaptables à différents types de bruit. De plus, leurs défauts diffèrent : l'effet de crénelage lié à la minimisation TV est situé sur les zones uniformes, où les NL-means fonctionnent bien, tandis que ces derniers souffrent davantage sur les zones singulières. Cela justifie la volonté de combiner ces deux méthodes, afin de limiter leurs défauts respectifs tout en tirant parti de leurs qualités. Nous proposons dans le chapitre suivant une régularisation adaptative des moyennes non locales qui associe les deux approches pour tirer le meilleur de chacune et réaliser un débruitage souple.

4.6 Annexes du chapitre : Éléments d'analyse convexe

4.6.1 Généralités

Soit \mathcal{X} un espace de Hilbert muni d'un produit scalaire $\langle \cdot, \cdot \rangle$ et de la norme associée $\| \cdot \|$. On désigne par C un sous-ensemble convexe, fermé, non vide de \mathcal{X} . On note Γ_0 l'espace des fonctions convexes de \mathcal{X} dans $] - \infty, +\infty]$, semi-continues inférieurement, non identiquement égales à $+\infty$.

Définition 5 (Projection) *La projection de $x \in \mathcal{X}$ sur C est l'unique point $\Pi_C x \in C$ qui vérifie $\|x - \Pi_C x\| = d_C x$. où $d_C x = \inf \|x - C\|$ est la distance de x à C . Le projeté vérifie la propriété fondamentale :*

$$\Pi_C x \in C \text{ et } \forall z \in C, \langle z - \Pi_C x | x - \Pi_C x \rangle \leq 0. \quad (4.61)$$

Définition 6 (Conjugué) *Soit $\varphi \in \Gamma_0$. Le conjugué de φ est donné pour tout $u \in \mathcal{X}$ par la relation :*

$$\varphi^*(u) = \sup_{x \in \mathcal{X}} \langle u, x \rangle - \varphi(x). \quad (4.62)$$

L'analyse convexe [Rockafellar 1970] nous fournit les propositions suivantes :

Proposition 4.6.1 *Soit $\varphi \in \Gamma_0$. Alors $\varphi^{**} = \varphi$.*

Proposition 4.6.2 *Soit $\varphi \in \Gamma_0$ une fonction homogène de degré 1, c'est-à-dire telle que $\varphi(\lambda u) = \lambda \varphi(u)$ pour tout $\lambda > 0$. Alors il existe un convexe fermé \mathcal{K} tel que*

$$\varphi^*(v) = \chi_{\mathcal{K}}(v) = \begin{cases} 0 & \text{si } v \in \mathcal{K} \\ +\infty & \text{sinon} \end{cases}.$$

Corollaire 4.6.3 *Soit $\varphi \in \Gamma_0$. Si φ est homogène de degré 1, alors :*

$$\varphi(x) = \varphi^{**}(x) = \sup_{u \in \mathcal{X}} \langle u, x \rangle - \varphi^*(u) = \sup_{u \in \mathcal{X}} \langle u, x \rangle - \chi_{\mathcal{K}}(u) = \sup_{u \in \mathcal{K}} \langle u, x \rangle.$$

Définition 7 (Sous-différentielle) *La sous-différentielle d'une fonction $\varphi \in \Gamma_0$ est donnée par :*

$$\begin{aligned} \partial\varphi(x) &= \{u \in \mathcal{X}, \forall y \in \mathcal{X}, \langle y - x, u \rangle + \varphi(x) \leq \varphi(y)\} \\ &= \{u \in \mathcal{X}, \varphi^*(u) + \varphi(x) \leq \langle x, u \rangle\}. \end{aligned}$$

La seconde identité s'obtient en passant au sup sur les y . Nous avons la proposition fondamentale suivante :

Proposition 4.6.4 *Soit $\varphi \in \Gamma_0$. Pour tout $x \in \mathcal{X}$, l'équivalence suivante est vérifiée :*

$$x \in \text{inf } \varphi \Leftrightarrow 0 \in \partial\varphi(x). \quad (4.63)$$

De plus, la définition du conjugué nous dit que pour tout $x \in \mathcal{X}$,

$$\varphi^*(u) \geq \langle u, x \rangle - \varphi(x). \quad (4.64)$$

Et d'après la définition de la sous-différentielle de φ permet d'écrire que :

$$u \in \partial\varphi(x) \Leftrightarrow \varphi^*(u) \leq \langle u, x \rangle - \varphi(x). \quad (4.65)$$

Ainsi, les relations 4.64 et 4.65 fournissent l'identité de Legendre-Fenchel :

Proposition 4.6.5 *Soit φ une fonction convexe. Alors*

$$u \in \partial\varphi(x) \Leftrightarrow \varphi^*(u) = \langle u, x \rangle - \varphi(x). \quad (4.66)$$

*De plus, si $\varphi \in \Gamma_0$, comme $\varphi = \varphi^{**}$, alors*

$$u \in \partial\varphi(x) \Leftrightarrow x \in \partial\varphi^*(u). \quad (4.67)$$

4.6.1.1 Opérateur proximal

Par analogie avec la projection $\Pi_C(x)$ de $x \in \mathcal{X}$ sur le convexe C qui peut s'écrire sous la forme

$$\Pi_C x = \arg \min_{y \in \mathcal{X}} \chi_C(y) + \frac{1}{2} \|x - y\|^2,$$

on définit l'opérateur proximal de la façon suivante.

Définition 8 (Opérateur proximal) *Soit $\varphi \in \Gamma_0$. L'opérateur proximal associé à la fonction φ , noté $\text{prox}_\varphi x$, est défini par :*

$$\text{prox}_\varphi x = \arg \min_{y \in \mathcal{X}} \varphi(y) + \frac{1}{2} \|x - y\|^2, \quad (4.68)$$

et de façon étendue

$$\text{prox}_{\gamma\varphi} x = \arg \min_{y \in \mathcal{X}} \varphi(y) + \frac{1}{2\gamma} \|x - y\|^2. \quad (4.69)$$

D'après la formule 4.69:

$$\begin{aligned} \text{prox}_{\gamma\varphi} x &= \arg \min_{y \in \mathcal{X}} \varphi(y) + \frac{1}{2\gamma} \|x - y\|^2 \\ \Leftrightarrow 0 &\in \partial \left(\varphi + \frac{1}{2\gamma} \|x - \cdot\|^2 \right) (y) \\ \Leftrightarrow 0 &\in \gamma \partial\varphi(y) + y - x, \end{aligned}$$

en utilisant le fait que $\partial(F + G) = \partial F + \partial G$, dès lors que F ou G est différentiable. Il vient donc que :

$$x - \text{prox}_{\gamma\varphi} x \in \gamma \partial\varphi(\text{prox}_{\gamma\varphi} x), \quad (4.70)$$

que l'on note

$$\text{prox}_{\gamma\varphi} x = (I + \gamma \partial\varphi)^{-1}(x). \quad (4.71)$$

Et en utilisant la formulation duale, on montre l'identité de Moreau :

Proposition 4.6.6 Soient $\varphi \in \Gamma_0$ et $x \in \mathcal{X}$. Alors :

$$x = (I + \gamma \partial \varphi)^{-1}(x) + \gamma \left(I + \frac{1}{\gamma} \partial \varphi^* \right)^{-1} \left(\frac{x}{\gamma} \right), \quad (4.72)$$

et en particulier avec $\gamma = 1$:

$$x = (I + \partial \varphi)^{-1}(x) + (I + \partial \varphi^*)^{-1}(x). \quad (4.73)$$

4.6.2 Méthode de Chambolle

Afin de conserver les algorithmes présentés ci-dessous dans une forme plus générale, nous notons désormais la fonction de régularisation $J \in \gamma_0$. Elle peut désigner la variation totale TV mais également d'autres fonctions de régularisation adaptées à d'autres types de problèmes. L'algorithme suivant a été proposé par Chambolle en 2004 dans [Chambolle 2004]. Il est possible de reformuler le problème (4.17) sous forme duale. En effet, il se caractérise de la façon suivante :

$$\begin{aligned} u &= \arg \min_{u \in \mathbb{R}^N} \lambda \frac{\|u - g\|_2^2}{2\sigma^2} + J(u) \\ \Leftrightarrow 0 &\in \frac{\lambda}{\sigma^2}(u - g) + \partial J(u) \\ \Leftrightarrow u &\in \partial J^* \left(\frac{\lambda}{\sigma^2}(g - u) \right) \\ \Leftrightarrow \frac{\lambda}{\sigma^2}g &\in \frac{\lambda}{\sigma^2}(g - u) + \frac{\lambda}{\sigma^2} \partial J^* \left(\frac{\lambda}{\sigma^2}(g - u) \right) \\ \Leftrightarrow 0 &\in w - \frac{\lambda}{\sigma^2}g + \frac{\lambda}{\sigma^2} \partial J^*(w), \text{ avec } w = \frac{\lambda}{\sigma^2}(g - u). \end{aligned}$$

Cela signifie que w minimise le problème suivant :

$$w = \arg \min_w F^*(w) = \arg \min_w \frac{1}{2} \left\| w - \frac{\lambda}{\sigma^2}g \right\|^2 + \frac{\lambda}{\sigma^2} J^*(w). \quad (4.74)$$

De plus, lorsque J désigne la variation totale, J^* peut se calculer explicitement. En effet, d'après [Chambolle 2004] :

$$\begin{aligned} J(u) &= TV(u) = \sum_{i=1}^N \|(\nabla u)_i\| = \|\nabla u\|_{2,1} = \sup_{p \in \mathbb{R}^{N,2}} \{ \langle p, \nabla u \rangle, |p_i| \leq 1 \} \\ &= \sup_{p \in \mathbb{R}^{N,2}} \{ -\langle \operatorname{div} p, u \rangle, |p_i| \leq 1 \} = \sup_{\xi \in \mathbb{R}^N} \langle u, \xi \rangle - \chi_{\mathcal{K}}(\xi). \end{aligned} \quad (4.75)$$

avec $\mathcal{K} = \{ \xi = \operatorname{div} p, p \in \mathbb{R}^{N,2}, |p_i| \leq 1 \}$, et $\chi_{\mathcal{K}}$ la fonction caractéristique du convexe \mathcal{K} qui vaut 0 sur \mathcal{K} et $+\infty$ en dehors de \mathcal{K} . On en déduit d'après la définition du conjugué que $J^* = \chi_{\mathcal{K}}$, et que w n'est autre que la projection sur \mathcal{K} de $\frac{\lambda}{\sigma^2}g$: $w = \Pi_{\mathcal{K}}(\frac{\lambda}{\sigma^2}g)$.

La minimisation du problème original (4.17) se ramène donc à un calcul de projection sur un convexe \mathcal{K} , puis

$$u = g - \frac{\sigma^2}{\lambda} w = g - \Pi_{\sigma^2/\lambda\mathcal{K}}(g). \quad (4.76)$$

Et finalement, minimiser (4.74) dans le cas de la variation totale se traduit par :

$$\begin{aligned} & \arg \min_{w \in \mathbb{R}^N} \frac{1}{2} \left\| w - \frac{\lambda}{\sigma^2} g \right\|^2 + \frac{\lambda}{\sigma^2} J^*(w) \\ &= \arg \min_{w \in \mathbb{R}^N} \frac{1}{2} \left\| w - \frac{\lambda}{\sigma^2} g \right\|^2 + \frac{1}{\lambda} \chi_{\mathcal{K}}(w) \\ &= \arg \min_{w \in \mathcal{K}} \frac{1}{2} \left\| w - \frac{\lambda}{\sigma^2} g \right\|^2 \\ &= \arg \min_{|p_i| \leq 1} \frac{1}{2} \left\| \operatorname{div} p - \frac{\lambda}{\sigma^2} g \right\|^2. \end{aligned} \quad (4.77)$$

Puis on récupère u à partir du w obtenu par $u = g - \frac{\sigma^2}{\lambda} w = g - \frac{\sigma^2}{\lambda} \operatorname{div} p$.

Le problème à résoudre se pose donc sous la forme suivante :

$$\min_{p \in \mathbb{R}^{N,2}} \left\{ \left\| \frac{\sigma^2}{\lambda} \operatorname{div} p - g \right\|^2, |p_i|^2 \leq 1 \right\}. \quad (4.78)$$

Les conditions de Karush-Kuhn-Tucker exhibent des multiplicateurs de Lagrange $\alpha_i \geq 0$ associés à chaque contrainte de l'équation (4.78) tels que :

$$\begin{cases} -\nabla \left(\frac{\sigma^2}{\lambda} \operatorname{div} p - g \right)_i + \alpha_i p_i = 0 \\ \alpha_i (|p_i|^2 - 1) = 0 \end{cases}. \quad (4.79)$$

Deux cas se distinguent : si $\alpha_i > 0$ alors $|p_i|^2 = 1$, ou alors $\alpha_i = 0$ et $|p_i|^2 < 1$. Dans chacun des cas, $\alpha_i = \left| \nabla \left(\frac{\sigma^2}{\lambda} \operatorname{div} p - g \right)_i \right|$. Cela conduit à un algorithme de descente de gradient semi-implicite (ou point fixe) :

$$p_i^{k+1} = p_i^k + \tau \left(\nabla \left(\frac{\sigma^2}{\lambda} \operatorname{div} p^k - g \right)_i - \left| \nabla \left(\frac{\sigma^2}{\lambda} \operatorname{div} p^k - g \right)_i \right| p_i^{k+1} \right),$$

avec $\tau > 0$ le pas de l'itération. L'algorithme se résout donc à l'aide des itérations suivantes :

$$p_i^{k+1} = \frac{p_i^k + \tau \left(\nabla \left(\frac{\sigma^2}{\lambda} \operatorname{div} p^k - g \right)_i \right)}{1 + \tau \left| \nabla \left(\frac{\sigma^2}{\lambda} \operatorname{div} p^k - g \right)_i \right|}. \quad (4.80)$$

4.6.3 Algorithme forward-backward

L'algorithme forward-backward [Combettes 2005] permet de résoudre des problèmes de minimisation de type

$$\min_x F(x) + G(x) \quad (4.81)$$

avec F et G deux fonctions convexes semi-continues inférieurement, avec G différentiable, et ∇G Lipschitz de constante $1/\beta$.

Le problème (4.81) peut se récrire de la façon suivante :

$$\begin{aligned} x \text{ solution de (4.81)} &\Leftrightarrow 0 \in \partial(F + G)(x) = \partial F(x) + \partial G(x) = \partial F(x) + \{\nabla G(x)\} \\ &\Leftrightarrow -\nabla G(x) \in \partial F(x) \\ &\Leftrightarrow (x - \tau \nabla G(x)) - x \in \tau \partial F(x) \text{ et d'après la formule (4.70):} \\ &\Leftrightarrow x = \text{prox}_{\tau F}(x - \tau \nabla G(x)). \end{aligned}$$

Cela permet de mettre en place une résolution itérative :

$$x^{k+1} = \text{prox}_{\tau_k F}(x^k - \tau_k \nabla G(x^k)) = (I + \tau_k \partial F)^{-1}(x^k - \tau_k \nabla G(x^k)). \quad (4.82)$$

Cet algorithme fait donc apparaître deux étapes :

1. Étape forward (explicite): calcul de $y^k = x^k - \tau_k \nabla G(x^k)$
2. Étape backward (implicite): $x^{k+1} = \text{prox}_{\tau_k F}(y^k)$

Ces deux étapes se résument sous la notation :

$$x^{k+1} = (I + \tau_k \partial F)^{-1}(x^k - \tau_k \nabla G(x^k)). \quad (4.83)$$

On peut enfin mettre en place l'algorithme de forward-backward avec une relaxation γ_k [Combettes 2010]:

$$\begin{cases} y^k &= x^k - \tau_k \nabla G(x^k) \\ x^{k+1} &= x^k + \gamma_k (\text{prox}_{\tau_k F} y^k - x^k) \end{cases} \quad (4.84)$$

Des accélérations proposées notamment par Beck et Teboulle [Beck 2010] permettent d'accélérer la convergence de l'algorithme, en optimisant en particulier le choix et la mise à jour des pas à chaque itération. L'algorithme se présente alors de la façon suivante :

$$\begin{cases} x^k &= (I + \tau \partial F)^{-1}(y^k - \tau \nabla G(y^k)), \\ t^{k+1} &= \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}, \\ y^{k+1} &= x^k + \frac{t^k - 1}{t^{k+1}}(x^k - x^{k+1}). \end{cases} \quad (4.85)$$

4.6.4 Generalized forward-backward

Une adaptation de l'algorithme forward-backward par Raguet *et al.* [Raguet 2013] permet de gérer le cas d'un problème sous la forme :

$$\min_u G(u) + \sum_{l=1}^{N_l} F_l(u). \quad (4.86)$$

La fonction G doit être différentiable à gradient Lipschitz et les fonctions F_l doivent être simples, c'est-à-dire que leur opérateur proximal peut être calculé explicitement. L'algorithme se présente sous la forme suivante :

$$\begin{aligned} \text{Pour } l \in [1, N_l], \quad z_l &= z_l + \tau \left(\text{prox}_{\frac{\gamma}{w_l} F_l} (2x - z_l - \gamma \nabla G(x)) - x \right), \\ x &= \sum_{l=1}^{N_l} w_l z_l. \end{aligned}$$

Dans le cas du problème de débruitage avec régularisation TV, on peut réécrire le problème 4.17 de la manière suivante :

$$\begin{aligned} & \arg \min_u \lambda \frac{\|u - g\|_2^2}{2\sigma^2} + \text{TV}(u) \\ &= \arg \min_u G(u) + \|\nabla u\|_{1,2} \\ &= \arg \min_u G(u) + \|x\|_{1,2} + \chi_C(u, x) \quad \text{où } C = \{(u, x), x = \nabla u\} \\ &= \arg \min_u G(u) + F(u, x) + H(u, x), \end{aligned} \tag{4.87}$$

avec $G(u) = \lambda \frac{\|u - g\|_2^2}{2\sigma^2}$ l'attache aux données (différentiable), $F(u, x) = \|x\|_1 = \sqrt{x_1^2 + x_2^2}$ et $H(u, x) = \chi_C(u, x)$ avec $C = \{(u, x), x = \nabla u\}$. On note que cela nécessite l'introduction d'une variable auxiliaire x .

L'opérateur proximal associé à la fonction F n'est autre qu'un seuillage doux :

$$\text{prox}_{\gamma F}(u, x) = (u, ST_\gamma(x)) \quad \text{avec} \quad ST_\gamma(x) = x - x \times \min\left(\frac{\gamma}{|x|}, 1\right) \tag{4.88}$$

L'opérateur proximal associé à la fonction $H(u, x) = \chi_C(u, x)$ se calcule également explicitement de la façon suivante :

$$\begin{aligned} \text{prox}_{\gamma H}(\tilde{u}, \tilde{x}) &= \arg \min_{u, x} \frac{\|u - \tilde{u}\|^2 + \|x - \tilde{x}\|^2}{2\lambda} + \chi_C(u, x) \\ &= \arg \min_{u, x = \nabla u} \frac{\|u - \tilde{u}\|^2 + \|\nabla u - \tilde{x}\|^2}{2\lambda}. \end{aligned}$$

En dérivant afin d'obtenir le minimiseur :

$$(u - \tilde{u}) + \nabla^t (\nabla u - \tilde{x}) = 0,$$

on obtient la solution suivante :

$$\begin{cases} u &= (I + \nabla^t \nabla)^{-1} (\tilde{u} + \nabla^t \tilde{x}) \\ x &= \nabla u \end{cases} . \tag{4.89}$$

L'inversion de la matrice $(I + \nabla^t \nabla)$ se calcule dans le domaine de Fourier. Cette technique est cependant assez lente à converger.

4.6.5 Algorithme primal-dual

L'algorithme de Chambolle-Pock [Chambolle 2011] permet de résoudre un problème de la forme

$$\min_x F(Kx) + G(x), \quad (4.90)$$

où F et G sont deux fonctions convexes de Γ_0 . En utilisant la dualité et le fait que $F(Kx) = F^{**}(Kx) = \sup_y \langle Kx, y \rangle - F^*(y)$, le problème se réécrit sous forme primale-duale [Chambolle 2011]:

$$\min_x \max_y \langle Kx, y \rangle - F^*(y) + G(x) \quad (4.91)$$

La formulation (4.91) peut également se traduire sous forme duale exclusivement :

$$\begin{aligned} & \min_x \max_y \langle Kx, y \rangle - F^*(y) + G(x) \\ &= \min_x \max_y -\langle -Kx, y \rangle + G(x) - F^*(y) \\ &= \max_y \min_x -\langle x, -K^*y \rangle + G(x) - F^*(y) \\ &= \max_y -G^*(-K^*y) - F^*(y) \end{aligned}$$

Le problème (4.91) se décompose alors en deux parties :

— Problème en y :

$$\begin{aligned} & \max_y \langle Kx, y \rangle - F^*(y) \\ & \Leftrightarrow \langle Kx, y \rangle - F^*(y) \geq \langle Kx, u \rangle - F^*(u) \quad \forall u \\ & \Leftrightarrow F^*(u) \geq \langle Kx, u - y \rangle + F^*(y), \text{ donc par définition de la sous-différentielle :} \\ & \Leftrightarrow Kx \in \partial F^*(y). \end{aligned} \quad (4.92)$$

— Problème en x :

$$\begin{aligned} & \min_x \langle Kx, y \rangle + G(x) \\ & \Leftrightarrow \langle Kx, y \rangle + G(x) \leq \langle Ku, y \rangle + G(u) \quad \forall u \\ & \Leftrightarrow \langle x, K^*y \rangle + G(x) \leq \langle u, K^*y \rangle + G(u) \quad \forall u \\ & \Leftrightarrow G(u) \geq G(x) + \langle -K^*y, u - x \rangle, \text{ et par définition de la sous-différentielle :} \\ & \Leftrightarrow -K^*y \in \partial G(x). \end{aligned} \quad (4.93)$$

Il vient une résolution itérative en deux étapes, primales-duales [Chambolle 2011]:

1. Étape duale : maximisation de (4.92)

$$y^{k+1} = (I + \gamma \partial F^*)^{-1} (y^k + \gamma K \bar{x}), \quad (4.94)$$

2. Étape primale : minimisation de (4.93)

$$x^{k+1} = (I + \tau \partial G)^{-1} (x^k - \tau K^* y^{k+1}), \quad (4.95)$$

3.

$$\bar{x}^{k+1} = x^{k+1} + \theta (x^{k+1} - x^k) \text{ avec } \theta \in [0, 1]. \quad (4.96)$$

Régularisation adaptative des moyennes non locales

5.1 Introduction

La simplicité des approches variationnelles et non locales étudiées dans le chapitre précédent ainsi que leurs qualités et défauts qui diffèrent les poussent à être combinées afin de tirer le meilleur de chacune.

Les méthodes variationnelles et non locales ont par exemple été associées en définissant des termes de régularisation non locale [Kindermann 2005, Gilboa 2008]. Ces approches imposent une régularité sur des ensembles de voisins formés par similarité entre patches. Cela permet de traiter séparément les zones texturées et les zones uniformes. Les auteurs de [Gilboa 2008] effectuent une régularisation non locale en définissant un gradient non local, qui permet de lisser les surfaces lisses tout en préservant les structures fines. Cette approche permet de s'affranchir de l'effet de crénelage associé à la minimisation de la variation totale, mais souffre néanmoins de l'effet de patch rare. La régularisation non locale a ensuite été entendue à de nombreux problèmes inverses, par exemple pour des problèmes d'inpainting et de compressive sensing dans [Peyré 2011] ou pour de la déconvolution dans [Zhang 2010b].

Les auteurs de [Louchet 2011] ont également proposé d'associer les NL-means à la régularisation TV dans l'algorithme TV-means. La variation totale est adaptée pour agir de manière plus locale, ce qui réduit l'effet de crénelage, puis la variation totale locale est utilisée dans la sélection de patches pour les moyennes non locales afin de réduire de réduire l'effet de patch rare, le tout sous un schéma itératif.

Les méthodes variationnelles et non locales ont également été associées pour résoudre des problèmes de super-résolution, par exemple dans [Protter 2009, d'Angelo 2011]. Au lieu d'une régularisation non locale, une énergie est minimisée à partir de la variation totale et d'une attache aux données non locales, pondérée par les poids issus des NL-means. Cela revient à minimiser la variation totale de la solution des NL-means. Cette approche a été appliquée à des problèmes de déconvolution dans [Mignotte 2008].

Les travaux présentés dans ce chapitre suivent une idée similaire et associent la variation totale aux moyennes non locales afin de réduire les défauts respectifs des deux approches. Dans un premier temps, l'effet de *jittering* est corrigé au sein de l'algorithme des NL-means, puis une régularisation localement adaptative est appliquée là où les NL-means n'ont pas pu assurer un débruitage suffisant, réduisant ainsi l'effet de patch rare. La première étape de *dejittering* réduit le biais au prix

d'une variance accrue. Cela garantit que le débruitage effectué par les NL-means est fiable, dans le sens où la sélection de candidats non pertinents est limitée. Un indice de confiance est alors basé sur le niveau de bruit résiduel après *dejittering*. Cette mesure de confiance vient pondérer la régularisation TV afin de créer un modèle adaptatif, conformément à ce qui est suggéré dans [Gilboa 2006]. Contrairement au modèle issu de [Protter 2009] et [d'Angelo 2011], la méthode proposée ici préserve la solution des NL-means quand celle-ci est estimée correcte, ce qui permet d'éviter d'introduire les phénomènes de crénelage, sur-lissage ou perte de contraste dont souffre la régularisation TV non adaptative. Sur les structures singulières en revanche, cette régularisation adaptative vient réduire le bruit résiduel. De plus, le modèle s'étend naturellement à différents modèles de bruit, grâce aux bonnes propriétés des NL-means et des méthodes variationnelles.

5.2 *Dejittering* des NL-means (NLDJ)

L'effet de *jittering* se manifeste par un sur-lissage lié à une trop forte réduction du bruit due à des mélanges de populations dans la sélection des candidats non locaux. La réduction de l'effet de *jittering* consiste donc à réduire le biais introduit par ce sur-lissage, c'est-à-dire à rétablir le compromis biais-variance. De façon générale, le debiasing peut être traité de façon itérative et réinjectant à la solution \hat{u} une version filtrée du résidu $g - \hat{u}$. Cette idée remonte à la méthode de *twicing* introduite dans [Tukey 1977], et elle a été étudiée dans [M. 2006, Cornillon 2013, Milanfar 2013], et appliquée aux NL-means dans [Talebi 2013].

Kervrann et Boulanger ont proposé de traiter l'effet de *jittering* associé aux NL-means dans [Kervrann 2008]. Ils utilisent pour cela des fenêtres de recherche adaptatives dont la taille est automatiquement adaptée au contenu local de l'image, ce qui permet de réduire le nombre de mauvais candidats potentiels. La taille de la fenêtre est ajustée localement à l'aide du principe de compromis biais-variance. En effet, la variance résiduelle en un pixel i de l'image estimée u^{NL} peut être approchée par

$$(\hat{\sigma}_i^{\text{résiduel}})^2 = \left[\sum_{j \in \Omega} (w_{i,j}^{\text{NL}})^2 \right] (\sigma_i^{\text{bruit}})^2. \quad (5.1)$$

où $(\sigma_i^{\text{bruit}})^2$ désigne la variance du bruit, supposée constante sur le voisinage non local du pixel i , et les poids $w_{i,j}$ sont supposés constants par rapport à la réalisation g_i (ou du moins, leur dépendance peut être négligée en raison de la taille de patch suffisamment grande utilisée en pratique). La quantité $(\hat{\sigma}_i^{\text{résiduel}})^2$ joue un rôle important en tant qu'indicateur de la quantité totale de bruit qui a été ôté pour le pixel i . Cependant, cette variance résiduelle ne peut pas à elle seule traduire la qualité du débruitage effectué. En effet, l'effet de *jittering* est provoqué par de fortes réductions du bruit liées à des mélanges de différentes populations qui introduisent du biais. Dans [Kervrann 2008], le biais est détecté de façon itérative en augmentant la taille de la fenêtre de recherche et en construisant des intervalles de confiance basés sur (5.1). La taille de fenêtre retenue est la plus grande possible qui fournit une

estimation u_i^{NL} incluse dans tous les intervalles de confiance issus des fenêtres plus petites. Cette méthode réduit efficacement l'effet de *jittering*, cependant l'utilisation de fenêtres de tailles potentiellement différentes pour des pixels voisins a tendance à créer des effets de choc. De plus, elle ne s'étend pas directement à des modèles non gaussiens.

L'effet de *jittering* peut être traité dans un cadre général et unifié pour différents types de bruit en limitant le débruitage effectué par les NL-means si celui-ci est jugé biaisé. Le principe reprend l'idée de [Lee 1981, Kuan 1985] proposé à l'origine pour du filtrage spatial adaptatif puis étendu filtrage non local adaptatif dans [Deledalle 2014].

On considère qu'au sein du voisinage non local du pixel i , les observations g_j sont toutes des réalisations de la même variable aléatoire $G_i = f_i + \varepsilon_i$, où f_i et ε_i sont deux variables aléatoires indépendantes. La quantité f_i représente les variations de signal, et ε_i les fluctuations liées au bruit. Le signal f_i est supposé avoir pour moyenne la valeur u_i^{NL} et pour écart type une quantité σ_i^{signal} , tandis que ε_i est de moyenne nulle et d'écart type σ_i^{bruit} . Une valeur σ_i^{signal} élevée est synonyme de *jittering* : en effet, une variance de signal importante traduit le fait que les observations au sein du voisinage non local de i sont issues de populations différentes.

Conformément à la stratégie du LLMMSE (*Local Linear Minimum Mean Square Estimator*) [Lee 1981, Kuan 1985], une combinaison convexe entre l'estimation non locale u^{NL} et l'observation bruitée g est effectuée à partir de la formule suivante :

$$u_i^{\text{NLDJ}} = (1 - \alpha_i)u_i^{\text{NL}} + \alpha_i g_i, \quad (5.2)$$

où α_i est un indice de *jittering* donné par :

$$\alpha_i = \frac{(\sigma_i^{\text{signal}})^2}{(\sigma_i^{\text{signal}})^2 + (\sigma_i^{\text{bruit}})^2} \approx \frac{|\hat{\sigma}_i^{\text{NL}}|^2 - (\sigma_i^{\text{bruit}})^2}{|\hat{\sigma}_i^{\text{NL}}|^2 - (\sigma_i^{\text{bruit}})^2 + (\sigma_i^{\text{bruit}})^2}. \quad (5.3)$$

Dans le cas gaussien, la variance du bruit est supposée connue et constante sur toute l'image ; elle correspond au niveau de bruit σ^2 associé au modèle (2.3). Lorsqu'il s'agit de bruit dépendant du signal, la variance non locale du bruit est estimée à partir de l'estimation u_i^{NL} . Les calculs seront détaillés pour chaque cas dans la section 5.5. L'approximation par le membre de droite vient de l'hypothèse d'indépendance non locale entre f_i et ε_i , ce qui conduit à $\text{Var}[g_i] = (\sigma_i^{\text{signal}})^2 + (\sigma_i^{\text{bruit}})^2$. La variance du signal $\text{Var}[f_i]$ peut donc être estimée directement à partir des données g_j situées dans le voisinage non local de i à partir de la formule suivante :

$$(\hat{\sigma}_i^{\text{NL}})^2 = \sum_{j \in \Omega} w_{i,j}^{\text{NL}} g_j^2 - \left(\sum_{j \in \Omega} w_{i,j}^{\text{NL}} g_j \right)^2. \quad (5.4)$$

Au sein d'un même voisinage non local, les pixels sont censés appartenir à la même population, aussi l'écart type estimé $\hat{\sigma}_i^{\text{NL}}$ devrait être proche de celui du bruit σ^{bruit} (et ne pas être corrompu par des fluctuations de signal). Ainsi, si l'écart type estimé $\hat{\sigma}_i^{\text{NL}}$ est proche de la valeur attendue σ_i^{bruit} , l'indice de *jittering* α_i est proche de 0. D'après l'équation (5.2), l'estimation u_i^{NL} est alors inchangée. En revanche si

la variance non locale $\hat{\sigma}_i^{\text{NL}}$ diffère de la valeur attendue, traduisant la présence de fluctuations de signal au sein d'un même voisinage, alors l'indice de *jittering* α_i se rapproche de 1. D'après l'équation (5.2), cela conduit à réinjecter de la donnée bruitée, afin de rééquilibrer le compromis biais-variance.

On remarque de plus que la solution u^{NLDJ} peut également s'écrire sous la forme d'une somme pondérée :

$$u_i^{\text{NLDJ}} = \sum_{j \in \Omega} w_{i,j}^{\text{NLDJ}} g_j \quad (5.5)$$

où $w_{i,j}^{\text{NLDJ}} = (1 - \alpha_i) w_{i,j}^{\text{NL}} + \alpha_i \delta_{i,j}$,

avec $\delta_{i,j}$ le symbole de Kronecker tel que $\delta_{i,j} = 1$ si $i = j$, et 0 sinon. Ainsi, de façon analogue à l'équation (5.1) pour u^{NL} , la variance résiduelle de la solution *dejittered* u^{NLDJ} au pixel i peut être estimée par :

$$(\hat{\sigma}_i^{\text{résiduel}})^2 = \left[\sum_{j \in \Omega} (w_{i,j}^{\text{NLDJ}})^2 \right] (\sigma_i^{\text{bruit}})^2. \quad (5.6)$$

La quantité $\hat{\sigma}_i^{\text{résiduel}}$ fournit un indicateur de la variance résiduelle au pixel i obtenue après l'étape de *dejittering*, c'est-à-dire après l'obtention d'un compromis biais-variance. Ainsi, contrairement au cas associé à l'équation (5.1), cet indicateur traduit la qualité du débruitage obtenu : il reflète à la fois les performances en terme de biais et de variance car la présence de biais dans l'estimation u^{NL} entraîne de la variance résiduelle dans la solution u^{NLDJ} .

Cet indicateur est au cœur de la régularisation adaptative des NL-means proposée dans la section suivante. Dans ce qui suit, et dans la mesure où l'étape de *dejittering* est effectuée au sein de l'algorithme des NL-means, la solution des NL-means (*dejittered* ou non) sera désignée par $u_i^{\text{NL}} = \sum_{j \in \Omega} w_{i,j} g_j$, où les poids $w_{i,j}$ sont ceux définis dans (5.5).

5.3 NL-means régularisés (R-NL)

La méthode décrite ci-dessous vise à combiner les NL-means avec la minimisation TV afin de réduire les défauts observés dans chacune des approches, en particulier l'effet de *jittering* et l'effet de patch rare issus des moyennes non locales ainsi que l'effet de crénelage et la baisse de contraste liés à la minimisation TV. L'idée consiste à réaliser une minimisation TV avec une attache aux données non locale comme ceci :

$$u^{\text{R-NL}} = \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \lambda_i \sum_{j \in \Omega} w_{i,j} (g_j - u_i)^2 + \text{TV}(u), \quad (5.7)$$

où les $\lambda_i > 0$ sont des paramètres de régularisation qui varient spatialement. En prenant des poids $w_{i,j} = \delta_{i,j}$, la solution est bien celle du modèle de ROF défini dans (4.17). Avec les poids non locaux définis dans la section 4.4 et $\lambda_i = +\infty$, la solution de (5.7) s'apparente à la solution des NL-means présentée dans les équations (4.44) et (5.5).

De plus, le problème (5.7) est équivalent au problème suivant :

$$\arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \lambda_i (u_i - u_i^{\text{NL}})^2 + \text{TV}(u). \quad (5.8)$$

Une telle équivalence s'obtient en développant le terme d'attache aux données des équations (5.7) et (5.8) en utilisant la définition de u^{NL} donnée dans (4.44) ou (5.5), puis en ne gardant que les termes qui jouent un rôle dans la minimisation sur u . Grâce à cette formulation, $u^{\text{R-NL}}$ peut s'interpréter comme une solution régularisée de u^{NL} , proche de u^{NL} quand les λ_i sont élevés.

D'après la section précédente, un indicateur de la qualité du débruitage non local en u_i^{NL} est donné par la quantité $\sigma_i^{\text{résiduel}}$ définie dans l'équation (5.6) (plus la variance résiduelle est faible, meilleur est le débruitage). Ainsi, le paramètre λ_i peut être choisi comme une fonction (positive) décroissante de $\sigma_i^{\text{résiduel}}$. Le choix de cette fonction de régularisation est crucial afin de garantir une bonne restauration générale sans pour autant dégrader les contours et les textures. Une simple analyse dimensionnelle de (4.17) montre que λ devrait être choisi inversement proportionnel à $\sigma^{\text{résiduel}}$. Une telle relation est également suggérée dans [Chambolle 1997], ainsi que dans le cas d'autres régularisations de type ℓ_1 , par exemple dans [Donoho 1994].

L'expérience suivante illustre ce point. À partir d'une image u , plusieurs versions bruitées sont générées avec des variances croissantes. Ces images bruitées représentent dans le contexte présent la solution u^{NL} dans le problème (5.8) avec des niveaux croissants de bruit résiduel (constant spatialement) $\hat{\sigma}^{\text{résiduel}}$. La solution du problème (5.8) est alors évaluée pour différentes valeurs de λ . Pour chaque niveau de bruit, le meilleur paramètre λ , c'est-à-dire celui qui minimise l'erreur quadratique moyenne (connue dans cette simulation) est retenu. À l'aide d'un schéma de Monte-Carlo, une valeur optimale de λ est obtenue pour chaque niveau de bruit. La figure 5.1 montre l'évolution de la valeur du λ optimal moyen en fonction du niveau de bruit résiduel $\hat{\sigma}^{\text{résiduel}}$. Cette expérience suggère elle aussi que la valeur du paramètre de régularisation λ devrait être inversement proportionnelle à l'écart type $\hat{\sigma}^{\text{résiduel}}$. Plus précisément, λ_i doit en fait être proportionnel à l'estimation de la réduction de bruit effectuée, c'est-à-dire :

$$\lambda_i = \gamma \left(\frac{\hat{\sigma}_i^{\text{résiduel}}}{\sigma_i^{\text{bruit}}} \right)^{-1} = \gamma \left(\sum_{j \in \Omega} w_{i,j}^2 \right)^{-1/2}. \quad (5.9)$$

où $\gamma > 0$ est une constante qui règle la force de la régularisation adaptative. La relation entre la réduction de bruit et les poids non locaux provient de l'équation (5.6). Tandis que le paramètre λ de l'équation (4.17) règle la régularité de façon globale sur toute l'image, ici le paramètre γ n'a d'influence que localement là où les NL-means n'ont pas pu réduire le bruit de façon significative.

Ce modèle permet ainsi aux NL-means et à la variation totale de se compléter : sur les zones pour lesquelles la redondance est élevée, comme les zones homogènes, les moyennes non locales parviennent à sélectionner suffisamment de (bons) candidats pour réaliser un débruitage efficace, donc la variance résiduelle est faible. Au sein

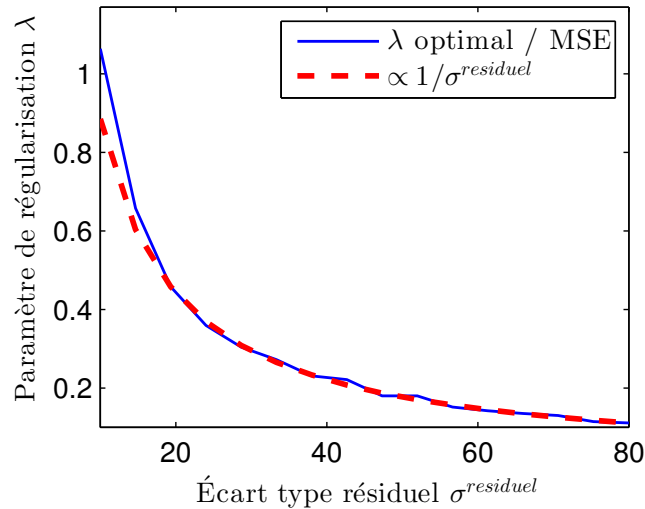


FIGURE 5.1 – Évolution du paramètre de régularisation optimal λ (en bleu) choisi à partir de l’erreur quadratique minimale (MSE) en fonction de l’écart type du bruit résiduel $\sigma^{\text{résiduel}}$. Celui-ci se révèle bien être proportionnel à l’inverse de l’écart type du bruit résiduel $(\sigma^{\text{résiduel}})^{-1}$ (ligne pointillée rouge).

de l’énergie à minimiser (5.8), le paramètre λ est alors élevé, ce qui rend le terme d’attache aux données prépondérant sur le terme de régularisation. Ainsi, la solution obtenue est proche de la solution des NL-means. Cela garantit un lissage efficace et permet d’éviter l’effet de crénelage qui serait observé si ces zones lisses étaient traitées par minimisation TV. Autour des structures singulières et des contours où la redondance est plus faible, peu de candidats interviennent dans les moyennes non locales donc la variance résiduelle est plus élevée. C’est également le cas si de mauvais candidats ont été sélectionnés en premier lieu, car l’étape de *dejittering* réintroduit du bruit et conduit donc à une variance résiduelle plus élevée. Dans l’énergie (5.8), le paramètre λ est alors faible, donc le terme de régularisation devient prépondérant sur le terme d’attache aux données, ce qui conduit à minimiser la variation totale de la solution. Cela permet ainsi de réduire le bruit résiduel et donc l’effet de patch rare tout en préservant les contours de la solution.

Si comme il a été évoqué précédemment les paramètres de régularisation du modèle TV et les paramètres des NL-means sont fortement dépendants de l’image à traiter, ici les paramètres optimaux sont moins fortement influencés par le contenu de l’image. En effet, le modèle s’adapte intrinsèquement au contenu local de l’image, à l’aide de la régularisation localement adaptative, aussi les paramètres ne nécessitent pas de réglage spécifique : un débruitage équivalent peut être obtenu pour des valeurs de paramètres fixées indépendantes des images, comme c’est le cas pour les résultats exposés dans la section 5.6.

La méthode est plutôt intuitive, dans la mesure où elle s’appuie sur les forces et les faiblesses des moyennes non locales et de la minimisation TV. Une implémentation simple et directe est donnée dans un cadre général dans la section 5.5.

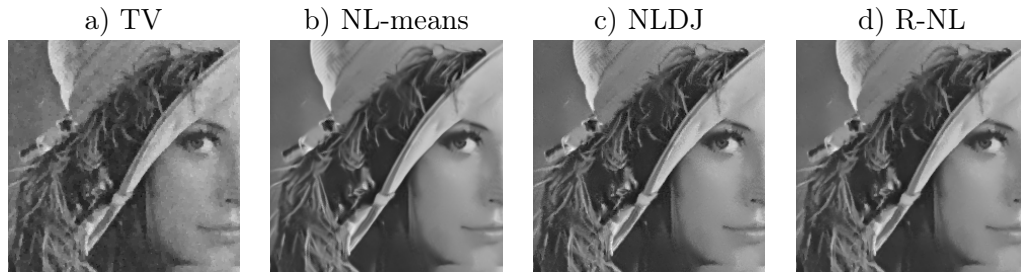


FIGURE 5.2 – Débruitage de bruit gaussien d'écart type $\sigma = 20$. Les paramètres utilisés pour toutes les expériences sont détaillés dans la section 5.6.1. Sur l'image TV (a), l'effet de crénelage peut être observé sur le visage de Lena et l'arrière plan. Les NL-means (b) souffrent de sur-lissage, au niveau des plumes par exemple, traduisant l'effet de *jittering*, corrigé par la version *dejittered* NLDJ (c). Les R-NL (d) combinent les trois méthodes pour permettre un débruitage efficace exempt des effets de crénelage, de *jittering*, et de patch rare.

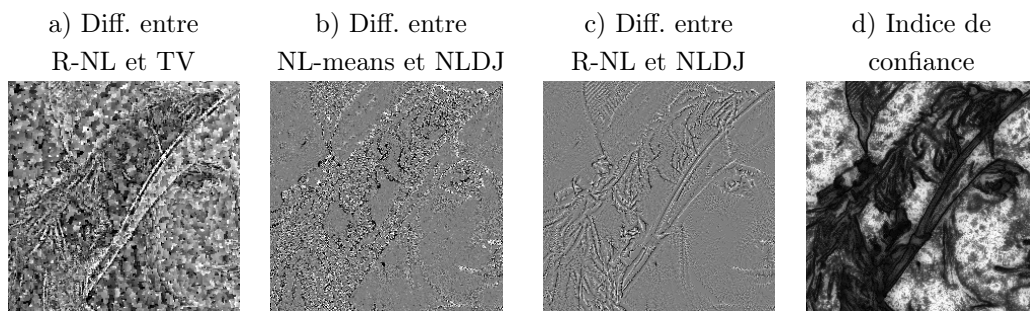


FIGURE 5.3 – a) Différence entre R-NL et TV. Les deux défauts associés à la minimisation TV sont mis en avant : les surfaces homogènes sont plus lisses avec R-NL car ce dernier ne souffre pas de l'effet de crénelage, et les différences d'intensité visibles sur la carte des différences révèlent la perte de contraste dont souffre l'image TV. b) Différence entre les NL-means standards et la version *dejittered*. La différence est visible au niveau des plumes et du chapeau où du bruit a été réinjecté afin de corriger le sur-lissage. c) Différence entre NLDJ et R-NL. Ici, les changements sont très localisés puisque la régularisation TV ne s'applique que localement où le bruit résiduel persiste. d) Carte de l'indice de confiance λ_i . Conformément à l'intuition, la confiance dans les NL-means est moins bonne au niveau des contours où le nombre de candidats est limité ainsi que sur les structures fines où le *jittering* a été corrigé.

Sur la figure 5.2 sont illustrées les solutions du débruitage obtenu par minimisation TV, par l'algorithme des NL-means, avant et après l'étape de *dejittering*, puis par l'algorithme régularisé R-NL.

La figure 5.3 illustre l'effet de l'étape de *dejittering* et de régularisation adaptative par rapport aux solutions des NL-means et TV classiques. Sur la figure 5.3-(a)

est affichée la différence entre la solution de TV et de R-NL. Deux défauts associés à la minimisation TV et corrigés par R-NL sont ici mis en avant : l'effet de crénelage et la perte de contraste générale. La figure 5.3-(b) montre la différence entre les moyennes non locales avant et après *dejittering*. Du bruit a visiblement été ré-introduit sur le chapeau et les plumes, qui avaient au préalable été trop lissés par les NL-means. La différence entre les moyennes non locales *dejittered* et la solution régularisée R-NL est illustrée sur la figure 5.3-(c). Ici, la différence est bien plus localisée, autour des contours notamment où la régularisation TV a été appliquée de façon sélective. Enfin, sur la figure 5.3-(d) sont affichées les valeurs de l'indice de confiance λ_i qui traduit la réduction de bruit et pondère le terme d'attache aux données dans le processus de régularisation. Conformément à la figure précédente, l'indice de confiance est élevé sur les zones uniformes, dans lesquelles peu ou pas de régularisation n'est appliquée sur la solution des NL-means jugée de bonne qualité. Sur les contours en revanche, où les moyennes non locales n'ont pas trouvé suffisamment de candidats pour réduire de façon significative le niveau de bruit, l'indice de confiance se trouve réduit. C'est également le cas au niveau des structures fines comme sur le côté du chapeau où l'étape de *dejittering* a augmenté la variance résiduelle, corrigeant la prise en compte de mauvais candidats. Ces différentes figures illustrent comment l'association des moyennes non locales à une minimisation TV adaptative peut permettre de réduire leurs défauts respectifs.

5.4 D'autres coopérations des modèles variationnels et non locaux

L'association des moyennes non locales aux méthodes variationnelles, ou le fait d'interpréter les NL-means d'un point de vue variationnel, a été étudié dans différentes approches [Awate 2006, Brox 2007, Pizarro 2010]. De plus, dans la mesure où les NL-means offrent de belles performances en débruitage, ces derniers ont souvent été adaptés à d'autres gammes de problèmes comme par exemple pour de la déconvolution, de l'inpainting ou de la super-résolution [Buades 2006, Mignotte 2008, Arias 2012]. Ces adaptations ont souvent été réalisées à l'aide d'une approche basée sur un principe de minimisation et des propriétés non locales.

L'une des méthodes hybrides les plus répandues est la TV non locale (NL-TV) proposée par Gilboa et Osher dans [Gilboa 2008]. À partir des travaux sur le laplacien d'un graphe de Zhou et Scholkopf [Zhou 2004] et Bougleux *et al.* [Bougleux 2007], ainsi que de la définition de termes de régularisation non locaux de Kindermann *et al.* [Kindermann 2005], un gradient non local est défini de la façon suivante :

$$(\nabla_w u)_{i,j} = (u_i - u_j) \sqrt{w_{i,j}}, \quad (5.10)$$

où $w_{i,j}$ est un poids qui mesure la similarité entre les pixels i et j . Cela permet la définition d'un cadre de travail non local, en particulier le modèle de ROF non local

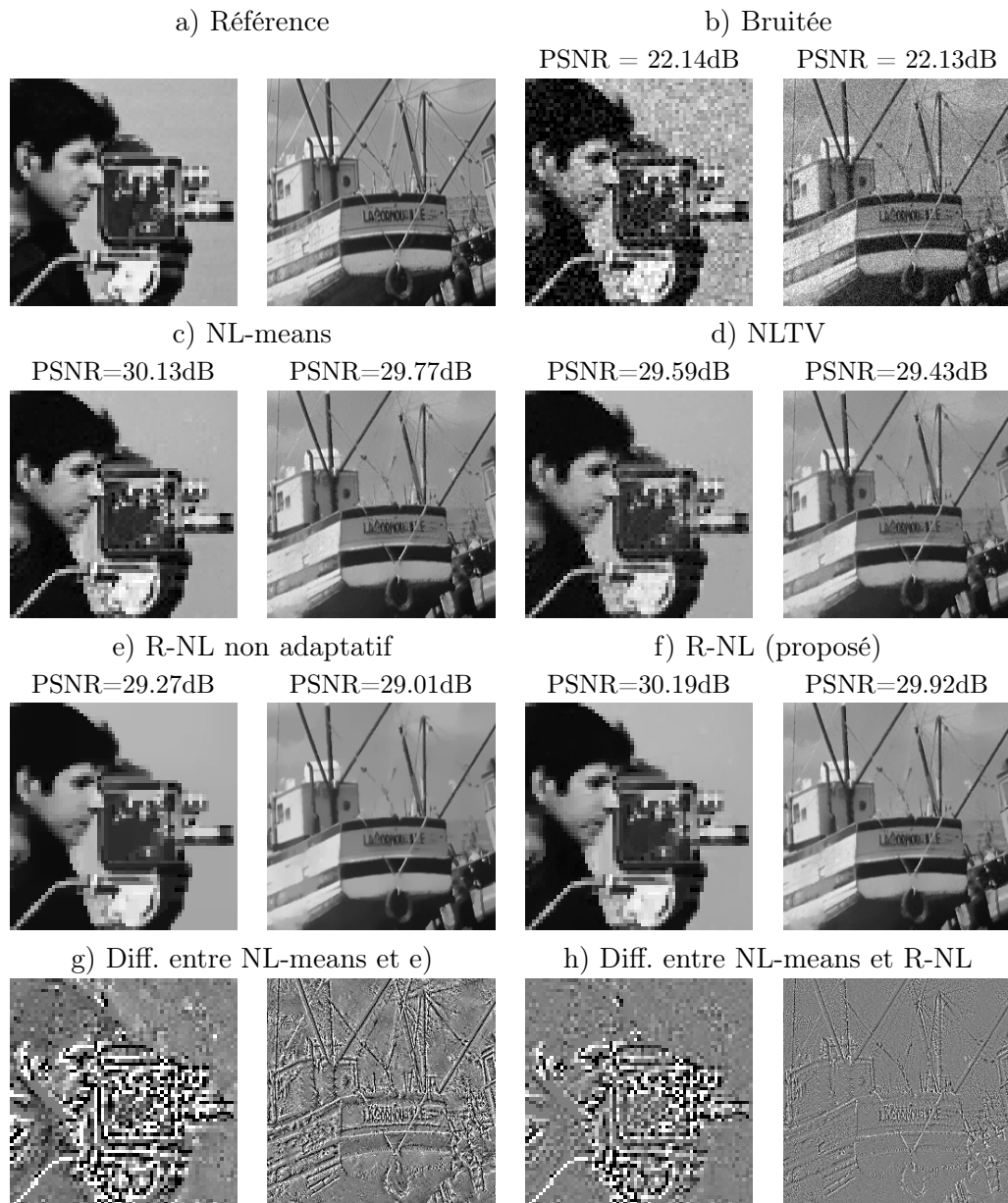


FIGURE 5.4 – Débruitage de bruit gaussien, écart type $\sigma = 20$. De haut en bas, et de gauche à droite : Images a) originales et b) bruitées de *Cameraman* et *Boat*, traitées avec c) NL-means, d) NL-TV, e) R-NL non adaptatif et f) R-NL. g) Différence entre les solutions des NL-means et de e) et h) Différence entre les solutions des NL-means et de f). L'association des NL-means aux méthodes variationnelles permet un débruitage efficace tout en limitant l'effet de crénelage et/ou de patch rare. Les paramètres sont explicités dans la section 5.6.1.

:

$$u^{\text{NLTV}} = \arg \min_u \|u - g\|^2 + \lambda \sum_{i \in \Omega} \|(\nabla_w u)_i\| \quad (5.11)$$

avec $\sum_{i \in \Omega} \|(\nabla_w u)_i\| = \sum_{i \in \Omega} \sqrt{\sum_{j \in \Omega} (u_i - u_j)^2 w_{i,j}}$.

Ce type de modèle a été introduit afin de traiter séparément les textures et les zones lisses. Il a ensuite été adapté à des problèmes de déconvolution, inpainting ou compressive sensing [Peyré 2011, Zhang 2010b]. Gilboa et Osher ont également étendu cette régularisation non locale à de la diffusion non locale [Gilboa 2007], qui offre de belles performances de débruitage. Sur la figure 5.4-(d) sont illustrés des résultats de débruitage avec NL-TV. Les structures comme les câbles et les inscriptions sur le bateau sont préservées et l'effet de crénelage est bien réduit sur les zones lisses, grâce à la régularisation non locale. Si une telle approche paraît similaire au modèle R-NL, en réalité la philosophie est opposée. NL-TV réalise une régularisation non locale afin de préserver les textures et de réduire l'effet de crénelage sous-jacent à la minimisation TV. Il est de plus possible d'ajouter une étape qui réduit l'effet de patch rare, mais le modèle NL-TV lui-même ne traite pas ce problème. Au contraire, l'approche R-NL applique très localement une régularisation TV à l'aide d'une attache aux données qui est elle non locale, et dont le but principal est de réduire l'effet de patch rare et le *jittering* dont souffrent les moyennes non locales. Un second effet positif est que l'approche ne souffre pas non plus de l'effet de crénelage, car la régularisation TV est appliquée de façon très locale et sélective.

Louchet et Moisan [Louchet 2011] ont également associé les NL-means aux méthodes variationnelles afin de réduire à la fois l'effet de patch rare et l'effet de crénelage. Dans un premier temps, ils redéfinissent un gradient afin de créer une variation totale locale qui permet de limiter l'effet de crénelage. Ensuite, l'effet de patch rare issu des moyennes non locales est corrigé grâce à la minimisation TV locale : lorsque les NL-means ne parviennent pas à sélectionner assez de candidats pour effectuer un débruitage satisfaisant, les patchs sont régularisés localement à l'aide de la régularisation TV locale afin de garantir une similarité suffisante. Cela revient à appliquer une régularisation TV adaptative avant d'avoir recours aux NL-means afin de garantir que le nombre de candidats intervenant dans le processus de moyennage est optimal. Cette fois également, le modèle R-NL fonctionne dans l'autre sens, puisqu'il part de la solution des NL-means pour ensuite appliquer localement une régularisation TV adaptative là où le débruitage est jugé trop faible, en se basant sur l'indice de confiance de réduction de bruit.

Les approches variationnelles et non locales ont également été associées dans le contexte de la super-résolution. Protter *et al.* [Protter 2009] ainsi que d'Angelo et Vanderghyest [d'Angelo 2011] utilisent un terme d'attache aux données non local, et combiné à une minimisation TV, afin d'obtenir une image de haute qualité à partir d'une séquence d'images basse résolution. Si la formulation est assez proche du modèle R-NL, la philosophie reste bien distincte. La figure 5.4-(e) illustre les

résultats d'un tel algorithme, mais dans un contexte de débruitage. Cela équivaut à appliquer une régularisation TV globale sur la solution des NL-means. Cela permet effectivement de réduire l'effet de patch rare par rapport à la solution des NL-means, en régularisant le bruit résiduel. Cependant, comme la régularisation n'est pas adaptative, la suppression du bruit résiduel requiert d'appliquer un lissage important qui entraîne une perte de détail et de contraste. Le modèle R-NL, lui, possède une attache aux données pondérée par l'indice de confiance, ce qui permet à la régularisation TV de ne s'appliquer que localement et sélectivement.

La figure 5.4 illustre les résultats de débruitage obtenus avec les différentes méthodes qui combinent les méthodes non locales et variationnelles décrites dans cette section. La variation totale non locale NL-TV parvient à préserver efficacement les structures fines comme les câbles du bateau et réduit l'effet de crénelage, en revanche les zones uniformes telles que le ciel autour de la tête de *Cameraman* ne sont pas autant lissées que par les NL-means. La méthode basée sur l'algorithme de super-résolution issu de [d'Angelo 2011] permet elle de réduire l'effet de patch rare grâce à la régularisation TV, mais le compromis entre régularisation et préservation des contours est difficile à atteindre : par exemple, les câbles du bateau doivent être sur-lissés pour permettre au bruit résiduel sur les inscriptions du bateau d'être réduit efficacement. De plus, la carte de différence entre cette méthode et les NL-means sur la figure 5.4-(g) révèle la perte de contraste qui affecte la solution régularisée. La méthode R-NL, figure 5.4-(f), offre une alternative simple pour gérer le compromis entre débruitage performant et préservation des textures, grâce à l'information fournie par l'estimation du bruit résiduel qui joue le rôle d'indicateur de confiance. La différence entre les NL-means et R-NL est représentée sur la figure 5.4-(h). La régularisation TV adaptative est appliquée localement seulement sur les contours et les structures singulières, conformément à l'évaluation de la confiance des NL-means ; ce qui permet d'empêcher le sur-lissage, le crénelage et la perte de contraste.

5.5 R-NL pour la famille exponentielle

La minimisation TV et les NL-means sont deux méthodes qui s'adaptent à différents types de bruits, conformément à ce qui a été présenté dans les sections 4.3 et 4.4. Ainsi, l'approche R-NL peut elle aussi naturellement s'étendre à d'autres types de bruit non corrélés à l'aide d'un terme d'attache aux données pondéré de la forme $-\sum \sum w_{i,j} \log p(g_j|u_i)$, d'après l'idée issue de [Polzehl 2006], et de poids adaptés conformément à la section 4.4.

Ce modèle étendu peut alors se résoudre de façon efficace dans le cadre général de la famille exponentielle, qui inclut notamment le bruit blanc additif gaussien, le bruit de Poisson, ainsi que des bruits de type multiplicatif.

Une loi de probabilité appartient à la famille exponentielle [Collins 2001] si elle peut s'écrire sous la forme suivante :

$$p(T(g)|u) = c(g) \exp(\eta(u)T(g) - A(u)) \quad (5.12)$$

où c , T , η et A sont des fonctions connues. Le modèle étendu est alors le suivant :

$$u^{\text{R-NL}} = \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \lambda_i \sum_{j \in \Omega} w_{i,j} [A(u_i) - \eta(u_i)T(g_i)] + \text{TV}(u), \quad (5.13)$$

où $\lambda_i = \gamma \left(\sum_{j \in \Omega} w_{i,j}^2 \right)^{-1/2}$. De façon analogue au cas gaussien, le problème peut se reformuler à l'aide d'une attache aux données pondérée par rapport à la solution des NL-means :

$$u^{\text{R-NL}} = \arg \min_{u \in \mathbb{R}^N} - \sum_{i \in \Omega} \lambda_i \log p(u_i^{\text{NL}} | u_i) + \text{TV}(u), \quad (5.14)$$

où $\mathcal{L}_w(u_i) = -\lambda_i \log p(u_i^{\text{NL}} | u_i)$ désigne la log-vraisemblance pondérée, et $u_i^{\text{NL}} = \sum_{j \in \Omega} w_{i,j} T(g_j)$ et λ_i sont calculés à l'aide d'une implémentation rapide des NL-means. Une méthode rapide pour le calcul des poids est notamment détaillée dans [Darbon 2008]. L'étape de minimisation est alors effectuée à l'aide des algorithmes standards de minimisation, en fonction du type de bruit impliqué, d'après les résultats présentés dans la section 4.3. Une implémentation générale de l'algorithme des R-NL est donnée dans l'algorithme 6. Les détails associés aux caractéristiques propres à chaque type de bruit sont fournis dans les sections à venir.

5.5.1 Cas du bruit gaussien

Dans le cas standard où l'image bruitée g est corrompue par un bruit blanc additif gaussien de variance σ^2 , la vraisemblance de l'image sous-jacente f sachant les observations g est donnée par :

$$\begin{aligned} p(g|f) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\|g-f\|^2}{2\sigma^2}\right) \\ &= \sum_{i \in \Omega} \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{g_i^2}{2\sigma^2}\right)}_{c(g_i)} \exp\left(\underbrace{\frac{(g_i f_i)}{\sigma^2}}_{T(g_i)=g_i, \eta(f_i)=\frac{f_i}{\sigma^2}} - \underbrace{\frac{f_i^2}{2\sigma^2}}_{A(f_i)=\frac{f_i^2}{2\sigma^2}}\right), \end{aligned} \quad (5.15)$$

ce qui fait bien du bruit gaussien une loi de la famille exponentielle.

Le bruit blanc additif gaussien est dit homoscédastique ou indépendant du signal, car l'amplitude du bruit est constante sur toute l'image quelle que soit la valeur sous-jacente du signal, avec une variance égale à σ^2 . Ainsi, la variance non locale attendue lors de l'étape de *dejittering* est bien $(\sigma_i^{\text{bruit}})^2 = \sigma^2$.

Le problème (5.14) est alors équivalent au problème suivant :

$$\arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \lambda_i \frac{(u_i - u_i^{\text{NL}})^2}{2\sigma^2} + \text{TV}(u), \quad (5.16)$$

qui correspond à une reparamétrisation près au modèle défini dans la section 5.3, eq. (5.7). En pratique, la prise en compte des constantes multiplicatives qui interviennent dans la log-vraisemblance négative permet de considérer la même constante

Algorithme 6 R-NL**Précondition :** g : image initiale bruitée, h : paramètre de filtrage, $|P|$: taille de patch, \mathcal{N} : taille de la fenêtre de recherche γ : paramètre de régularisation**pour** $i \in \Omega$ **faire****Étape NL-means**Calculer $w_{i,j} \leftarrow \varphi_h[d(g(P_i), g(P_j))], \forall j \in \mathcal{N}_i$ Normaliser $w_{i,j} \leftarrow w_{i,j} / \sum_j w_{i,j}, \forall j \in \mathcal{N}_i$ Calculer $u_i^{\text{NL}} \leftarrow \sum_j w_{i,j} T(g_j)$ Calculer $(\hat{\sigma}_i^{\text{NL}})^2 \leftarrow \sum_j w_{i,j} T(g_j)^2 - (u_i^{\text{NL}})^2$ Calculer $(\sigma_i^{\text{bruit}})^2$ adapté au type de bruit**Étape de de jittering**Calculer $\alpha_i \leftarrow \frac{|(\hat{\sigma}_i^{\text{NL}})^2 - (\sigma_i^{\text{bruit}})^2|}{|(\hat{\sigma}_i^{\text{NL}})^2 - (\sigma_i^{\text{bruit}})^2| + (\sigma_i^{\text{bruit}})^2}$ Mise à jour de $u_i^{\text{NL}} \leftarrow (1 - \alpha_i)u_i^{\text{NL}} + \alpha_i g_i$ Mise à jour de $w_{i,j} \leftarrow (1 - \alpha_i)w_{i,j} + \alpha_i \delta_{i,j}$ Calculer $\lambda_i \leftarrow \gamma \left(\sum_j w_{i,j}^2 \right)^{-1/2}$ **Étape de minimisation**

$$u^{\text{R-NL}} = \arg \min_u \sum_{i \in \Omega} \lambda_i [A(u_i) - \eta(u_i)u_i^{\text{NL}}] + \text{TV}(u)$$

retour $u^{\text{R-NL}}$

γ (qui règle la force de la régularisation) pour tous les types de bruit, notamment pour tous les résultats présentés dans la section 5.6.

De nombreuses solutions permettent de résoudre ce problème, dont certaines sont décrites dans la section 4.3. La méthode utilisée ici est l'algorithme de Chambolle-Pock [Chambolle 2011], dont les détails sont donnés dans l'algorithme 5.

Dans ce cas, l'opérateur proximal associé à l'attache aux données pondérée \mathcal{L}_w dépend du paramètre de régularisation adaptatif et il est donné par :

$$u = (I + \tau \partial \mathcal{L}_w)^{-1}(\tilde{u}) \Leftrightarrow u_i = \frac{\tilde{u}_i + \tau / \sigma^2 \lambda_i u_i^{\text{NL}}}{1 + \tau / \sigma^2 \lambda_i}. \quad (5.17)$$

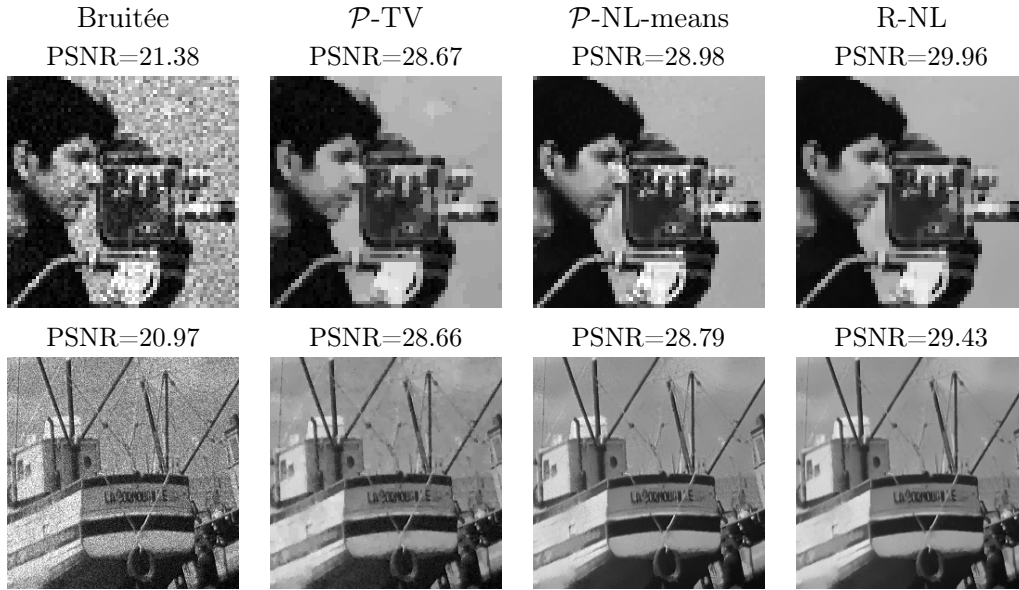


FIGURE 5.5 – Débruitage de bruit de Poisson, $Q = 4$. Les structures fines telles que les inscriptions et les câbles du bateau sont mieux préservées par R-NL.

5.5.2 Cas du bruit de Poisson

Dans le cas du bruit de Poisson, la densité de probabilité pour l'intensité observée g_i sachant le nombre de photons $f_i > 0$ et le paramètre Q est donnée par

$$p(g_i|f_i) = \frac{\frac{f_i}{Q}^{g_i}}{(g_i/Q)!} \exp\left(-\frac{f_i}{Q}\right) = \frac{1}{(g_i/Q)!} \exp\left(\underbrace{\frac{g_i}{Q}}_{T(g_i)=g_i/Q, \eta(f_i)=\ln f_i/Q} \underbrace{\ln \frac{f_i}{Q}}_{A(f_i)=f_i/Q} - \underbrace{\frac{f_i}{Q}}_{A(f_i)=f_i/Q}\right), \quad (5.18)$$

ce qui en fait bien une loi de la famille exponentielle. La log-vraisemblance négative est de la forme :

$$\mathcal{L}(g|f) = \sum_{i \in \Omega} \frac{f_i}{Q} - \frac{g_i}{Q} \log \frac{f_i}{Q} + \log \frac{g_i!}{Q}, \quad (5.19)$$

où g_i/Q désigne un entier positif, et on rappelle que l'on a $\mathbb{E}[g_i] = f_i$ et $\text{Var}[g_i] = Qf_i$. Ainsi, la variance non locale impliquée dans l'étape de *dejittering* est donnée par :

$$(\sigma_i^{\text{bruit}})^2 = Qu_i^{\text{NL}}, \quad (5.20)$$

et la solution des NL-means ainsi que le paramètre de régularisation adaptative λ_i peuvent être calculés et le problème variationnel devient :

$$u^{\text{R-NL}} = \arg \min_{u \geq 0} \sum_{i \in \Omega} \lambda_i \left[\frac{u_i}{Q} - \frac{u_i^{\text{NL}}}{Q} \log \left(\frac{u_i}{Q} \right) \right] + \text{TV}(u). \quad (5.21)$$

De façon analogue à la section 4.3.3.1, ce problème de minimisation se résout à l'aide de l'algorithme de Chambolle-Pock appliqué au bruit de Poisson, dont la

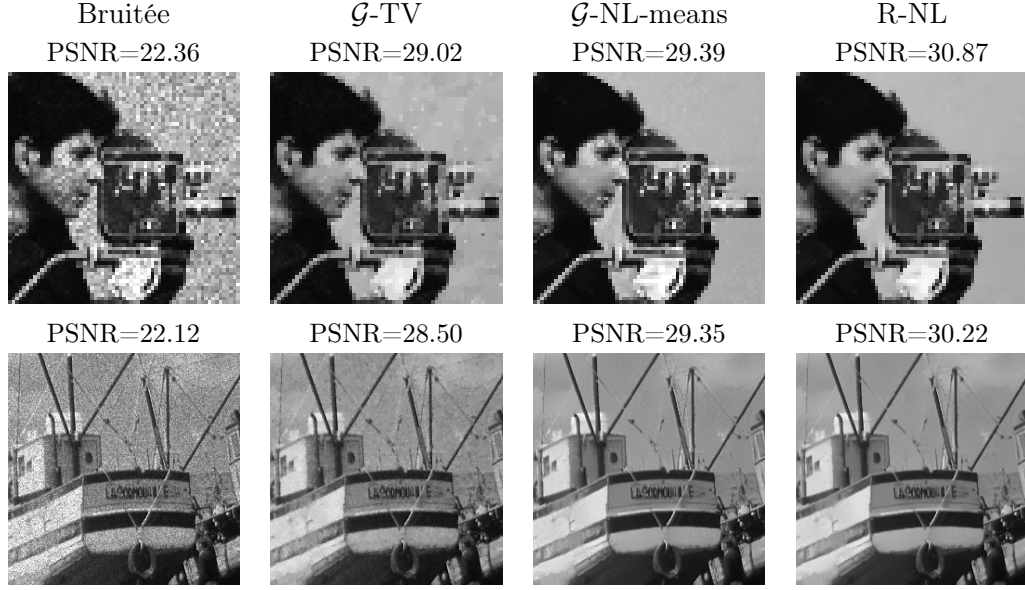


FIGURE 5.6 – Débruitage de bruit gamma, $L = 12$. Le bruit résiduel observé sur le visage de *Cameraman* et les inscriptions du bateau avec les NL-means est corrigé par la régularisation TV sur le résultat par R-NL. Les surfaces lisses, bien gérées par les NL-means, ne souffrent pas pour autant de l'effet de crénelage associé au débruitage TV.

forme générale est rappelée dans l'algorithme 5, et le calcul de l'opérateur proximal associé à l'attache aux données pondérée \mathcal{L}_w est donné par :

$$\begin{aligned}
 u &= (I + \tau \partial \mathcal{L}_w)^{-1}(\tilde{u}) \Leftrightarrow & (5.22) \\
 u_i &= \begin{cases} \frac{1}{2} \left(\tilde{u}_i - \tau \frac{\lambda_i}{Q} + \sqrt{\left(\tilde{u}_i - \tau \frac{\lambda_i}{Q} \right)^2 + 4\tau \frac{\lambda_i}{Q} u_i^{\text{NL}}} \right) & \text{si } u_i^{\text{NL}} > 0, \\ \max(\tilde{u}_i - \tau \frac{\lambda_i}{Q}, 0) & \text{sinon.} \end{cases}
 \end{aligned}$$

5.5.3 Cas du bruit gamma

On rappelle que la perturbation d'un signal bruité selon une loi gamma est modélisée par $g = f \times b$, où $b \sim \mathcal{G}(L)$. L est le paramètre appelé nombre de vues, et détermine la puissance du bruit, le pire cas étant celui où $L = 1$. La vraisemblance de l'image f étant données les observations g est alors donnée par :

$$p(g|f) = \sum_{i \in \Omega} \frac{L^L g_i^{L-1} e^{-L \frac{g_i}{f_i}}}{\Gamma(L) f_i^L} = \underbrace{\frac{L^L}{\Gamma(L)} g_i^{L-1}}_{C(g_i)} \exp \left(\underbrace{-L \frac{g_i}{f_i}}_{\eta(f_i) = -L/f_i} - \underbrace{L \ln f_i}_{A(f_i)} \right). \quad (5.23)$$

$T(g_i) = g_i$

La log-vraisemblance négative de $f > 0$ pour une image observée $g > 0$ est donc :

$$\mathcal{L}(g|f) = \sum_{i \in \Omega} \frac{Lg_i}{f_i} + L \log f_i + \log \Gamma(L) - L \log L - (L - 1) \log g_i, \quad (5.24)$$

De même que pour le bruit de Poisson, les perturbations de type gamma sont dépendantes du signal, et on a $\mathbb{E}[g_i] = f_i$ et $\text{Var}[g_i] = \frac{f_i^2}{L}$. Ainsi, la variance non locale attendue lors de l'étape de *dejittering* est :

$$(\sigma_i^{\text{bruit}})^2 = (u_i^{\text{NL}})^2 / L \quad (5.25)$$

Après avoir évalué la solution des NL-means ainsi que le paramètre de régularisation adaptative λ_i , le problème variationnel devient :

$$u^{\text{R-NL}} = \arg \min_{u > 0} \sum_{i \in \Omega} \lambda_i \left[L \log(u_i) + L \frac{u_i^{\text{NL}}}{u_i} \right] + \text{TV}(u). \quad (5.26)$$

Bien que la fonctionnelle ne soit pas convexe, la section 4.3.3.2 fournit des éléments de résolution et propose d'utiliser la méthode forward-backward [Combettes 2005] dont une version accélérée [Beck 2010] est détaillée dans l'algorithme 4.

Les figures 5.2, 5.4, 5.5 et 5.6 illustrent et comparent les performances de débruitage par l'algorithme des R-NL aux méthodes standards telles que les NL-means et la minimisation TV, sur des images en niveaux de gris entre 0 et 255. Dans le cas du bruit gaussien, un bruit additif d'écart type $\sigma = 20$ a été simulé. Pour les bruits de type Poisson et gamma, les images originales ont été corrompues afin d'atteindre un PSNR initial d'environ 22dB. Les paramètres utilisés pour l'ensemble des expériences sont donnés dans la section 5.6.1.

5.5.4 Cas d'un bruit hybride

Suivant le modèle de bruit proposé dans la partie sur l'estimation de bruit, on peut également s'intéresser à l'adaptation de l'algorithme R-NL au cas d'un bruit hybride. Pour cela, on fait l'hypothèse que le bruit est additif, gaussien, dépendant du signal, et que sa variance est une fonction polynomiale d'ordre deux, de sorte que l'image bruitée g est une représentation de la variable aléatoire G donnée par :

$$G = f + \text{NLF}_{(a,b,c)}(f) \cdot \varepsilon, \quad \text{où} \quad \text{NLF}_{(a,b,c)}(f) = af^2 + bf + c \quad \text{et} \quad \varepsilon \sim \mathcal{N}(0, 1). \quad (5.27)$$

L'algorithme des NL-means qui suit, baptisé NLF-means, prend en compte la dépendance au signal du bruit, sans connaissance de la distribution exacte du bruit, mais seulement des paramètres (a, b, c) de la NLF estimée. La mesure de dissimilarité d est alors adaptée comme suit :

$$d(g^{\rho_i}, g^{\rho_j}) = \frac{1}{|\rho|} \sum_{k=1}^{|\rho|} \frac{(g_k^{\rho_i} - g_k^{\rho_j})^2}{\text{NLF}_{\theta}(g_k^{\rho_i}) + \text{NLF}_{\theta}(g_k^{\rho_j})}, \quad (5.28)$$

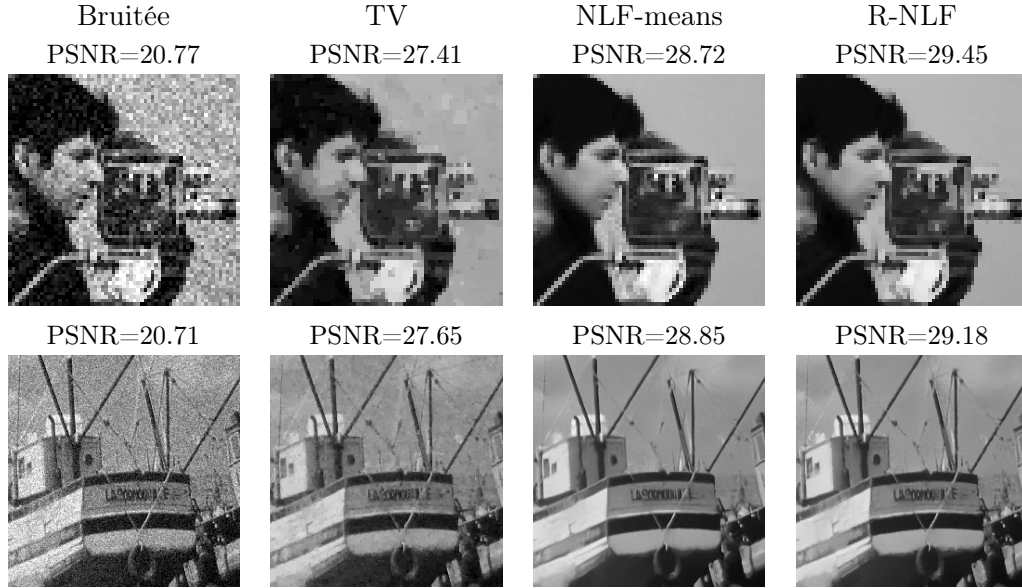


FIGURE 5.7 – Débruitage de bruit hybride, avec des paramètres de bruit $(a, b, c) = (0.0312, 1.875, 100)$. On observe le même comportement que pour les bruits traditionnels, et en particulier la correction par R-NLF de l'effet de patch rare que l'on observe par exemple autour de la caméra, ainsi que meilleure préservation des structures fines comme les câbles du bateau garantie par l'association du *dejittering* et de la régularisation adaptative.

où m_d^ρ et s_d^ρ sont respectivement la valeur moyenne et l'écart type de la dissimilarité d calculée entre deux patches de taille ρ identiquement distribués, conformément à la section 4.4.2.

Reste à adapter les étapes de *dejittering* et de régularisation adaptative.

La mise en place de l'étape de *dejittering* est directe ; elle repose sur l'estimation du paramètre α_i , qui dépend d'une part de la variance non locale $(\hat{\sigma}_i^{\text{NL}})^2$ et de la variance du bruit $(\sigma_i^{\text{bruit}})^2$. Cette dernière est calculée à partir de la NLF de la façon suivante :

$$(\sigma_i^{\text{bruit}})^2 = \text{NLF}_{(a,b,c)}(u_i^{\text{NL}}) = a(u_i^{\text{NL}})^2 + b(u_i^{\text{NL}}) + c. \quad (5.29)$$

Concernant la régularisation adaptative, en s'appuyant sur le modèle gaussien à variance polynomiale, le problème (5.14) s'écrit de la façon suivante :

$$u^{\text{R-NLF}} = \arg \min_{u \in \mathbb{R}^N} \sum_{i \in \Omega} \lambda_i \frac{(u_i - u_i^{\text{NL}})^2}{2 \text{NLF}_{(a,b,c)}(u_i^{\text{NL}})} + \text{TV}(u). \quad (5.30)$$

On note la dépendance locale de la variance en u_i^{NL} ; dans le modèle gaussien il s'agit de la variance du bruit σ^2 . Dans le cas présent, la variance du bruit correspondrait normalement en chaque pixel à $\text{NLF}_{(a,b,c)}(f_i)$, où f_i est l'intensité sous-jacente. On ne dispose pas de cette information, mais de l'estimation u_i^{NL} , qui devrait vraisemblablement permettre une bonne estimation de la variance.

Ainsi, de façon similaire au cas du bruit blanc gaussien, on peut utiliser pour la résolution l'algorithme primal-dual. L'opérateur proximal associé à la variation totale est inchangé, tandis que celui associé à l'attache aux données est donné par :

$$u = (I + \tau \partial \mathcal{L}_w)^{-1}(\tilde{u}) \Leftrightarrow u_i = \frac{\tilde{u}_i + \frac{\tau \lambda_i}{\text{NLF}_{(a,b,c)}(u_i^{\text{NL}})} u_i^{\text{NL}}}{1 + \frac{\tau \lambda_i}{\text{NLF}_{(a,b,c)}(u_i^{\text{NL}})}}. \quad (5.31)$$

La figure 5.7 illustre l'application de R-NLF au débruitage de bruit hybride, comparé également à l'algorithme des NLF-means ainsi qu'au débruitage TV classique adapté au bruit hybride (la même attache aux données que dans (5.30), mais évaluée par rapport à la donnée bruitée g et avec une régularisation non adaptative). On observe les mêmes comportements que dans le cas des bruits classiques, en particulier la réduction de l'effet de patch rare (autour de la caméra par exemple) ainsi que la meilleure préservation des structures fines comme les câbles du bateau, grâce à la coopération entre le *dejittering* et la régularisation adaptative.

5.6 Résultats et discussion

5.6.1 Sélection des paramètres

L'ensemble des algorithmes décrits ci-dessus ont été soigneusement réglés, notamment concernant les constantes de normalisation qui interviennent dans les vraisemblances, afin de limiter leur variabilité en fonction du type et du niveau de bruit. Pour l'ensemble des expériences effectuées, trois niveaux de bruit ont été considérés : faible, moyen et élevé. Dans le cas gaussien, cela correspond à des valeurs d'écart type respectives $\sigma = 20, 30, 40$. Les paramètres Q et L des distributions de Poisson et de gamma ont ensuite été sélectionnés afin d'atteindre des PSNR initiaux équivalents, soit 22dB à bruit faible, 18db à bruit moyen et 16dB à bruit élevé.

Le noyau défini dans l'équation (4.47) pour le calcul des poids non locaux permet de limiter la sensibilité à la distribution et à la puissance du bruit. Des valeurs standards de tailles de patches et de fenêtre de recherche ont été utilisées, soit des patches de taille 7×7 et des fenêtres de recherche de taille 21×21 . De plus, les constantes issues des log-vraisemblances pour la minimisation TV permettent de fixer raisonnablement le paramètre de réglage γ indépendamment de la distribution du bruit. Celui-ci a été réglé à $1/0.015 \approx 66$, pour toutes les images quantifiées sur 8 bits.

5.6.2 Débruitage d'images

Quelques résultats numériques de comparaisons de R-NL aux autres approches hybrides non locales et/ou variationnelles sont présentés ici. Les tableaux 5.1-5.4 présentent les résultats en terme de PSNR et de SSIM [Wang 2004]. La supériorité numérique générale des méthodes hybrides par rapport aux NL-means ou TV

classiques confirme les observations issues des figures 5.2-5.6. Dans le cas des bruits de Poisson et de gamma, l'algorithme NL-TV n'a pas été introduit car celui-ci ne s'applique directement qu'à du bruit gaussien. Ces résultats illustrent également l'intérêt de la régularisation adaptative par rapport au modèle non adaptatif issu de [d'Angelo 2011]. Des extensions de ce modèle aux cas Poisson et gamma sont également présentés sous le nom *NA/R-NL*, pour *non-adaptive R-NL*. Conformément à l'état de l'art, les résultats obtenus avec l'algorithme BM3D [Dabov 2007b] figurent également dans le tableau. Celui-ci a été appliqué après stabilisation de variance à l'aide de la transformée d'Anscombe pour traiter du bruit de Poisson [Mäkitalo 2012], et après transformée logarithmique dans le cas de bruit gamma [Mäkitalo 2010]. Les résultats de R-NL ne sont certes pas aussi bons que ceux issus de BM3D en terme de PSNR, en revanche l'amélioration des NL-means grâce à l'interprétation de la réduction de variance comme mesure de confiance est justifiée, dans un modèle intuitif et simple à implémenter.

5.6.3 Discussion

D'une façon générale, l'approche étudiée ici repose sur une amélioration des méthodes non locales en deux étapes. Elle consiste d'une part à réduire l'effet de *jittering* qui se produit lorsque des candidats issus de population sous-jacentes différentes sont confondus. Ce phénomène est par exemple observé sur l'herbe de *Cameraman*, sur la figure 5.8. L'autre effet indésirable visé est l'effet de patch rare. Ces deux défauts proviennent de la difficulté à évaluer des poids non locaux fiables à partir de patches bruités et donc à trouver des candidats pertinents. Ils sont de fait intimement liés à l'aspect non local. La philosophie de l'approche étudiée ici repose sur l'estimation locale de la réduction de variance, utilisée comme indicateur de confiance pour corriger d'abord l'effet de *jittering* puis l'effet de patch rare. Le modèle appliqué ici au cas des NL-means peut donc s'étendre à d'autres méthodes non locales basées sur le calcul de poids non locaux, par exemple à BM3D [Dabov 2007b], ou aux améliorations des NL-means SAFIR [Kervrann 2008] et SAIF [Talebi 2013]. Pour chacune de ces méthodes, la variance résiduelle de la solution non locale peut être estimée et réinjectée dans les étapes de *dejittering* puis de régularisation. La figure 5.8 illustre cette application; des résultats de débruitage de bruit additif gaussien utilisant BM3D, SAFIR et SAIF sont présentés, et comparés aux versions régularisées associées R-BM3D, R-SAFIR et R-SAIF, qui bénéficient des étapes de *dejittering* et de régularisation adaptative. Si les fines structures comme l'herbe sur *Cameraman* et la paille sur *Man* ont tendance à être sur-lissées sur les versions débruitées d'origine, le phénomène est corrigé sur les versions régularisées. Certains artefacts, au niveau de la tour par exemple, sont également réduits. De plus, grâce à la sélectivité de l'approche, si l'image ne souffre pas de *jittering* ou de bruit résiduel, la réduction de variance est jugée satisfaisante et le résultat est inchangé.

TABLE 5.1 – PSNR et SSIM obtenus après débruitage d’images corrompues par du bruit gaussien.

	<i>House</i> (256)	<i>Peppers</i> (256)	<i>Cameraman</i> (256)	<i>Boat</i> (512)	<i>Lena</i> (512)	<i>Barbara</i> (512)	<i>Man</i> (512)
	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM
Bruit gaussien, $\sigma = 20$							
Bruitée	22.12 0.33	22.11 0.47	22.14 0.39	22.13 0.73	22.11 0.67	22.12 0.78	22.10 0.79
TV	30.48 0.68	29.23 0.78	28.57 0.70	28.94 0.87	30.22 0.87	26.66 0.87	28.42 0.89
NL-TV	27.04 0.73	29.99 0.81	29.59 0.79	29.43 0.87	29.16 0.89	27.92 0.90	28.05 0.88
NL-Means	32.23 0.77	29.87 0.82	29.01 0.78	29.30 0.85	31.53 0.89	30.09 0.92	28.48 0.87
NLDJ	32.31 0.77	30.45 0.83	30.13 0.81	29.77 0.88	31.73 0.91	29.98 0.93	29.10 0.90
NA/R-NL	32.16 0.77	30.11 0.82	29.27 0.78	29.01 0.84	31.33 0.89	28.11 0.89	28.38 0.85
R-NL	32.69 0.79	30.78 0.84	30.19 0.82	29.92 0.88	32.04 0.91	29.76 0.93	29.26 0.89
BM3D	33.70 0.81	31.32 0.86	30.49 0.83	30.85 0.91	33.03 0.92	31.74 0.95	29.81 0.92
Bruit gaussien, $\sigma = 30$							
Bruitée	18.60 0.24	18.59 0.35	18.60 0.30	18.57 0.61	18.57 0.55	18.59 0.66	18.59 0.68
TV	26.04 0.71	27.19 0.73	27.43 0.64	27.25 0.80	27.46 0.82	25.92 0.80	26.33 0.80
NL-TV	25.15 0.61	25.97 0.68	26.26 0.63	26.04 0.81	26.15 0.83	25.68 0.85	25.75 0.82
NL-Means	30.11 0.72	27.90 0.75	27.36 0.72	27.38 0.78	29.52 0.84	27.80 0.87	26.69 0.80
NLDJ	30.06 0.71	28.11 0.76	27.96 0.72	27.74 0.82	29.62 0.86	27.61 0.88	27.10 0.84
NA/R-NL	30.17 0.74	27.82 0.76	27.23 0.73	27.14 0.77	29.32 0.84	25.86 0.83	26.50 0.78
R-NL	30.70 0.75	28.53 0.78	28.12 0.75	27.98 0.82	30.10 0.86	27.50 0.88	27.36 0.83
BM3D	32.22 0.77	29.31 0.82	28.66 0.78	29.04 0.87	31.25 0.89	29.76 0.92	28.00 0.87

TABLE 5.2 – PSNR et SSIM obtenus après débruitage d’images corrompues par du bruit de Poisson.

	<i>House</i> (256)	<i>Peppers</i> (256)	<i>Cameraman</i> (256)	<i>Boat</i> (512)	<i>Lena</i> (512)	<i>Barbara</i> (512)	<i>Man</i> (512)
	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM
Bruit de Poisson, $Q = 4$							
Bruitée	20.70 0.31	21.23 0.45	21.38 0.84	20.97 0.70	21.21 0.64	21.40 0.76	22.59 0.84
\mathcal{P} -TV	30.31 0.72	29.25 0.78	28.67 0.73	28.66 0.86	30.35 0.87	25.99 0.86	28.91 0.92
\mathcal{P} -NL-Means	31.62 0.75	29.48 0.80	28.98 0.77	28.79 0.83	31.36 0.88	29.55 0.91	28.74 0.89
\mathcal{P} -NLDJ	31.68 0.75	29.91 0.81	29.91 0.80	29.28 0.87	31.58 0.90	29.44 0.92	29.49 0.92
NA/R-NL	31.43 0.76	29.44 0.80	29.00 0.78	28.40 0.81	31.22 0.88	27.19 0.88	28.82 0.88
R-NL	32.19 0.78	30.22 0.83	29.96 0.81	29.43 0.86	31.95 0.90	29.14 0.92	29.67 0.91
Ans. + BM3D	33.30 0.80	30.89 0.85	30.35 0.82	30.42 0.90	32.88 0.92	31.23 0.94	30.17 0.93
Bruit de Poisson, $Q = 8$							
Bruitée	17.71 0.23	18.22 0.35	18.36 0.33	17.96 0.59	18.18 0.53	18.39 0.67	19.63 0.76
\mathcal{P} -TV	28.74 0.69	27.49 0.74	26.89 0.67	27.12 0.81	28.82 0.83	24.43 0.81	27.23 0.88
\mathcal{P} -NL-Means	28.69 0.65	27.37 0.71	27.17 0.68	26.66 0.72	29.04 0.81	26.81 0.83	26.83 0.79
\mathcal{P} -NLDJ	29.57 0.69	28.20 0.76	28.18 0.73	27.67 0.81	29.82 0.86	27.47 0.88	27.87 0.87
NA/R-NL	29.67 0.74	27.82 0.76	27.56 0.75	27.03 0.75	29.71 0.85	25.71 0.83	27.30 0.83
R-NL	30.18 0.74	28.59 0.79	28.35 0.77	27.88 0.81	30.33 0.87	27.29 0.88	28.08 0.87
Ans. + BM3D	31.78 0.76	29.23 0.81	28.80 0.78	28.88 0.86	31.36 0.89	29.57 0.92	28.56 0.90

TABLE 5.3 – PSNR et SSIM obtenus après débruitage d’images corrompues par du bruit gaussien ou du bruit de Poisson forts.

	<i>House</i> (256)	<i>Peppers</i> (256)	<i>Cameraman</i> (256)	<i>Boat</i> (512)	<i>Lena</i> (512)	<i>Barbara</i> (512)	<i>Man</i> (512)
	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM
Bruit gaussien, $\sigma = 40$							
Bruitée	16.10 0.18	16.10 0.27	16.12 0.24	16.09 0.51	16.11 0.46	16.09 0.57	16.10 0.59
TV	24.71 0.60	23.08 0.67	22.18 0.57	23.82 0.73	24.53 0.76	21.96 0.73	22.75 0.73
NL-TV	24.87 0.50	25.45 0.61	25.81 0.54	25.99 0.76	25.99 0.76	24.45 0.80	25.13 0.78
NL-Means	28.25 0.66	26.32 0.70	26.08 0.67	26.03 0.72	28.10 0.80	26.07 0.82	25.38 0.73
NLDJ	28.28 0.64	26.43 0.71	26.54 0.67	26.34 0.76	28.22 0.82	25.92 0.84	25.74 0.78
NA/R-NL	27.89 0.71	25.09 0.70	24.71 0.67	25.16 0.68	27.05 0.78	23.99 0.76	24.60 0.69
R-NL	29.05 0.71	26.85 0.75	26.70 0.72	26.61 0.76	28.77 0.83	25.72 0.83	26.04 0.78
BM3D	30.65 0.74	27.77 0.78	27.27 0.73	27.67 0.82	29.86 0.86	27.96 0.89	26.72 0.83
Bruit de Poisson, $Q = 12$							
Bruitée	15.91 0.19	16.43 0.30	16.54 0.29	16.23 0.53	16.42 0.48	16.65 0.61	17.86 0.70
\mathcal{P} -TV	27.07 0.59	26.10 0.67	25.67 0.59	25.95 0.77	27.18 0.77	23.60 0.76	25.96 0.84
\mathcal{P} -NL-Means	27.76 0.62	26.71 0.70	26.36 0.67	26.15 0.71	28.54 0.80	26.26 0.82	26.47 0.82
\mathcal{P} -NLDJ	28.29 0.63	27.20 0.71	27.04 0.67	26.76 0.78	28.82 0.83	26.46 0.85	27.03 0.85
NA/R-NL	28.53 0.72	26.84 0.74	26.49 0.72	26.07 0.71	28.81 0.83	24.44 0.79	26.49 0.80
R-NL	29.12 0.71	27.75 0.76	27.23 0.73	27.04 0.78	29.49 0.85	26.21 0.85	27.31 0.85
Ans. + BM3D	30.75 0.74	28.29 0.78	27.73 0.74	27.93 0.83	30.38 0.86	28.46 0.89	27.73 0.87

TABLE 5.4 – PSNR et SSIM obtenus après débruitage d’images corrompues par du bruit gamma.

	<i>House</i> (256)	<i>Peppers</i> (256)	<i>Cameraman</i> (256)	<i>Boat</i> (512)	<i>Lena</i> (512)	<i>Barbara</i> (512)	<i>Man</i> (512)
	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM	PSNR SSIM
Bruit gamma, $L = 12$							
Image bruitée	21.68 0.37	22.33 0.52	22.36 0.51	22.12 0.74	22.44 0.70	22.68 0.82	24.39 0.88
\mathcal{G} -TV	29.36 0.65	29.13 0.76	29.02 0.71	28.50 0.86	29.84 0.84	27.18 0.89	29.65 0.93
\mathcal{G} -NL-Means	32.24 0.77	29.72 0.82	29.39 0.80	29.35 0.85	32.22 0.91	30.22 0.93	29.79 0.92
\mathcal{G} -NLDJ	32.59 0.77	30.69 0.83	30.68 0.83	30.01 0.89	32.58 0.92	30.42 0.94	30.82 0.94
NA/R-NL	32.53 0.78	30.65 0.83	30.10 0.82	29.40 0.84	32.67 0.91	28.59 0.91	30.53 0.92
R-NL	33.09 0.80	31.12 0.85	30.87 0.84	30.22 0.88	33.05 0.92	30.33 0.94	31.03 0.94
log-BM3D	34.01 0.82	31.70 0.87	30.99 0.84	31.10 0.91	33.81 0.93	32.01 0.95	31.45 0.95
Bruit gamma, $L = 4$							
Image bruitée	16.94 0.24	17.60 0.37	17.66 0.40	17.40 0.59	17.75 0.55	17.95 0.68	19.65 0.78
\mathcal{G} -NL-Means	29.20 0.69	27.19 0.74	26.81 0.71	26.96 0.75	29.62 0.84	27.20 0.86	27.21 0.85
\mathcal{G} -NLDJ	29.20 0.66	27.56 0.74	27.65 0.72	27.48 0.81	29.78 0.86	27.15 0.88	28.11 0.89
NA/R-NL	30.03 0.74	27.78 0.77	27.49 0.75	27.21 0.76	30.17 0.87	25.86 0.84	28.39 0.88
R-NL	30.15 0.74	28.09 0.78	27.99 0.76	27.86 0.81	30.46 0.87	27.18 0.88	28.49 0.89
log-BM3D	31.41 0.76	28.79 0.80	28.50 0.78	28.60 0.85	31.15 0.88	29.21 0.91	28.81 0.91

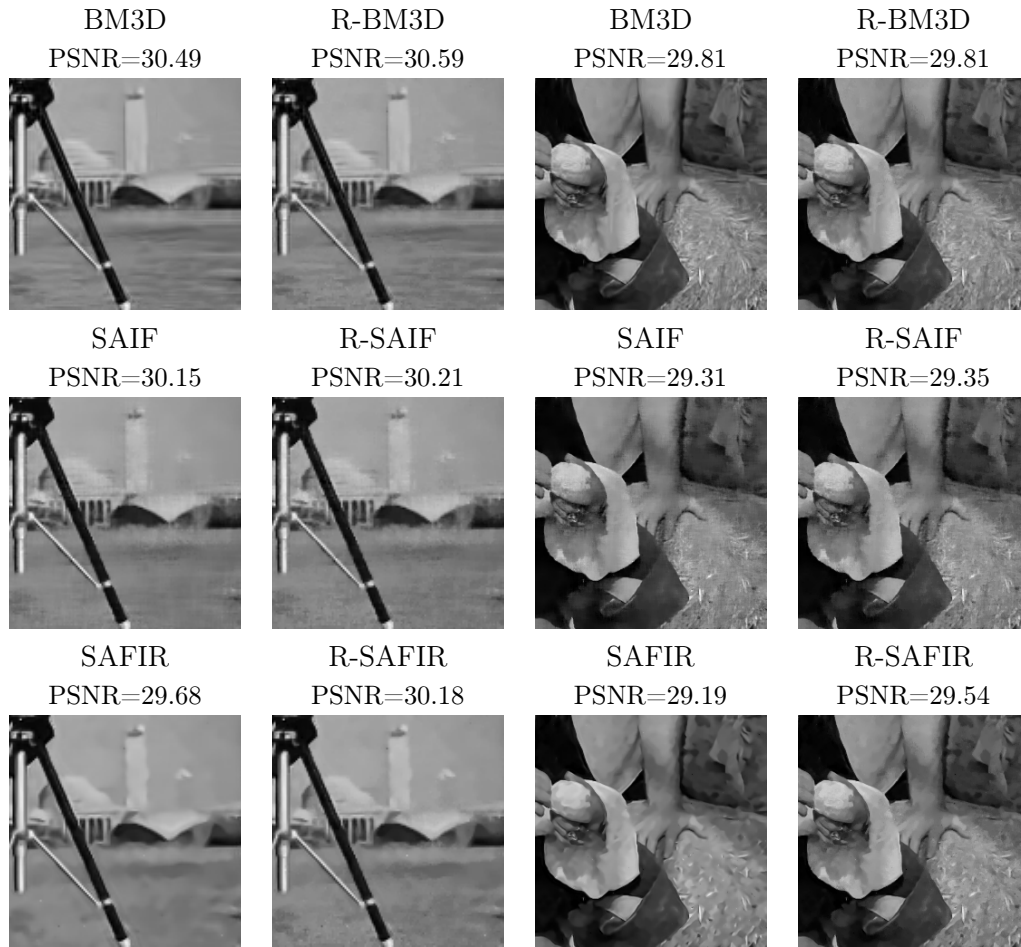


FIGURE 5.8 – Débruitage de bruit gaussien, écart type $\sigma=20$, à l'aide de BM3D, SAFIR et SAIF et les versions régularisées associées R-BM3D, R-SAFIR et R-SAIF qui bénéficient du *dejittering* et de la régularisation adaptative. Les artefacts observés par exemple sur l'herbe et au niveau de la tour sur *Cameraman* ou sur la paille dans *Man* sont réduits, mais grâce à la sélectivité du modèle ces étapes supplémentaires ne dégradent pas le résultat quand celui-ci ne présente pas ces défauts particuliers, par exemple pour le débruitage de *Man* par BM3D qui ne présente ni *jittering* ni bruit résiduel.

5.7 Conclusion

Nous avons présenté dans ce chapitre une coopération entre les méthodes variationnelles et non locales basée sur une régularisation adaptative des NL-means. La première étape consiste à réduire le *jittering* des NL-means en rétablissant le compromis biais-variance, ce qui assure que la réduction de bruit est pertinente. Puis la régularisation adaptative est effectuée à l'aide d'une attache aux données non locale, pondérée par un indice de confiance associé au débruitage effectué par

les NL-means. Cette régularisation permet de corriger l'effet de patch rare lié à la présence de bruit résiduel, mais son caractère local et adaptatif n'introduit pas les défauts classiquement associés à la minimisation TV, à savoir l'effet de crénelage et la perte de résolution et de contraste.

L'implémentation de la méthode s'applique à de nombreux types de bruit, en particulier au cas d'un bruit dépendant du signal à variance polynomiale, ce qui la rend apte à traiter le bruit qui corrompt les images BNL.

La dernière étape consiste donc à exploiter la redondance spatiale du flux vidéo pour adapter l'algorithme R-NL au traitement des séquences d'images, ce qui s'effectue naturellement grâce aux bonnes propriétés des NL-means.

R-NL pour le débruitage vidéo

6.1 Introduction

Afin de traiter les séquences d'images BNL en exploitant la redondance d'information fournie par le flux vidéo, l'algorithme développé dans le chapitre précédent est adapté au débruitage vidéo. En effet, et cela a entre autres justifié notre choix de les exploiter, les NL-means offrent un débruitage vidéo efficace sans compensation de mouvement préalable [Buades 2005a]. Il est d'ailleurs même suggéré que la compensation de mouvement peut se montrer contre-productive, à vouloir limiter le nombre de candidats au lieu de chercher à bénéficier au maximum de la redondance d'information. De plus, l'utilisation de patchs spatio-temporels, suggérée dans [Boulanger 2007, Wexler 2007], permet d'assurer la cohérence temporelle d'une trame à l'autre. Au lieu de comparer des patchs "plats" issus d'une seule trame dans une fenêtre de recherche spatio-temporelle, l'utilisation de patchs tridimensionnels et la prise en compte de l'information temporelle permet de garantir la cohérence dans le temps. Si les candidats sélectionnés sont plus robustes ils sont cependant moins nombreux, ce qui se traduit par un effet de patch rare accru. Ce dernier est alors corrigé par la régularisation TV adaptative, appliquée elle aussi sur les trois dimensions spatio-temporelles. Ainsi, le modèle combiné R-NL s'adapte naturellement au cas de la vidéo, à l'aide des NL-means utilisant des patchs spatio-temporels associés à une régularisation TV adaptative elle aussi spatio-temporelle.

Des résultats sur des séquences synthétiques illustrent la performance de la méthode proposée, notamment en termes de régularité temporelle de la solution. Enfin, l'algorithme est adapté au débruitage des séquences d'images BNL à l'aide de l'estimation des paramètres de la NLF réalisée dans la partie précédente.

6.2 Adaptation 3D

Les moyennes non locales ont naturellement été adaptées dans [Buades 2008] au débruitage de séquences d'images. Des améliorations ont par la suite été envisagées, par exemple dans [Boulanger 2007, Mahmoudi 2005]. Les avantages qu'apportent R-NL aux NL-means peuvent également se porter au débruitage vidéo. L'algorithme des NL-means dit temporel, c'est-à-dire qui traite des séquences d'images, applique un filtrage spatio-temporel sans recalage préalable des trames les unes par rapport aux autres, c'est-à-dire sans compensation de mouvement. En effet, la compensation de mouvement est une étape sensible et l'un des verrous du débruitage spatio-temporel. Elle peut de plus se révéler difficile à résoudre sur des zones constantes

où le problème d'ouverture est trop élevé, et donc la mise en correspondance des structures est délicate. Buades *et al.* ont en fait montré que dans ce cas, la compensation de mouvement se révèle contre-productive : lors du débruitage de séquences d'images, les moyennes non locales utilisent donc des fenêtres de recherche spatio-temporelles en trois dimensions, la troisième dimension correspondant à l'évolution des trames dans le temps. La compensation de mouvement cible les structures et donc réduit le nombre de candidats potentiels, quand finalement pour les NL-means plus on en a, mieux c'est. Pire, si la compensation de mouvement est inexacte elle peut nuire à la qualité du débruitage et introduire du flou.

Dans l'implémentation vidéo classique des NL-means (NL-means-2D) qui utilise une fenêtre de recherche spatio-temporelle mais des patches bidimensionnels et dans la version régularisée spatialement (R-NL-2D) associée, si le nombre de trames utilisées dans le calcul de la solution est relativement faible, ce qui est en général le cas afin de réduire les temps de calcul et la mémoire, aucune stabilité temporelle n'est garantie. En effet, la compensation de mouvement est basée sur l'hypothèse lambertienne qui suppose que le long de sa trajectoire, un pixel conserve son niveau de gris, ce qui garantit une certaine régularité temporelle. Dans le cas des NL-means-2D, cette hypothèse n'est plus vérifiée, donc la valeur d'un pixel peut changer d'une trame à l'autre, même si celui-ci ne s'est pas déplacé. Cela entraîne un effet de scintillement sur l'ensemble de la vidéo, même si celui-ci n'est pas quantifiable en terme de PSNR ou en observant une seule trame. Un tel phénomène n'est pas préjudiciable lorsque l'ensemble de l'image est en mouvement, contrairement au cas où seule une partie du champ est en mouvement tandis qu'une autre est statique. L'effet de scintillement, similaire à l'effet de patch rare dans le cas des images, révèle alors la conséquence indésirable de la variabilité de l'estimation.

L'utilisation de patches spatio-temporels, c'est-à-dire en trois dimensions, au lieu des patches bidimensionnels classiques, permet de réduire cet effet gênant. En effet, dans la version classique des NL-means adaptés à la vidéo, des patches 2D de l'image courante sont comparés à des patches 2D issus d'une fenêtre de recherche spatio-temporelle qui est elle en trois dimensions. D'une certaine façon, la stabilité temporelle d'une trame à l'autre n'est pas prise en compte, ce qui entraîne l'effet de scintillement. Au contraire, l'utilisation de patches tridimensionnels permet de comparer des voisinages spatio-temporels, et donc de prendre en compte les structures et les régularités à la fois spatiales et temporelles, afin de forcer une certaine cohérence dans le temps. Cette idée a été suggérée dans [Boulanger 2007], ainsi que dans [Boulanger 2010] pour le débruitage de séquences d'images 3D et dans [Wexler 2007] dans le contexte de l'inpainting.

Cependant, l'utilisation seule de patches 3D favorise l'effet de patch rare. En effet, les patches tridimensionnels rendent le critère beaucoup plus sélectif, ce qui abaisse le nombre de candidats et donc augmente le bruit résiduel. C'est alors la régularisation TV adaptative qui réduit ce défaut. La première idée consisterait à appliquer localement comme précédemment une régularisation TV uniquement spatiale, basée sur le paramètre de régularisation adaptatif λ . Cela permettrait effectivement de réduire le bruit résiduel, en revanche la stabilité temporelle est dégradée également, puisque



FIGURE 6.1 – Illustration du débruitage TV spatio-temporel.

la régularisation TV est appliquée indépendamment sur chaque trame, sans souci de cohérence temporelle. Pour remédier à cela, une régularisation TV spatio-temporelle est introduite. Une telle approche peut sembler contre-intuitive dans le contexte habituel, car l'hypothèse du modèle constant par morceaux le long de l'axe temporel ne semble pas raisonnable (cela étant, elle n'est pas particulièrement réaliste dans le domaine spatial non plus mais elle est largement utilisée pour autant). Illustré sur la figure 6.1, le débruitage vidéo par la seule régularisation TV spatio-temporelle présente les caractéristiques classiques de la variation totale, mais en trois dimensions : les zones homogènes dans le temps, c'est-à-dire immobiles, sont bien restaurées, et la stabilité temporelle est assurée, mais on observe une perte de contraste autour des structures en mouvement. Associé au cadre de travail R-NL en revanche, la régularisation TV est appliquée très localement et de façon sélective là où les NL-means n'ont pas pu lisser suffisamment, ce qui permet comme dans le cas du débruitage d'images de limiter l'apparition de l'effet de crénelage. La régularisation TV temporelle assure la cohérence temporelle quand celle-ci est appliquée, mais ne dégrade pas le contenu de la vidéo ailleurs. De plus, le passage en trois dimensions du calcul de la variation totale est direct : il nécessite simplement la mise en place d'un opérateur de gradient et de divergence en trois dimensions.

6.3 Restauration de vidéos

L'application de R-NL au débruitage de séquences d'images est illustrée ici à partir de trois séquences d'images : *Target*, *Tennis* et *Bicycle*. Les séquences d'images originales, bruitées ainsi que les résultats restaurés sont disponibles au téléchargement¹. Les figures 6.2 et 6.3 illustrent de façon qualitative l'intérêt des patches 3D et de la régularisation adaptative, sur les séquences *Target* et *Tennis*. On observe en particulier une meilleure préservation des textures permise par l'utilisation des

1. <http://image.math.u-bordeaux1.fr/RNL>

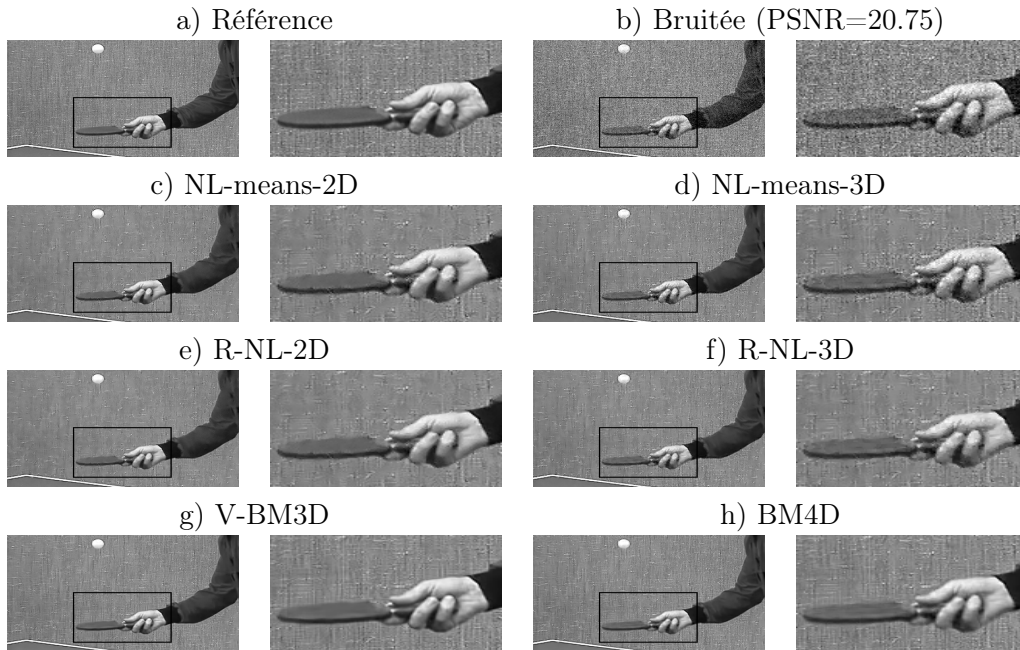


FIGURE 6.2 – Différence entre patches 2D et patches 3D. Extraits de séquences d’images débruitées par des méthodes à patches 2D, à savoir NL-means-2D, R-NL-2D et V-BM3D, et par des méthodes à patches 3D, NL-means-3D, R-NL-3D et BM4D. La texture du fond de l’image est perdue à cause de l’effet de *jittering* avec NL-means-2D et R-NL-2D. NL-means-3D préserve mieux le fond mais souffre de bruit résiduel autour du bras en mouvement, effet corrigé par la régularisation adaptative de R-NL-3D. Ce dernier a tendance à moins lisser que BM4D, comme on peut l’observer sur la main et le mur.

patches 3D. La régularité temporelle accrue n’est observable que sur les vidéos. La régularisation adaptative prend alors tout son sens combinée aux patches 3D, où elle permet de réduire le bruit résiduel. Le tableau 6.1 fournit les PSNR moyens des vidéos restaurées à l’aide des NL-means standards adaptés à la vidéo (NL-means 2D), de l’algorithme des R-NL adapté au débruitage vidéo (R-NL 2D), des NL-means utilisant des patches 3D (NL-means 3D), de R-NL utilisant des patches 3D (R-NL 3D) ainsi qu’avec les algorithmes de débruitage V-BM3D [Dabov 2007a] et BM4D [Maggioni 2012].

La cohérence temporelle obtenue avec chacun des algorithmes ci-dessus est également évaluée, afin de montrer la meilleure cohérence temporelle assurée par l’utilisation de patches 3D. Pour cela, la variance temporelle est calculée sur les séquences d’images afin d’illustrer l’hypothèse lambertienne : sur les zones qui sont immobiles au cours d’une partie de la séquence, la valeur des pixels devrait être inchangée d’une trame à l’autre, donc la variance temporelle doit être proche de zéro. À partir de la vérité terrain fournie par les séquences d’origine, des zones fixes dans le temps sont sélectionnées, et la variance temporelle est calculée pour chacun des résultats de dé-

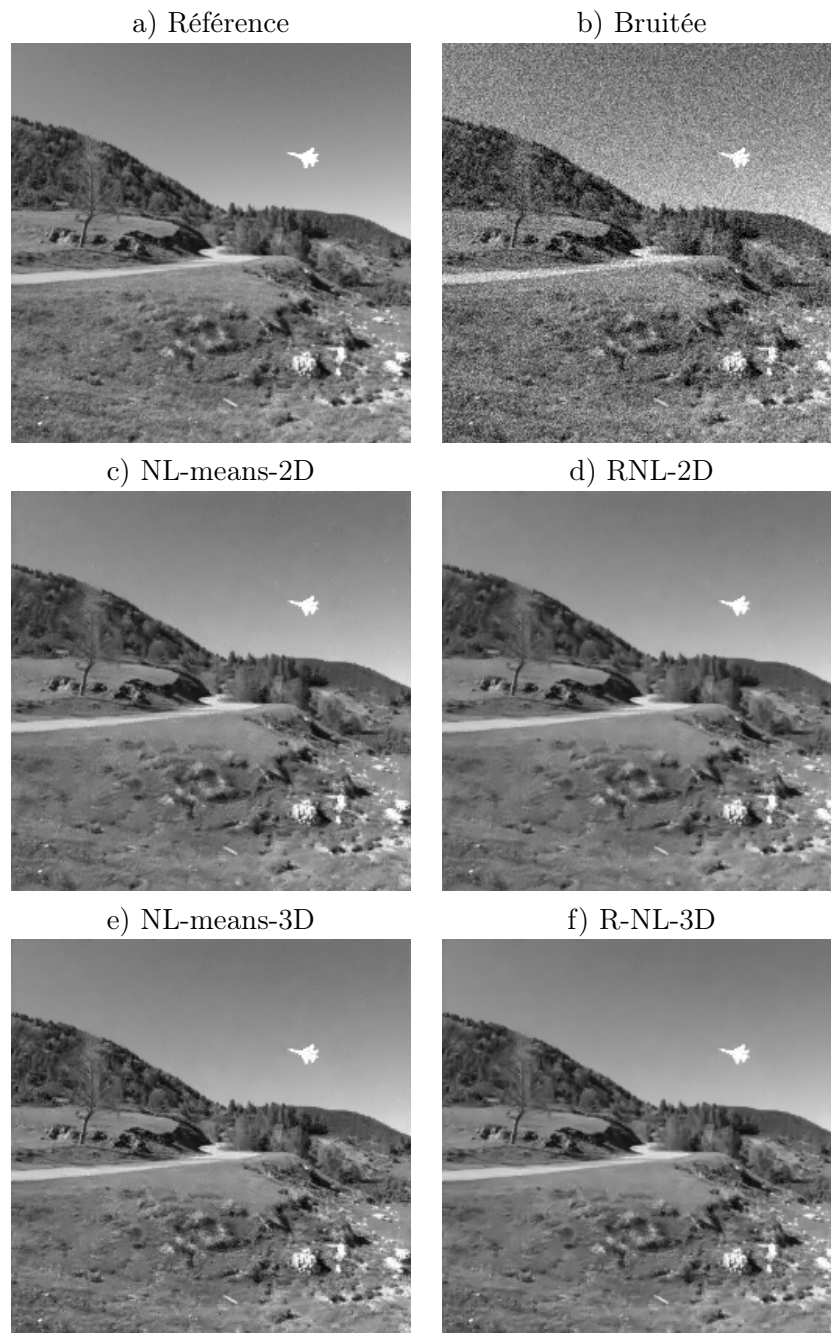


FIGURE 6.3 – Extraits de la séquence d'images *Target*. a) Originale, b) Image corrompue par un bruit blanc gaussien d'écart type 20, PSNR initial égal à 22.11dB. c) Débruitage obtenu par les NL-means à patches 2D et d) Régularisation adaptative. Malgré le *dejittering*, la texture de l'herbe a tendance à être sur-lissée. e) NL-means à patches 3D. L'herbe est mieux préservée mais du bruit résiduel est présent au niveau des arbres. f) R-NL-3D, qui combine la réduction de l'effet de patch rare avec une bonne préservation des textures.

TABLE 6.1 – PSNR moyen obtenu sur des séquences d’images débruitées par les NL-means-2D, R-NL-2D et V-BM3D (patchs 2D), ainsi que NL-means-3D, R-NL-3D et BM4D (patchs 3D).

	<i>Target</i>	<i>Tennis</i>	<i>Bicycle</i>
NL-means 2D [Buades 2008]	30.84	30.07	<i>32.24</i>
R-NL-2D	31.27	30.16	32.29
V-BM3D [Dabov 2007a]	30.21	29.79	32.92
NL-means 3D	32.47	30.52	31.72
R-NL-3D	<i>33.35</i>	<i>30.60</i>	32.16
BM4D [Maggioni 2012]	34.53	31.06	33.37

TABLE 6.2 – Écart type temporel mesuré sur les séquences débruitées à l’aide des NL-means et R-NL à patchs 2D, V-BM3D, NL-means et R-NL à patchs 3D, et BM4D. L’utilisation de patchs 3D chez R-NL-3D et BM4D permet d’assurer une meilleure cohérence temporelle et plus de confort visuel.

	<i>Target</i>	<i>Tennis</i> (1-24)	<i>Tennis</i> (90-148)	<i>Bicycle</i>
NL-means 2D [Buades 2008]	5.45	4.27	4.36	1.64
R-NL-2D	4.47	3.34	3.57	1.25
V-BM3D [Dabov 2007a]	5.25	3.66	4.60	1.76
NL-means 3D	5.20	4.04	4.15	1.37
R-NL-3D	3.85	2.82	3.09	0.90
BM4D [Maggioni 2012]	3.67	2.78	3.14	0.94

bruitage obtenus. Le tableau 6.2 révèle l’écart-type temporel obtenu dans chaque cas, à partir des trois mêmes séquences. Il permet d’établir que la cohérence temporelle est mieux préservée pour R-NL-3D et BM4D, en particulier par rapport respectivement à R-NL-2D et V-BM3D, grâce à l’utilisation des patchs 3D. La différence de résultats entre R-NL-3D (resp. R-NL-2D) et NL-means-3D (resp. NL-means-2D) montre également que la régularisation TV spatio-temporelle adaptative permet de réduire l’effet de scintillement indésirable.

6.4 Application au débruitage de vidéos BNL

L’algorithme décrit ci-dessus dans le cas de la vidéo peut s’appliquer au débruitage des séquences d’images BNL. Cependant, les contraintes opérationnelles limitent la taille mémoire disponible, et donc en particulier le nombre de trames que peut utiliser l’algorithme pour la recherche spatio-temporelle, ainsi que les temps de calcul, ce qui se répercute notamment sur la taille de la fenêtre de recherche et la dimension des patchs. Il est évident également que dans le cadre d’une implémentation embarquée qui traite le flux vidéo au fur et à mesure de son acquisition, on ne peut se reposer pour la fenêtre de recherche temporelle que sur les trames

TABLE 6.3 – Sélection des paramètres optimaux pour le débruitage des séquences d’images BNL.

		R-NL-2D	R-NL-3D
Patches ρ	Rayon spatial r_s	$r_s = 3$	
	Taille temporelle r_t	$r_t = 0$	$r_t = 2$
Fenêtre de recherche W	Rayon spatial w_s	$w_s = 3$	
	Taille temporelle w_t	$w_t = 4$	
Paramètre de régularisation γ		$\gamma = 200$	$\gamma = 100$
Itération (TV)		$T = 100$	

TABLE 6.4 – Complexité de deux implémentations des NL-means : implémentation classique et implémentation rapide de [Darbon 2008] qui fournit une implémentation indépendante de la taille des patches, dans le cas de patches 2D ou 3D.

NL-means Patches 2D		NL-means Patches 3D	
Classique	Rapide	Classique	Rapide
$ \Omega W_s ^2 W_t \rho_s ^2$	$ \Omega W_s ^2 W_t 2^2$	$ \Omega W_s ^2 W_t \rho_s ^2 \rho_t $	$ \Omega W_s ^2 W_t 2^3$

antérieures, à moins d’introduire une latence.

Un stage de master a été défini et co-encadré avec Thales afin d’étudier les différentes versions des algorithmes et de dégager d’une part les paramètres mais aussi les choix d’implémentation optimaux, en terme de compromis entre performance et complexité calculatoire. Anne-Sophie Auger, étudiante en master à l’école de Télécommunications de Lille, s’est intéressée en particulier au choix d’utiliser des patches bi ou tridimensionnels, ainsi qu’à l’influence de la fenêtre de recherche temporelle. En effet, comme il été évoqué dans la section précédente, les patches spatio-temporels permettent de garantir une meilleure stabilité temporelle du débruitage, en particulier sur les zones qui ne changent pas le temps pour lesquelles l’hypothèse lambertienne est primordiale. Sur des séquences acquises en vol néanmoins, la scène globale est toujours en mouvement, ce qui peut limiter l’intérêt des patches 3D par rapport au surcoût calculatoire qu’ils engendrent. L’influence de la fenêtre de recherche, en particulier temporelle ici, est également à prendre en considération. Augmenter la taille de la fenêtre de recherche permet a priori d’augmenter la redondance d’information et devrait donc garantir un débruitage plus efficace. Cependant, outre les temps de calcul accrus, cela introduit également une latence, et en cas de mouvements importants entre les trames augmenter la taille de la fenêtre temporelle peut même se montrer contre-productif. En effet si dans des trames trop lointaines les structures que l’on cherche à débruiter (par exemple l’avion dans la séquence *Target*) ont disparu de la fenêtre, on risque de sélectionner de mauvais candidats et introduire du *jittering*. Un compromis doit donc être établi.

6.4.1 Choix algorithmiques et sélection des paramètres

Le tableau 6.3 fournit les paramètres qui ont été retenus à l'issue de l'étude du stage. Les tailles de patches r_s et de fenêtre de recherche temporelles w_s correspondent en fait au rayon en nombre de pixels, de sorte que le nombre de pixels total du patch (respectivement de la fenêtre de recherche) dans la dimension spatiale est $|\rho_s|^2 = (2r_s + 1)^2$ (respectivement $|W_s|^2 = (2w_s + 1)^2$). Et puisque l'on ne considère que les trames antérieures, les tailles temporelles r_t et w_t correspondent au nombre de trames antérieurs considérés en plus du trame actuel, de sorte que le nombre de pixels total du patch (respectivement de la fenêtre de recherche) dans la dimension temporelle est $|\rho_t| = r_t + 1$ (respectivement $|W_t| = w_t + 1$). Ainsi, la taille totale des patches (respectivement de la fenêtre de recherche) est $|\rho| = |\rho_s|^2 |\rho_t|$ (respectivement $|W| = |W_s|^2 |W_t|$). On note que quatre trames antérieures ont été retenues pour la fenêtre de recherche temporelle ; il s'agit d'après le stage effectué d'un bon compromis entre taille mémoire, redondance et latence.

Le tableau 6.4 compare les complexités associées à deux implémentations. L'implémentation classique pour le traitement d'une image a une complexité qui dépend de la taille de l'image $|\Omega|$, de la taille de la fenêtre de recherche ainsi que de la taille des patches. L'implémentation rapide proposée dans [Darbon 2008] fournit une méthode de calcul rapide des poids qui permet d'obtenir une complexité indépendante de la taille des patches. Les complexités fournies dans le tableau 6.4 sont valables pour le débruitage d'une image, ou d'une vidéo si l'on traite simultanément toutes les trames sous forme d'un bloc de taille $|\Omega|$ (où $|\Omega|$ serait un domaine de dimension trois). Si l'on traite le flux vidéo au fur et à mesure de l'acquisition, pour chaque trame à traiter il faut manipuler un bloc composé de $|W_t|$, mais un raffinement est possible pour ne pas calculer toutes les similarités du bloc mais seulement celles impliquant la trame à traiter. De plus, une parallélisation est possible.

6.4.2 Compensation de mouvement

L'un des arguments en faveur des NL-means pour le débruitage de séquences d'images repose sur le fait que l'on peut s'affranchir de la compensation de mouvement préalable entre les trames [Buades 2005a]. Au contraire dans [Boulanger 2007, Liu 2010], les auteurs justifient que la compensation de mouvement est bénéfique, même dans un contexte non local, car elle permet de cibler le déplacement et de maximiser les chances de trouver de bons candidats. Ils intègrent pour cela une estimation par flot optique au sein des NL-means afin de permettre un débruitage plus efficace tout en assurant une meilleure stabilité temporelle. Cette opération est malheureusement coûteuse dans un contexte temps réel.

Cependant, nous disposons en pratique d'une estimation du mouvement de la caméra. En effet, le casque sur lequel est intégré le dispositif d'acquisition est équipé d'un détecteur de posture qui fournit des informations sur le mouvement de la tête du pilote, et les systèmes inertiels de l'hélicoptère permettent d'évaluer le mouvement global de l'appareil. Cela permet donc d'effectuer une compensation de mouvement

grossière et à bas coût, maximisant ainsi les chances de trouver de bons candidats pour effectuer un débruitage adapté. Notons que cette prise en compte du mouvement peut s'effectuer en recalant les trames par rapport à la trame actuelle (ce que nous appliquons en pratique), ou bien de façon plus sommaire en orientant la fenêtre de recherche temporelle dans la direction du mouvement.

6.4.3 Illustrations sur données réelles

Le stage de master que nous avons encadré s'est consacré aux NL-means (en mettant de côté la régularisation adaptative) et a analysé les performances et les défauts associés à l'utilisation de patches 2D ou 3D. La figure 6.4 illustre les différences de performances liées aux patches 2D, aux patches 3D ainsi qu'aux patches 3D après recalage temporel des trames. Ce dernier a été obtenu ici à l'aide de l'algorithme Motion 2D [Odobez 1995] qui permet de simuler l'estimation de mouvement fournie par les capteurs embarqués. L'algorithme se base sur l'estimation automatique préalable du bruit, et nous avons utilisé pour ces résultats la version hybride NLF-means. Les séquences complètes peuvent être consultées à <http://image.math.u-bordeaux1.fr/RNL>. Comme il a été mentionné dans le cas des images synthétiques, les patches 2D souffrent avant tout d'un manque de régularité temporelle qui se manifeste par un effet de scintillement, en particulier sur les zones uniformes. On observe également un effet de *jittering* lié à la difficulté de sélectionner des candidats pertinents, notamment sur la figure de gauche où le rapport signal à bruit est très faible. Ces deux effets indésirables sont corrigés par l'utilisation de patches 3D, mais au prix d'un effet de patch rare accru, autour des structures fines comme les arbres. Ici, plutôt que d'effectuer une régularisation adaptative, la régularité est assurée par le recalage préalable des trames. Ce dernier permet en effet de maximiser la redondance et donc le nombre de candidats pertinents, ce qui réduit à la fois le *jittering* et l'effet de patch rare.

6.5 Conclusion

Ce chapitre a étendu la régularisation adaptative des NL-means au débruitage de séquences d'images. Pour cela, nous avons associé les NL-means vidéo à fenêtre de recherche spatio-temporelle à l'utilisation de patches 3D, combinés à une régularisation TV spatio-temporelle adaptative. Cela permet d'obtenir des séquences restaurées exemptes des effets indésirables de *jittering*, de bruit résiduel et de scintillement temporel. Associé à l'estimation de la NLF réalisée en première partie, l'algorithme peut ainsi s'appliquer au débruitage des séquences d'images BNL.

En résumé, cette deuxième partie s'est consacrée au problème de débruitage des images et des séquences d'images, pour permettre de restaurer les images IL. Nous avons étudié en particulier deux méthodes classiques de débruitage, les moyennes non locales et la minimisation TV. Les défauts respectifs de chacune des méthodes

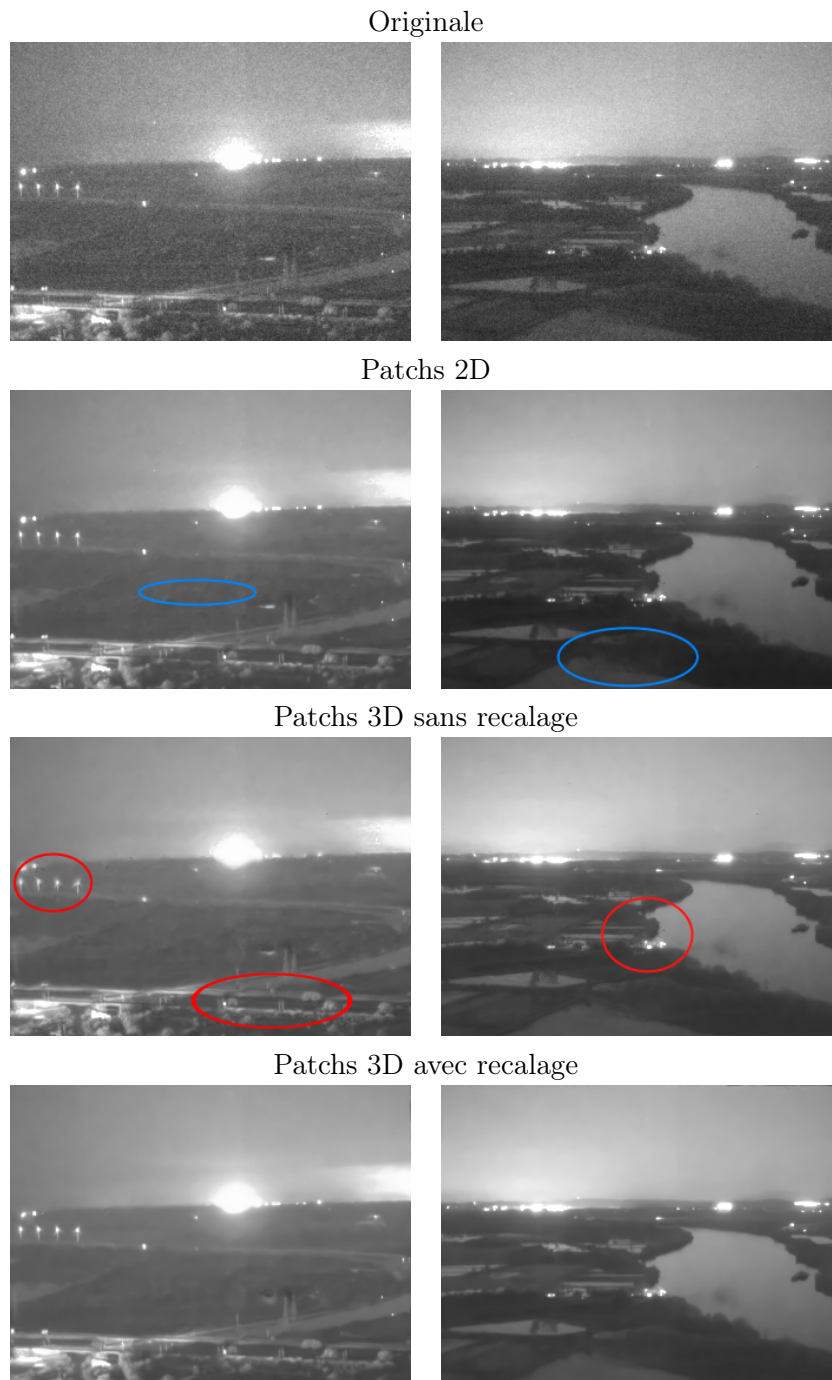


FIGURE 6.4 – Comparaison du débruitage de séquences réelles obtenu avec des patchs 2D et des patchs 3D, avec et sans recalage. Illustration du *jittering* (en bleu) avec les patchs 2D, et de l'effet de patch rare (en rouge) accru par les patchs 3D. Le recalage permet de corriger ces deux défauts. Cette figure est optimisée pour un affichage numérique en couleur. Les défauts illustrés, ainsi que l'effet de scintillement associé aux patchs 2D, sont observables sur les séquences vidéos.

ont été mis en avant afin de les combattre, et une méthode qui combine les deux approches a été développée. Le principe repose sur une minimisation TV adaptative basée sur une attache aux données non locale, pondérée par un indice de confiance qui traduit la qualité du débruitage effectué par les NL-means. Une mesure de la réduction de variance opérée par les NL-means offre un indicateur de qualité qui est ensuite utilisé localement pour ajuster la régularisation TV. Cela permet de définir un algorithme flexible qui utilise à la fois les propriétés de redondance et de régularité des images naturelles pour permettre une réduction efficace des effets de *jittering* et de patch rare dont souffrent les NL-means, ainsi que de l'effet de crénelage associé à la variation totale. Le modèle s'accompagne d'une interprétation intuitive qui s'adapte à de nombreux modèles de bruit, et une résolution rapide peut être obtenue pour des bruits issus de la famille exponentielle.

Si la méthode a été développée dans le contexte des NL-means et de la régularisation TV, le modèle peut s'étendre naturellement à d'autres termes de régularisation en fonction des a priori disponibles sur les images à traiter. Au delà des problèmes de débruitage, ce modèle pourrait s'étendre également à d'autres problèmes basés sur des moyennes pondérées.

La souplesse de la méthode permet de l'exploiter pour le débruitage des séquences d'images BNL, en utilisant l'estimation des paramètres de la fonction de niveau de bruit effectuée dans la partie précédente. La dernière étape consisterait à optimiser et paralléliser les algorithmes afin d'envisager une implémentation embarquée qui offrirait aux pilotes un débruitage temps réel des images projetées sur leur visière.

D'autres traitements pourraient ensuite être réalisés à partir de ces images débruitées, notamment du rehaussement de contraste, la correction des artefacts comme la suppression des reflets et des halos, ou encore du rehaussement de contours. Cette dernière tâche peut être contournée à l'aide des images infrarouge, dont les contours sont souvent mieux prononcés. Réaliser la fusion des deux modalités permettrait donc entre autres de rehausser artificiellement les contours de l'image IL à l'aide de ceux de l'image IR. Cependant, comme les images sont acquises à l'aide de caméras localisées à des emplacements différents, l'angle de vue varie, et donc un recalage des deux modalités est nécessaire avant de pouvoir effectuer la fusion des capteurs. Le recalage des images optiques et infrarouge fait l'objet de la partie suivante.

Troisième partie

Recalage multimodal

Table des matières

7	Critère de recalage basé sur l’alignement des contours	187
7.1	Introduction	187
7.2	Définition du critère	193
7.3	Analyse du modèle proposé	194
7.4	Extension aux transformations projectives	206
7.5	Conclusion	208
8	Schéma d’optimisation du critère de recalage	211
8.1	Introduction	211
8.2	Montée de gradient	211
8.3	Implémentation temporelle	215
8.4	Analyse du schéma proposé	217
8.5	Résultats	220
8.6	Conclusion	227
9	Conclusion et perspectives	229
	Bibliographie	237

Critère de recalage basé sur l’alignement des contours

7.1 Introduction

7.1.1 Contexte opérationnel et motivations

Le recalage d’images consiste à aligner dans un même plan plusieurs images d’une même scène prises sous différents angles de vues ou à différents instants, ou bien des images issues de différents capteurs. Dans le premier cas le recalage est dit unimodal car les images sont issues de la même modalité, et dans le second cas le recalage est dit multimodal. Le recalage multimodal est particulièrement utilisé en imagerie médicale, par exemple pour comparer des images du cerveau obtenues par tomodensitométrie (CT, pour *Computer Tomography*) à des images acquises par tomographie à émission de positrons (PET, pour *Positron Emission Tomography*) ou par imagerie à résonance magnétique (IRM). Les différents capteurs permettent de fournir des informations diverses, en révélant notamment différents tissus, ce qui apporte une aide précieuse au diagnostic. Le recalage multimodal est également beaucoup utilisé en télédétection, pour associer par exemple des images SAR (pour *Synthetic Aperture Radar*) à des images optiques aériennes conventionnelles. Il constitue souvent une étape préliminaire essentielle avant d’effectuer des analyses de haut niveau comme de la fusion de capteurs, de la détection de changements, de la réalité augmentée, etc. Un descriptif des méthodes de recalage usuelles est effectué dans [Brown 1992, Zitova 2003].

Le but ici est d’aligner des séquences d’images optiques issues d’un dispositif de vision nocturne à des séquences d’images infrarouge acquises pendant le vol de nuit d’un hélicoptère, pour permettre la fusion ultérieure de ces deux modalités. Conformément au dispositif décrit dans la première partie, les images optiques sont issues d’un intensificateur de lumière (IL) qui multiplie artificiellement les photons afin d’amplifier la luminosité. Ce dernier est couplé à une caméra CCD (*Charge Coupled Device*) pour permettre l’obtention d’images numériques. Le dispositif IL est situé au dessus de la visière de casque du pilote, et la séquences d’images générée est directement projetée sur la visière. Les images issues d’un tel dispositif sont plutôt conformes à la perception visuelle, quoique souvent vertes ou en niveau de gris, et souffrent des défauts classiques associés à la vision de nuit : elles sont dégradées par un fort bruit non gaussien lié au faible nombre de photons présents, elles sont peu contrastées malgré une très grande dynamique, et elles souffrent de nombreux arte-

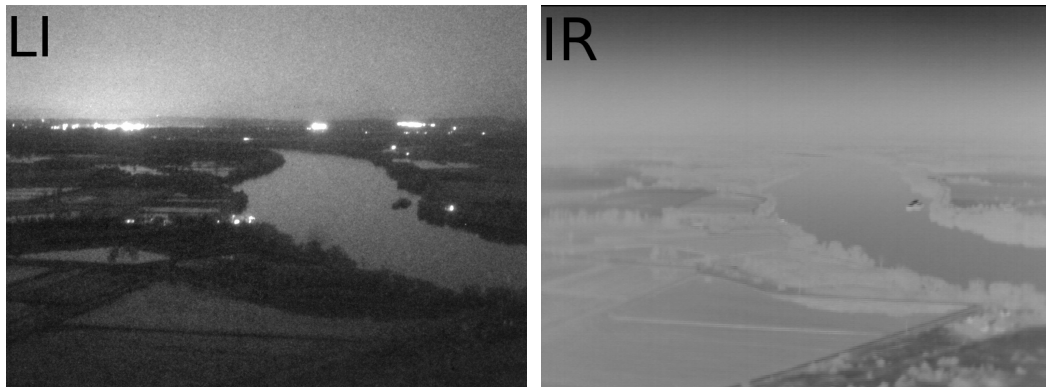


FIGURE 7.1 – Exemple d’une image optique (IL, à gauche) et d’une image infrarouge (IR, à droite) acquises simultanément durant le vol d’un hélicoptère. La résolution et le type d’information révélé, ainsi que les distributions des intensités, sont très différentes d’une modalité à l’autre.

faits tels que la présence de maillages, la saturation autour des sources lumineuses ou des différences d’illumination. Un exemple est représenté sur la figure 7.1. D’autre part, les images infrarouge (IR) reflètent la température de la scène. Elles sont plus difficiles à interpréter car elles ne présentent pas des caractéristiques du domaine visible, comme on peut l’observer sur la figure 7.1. Cependant, elles peuvent fournir des informations utiles concernant la présence de véhicules, de routes ou de bâtiments, qui sont des sources chaudes par rapport au sol et au reste de la scène. La caméra infrarouge est localisée sous l’hélicoptère, et peut être dirigée par le pilote. Les deux caméras impliquées observent donc la scène sous un angle différent, et se déplacent indépendamment l’une de l’autre, d’où la nécessité d’effectuer un recalage soigneux qui tient compte des différences de perspective et du mouvement relatif des deux dispositifs avant de pouvoir combiner les deux sources d’informations.

Dans le cadre de cette étude, les deux modalités optiques (IL) et infrarouge (IR) sont obtenues avec la même fréquence d’acquisition. Pour chaque couple d’images (IL, IR) générées, le but est d’aligner l’image infrarouge IR dans le plan de l’image optique IL. Comme l’image infrarouge est faiblement résolue, il est plus judicieux en pratique d’appliquer une transformation à l’image IL. On cherche donc la transformation spatiale globale T qui associe à chaque pixel de l’image IL (appelée alors image courante, et désignée par u) son emplacement correspondant dans le plan de l’image IR (appelée alors image de référence, et désignée par v).

7.1.2 État de l’art

Les techniques de recalage ont beaucoup été inspirées par des applications en imagerie médicale, en télédétection, ou encore en vision par ordinateur. En vision par ordinateur, le recalage est souvent unimodal. Il peut s’effectuer par minimisation du flot optique [Horn 1981, Lefébure 2001]. L’alignement peut égale-

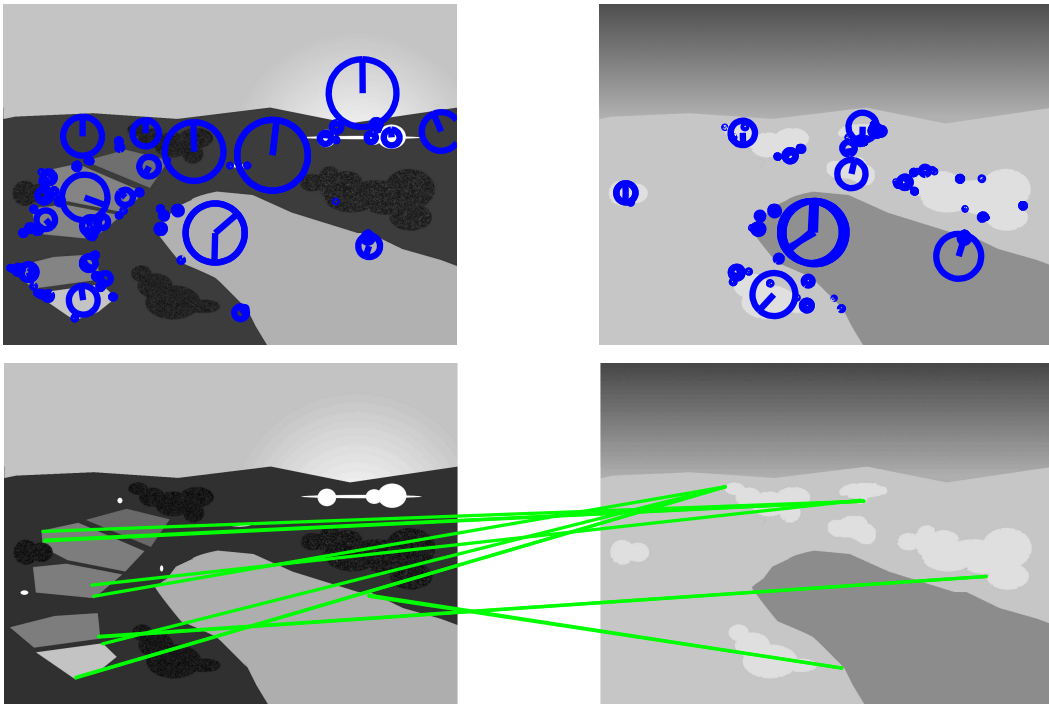


FIGURE 7.2 – Détection des points d'intérêt SIFT et des descripteurs associés pour chacune des modalités synthétiques, et échec de l'appariement des points.

ment s'effectuer via la mise en correspondance de points d'intérêt. En particulier, la méthode des descripteurs SIFT (pour *Scale Invariant Feature Transform*) définis dans [Lowe 2004] utilise un détecteur de régions invariant aux changements d'échelle et fournit des descripteurs basés sur la distribution du gradient dans les régions détectées. Associés à des algorithmes d'appariement comme l'algorithme RANSAC [Fischler 1981], ils permettent alors d'estimer la transformation optimale entre deux images. De tels descripteurs ont conduit à de nombreuses études et adaptations [Bay 2008, Mikolajczyk 2005], et ils renforcent l'estimation du flot optique pour de grands déplacements dans [Brox 2011]. Les descripteurs SIFT ont ensuite été étendus pour le recalage de données de type CT et IRM dans [Allaire 2008, Toews 2013, Paganelli 2012], ou pour des images SAR dans [Dellinger 2012], mais la méthode des points d'intérêt de type SIFT ne permet pas intrinsèquement de traiter le cadre multimodal. En effet, même si chaque modalité bénéficie de ses propres descripteurs, l'appariement de ces derniers peut se révéler difficile en raison des informations différentes portées par chacune des modalités [Yu 2008]. Un tel problème est donc notamment très dépendant des modalités impliquées et est peu généralisable. La figure 7.2 illustre sur des images synthétiques la difficulté que peut présenter l'appariement des descripteurs associés aux différents capteurs : bien que chacune des images bénéficie de points d'intérêts, les descripteurs associés ne permettent pas de réaliser l'appariement nécessaire à l'estimation d'une

transformation.

Une alternative pour gérer le cas de plusieurs modalités repose sur la maximisation d'une énergie, dont la solution fournit la transformation de domaine optimale T . La mesure de corrélation croisée, ou *cross-correlation* [Pratt 1978, Roche 1998, Roshni 2008], évalue la corrélation entre les valeurs de l'image courante transformée, $u(T)$, et celles de l'image de référence v à l'aide de la formule suivante :

$$CC(T) = \frac{\int_{\Omega_c} u_0(T(\mathbf{x})) \cdot v_0(\mathbf{x}) d\mathbf{x}}{\sqrt{\int_{\Omega_c} u_0(T(\mathbf{x}))^2 d\mathbf{x}} \cdot \sqrt{\int_{\Omega_c} v_0(\mathbf{x})^2 d\mathbf{x}}}, \quad (7.1)$$

où Ω_c désigne le domaine (continu) de l'image, et $u_0(T(\mathbf{x}))$ et $v_0(\mathbf{x})$ sont les images centrées, c'est-à-dire la différence entre l'image et sa valeur moyenne $\bar{u}(T)$ ou \bar{v} : $u_0(T(\mathbf{x})) = u(T(\mathbf{x})) - \bar{u}(T)$ et $v_0(\mathbf{x}) = v(\mathbf{x}) - \bar{v}$. La maximisation de la corrélation croisée $CC(T)$ fournit la transformation T qui maximise la corrélation entre les deux images u et v , donc favorise leur alignement. La corrélation croisée quadratique [Hermosillo 2002] peut se révéler plus adaptée encore au recalage multimodal, cependant les deux métriques sont basées sur l'hypothèse que les intensités des deux images à recaler sont liées par une relation affine, ce qui n'est pas le cas pour les images IL et IR à traiter.

L'information mutuelle [Viola 1997, Maes 1997] reflète elle la relation entre les distributions des intensités des deux images, mais sans hypothèse concernant la nature de cette relation. L'information mutuelle est issue de la théorie de l'information et de la notion d'entropie. Elle mesure la dépendance statistique entre les deux variables $u(T)$ et v [Pluim 2003]:

$$I(u, v) = H(u(T)) - H(u(T)|v) = H(v) - H(v|u(T)) = H(u(T)) + H(v) - H(u(T), v), \quad (7.2)$$

où $H(u(T))$ et $H(v)$ désignent respectivement l'entropie de $u(T)$ et v , et $H(u(T), v)$ l'entropie jointe. L'entropie est donnée par la formule suivante :

$$H(v) = - \sum_v p(v) \log p(v), \quad (7.3)$$

où $p(v)$ désigne la distribution des intensités de l'image v , estimée empiriquement à partir de l'histogramme issu des valeurs des pixels de l'image v : $p(v = q) = \frac{1}{\#\Omega} \#\{i \in \Omega \mid v_i = q\}$. Cette mesure peut également s'interpréter via la divergence de Kullback-Leibler :

$$I(T) = \int_{\Omega_c^2} p(u(T(\mathbf{x})), v(\mathbf{y})) \log \frac{p(u(T(\mathbf{x})), v(\mathbf{y}))}{p(u(T(\mathbf{x}))p(v(\mathbf{y}))} d\mathbf{x}d\mathbf{y}, \quad (7.4)$$

où $p(u, v)$ est la distribution de probabilité jointe de u et v et $p(u)$ et $p(v)$ sont les distributions marginales. Elle mesure la quantité d'information que l'une des variables contient sur l'autre : plus u et v sont indépendantes, plus la distribution jointe $p(u, v)$ est proche de la distribution $p(u) \times p(v)$. Le recalage est alors effectué en cherchant la transformation T qui maximise l'information mutuelle entre les images,

c'est-à-dire celle qui maximise leur degré de dépendance. Cependant, un tel modèle nécessite que les distributions des intensités des deux images soient proches d'être en bijection, ce qui n'est pas vérifié pour les modalités nocturnes impliquées : en effet, certaines zones hautement texturées sur l'image optique sont lisses sur l'image IR, tandis que des zones constantes sur l'image IL (comme le ciel ou la rivière, voir figure 7.1) présentent un dégradé sur l'image IR.

L'utilisation de modèles de représentation des images peut permettre également le recalage d'images. Dans [Irani 1998], les auteurs définissent un modèle de représentation des images basé sur quatre filtres dérivatifs directionnels qui extraient les contours selon quatre directions. Le recalage est alors effectué simultanément selon les quatre directions à l'aide d'une mesure de corrélation normalisée locale. De façon générale, l'information des contours est un outil répandu pour l'alignement d'images. Dans [Sun 2004], une métrique basée sur les contours des images est définie afin de mesurer l'alignement entre le gradient ∇v de l'image de référence et le gradient $\nabla u(T)$ de l'image courante transformée, où T est la transformation testée. L'alignement des contours est mesuré à l'aide du critère suivant :

$$C_S(T) = \frac{\int_{\Omega_c} w_T(\mathbf{x}) \cos(2\Delta\theta_T(\mathbf{x})) \, d\mathbf{x}}{\int_{\Omega_c} w_T(\mathbf{x}) \, d\mathbf{x}}, \quad (7.5)$$

où $w_T(\mathbf{x})$, $\Delta\theta_T(\mathbf{x})$ sont définis à partir de l'amplitude M et de l'orientation θ des gradients des images localisés en \mathbf{x} :

$$\begin{aligned} w_T(\mathbf{x}) &= M_u(T(\mathbf{x}))M_v(\mathbf{x}), \\ \Delta\theta_T(\mathbf{x}) &= \theta_u(T(\mathbf{x})) - \theta_v(\mathbf{x}). \end{aligned} \quad (7.6)$$

La présence du $\cos(2\Delta\theta_T(\mathbf{x}))$ dans l'équation (7.5) favorise les transformations qui alignent les contours, indépendamment de la direction du gradient, c'est-à-dire des inversions de contraste. De plus, du fait de traiter des images issues de modalités différentes, certains contours peuvent n'apparaître que dans une des deux modalités. Pour traiter cela, le poids $w_T(\mathbf{x})$ permet de favoriser l'alignement des contours les plus forts, et présents dans les deux images, sans pénaliser le non-alignement des contours qui ne sont présents que dans une seule des deux modalités.

Une métrique similaire basée également sur les contours est définie dans [Haber 2006] :

$$\omega_T(\mathbf{x}) = \left\langle \frac{\nabla u(T(\mathbf{x}))}{\|\nabla u(T(\mathbf{x}))\|_\varepsilon}, \frac{\nabla v(\mathbf{x})}{\|\nabla v(\mathbf{x})\|_\varepsilon} \right\rangle^2, \quad (7.7)$$

où $\|\nabla v(\mathbf{x})\|_\varepsilon = \sqrt{\nabla v(\mathbf{x})^T \nabla v(\mathbf{x}) + \varepsilon^2}$, et ε est déterminé automatiquement à partir de l'estimation de la variance du bruit pour permettre de distinguer les fluctuations liées au bruit de celles liées aux contours. La première partie de cette thèse a néanmoins illustré la difficulté associée à la distinction du signal et du bruit. La métrique NGF (pour *Normalized Gradient Fields*) est alors définie ainsi :

$$C_H(T) = \int_{\Omega_c} \omega_T(\mathbf{x}) \, d\mathbf{x}$$

$$= \int_{\Omega_c} \cos^2(\Delta\theta_T(\mathbf{x})) \, d\mathbf{x} \quad (7.8)$$

Les gradients utilisés ci-dessus sont normalisés ; la métrique peut s'interpréter comme un produit scalaire ou comme le cosinus de l'angle formé par les contours, pourvu que l'amplitude de ces derniers soit suffisamment grande devant le paramètre de bruit ε .

Cette métrique basée sur les contours s'inspire de [Pluim 2000], où l'information géométrique des contours est combinée à l'information mutuelle. De plus, afin de ne considérer que les contours présents dans les deux modalités, le produit scalaire $\omega_T(\mathbf{x})$ est pondéré en chaque point par le minimum des amplitudes des gradients :

$$G(T) = \int_{\Omega_c} \omega_T(\mathbf{x}) \min(|\nabla u(T(\mathbf{x}))|, |\nabla v(\mathbf{x})|) \, d\mathbf{x}, \quad (7.9)$$

puis $G(T)$ est associé à l'information mutuelle pour combiner les informations géométriques à la distribution statistique des intensités :

$$\text{MI-G}(T) = I(T) \times G(T). \quad (7.10)$$

Enfin, dans [Heinrich 2012], des descripteurs locaux invariants au type de modalité basés sur la notion d'autosimilarité issue de l'algorithme des NL-means [Buades 2005b] sont proposés. À chaque point \mathbf{x} du domaine de l'image u est associé un descripteur basé sur les similarités locales :

$$\text{MIND}(u, \mathbf{x}, r) = \frac{1}{Z} \exp\left(-\frac{D_p(u, \mathbf{x}, \mathbf{x} + r)}{V(u, \mathbf{x})}\right), \quad (7.11)$$

où Z est une constante de normalisation, $r \in R$ définit la zone de recherche sur laquelle les patchs sont comparés, D_p est la mesure de distance entre les patchs de taille $|\rho|$, en générale la distance euclidienne point à point, et $V(u, \mathbf{x})$ est une estimation de la variance locale visant à tenir compte de la présence de bruit. Ce descripteur MIND associe ainsi à chaque pixel de l'image u un vecteur de similarités de taille $|R|$, puis les descripteurs issus de chaque modalité sont comparés via la métrique suivante :

$$S(\mathbf{x}, T) = \frac{1}{|R|} \sum_{r \in R} |\text{MIND}(u(T), \mathbf{x}, r) - \text{MIND}(v, \mathbf{x}, r)|. \quad (7.12)$$

Minimiser la distance globale $\int_{\Omega_c} S(\mathbf{x}, T) d\mathbf{x}$ renvoie la transformation optimale T qui met en correspondance les cartes de similarité entre les patchs issus des deux modalités, autrement dit les cartes de poids non locaux (voir partie II). Le recalage se base ainsi sur des informations structurelles plutôt que colorimétriques. Cependant, cette méthode souffre des limites des méthodes non locales : elle est sensible à de nombreux paramètres comme l'estimation de la variance locale ou la taille des patchs et de la fenêtre de recherche, et est affectée par les faibles taux de contraste. De plus, si l'information structurelle change d'une modalité à l'autre, par exemple

si une surface lisse dans l'une est texturée dans l'autre, alors les cartes de similarités sont différentes et donc les descripteurs ne sont plus fiables.

Dans le contexte de la vision nocturne, le recalage multimodal doit se baser sur une information invariante d'une modalité à l'autre. D'après les caractéristiques des images IL et IR, une telle invariance est observée dans une certaine mesure au sein des contours. Le recalage de ces deux modalités peut donc s'effectuer en adaptant les métriques définies dans les équations (7.5) et (7.8). Il s'agit alors de développer un critère qui prenne en compte à la fois l'amplitude et l'orientation des contours, mais qui soit robuste aux différences de contours d'une modalité à l'autre ainsi qu'au inversions d'intensité.

7.2 Définition du critère

L'image courante u est recalée sur la position de référence de l'image v comme suit. Soient M_v et θ_v l'amplitude et l'orientation du gradient de l'image de référence v , obtenus à l'aide d'un détecteur de Sobel [Spontón 2012]. M_u et θ_u sont définis de façon analogue pour l'image courante u .

Le but est de définir une métrique basée sur les contours adaptée au cadre de la vision nocturne et aux caractéristiques de chacune des modalités, et qui peut s'utiliser dans un contexte opérationnel embarqué. Nous définissons le critère suivant :

$$C(T) = \int_{\Omega_c} |\nabla u(T(\mathbf{x})) \cdot \nabla v(\mathbf{x})| \, d\mathbf{x}, \quad (7.13)$$

qui peut également s'écrire sous la forme suivante :

$$C(T) = \int_{\Omega_c} w_T(\mathbf{x}) |\cos(\Delta\theta_T(\mathbf{x}))| \, d\mathbf{x}, \quad (7.14)$$

où $w_T(\mathbf{x})$ et $\Delta\theta_T(\mathbf{x})$ désignent l'amplitude et l'orientation définies dans l'équation (7.6). Ce critère favorise l'alignement des contours forts, grâce à la pondération par les amplitudes, et il est insensible aux inversions de contraste, c'est-à-dire à la direction du gradient : la valeur absolue du cosinus permet de considérer que des contours parallèles ou antiparallèles sont alignés. Cette distinction est tout particulièrement adaptée à la multimodalité du problème, dans lequel les changements d'intensité peuvent être inversés, mais de façon non systématique. De plus, contrairement à la métrique initialement proposée dans [Sun 2004], ce critère est non normalisé. Comme nous allons le voir par la suite, cela le rend plus stable donc plus facile à manipuler, et plus sensible au nombre de contours qui ont pu être mis en correspondance. En effet, le critère normalisé réalise une moyenne pondérée du score réalisé en chaque contour, donc il mesure l'alignement moyen réalisé à partir de l'ensemble des contours présents sur les deux modalités. Inversement, le critère non normalisé ajoute les scores de chaque paire de contours alignés, ce qui permet d'augmenter le score en fonction du nombre de contours mis en correspondance. Le critère normalisé peut favoriser

des alignements plus précis, indépendamment du nombre d'occurrences, et pénaliser les non-correspondances, tandis que le critère non normalisé aurait tendance à préférer des associations moins exactes, pourvu que le nombre d'associations soit satisfaisant.

7.2.1 Le modèle de transformation

Dans le modèle d'origine de Sun *et al.* [Sun 2004], qui effectue le recalage d'images de reins, les transformations de recalage envisagées consistent en une translation horizontale et/ou verticale. Le critère (7.5) peut donc être optimisé à l'aide d'une recherche exhaustive sur l'ensemble de l'espace des paramètres de transformation, à deux dimensions. Dans le cadre de l'application à la vision nocturne, les transformations de recalage envisagées dans un premier temps consistent en une translation dans chaque direction (horizontale et verticale), ainsi qu'un zoom uniforme sur toute l'image. Le paramètre de zoom est noté z , et les translations selon les directions horizontales et verticales sont respectivement notées t_1 et t_2 . Si $\mathbf{x} = (x_1 \ x_2 \ 1)^t$ désigne les coordonnées de l'image à recaler (notées également sous la forme concise $\mathbf{x} = (x_1 \ x_2)^t$), la matrice de transformation $T = T_{t_1, t_2, z}$ peut s'écrire :

$$\begin{aligned} T_{t_1, t_2, z}(\mathbf{x}) &= \begin{pmatrix} 1 & 0 & t_1 \\ 0 & 1 & t_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1+z & 0 & 0 \\ 0 & 1+z & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x} \\ &= \begin{pmatrix} 1+z & 0 & t_1 \\ 0 & 1+z & t_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}. \end{aligned} \quad (7.15)$$

7.3 Analyse du modèle proposé

7.3.1 Analyse théorique

Le but de cette section est d'étudier la méthode de recalage proposée d'un point de vue mathématique afin de montrer que la maximisation du critère $T \mapsto C(T)$ permet effectivement de recouvrer les paramètres optimaux de recalage.

On se limite ici au cas unidimensionnel et on cherche à aligner deux contours lorsque le signal est soumis à une translation et un changement d'échelle. On remarque que le recalage d'un seul contour est soumis à un problème d'ouverture, dans le sens où un unique contour, quel que soit le niveau d'échelle auquel il est observé, reste inchangé, ce qui rend impossible l'estimation du paramètre de zoom. D'où la nécessité de considérer deux contours, pour appréhender la distance entre les discontinuités. Le signal de référence est donc défini par une fonction boîte :

$$v_0(x) = 1 \text{ si } x \in [-1, 1], 0 \text{ sinon.} \quad (7.16)$$

Ce dernier est représenté sur la figure 7.4. Le signal à recaler, soumis à une translation et à un changement d'échelle, est défini par

$$u_0(x) = v_0(ax - b), \quad (7.17)$$

où $a > 0$ désigne le paramètre de zoom (et correspond au facteur $1 + z$ défini dans la section 7, équation (7.15)) et b désigne le vecteur de translation.

Calcul heuristique

Le signal de référence v_0 défini ci-dessus n'est pas différentiable en ± 1 , ce qui ne permet pas de décrire ses contours. Cependant, sa dérivée ∇v_0 peut être représentée sous la forme de la somme de deux Diracs situés en ± 1 :

$$\nabla v_0(x) = \delta_{-1} - \delta_1. \quad (7.18)$$

La dérivée du signal à recaler peut alors s'écrire

$$\nabla u_0(x) = a \cdot \nabla v(ax - b) = a \left(\delta_{\frac{-1+b}{a}} - \delta_{\frac{1+b}{a}} \right), \quad (7.19)$$

et donc la fonctionnelle à maximiser peut s'exprimer de façon heuristique sous la forme :

$$\begin{aligned} C(a, b) &= \int_{\mathbb{R}} |\nabla u_0(x) \cdot \nabla v_0(x)| dx \\ &= \int_{\mathbb{R}} a |\nabla v_0(ax - b) \cdot \nabla v_0(x)| dx \\ &= \int_{\mathbb{R}} a \left| \left(\delta_{\frac{-1+b}{a}} - \delta_{\frac{1+b}{a}} \right) \cdot (\delta_{-1} - \delta_1) \right| dx \\ &= \int_{\mathbb{R}} a \left(\delta_{\frac{-1+b}{a}} + \delta_{\frac{1+b}{a}} \right) \cdot (\delta_{-1} + \delta_1) dx. \end{aligned} \quad (7.20)$$

Bien qu'il ne soit pas formellement correct de manipuler des produits de fonction de Dirac, il est possible de prédire intuitivement le comportement de la fonctionnelle grâce à cette formulation :

— Alignement parfait des deux paires :

Afin que les deux paires de Diracs coïncident ensemble, a et b doivent satisfaire les conditions suivantes :

$$\begin{cases} \frac{-1+b}{a} = -1 \\ \frac{1+b}{a} = 1 \end{cases} \Leftrightarrow \begin{cases} b = 0 \\ a = 1 \end{cases} \quad (7.21)$$

— Alignement d'une seule paire :

Pour que seul l'un des deux pics de ∇u_0 ne coïncide avec un pic de ∇v_0 , a et b doivent satisfaire l'une des conditions suivantes :

$$\begin{aligned} \frac{1+b}{a} &= 1 \Leftrightarrow a - b = 1 \\ \frac{-1+b}{a} &= -1 \Leftrightarrow a + b = 1 \\ \frac{1+b}{a} &= -1 \Leftrightarrow a + b = -1 \\ \frac{-1+b}{a} &= 1 \Leftrightarrow b - a = 1 \end{aligned} \quad (7.22)$$

— Dans tous les autres cas, les deux paires sont disjointes, ce qui conduit à une fonctionnelle nulle.

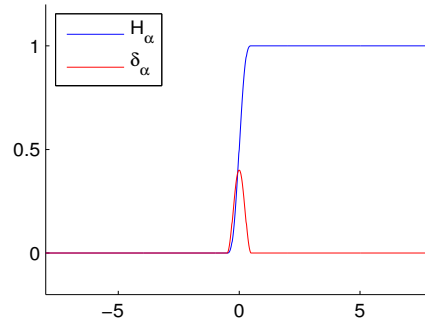


FIGURE 7.3 – Approximations δ_α de la fonction de Dirac et H_α de la fonction de Heaviside.

Cette étude heuristique permet de mettre en avant trois configurations :

- Un unique cas ($a = 1$ et $b = 0$) pour lequel les deux paires de pics de Dirac sont parfaitement alignés, ce qui signifie que les deux contours sont correctement recalés. Intuitivement, cela correspond au maximum de la fonctionnelle, bien qu'on ne puisse pas l'évaluer formellement en raison du produit de fonctions de Dirac.
- Quatre relations linéaires entre a et b pour lesquelles les deux signaux sont alignés selon un seul des deux contours. Ces droites reflètent la présence d'une infinité de maxima locaux, pour lesquelles la valeur de la fonctionnelle est cependant supposée plus faible que dans le cas d'un alignement parfait.
- Aucun alignement, auquel cas la fonctionnelle est nulle.

Une approximation du problème par des signaux différentiables qui approchent de façon lisse la fonction boîte permet de décrire plus formellement le comportement de la fonctionnelle, afin de distinguer le maximum global des maxima locaux,

Recalage théorique

Afin de manipuler des signaux différentiables, nous utilisons l'approximation de la fonction de Heaviside suivante [Aujol 2003]:

$$H_\alpha(x) = \begin{cases} \frac{1}{2}(1 + \frac{x}{\alpha} + \frac{1}{\pi} \sin \frac{\pi x}{\alpha}) & \text{si } |x| \leq \alpha \\ 1 & \text{si } x > \alpha \\ 0 & \text{si } x < -\alpha \end{cases} \quad (7.23)$$

Cette fonction est différentiable, et sa dérivée est donnée par :

$$\delta_\alpha(x) = \begin{cases} \frac{1}{2\alpha}(1 + \cos \frac{\pi x}{\alpha}) & \text{si } |x| \leq \alpha \\ 0 & \text{si } |x| > \alpha \end{cases} \quad (7.24)$$

Sur la figure 7.3 sont représentés les graphes de fonction H_α ainsi que de sa dérivée δ_α . Quand $\alpha \rightarrow 0$, $\delta_\alpha \rightarrow \delta$ et $H_\alpha \rightarrow H$, où δ désigne une distribution de

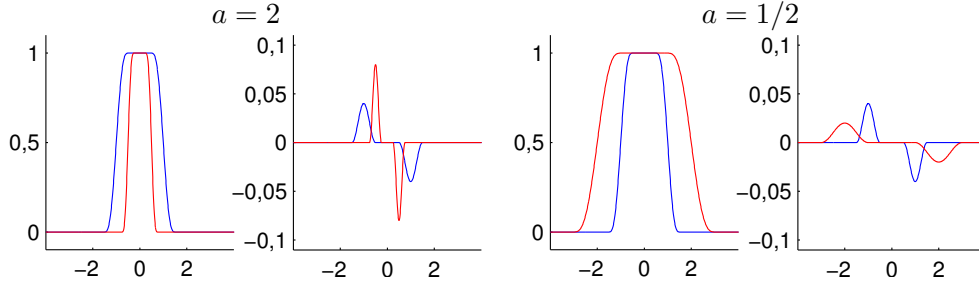


FIGURE 7.4 – Approximation avec $\alpha = 0.5$ de la fonction boîte (à gauche, en bleu) et de sa dérivée (à droite, en bleu) et zoom pour différentes valeurs du facteur a (en rouge).

Dirac et H la fonction de Heaviside. En pratique, on considère que α tend vers 0 pour simuler un contour discret.

À partir de la définition du signal de référence v_0 , son approximation continue est donnée par :

$$v(x) = H_\alpha(x + 1) - H_\alpha(x - 1), \quad (7.25)$$

qui correspond donc à une fonction boîte sur l'intervalle $[-1 - \alpha; 1 + \alpha]$. Sa dérivée ∇v est donnée par :

$$\nabla v(x) = \delta_\alpha(x + 1) - \delta_\alpha(x - 1) \quad (7.26)$$

De même, le signal à recaler ainsi que sa dérivée sont :

$$\begin{aligned} u_{a,b}(x) &= v(ax - b) \\ &= H_\alpha(ax - b + 1) - H_\alpha(ax - b - 1), \\ \nabla u_{a,b}(x) &= a(\delta_\alpha(ax - b + 1) - \delta_\alpha(ax - b - 1)). \end{aligned}$$

On remarque que lorsque $a > 1$, il s'agit d'un zoom négatif, dans le sens où la taille du support est réduite par un facteur a , tandis que la hauteur des pics est multipliée par le facteur a , donc augmentée. Au contraire lorsque $a < 1$ le zoom est positif : le support est étiré d'un facteur $1/a$ et la hauteur des pics réduite. La figure 7.4 illustre l'effet du zoom ($a > 1$ et $a < 1$) sur la fonction u et sa dérivée.

La fonctionnelle à maximiser devient :

$$\begin{aligned} C(a, b) &= \int_{\mathbb{R}} |\nabla u_{a,b}(x) \cdot \nabla v(x)| \, dx \\ &= a \int_{\mathbb{R}} (\delta_\alpha(x + 1) + \delta_\alpha(x - 1)) \cdot (\delta_\alpha(ax - b + 1) + \delta_\alpha(ax - b - 1)) \, dx \\ &= a \int_{\mathbb{R}} \delta_\alpha(x + 1) \delta_\alpha(ax - b + 1) + \delta_\alpha(x + 1) \delta_\alpha(ax - b - 1) \\ &\quad + \delta_\alpha(x - 1) \delta_\alpha(ax - b + 1) + \delta_\alpha(x - 1) \delta_\alpha(ax - b - 1) \, dx \\ &= C_1(a, b) + C_2(a, b) + C_3(a, b) + C_4(a, b). \end{aligned} \quad (7.27)$$

avec :

$$\begin{aligned}
 C_1(a, b) &= a \int_{\mathbb{R}} \delta_{\alpha}(x + 1) \delta_{\alpha}(ax - b + 1) \, dx, \\
 C_2(a, b) &= a \int_{\mathbb{R}} \delta_{\alpha}(x + 1) \delta_{\alpha}(ax - b - 1) \, dx, \\
 C_3(a, b) &= a \int_{\mathbb{R}} \delta_{\alpha}(x - 1) \delta_{\alpha}(ax - b + 1) \, dx, \\
 C_4(a, b) &= a \int_{\mathbb{R}} \delta_{\alpha}(x - 1) \delta_{\alpha}(ax - b - 1) \, dx.
 \end{aligned} \tag{7.28}$$

Chacune des sous-fonctionnelles C_1, \dots, C_4 peut être étudiée indépendamment afin de déterminer les conditions sur a et b pour que les intégrales soient maximales, ainsi qu'une forme explicite pour $C(a, b)$.

Proposition 7.3.1 *La fonctionnelle C peut s'écrire de la façon suivante :*

$$C(a, b) = \begin{cases} \frac{3}{2\alpha} \text{ si } a = 1 \text{ et } b = 0, \\ \frac{1}{2\alpha} \left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \right) \\ \text{si } a > 1 \text{ et } \begin{cases} a + b = 1 \\ \text{ou } a + b = -1 \\ \text{ou } a - b = 1 \\ \text{ou } a - b = -1 \end{cases} \\ \frac{1}{2\alpha} \left(a + \frac{1}{\pi(1+a)(1-a)} \sin \pi a \right) \\ \text{si } a < 1 \text{ et } \begin{cases} a + b = 1 \\ \text{ou } a + b = -1 \\ \text{ou } a - b = 1 \\ \text{ou } a - b = -1 \end{cases} \\ 0 \text{ sinon.} \end{cases} \tag{7.29}$$

De plus, elle atteint son maximum global égal à $\frac{3}{2\alpha}$ pour $a = 1$ et $b = 0$.

Preuve. La fonctionnelle $C(a, b)$ s'exprime comme la somme des quatre sous-fonctionnelles C_1, C_2, C_3 et C_4 définies dans (7.28). Chaque sous-fonctionnelle peut être évaluée séparément afin de déterminer les conditions sur a et b pour que chacune d'elles soit maximale.

Rappelons que le paramètre α qui détermine la largeur des pics est destiné à tendre vers 0. Les deux pics issus de la dérivée du signal de référence v sont situés en $+1$ et -1 , et leur support est donné par $[\pm 1 - \alpha, \pm 1 + \alpha]$. En ce qui concerne le signal transformé $u_{a,b}$, les pics sont localisés en $\frac{\pm 1 + b}{a}$ et leur support est $[\frac{\pm 1 + b - \alpha}{a}, \frac{\pm 1 + b + \alpha}{a}]$, de demie largeur α/a . Ainsi, lorsque α tend vers 0, la largeur de chaque pic tend vers 0. Cela implique pour le signal transformé que $a > 0$, ce qui est pertinent en pratique.

Cette remarque permet de simplifier le problème : il est possible de considérer que si les centres des pics ne sont pas parfaitement alignés, alors pour un α suffisamment petit les supports finissent par être disjoints.

Ainsi, la preuve est divisée en cinq étapes et l'étude des quatre sous-fonctionnelles est limitée aux conditions sur a et b pour que les centres des pics soient alignés.

Étape 1 : $C_1(a, b) = a \int_{\mathbb{R}} \delta_{\alpha}(x+1)\delta_{\alpha}(ax-b+1) dx$

$\delta_{\alpha}(x+1)$ a pour support $[-1-\alpha; -1+\alpha]$, centré en -1 , et $\delta_{\alpha}(ax-b+1)$ a pour support $[\frac{-1+b-\alpha}{a}; \frac{-1+b+\alpha}{a}]$, centré en $-1+b/a$.

Pour que la fonction C_1 ne soit pas nulle, il faut résoudre :

$$\frac{-1+b}{a} = -1 \Leftrightarrow a+b=1 \quad (7.30)$$

De plus, lorsque la condition (7.30): $a+b=1$ est vérifiée, C_1 s'exprime sous forme explicite :

- $a > 1$:

Si $a > 1$, la demie largeur du support de $u_{a,b}$ est $\alpha/a < \alpha$, donc le calcul de C_1 est restreinte à l'intervalle $[-\alpha/a, \alpha/a]$, et on a :

$$\begin{aligned} C_1(a, b) &= a \int_{\mathbb{R}} \delta_{\alpha}(x+1)\delta_{\alpha}(ax-b+1) dx, \\ &= \frac{a}{4\alpha^2} \int_{-\frac{\alpha}{a}}^{\frac{\alpha}{a}} \left(1 + \cos \frac{\pi x}{\alpha}\right) \cdot \left(1 + \cos \frac{\pi ax}{\alpha}\right) dx, \\ &= \frac{1}{2\alpha} \left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a}\right). \end{aligned} \quad (7.31)$$

- $a < 1$:

Si $a < 1$, la demie largeur du support de $u_{a,b}$ est $\alpha/a > \alpha$, donc le calcul de C_1 est restreinte à l'intervalle $[-\alpha, \alpha]$, et on a :

$$\begin{aligned} C_1(a, b) &= a \int_{\mathbb{R}} \delta_{\alpha}(x+1)\delta_{\alpha}(ax-b+1) dx, \\ &= \frac{a}{4\alpha^2} \int_{-\alpha}^{\alpha} \left(1 + \cos \frac{\pi x}{\alpha}\right) \cdot \left(1 + \cos \frac{\pi ax}{\alpha}\right) dx, \\ &= \frac{1}{2\alpha} \left(a + \frac{1}{\pi(1+a)(1-a)} \sin \pi a\right). \end{aligned} \quad (7.32)$$

- $a = 1$:

Si $a = 1$, la condition (7.30): $a+b=1$ implique que $b=0$, donc $u=v$, et :

$$C_1(1, 0) = \int_{-\alpha}^{\alpha} \left[\frac{1}{2\alpha} \left(1 + \cos \frac{\pi x}{\alpha}\right)\right]^2 dx = \frac{3}{4\alpha}. \quad (7.33)$$

Conclusion :

$$C_1(a, b) = \begin{cases} \frac{3}{4\alpha} \\ \text{si } a = 1 \text{ et } b = 0 \\ \frac{1}{2\alpha} \left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \right) \\ \text{si } a + b = 1 \text{ et } a > 1 \\ \frac{1}{2\alpha} \left(a + \frac{1}{\pi(1+a)(1-a)} \sin \pi a \right) \\ \text{si } a + b = 1 \text{ et } a < 1 \\ 0 \text{ sinon.} \end{cases} \quad (7.34)$$

Étape 2 : $C_2(a, b) = a \int_{\mathbb{R}} \delta_{\alpha}(x+1) \delta_{\alpha}(ax-b-1) dx$

Une étude analogue du support de $\delta_{\alpha}(x+1)$ et $\delta_{\alpha}(ax-b-1)$ met en avant les conditions suivantes sur a et b pour que les supports s'intersectent :

$$\frac{1+b}{a} = -1 \Leftrightarrow a+b = -1 \quad (7.35)$$

Alors, quand la condition (7.35): $a+b = -1$ est satisfaite, C_2 s'exprime de façon analogue à C_1 :

$$C_2(a, b) = \begin{cases} \frac{3}{4\alpha} \\ \text{si } a = 1 \text{ et } b = -2 \\ \frac{1}{2\alpha} \left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \right) \\ \text{si } a + b = -1 \text{ et } a > 1 \\ \frac{1}{2\alpha} \left(a + \frac{1}{\pi(1+a)(1-a)} \sin \pi a \right) \\ \text{si } a + b = -1 \text{ et } a < 1 \\ 0 \text{ sinon.} \end{cases} \quad (7.36)$$

Étape 3 : $C_3(a, b) = a \int_{\mathbb{R}} \delta_{\alpha}(x-1) \delta_{\alpha}(ax-b+1) dx$

de façon similaire à nouveau, on résout :

$$\frac{-1+b}{a} = 1 \Leftrightarrow a-b = -1 \quad (7.37)$$

et on obtient l'expression suivante :

$$C_3(a, b) = \begin{cases} \frac{3}{4\alpha} \\ \text{si } a = 1 \text{ et } b = 2 \\ \frac{1}{2\alpha} \left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \right) \\ \text{si } a - b = -1 \text{ et } a > 1 \\ \frac{1}{2\alpha} \left(a + \frac{1}{\pi(1+a)(1-a)} \sin \pi a \right) \\ \text{si } a - b = -1 \text{ et } a < 1 \\ 0 \text{ sinon.} \end{cases} \quad (7.38)$$

Étape 4 : $F_4(a, b) = a \int_{\mathbb{R}} \delta_\alpha(x-1) \delta_\alpha(ax-b-1) dx$

On résout :

$$\frac{1+b}{a} = 1 \Leftrightarrow a-b=1, \quad (7.39)$$

et on obtient :

$$C_4(a, b) = \begin{cases} \frac{3}{4\alpha} & \text{si } a=1 \text{ et } b=0 \\ \frac{1}{2\alpha} \left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \right) & \text{si } a-b=1 \text{ et } a > 1 \\ \frac{1}{2\alpha} \left(a + \frac{1}{\pi(1+a)(1-a)} \sin \pi a \right) & \text{si } a-b=1 \text{ et } a < 1 \\ 0 & \text{sinon.} \end{cases} \quad (7.40)$$

Étape 5: retour à la fonctionnelle C

L'étude de chacune des sous-fonctionnelles a permis de mettre en avant les conditions pour lesquelles deux pics sont alignés. La mise en commun des résultats conduit à chercher les conditions pour que les deux paires de pics soient alignées en même temps. En comparant les conditions sur a et b pour chacune des sous-fonctionnelles, la seule association simultanée possible est entre C_1 et C_4 , pour $a=1$ et $b=0$. Dans ce cas, on a alors $C(1, 0) = C_1(1, 0) + C_4(1, 0) = \frac{3}{4\alpha} + \frac{3}{4\alpha} = \frac{3}{2\alpha}$.

Conclusion :

$$C(a, b) = \begin{cases} \frac{3}{2\alpha} & \text{si } a=1 \text{ et } b=0, \\ \frac{1}{2\alpha} \left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \right) & \text{si } a > 1 \text{ et } \begin{cases} a+b=1 \\ \text{ou } a+b=-1 \\ \text{ou } a-b=1 \\ \text{ou } a-b=-1 \end{cases} \\ \frac{1}{2\alpha} \left(a + \frac{1}{\pi(1+a)(1-a)} \sin \pi a \right) & \text{si } a < 1 \text{ et } \begin{cases} a+b=1 \\ \text{ou } a+b=-1 \\ \text{ou } a-b=1 \\ \text{ou } a-b=-1 \end{cases} \\ 0 & \text{sinon.} \end{cases} \quad (7.41)$$

Pour pouvoir conclure, il reste à montrer que $C(a, b) \leq \frac{3}{2\alpha}$ pour tout (a, b) , afin que le couple $(a, b) = (1, 0)$ soit bien l'optimum. Pour cela, on se focalise sur les fonctions $f_1(a) = \frac{1}{2\alpha} \left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \right)$ pour $a > 1$ et $f_2(a) = \frac{1}{2\alpha} \left(a + \frac{1}{\pi(1+a)(1-a)} \sin \pi a \right)$ pour $0 < a < 1$ et on divise la preuve en trois points :

- $0 < a < 1$

Pour pouvoir assurer que $f_2(a) \leq \frac{3}{2\alpha}$, on doit montrer que $\frac{1}{\pi(1+a)(1-a)} \sin \pi a \leq 2$. On a :

$$\frac{1}{\pi(1+a)(1-a)} \sin \pi a = \frac{1}{\pi(1+a)(1-a)} \sin \pi(1-a) \leq \frac{1}{(1+a)} \leq 1 \quad (7.42)$$

sur l'intervalle $[0; 1]$.

- $1 < a < 2$

Pour assurer que $f_1(a) \leq \frac{3}{2\alpha}$ sur l'intervalle $[1; 2]$, on vérifie que $\frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \leq 2$. On a :

$$\begin{aligned} \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} &= \frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi(a-1)}{a} & (7.43) \\ &\leq \frac{a^3}{\pi(a+1)(a-1)} \times \frac{\pi(a-1)}{a} \leq \frac{a^2}{a+1} \leq \frac{4}{3} \end{aligned}$$

sur l'intervalle $[1; 2]$.

- $a > 2$

$$\frac{a^3}{\pi(a+1)(a-1)} \sin \frac{\pi}{a} \leq \frac{a^2}{a^2-1} \leq \frac{4}{3} \quad (7.44)$$

sur l'intervalle $[2; +\infty]$.

Cela conclut la preuve en assurant que la fonctionnelle admet un unique maximum pour les paramètres recherchés $(a, b) = (1, 0)$. ■

Cette proposition confirme l'étude intuitive effectuée à partir des distributions de Dirac, et révèle bien l'existence d'un maximum global pour les paramètres de transformation $(a, b) = (1, 0)$, ainsi que des sous-espaces linéaires de maxima locaux.

7.3.2 Étude de la performance du critère proposé

Après avoir montré que le critère est capable d'un point de vue théorique de retrouver les paramètres optimaux de transformation pour le problème de recalage, cette section permet de valider l'étude en pratique. Les évaluations numériques s'appuient sur des images synthétiques qui reflètent (en les simplifiant) les caractéristiques des deux modalités impliquées, représentées sur la figure 7.5.

7.3.2.1 Étude des performances

Dans un premier temps, c'est la capacité du critère à atteindre son maximum pour les paramètres de transformation optimaux qui est évaluée, sur des images

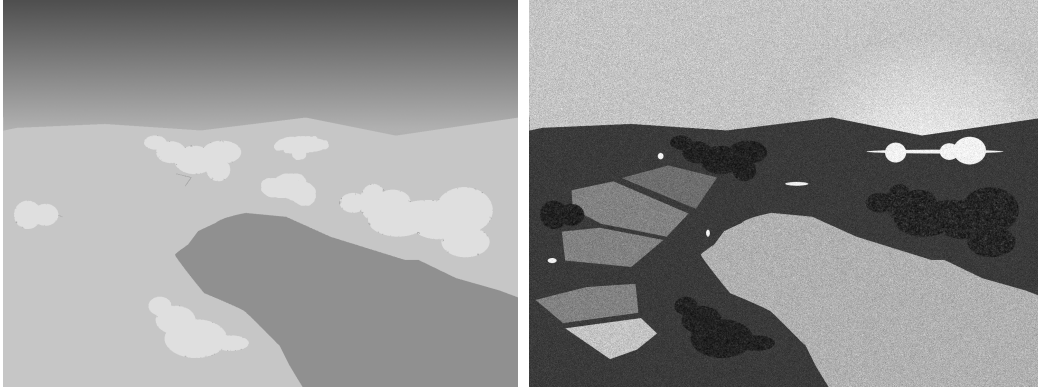


FIGURE 7.5 – Images synthétiques correspondant aux modalités IR (à gauche) et IL (à droite).

TABLE 7.1 – Estimation du pourcentage de transformations correctement estimées sur des images synthétiques avec la corrélation croisée (CC), l'information mutuelle (MI), l'information mutuelle combinée gradient (MI-G), le descripteur MIND, le cosinus pondéré, NGF et la métrique proposée. L'image IL est d'abord non bruitée puis corrompue par un bruit de Poisson, avec un PSNR initial autour de 18 puis 15dB.

Métrique	Images non bruitées	Bruit de Poisson, PSNR \approx 18dB	Bruit de Poisson, PSNR \approx 15dB
corrélation croisée	73%	77%	77%
Information mutuelle	77%	68%	61%
MI-G	91%	100%	100%
MIND	100%	52%	34%
Cos. pondéré	100%	100%	100%
NGF	100%	100%	95%
Proposée	100%	100%	100%

synthétiques. Une centaine de transformations, avec des paramètres de zoom et de translation aléatoires et distribués uniformément entre -40 et +40 pixels pour les translations et entre -0.4 et +0.4 pour le zoom, sont simulées. La maximisation de la métrique est ensuite effectuée à l'aide d'une recherche exhaustive sur l'espace tridimensionnel des paramètres de la transformation. Cela permet d'évaluer simplement que la maximisation du critère fournit le résultat attendu. Le nombre de transformations correctement estimées pour chacune des méthodes testées est comptabilisé, pour fournir un pourcentage de succès. On étudie également la robustesse au bruit, puisqu'en pratique les images optiques sont corrompues par un bruit non gaussien fort. Pour cela, du bruit de Poisson est ajouté aux images IL afin d'atteindre un PSNR autour de 18 puis 15dB, afin de refléter les dégradations naturelles observées sur ce type d'images en vision nocturne. Les performances des différentes méthodes présentées dans la section 7.1.2 sont évaluées : la corrélation croisée (CC) définie

dans l’équation (7.1), l’information mutuelle (MI) calculée à partir de (7.4), l’information mutuelle combinée gradient (MI-G) basée sur l’équation (7.10), le descripteur MIND (7.11), ainsi que les métriques basées contour : le cosinus pondéré présenté dans l’équation (7.5), NGF (7.8), et la méthode proposée.

Cela permet d’évaluer la capacité de chacune des métriques à atteindre son maximum pour les paramètres de transformation optimaux. Le tableau 7.1 révèle les pourcentages de transformations correctement estimées à partir de l’ensemble des transformations simulées, sur des images non bruitées puis corrompues par du bruit de Poisson. Les résultats montrent que d’une façon générale les méthodes basées contour offrent une estimation plus fiable, et plus robuste au bruit. Conformément à ce qui était annoncé dans la section 7.1.2, la corrélation croisée et l’information mutuelle ne sont pas adaptées aux caractéristiques multimodales des images IL/IR. Dans la version combinée de la métrique MI-G [Pluim 2000], l’information des contours permet un recalage plus exact, mais au point finalement où l’information mutuelle n’est plus vraiment nécessaire. En effet, lorsque la puissance du bruit augmente, la valeur minimale des gradients des images augmente également, ce qui d’après l’équation (7.10) accroît l’influence de l’information de contour par rapport à l’information mutuelle, expliquant ainsi les meilleurs résultats obtenus sur les images bruitées. Le descripteur MIND est lui bien adapté au caractère multimodal du problème, en revanche les descripteurs proposés dans [Heinrich 2012] ne sont pas adaptés pour mesurer des similarités sur des patches corrompus par un bruit dépendant du signal. Les trois méthodes basées contour sont les plus performantes ici, bien que NGF finisse pas décrocher à haut niveau de bruit. Cela peut s’expliquer par la normalisation des gradients qui rend le critère plus sensible au bruit, et la sélection du paramètre de bruit ε (issu de l’équation (7.7)) qui peut ne pas permettre de distinguer finement les contours d’un fort bruit.

7.3.2.2 Étude de la stabilité

Afin de pouvoir envisager d’utiliser une méthode d’optimisation pour la recherche de la transformation optimale, la stabilité de la métrique, c’est-à-dire sa sensibilité aux maxima locaux et son comportement autour du maximum, est un aspect crucial. En effet, le chapitre suivant sera consacré à la mise en place d’une montée de gradient fiable, pour laquelle il est nécessaire que l’énergie à maximiser soit aussi concave que possible, lisse, et non sujette à trop de maxima locaux qui masqueraient l’optimum.

L’expérience suivante consiste donc naturellement à évaluer le comportement des différentes métriques. À partir des images synthétiques sont générées des transformations connues, afin d’observer l’évolution du critère par rapport à la variation d’un seul des paramètres, tandis que les deux autres sont fixés aux valeurs optimales. L’évolution de la corrélation croisée, de l’information mutuelle, de l’information mutuelle combinée gradient [Pluim 2000], des descripteurs MIND [Heinrich 2012], de NGF [Haber 2006] et de la métrique proposée par rapport à la variation d’un paramètre sont représentées sur la figure 7.6. Sur la première ligne, l’image est corrompue par du bruit de Poisson (PSNR \approx 15dB), les paramètres optimaux de transforma-

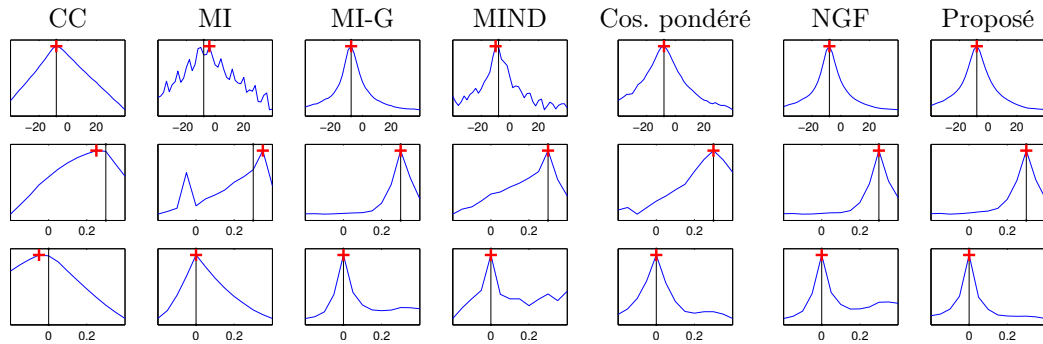


FIGURE 7.6 – Évolution de la métrique en fonction d’un seul des paramètres, les deux autres étant fixés à la valeur optimale. Sur la première ligne, les images sont corrompues par un bruit de Poisson (PSNR ≈ 15 dB), les paramètres optimaux sont $[t_1, t_2, z] = [8, -8, 0.35]$, et le paramètre observé est celui de translation horizontale. Sur la deuxième ligne, les images sont non bruitées, les paramètres optimaux sont $[t_1, t_2, z] = [24, -32, 0.3]$, et le paramètre observé est celui de zoom. Sur la troisième ligne, les images sont corrompues par un bruit de Poisson (PSNR ≈ 15 dB), les paramètres optimaux sont $[t_1, t_2, z] = [-4, 2, 0]$, et le paramètre observé est celui de zoom. Les métriques, de gauche à droite : corrélation croisée [Pratt 1978], Information mutuelle [Viola 1997], Information mutuelle combinée gradient [Pluim 2000], MIND [Heinrich 2012], Cosinus pondéré [Sun 2004], NGF [Haber 2006] et la métrique proposée (7.13).

tion sont $[t_1, t_2, z] = [8, -8, 0.35]$, et le paramètre qui varie est celui de translation horizontale t_2 . Sur la deuxième ligne, l’image est non bruitée, les paramètres optimaux de transformation sont $[t_1, t_2, z] = [24, -32, 0.3]$, et le paramètre qui varie est le coefficient de zoom z . Sur la dernière ligne, l’image est corrompue par du bruit de Poisson (PSNR ≈ 15 dB), les paramètres optimaux de transformation sont $[t_1, t_2, z] = [-4, 2, 0]$, et le paramètre qui varie est le coefficient de zoom z . Ces figures illustrent d’une part la capacité de chacune des métriques à atteindre son maximum à la valeur du paramètre attendue, et d’autre part le comportement de celle-ci autour du maximum. La corrélation croisée offre bien des courbes lisses, mais le maximum a tendance à être mal placé, sur les deux lignes inférieures par exemple. L’information mutuelle est instable et souffre de multiples maxima locaux, ce qui introduit des erreurs d’estimation (sur les deux première lignes). L’apport de l’information de contour dans MI-G stabilise le critère, cependant les métriques basées uniquement sur les contours sont aussi performantes, et plus efficaces en terme d’optimisation car il s’agit de métriques calculées sur les pixels directement plutôt qu’à partir d’histogrammes. Enfin, les trois métriques basées contour (cosinus pondéré, NGF, et la méthode proposée) offrent des courbes satisfaisantes, bien que le cosinus pondéré et NGF présentent moins de régularité en s’éloignant du maximum, comme on peut l’observer sur la première et la dernière ligne.

Les deux études sur la performance et la stabilité menées ici permettent ainsi de

confirmer l'étude théorique et de valider la métrique étudiée.

7.4 Extension aux transformations projectives

Lorsque l'hélicoptère vole à haute altitude, l'hypothèse qui stipule que la transformation entre les deux modalités peut être approchée par un zoom uniforme et une translation (ou de façon plus générale par une transformation affine) peut se vérifier. En revanche lorsque l'hélicoptère vole à plus faible altitude, la différence de perspective entre les caméras est plus grande, ce qui impose d'adopter un modèle projectif, non linéaire.

7.4.1 Un peu de géométrie projective

Une transformation projective [Hartley 2003] est caractérisée par la matrice d'homographie

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \quad (7.45)$$

et possède 8 degrés de liberté.

L'équation de la transformation associée est donnée par :

$$\begin{pmatrix} wx'_1 \\ wx'_2 \\ w \end{pmatrix} = H \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11}x + h_{12}y + h_{13} \\ h_{21}x + h_{22}y + h_{23} \\ h_{31}x + h_{32}y + 1 \end{pmatrix}, \quad (7.46)$$

et on retrouve les coordonnées x' et y' en normalisant par w :

$$\begin{cases} x'_1 &= \frac{h_{11}x_1 + h_{12}x_2 + h_{13}}{h_{31}x_1 + h_{32}x_2 + 1} \\ x'_2 &= \frac{h_{21}x_1 + h_{22}x_2 + h_{23}}{h_{31}x_1 + h_{32}x_2 + 1}. \end{cases} \quad (7.47)$$

On note pour simplifier :

$$\mathbf{x}' = H\mathbf{x} = \begin{pmatrix} \frac{h_{11}x_1 + h_{12}x_2 + h_{13}}{h_{31}x_1 + h_{32}x_2 + 1} \\ \frac{h_{21}x_1 + h_{22}x_2 + h_{23}}{h_{31}x_1 + h_{32}x_2 + 1} \end{pmatrix} \quad (7.48)$$

Ce type de transformation offre une généralisation du modèle affine, et inclut le modèle de transformation considéré jusque là, auquel s'ajoutent notamment les rotations et les changements de perspective.

7.4.2 Influence des paramètres

En étudiant l'influence de chacun des paramètres indépendamment, il est possible d'associer un comportement, ou un type de transformation (et une direction, verticale ou horizontale), à chacun des paramètres.

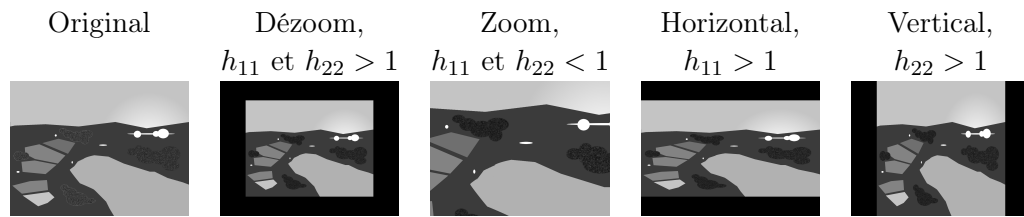


FIGURE 7.7 – Paramétrage des coefficients diagonaux de la matrice homographique, correspondant au zoom.

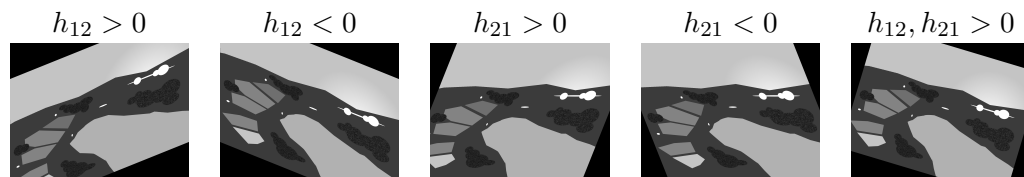


FIGURE 7.8 – Influence des coefficients anti-diagonaux de la matrice homographique, correspondant au *sheer*.

La façon dont la transformation est codée joue naturellement sur l'influence des paramètres et surtout le sens dans lequel ils varient. Dans les expériences qui sont présentées ici, la dimension x_1 fait référence aux lignes de l'image, et x_2 aux colonnes.

En ce qui concerne l'application aux images IL et IR, le modèle consiste à recaler l'image IL (optique) sur l'image IR. La transformation est donc appliquée à l'image IL afin de l'aligner avec l'image IR.

Paramètres diagonaux h_{11} , h_{22} : zoom

Les paramètres h_{11} et h_{22} sont responsables du zoom/dézoom de l'image. Avec des paramètres plus grands que 1, cela correspond à un dézoom de l'image, et avec des paramètres inférieurs à 1, il s'agit d'un zoom. En considérant chacun des paramètres indépendamment, le coefficient h_{11} est responsable du zoom horizontal (sur les lignes) et le coefficient h_{22} est lié au zoom vertical (sur les colonnes).

Paramètres anti-diagonaux h_{12} , h_{21} : rotation

Les paramètres h_{12} et h_{21} sont responsables de la rotation de l'image. Le paramètre h_{12} est responsable du déplacement sur les lignes, et h_{21} sur les colonnes. Le signe influe le sens de la rotation, horaire ou antihoraire.

Paramètres h_{13} , h_{23} : translation

Les paramètres h_{13} et h_{23} sont responsables de la translation de l'image. h_{13} correspond aux translations horizontales : $h_{13} > 0$ crée un déplacement de l'image

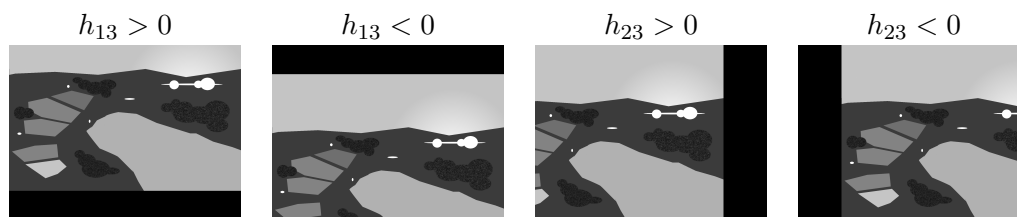


FIGURE 7.9 – Influence et paramétrage des coefficients de translation.

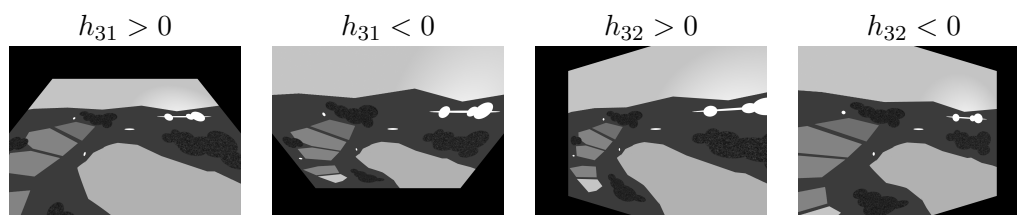


FIGURE 7.10 – Influence et paramétrage des coefficients de la matrice d'homographie responsables du caractère projectif de la transformation.

vers le haut, et $h_{13} < 0$ vers le bas. h_{23} correspond aux translations verticales : $h_{23} > 0$ crée un déplacement de l'image vers la gauche, et $h_{23} < 0$ vers la droite.

Paramètres h_{31} , h_{32} : projectif

Les deux paramètres h_{31} et h_{32} interviennent au dénominateur et ce sont ceux qui sont responsables de l'aspect projectif de la transformation. h_{31} est responsable de la projection horizontale (sur les lignes), et h_{32} de la projection verticale (sur les colonnes).

7.5 Conclusion

Dans ce chapitre a été proposée une méthode de recalage pour aligner les images optiques issues de l'intensificateur de lumière aux images infrarouge, en vue d'une ultérieure fusion. Pour cela, nous avons défini un critère qui mesure l'alignement des contours en se basant sur leur orientation et leur amplitude, afin de ne prendre en compte que les contours présents dans les deux modalités, indépendamment de la direction des gradients des intensités. La maximisation du critère par une recherche exhaustive sur l'espace des paramètres de la transformation permet d'estimer la transformation optimale qui aligne l'image courante sur l'image de référence. Cette transformation peut correspondre à l'association d'un changement d'échelle et de translations dans les deux directions, ou s'étendre au cas plus général des transformations projectives qui permettent entre autres de modéliser les changements de perspective. Une analyse théorique du critère ainsi que des résultats expérimentaux sur des images synthétiques ont démontré la capacité du critère à correctement es-

timer les paramètres de transformation optimaux, ainsi que sa robustesse au bruit et sa stabilité.

Cependant, en vue d'une implémentation embarquée et en particulier dans les cas des transformations projectives à huit paramètres, la recherche exhaustive peut se révéler trop lente en pratique. Elle est de plus limitée par l'espace de recherche prédéfini pour chacun des paramètres. Le chapitre suivant se consacre donc à la mise en place d'un schéma d'optimisation basé sur la montée de gradient du critère proposé.

Schéma d'optimisation du critère de recalage

8.1 Introduction

Le critère de recalage proposé dans le chapitre précédent offre de belles performances en terme de robustesse au bruit et de fiabilité, en revanche la résolution par recherche exhaustive sur l'espace des paramètres ne permet pas une implémentation embarquée et temps réel, en particulier dans le cas d'une transformation projective à huit degrés de liberté. De plus, effectuer une recherche exhaustive nécessite de définir au préalable un domaine et une grille de variation pour chacun des paramètres, aussi lorsque l'on sort de ce cadre le recalage n'est plus assuré. Il apparaît donc intéressant d'effectuer la maximisation du critère à l'aide d'une méthode d'optimisation. Cela repose dans un premier temps sur la mise en place d'une montée de gradient de la fonctionnelle à maximiser, et dans un second temps sur l'implémentation d'un schéma temporel qui assure que la méthode est fiable. En effet, comme le critère à maximiser n'est pas convexe, la montée de gradient n'est pas garantie de converger vers le maximum global de la fonctionnelle. Il faut donc se munir d'une bonne initialisation, raisonnablement proche du maximum global. Une telle initialisation est fournie par le flux vidéo. Il est ainsi possible d'initialiser le recalage pour un couple d'images en se basant sur les paramètres de recalage estimés à l'instant précédent. On peut même affiner cette prédiction en utilisant également les informations de déplacement de chacune des caméras indépendamment, grâce aux capteurs de position et aux systèmes inertiels embarqués. Ainsi, en propageant les paramètres de recalage obtenus à l'instant précédent à l'aide du déplacement estimé pour chacune des caméras, on obtient une prédiction fine qui fournit une bonne initialisation mais aussi un moyen de contrôler que la montée de gradient converge proprement.

Ce schéma temporel est validé sur des données synthétiques, par exemple dans des cas de mouvements soudains de la tête du pilote que la montée de gradient seule ne pourrait pas gérer, puis sur les données réelles.

8.2 Montée de gradient

8.2.1 Cas d'une transformation linéaire : zoom et translation

Grâce à la formulation proposée dans l'équation (7.13), un schéma d'optimisation explicite peut être développé pour maximiser la métrique de façon itérative, en

effectuant une montée de gradient sur les paramètres de la transformation $T_{t_1, t_2, z}$.

Proposition 8.2.1 *La maximisation du critère (7.13) peut s'effectuer à l'aide de la montée de gradient donnée pour chacun des paramètres (t_1, t_2, z) de la transformation $T_{t_1, t_2, z}$ par*

$$\begin{cases} t_1^{n+1} &= t_1^n + \lambda_1 \partial_{t_1} C(T_{t_1^n, t_2^n, z^n}) \\ t_2^{n+1} &= t_2^n + \lambda_2 \partial_{t_2} C(T_{t_1^n, t_2^n, z^n}) \\ z^{n+1} &= z^n + \lambda_3 \partial_z C(T_{t_1^n, t_2^n, z^n}) \end{cases}, \quad (8.1)$$

où les dérivées de la fonction $C(T_{t_1, t_2, z})$ sont données à chaque itération par :

$$\begin{aligned} \partial_{t_1} C(T_{t_1, t_2, z}) &= \int_{\Omega_c} \text{sign}(\nabla u(T_{t_1, t_2, z}(\mathbf{x})).\nabla v(\mathbf{x})) D^2 u(T_{t_1, t_2, z}(\mathbf{x})) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \nabla v(\mathbf{x}) d\mathbf{x}, \\ \partial_{t_2} C(T_{t_1, t_2, z}) &= \int_{\Omega_c} \text{sign}(\nabla u(T_{t_1, t_2, z}(\mathbf{x})).\nabla v(\mathbf{x})) D^2 u(T_{t_1, t_2, z}(\mathbf{x})) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \nabla v(\mathbf{x}) d\mathbf{x}, \\ \partial_z C(T_{t_1, t_2, z}) &= \int_{\Omega_c} \text{sign}(\nabla u(T_{t_1, t_2, z}(\mathbf{x})).\nabla v(\mathbf{x})) D^2 u(T_{t_1, t_2, z}(\mathbf{x}))(\mathbf{x}).\nabla v(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (8.2)$$

Preuve. À partir de la définition de la fonctionnelle, $C(T) = \int_{\Omega_c} |\nabla u(T(\mathbf{x})).\nabla v(\mathbf{x})| d\mathbf{x}$, où T est la transformation que l'on cherche à optimiser, en définissant un petit déplacement S alors :

$$C(T + S) = \int_{\Omega_c} |\nabla u(T(\mathbf{x}) + S(\mathbf{x})).\nabla v(\mathbf{x})| d\mathbf{x} \quad (8.3)$$

De plus, à partir de

$$\nabla u(T\mathbf{x} + S\mathbf{x}) = \nabla u(T\mathbf{x}) + D^2 u(T\mathbf{x})(S\mathbf{x}) + o(S). \quad (8.4)$$

et en utilisant l'approximation au premier ordre,

$$\begin{aligned} & \int_{\Omega_c} |\nabla u(T(\mathbf{x})).\nabla v(\mathbf{x}) + D^2 u(T\mathbf{x})(S\mathbf{x}).\nabla v(\mathbf{x})| d\mathbf{x} \\ &= \int_{\Omega_c} |\nabla u(T(\mathbf{x})).\nabla v(\mathbf{x})| \times \left| 1 + \frac{D^2 u(T\mathbf{x})(S\mathbf{x}).\nabla v(\mathbf{x})}{\nabla u(T(\mathbf{x})).\nabla v(\mathbf{x})} \right| d\mathbf{x} \\ &= \int_{\Omega_c} |\nabla u(T(\mathbf{x})).\nabla v(\mathbf{x})| \times \left(1 + \frac{D^2 u(T\mathbf{x})(S\mathbf{x}).\nabla v(\mathbf{x})}{\nabla u(T(\mathbf{x})).\nabla v(\mathbf{x})} \right) d\mathbf{x}, \end{aligned} \quad (8.5)$$

on a

$$C(T+S) = C(T) + \int_{\Omega_c} \text{sign}(\nabla u(T(\mathbf{x})).\nabla v(\mathbf{x})) D^2 u(T\mathbf{x})(S\mathbf{x}).\nabla v(\mathbf{x}) d\mathbf{x} + o(S). \quad (8.6)$$

On se focalise désormais sur la variation associée à chaque paramètre t_1, t_2, z en exprimant la fonctionnelle C en fonction de chacun des paramètres :

$$C(t_1 + \alpha, t_2, z) - C(t_1, t_2, z) = \alpha \partial_1 C(t_1, t_2, z) \quad (8.7)$$

et on désigne par T_α la perturbation α de T appliquée à la première coordonnée t_1 , soit :

$$T_\alpha = \begin{pmatrix} 1+z & 0 & (t_1 + \alpha) \\ 0 & 1+z & t_2 \\ 0 & 0 & 1 \end{pmatrix}, \text{ donc } T_\alpha \mathbf{x} = T\mathbf{x} + \alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \quad (8.8)$$

Ainsi,

$$\nabla u(T_\alpha \mathbf{x}) = \nabla u(T\mathbf{x}) + \alpha D^2 u(T\mathbf{x}) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + o(\alpha) \quad (8.9)$$

et :

$$C(t_1 + \alpha, t_2, z) - C(t_1, t_2, z) = \alpha \int_{\Omega_c} \sigma D^2 u(T\mathbf{x}) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \nabla v(\mathbf{x}) + o(\alpha) \quad (8.10)$$

et on en déduit que :

$$\partial_1 C(t_1, t_2, z) = \int_{\Omega_c} \sigma D^2 u(T\mathbf{x}) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \nabla v(\mathbf{x}) \quad (8.11)$$

Un résultat analogue est obtenu pour la seconde variable t_2 :

$$\partial_2 C(t_1, t_2, z) = \int_{\Omega_c} \sigma D^2 u(T\mathbf{x}) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \cdot \nabla v(\mathbf{x}) \quad (8.12)$$

Pour la variable associée au zoom z , on considère :

$$T_\gamma = \begin{pmatrix} 1+z+\gamma & 0 & t_1 \\ 0 & 1+z+\gamma & t_2 \\ 0 & 0 & 1 \end{pmatrix}, \text{ donc } T_\gamma \mathbf{x} = T\mathbf{x} + \gamma \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \quad (8.13)$$

Ainsi,

$$\nabla u(T_\gamma \mathbf{x}) = \nabla u(T\mathbf{x}) + \gamma D^2 u(T\mathbf{x}) \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} + o(\gamma) \quad (8.14)$$

et :

$$C(t_1, t_2, z + \gamma) - C(t_1, t_2, z) = \gamma \int_{\Omega_c} \sigma D^2 u(T\mathbf{x}) \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \cdot \nabla v(\mathbf{x}) + o(\gamma) \quad (8.15)$$

Et finalement :

$$\partial_3 C(t_1, t_2, z) = \int_{\Omega_c} \sigma D^2 u(T\mathbf{x}) \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \cdot \nabla v(\mathbf{x}). \quad (8.16)$$

En associant les trois différentielles, la forme explicite du gradient de la fonctionnelle est donnée par :

$$\nabla C(t_1, t_2, z) = \begin{pmatrix} \partial_1 C(t_1, t_2, z) \\ \partial_2 C(t_1, t_2, z) \\ \partial_3 C(t_1, t_2, z) \end{pmatrix}. \quad (8.17)$$

■

La fonctionnelle à maximiser n'est pas concave, elle est soumise à la présence de maxima locaux. L'initialisation de la montée de gradient est donc cruciale pour garantir la convergence vers une solution satisfaisante. Pour l'initialisation de la première trame de la séquence, il est possible d'avoir recours à une recherche exhaustive grossière, ou bien d'effectuer plusieurs initialisations et retenir celle qui fournit le meilleur score. Pour les trames suivantes, la séquence d'images fournit de la régularité temporelle ce qui permet d'initialiser la montée de gradient pour chaque trame à partir des paramètres obtenus à l'instant précédent.

8.2.2 Cas projectif

Bien qu'il soit possible de restreindre l'espace de recherche des paramètres en évaluant l'influence de chacun et son rôle par rapport aux deux caméras impliquées, la mise en place d'une recherche exhaustive fait intervenir un plus grand nombre de dimensions qui rend la tâche difficile à effectuer en temps réel. La mise en place d'une montée de gradient est donc d'autant plus judicieuse dans ce cas.

La fonctionnelle $C = C(H)$ à maximiser est régie par 8 paramètres, donc huit gradients partiels sont à évaluer.

On note H_{11} la variation sur le paramètre h_{11} :

$$H_{11} = \begin{pmatrix} h_{11} + \alpha & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}, \quad (8.18)$$

donc

$$H_{11}\mathbf{x} = \begin{pmatrix} \frac{(h_{11}+\alpha)x_1+h_{12}x_2+h_{13}}{h_{31}x_1+h_{32}x_2+h_{33}} \\ \frac{h_{21}x_1+h_{22}x_2+h_{23}}{h_{31}x_1+h_{32}x_2+h_{33}} \\ 0 \end{pmatrix} = H\mathbf{x} + \alpha \begin{pmatrix} \frac{x_1}{h_{31}x_1+h_{32}x_2+h_{33}} \\ 0 \\ 0 \end{pmatrix}. \quad (8.19)$$

Les variations H_{12}, \dots, H_{23} s'écrivent de façon similaire :

$$H_{12}\mathbf{x} = \begin{pmatrix} \frac{h_{11}x_1+(h_{12}+\alpha)x_2+h_{13}}{h_{31}x_1+h_{32}x_2+h_{33}} \\ \frac{h_{21}x_1+h_{22}x_2+h_{23}}{h_{31}x_1+h_{32}x_2+h_{33}} \\ 0 \end{pmatrix} = H\mathbf{x} + \alpha \begin{pmatrix} \frac{x_2}{h_{31}x_1+h_{32}x_2+h_{33}} \\ 0 \\ 0 \end{pmatrix}, \quad (8.20)$$

$$H_{13}\mathbf{x} = \begin{pmatrix} \frac{h_{11}x_1+h_{12}x_2+(h_{13}+\alpha)}{h_{31}x_1+h_{32}x_2+h_{33}} \\ \frac{h_{21}x_1+h_{22}x_2+h_{23}}{h_{31}x_1+h_{32}x_2+h_{33}} \\ 0 \end{pmatrix} = H\mathbf{x} + \alpha \begin{pmatrix} \frac{1}{h_{31}x_1+h_{32}x_2+h_{33}} \\ 0 \\ 0 \end{pmatrix}, \quad (8.21)$$

$$H_{21}\mathbf{x} = \begin{pmatrix} \frac{h_{11}x_1+h_{12}x_2+h_{13}}{h_{31}x_1+h_{32}x_2+h_{33}} \\ \frac{(h_{21}+\alpha)x_1+h_{22}x_2+h_{23}}{h_{31}x_1+h_{32}x_2+h_{33}} \end{pmatrix} = H\mathbf{x} + \alpha \begin{pmatrix} 0 \\ \frac{x_1}{h_{31}x_1+h_{32}x_2+h_{33}} \end{pmatrix}, \quad (8.22)$$

$$H_{22}\mathbf{x} = \begin{pmatrix} \frac{h_{11}x_1+h_{12}x_2+h_{13}}{h_{31}x_1+h_{32}x_2+h_{33}} \\ \frac{h_{21}x_1+(h_{22}+\alpha)x_2+h_{23}}{h_{31}x_1+h_{32}x_2+h_{33}} \end{pmatrix} = H\mathbf{x} + \alpha \begin{pmatrix} 0 \\ \frac{x_2}{h_{31}x_1+h_{32}x_2+h_{33}} \end{pmatrix}, \quad (8.23)$$

$$H_{23}\mathbf{x} = \begin{pmatrix} \frac{h_{11}x_1+h_{12}x_2+h_{13}}{h_{31}x_1+h_{32}x_2+h_{33}} \\ \frac{h_{21}x_1+h_{22}x_2+(h_{23}+\alpha)}{h_{31}x_1+h_{32}x_2+h_{33}} \end{pmatrix} = H\mathbf{x} + \alpha \begin{pmatrix} 0 \\ \frac{1}{h_{31}x_1+h_{32}x_2+h_{33}} \end{pmatrix}. \quad (8.24)$$

Pour les paramètres intervenant au dénominateur, il est nécessaire de linéariser :

$$H_{31}\mathbf{x} = \begin{pmatrix} \frac{h_{11}x_1+h_{12}x_2+h_{13}}{(h_{31}+\alpha)x_1+h_{32}x_2+h_{33}} \\ \frac{h_{21}x_1+h_{22}x_2+h_{23}}{(h_{31}+\alpha)x_1+h_{32}x_2+h_{33}} \end{pmatrix}, \quad (8.25)$$

donc

$$\begin{aligned} H_{31}\mathbf{x} &= H\mathbf{x} - \alpha \frac{x_1}{h_{31}x_1 + h_{32}x_2 + h_{33}} H\mathbf{x} + o(\alpha) \\ &= H\mathbf{x} \left(1 - \alpha \frac{x_1}{h_{31}x_1 + h_{32}x_2 + h_{33}} \right) + o(\alpha) \end{aligned}$$

Et de même :

$$\begin{aligned} H_{32}\mathbf{x} &= \begin{pmatrix} \frac{h_{11}x_1+h_{12}x_2+h_{13}}{h_{31}x_1+(h_{32}+\alpha)x_2+h_{33}} \\ \frac{h_{21}x_1+h_{22}x_2+h_{23}}{h_{31}x_1+(h_{32}+\alpha)x_2+h_{33}} \end{pmatrix} \\ &= H\mathbf{x} - \alpha \frac{x_2}{h_{31}x_1 + h_{32}x_2 + h_{33}} H\mathbf{x} + o(\alpha). \end{aligned}$$

Ces calculs de variations sont ensuite réinjectés dans le calcul de la métrique, conduisant à un gradient à huit dimensions.

Les transformations projectives sont assez instables en raison de la non linéarité du dénominateur, ce qui rend leur estimation délicate. Suivant l'inspiration de [Mann 1997], les premières trames peuvent être initialisées sans l'aspect projectif, ce qui permet d'assurer plus de stabilité et éventuellement d'utiliser une recherche exhaustive sur un espace à trois dimensions au lieu de huit, puis la dimension projective est incluse dans la montée de gradient pour les trames suivantes.

8.3 Implémentation temporelle

Afin d'accélérer la convergence, améliorer la fiabilité de l'algorithme et contrôler la performance du recalage, la montée de gradient est couplée à un schéma temporel

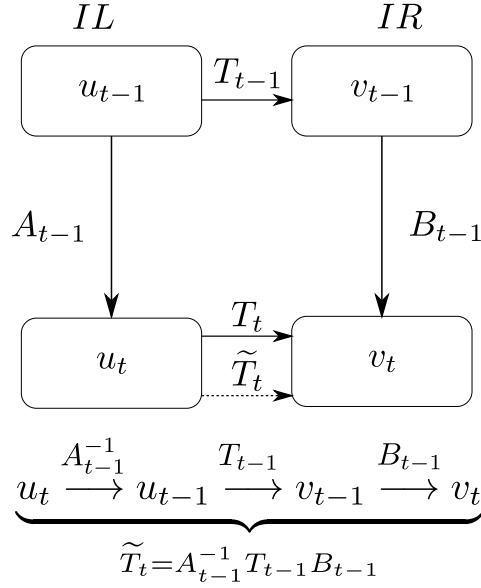


FIGURE 8.1 – Schéma temporel garantissant la convergence rapide de la montée de gradient, ainsi qu'un contrôle de l'erreur.

qui utilise l'information des trames précédemment recalées pour prédire et contrôler le recalage des trames suivantes. La figure 8.1 explicite les différentes étapes de l'implémentation temporelle qui permettent d'effectuer un recalage embarqué rapide.

À l'instant $t - 1$, les images IL et IR appelées respectivement u_{t-1} et v_{t-1} sont recalées à l'aide de la transformation T_{t-1} , de sorte que :

$$u_{t-1}(T_{t-1}\mathbf{x}) = v_{t-1}(\mathbf{x}). \quad (8.26)$$

De plus, en ne considérant qu'une seule des deux modalités indépendamment, le mouvement entre u_{t-1} et u_t d'une part ou v_{t-1} et v_t d'autre part peut être rapidement estimé. En effet, le casque est équipé d'un détecteur de posture qui fournit des informations sur le mouvement de la tête du pilote, tandis que les systèmes inertiels de l'hélicoptère permettent d'évaluer le mouvement global de l'appareil, et donc de la caméra infrarouge. Cela permet donc une estimation du mouvement temporel des deux caméras à bas coût. Le recalage temporel peut donc être estimé séparément sur chaque modalité, en estimant respectivement les transformations A_{t-1} entre u_{t-1} et u_t et B_{t-1} entre v_{t-1} et v_t , de sorte que :

$$\begin{cases} u_{t-1}(A_{t-1}\mathbf{x}) = u_t(\mathbf{x}), \\ v_{t-1}(B_{t-1}\mathbf{x}) = v_t(\mathbf{x}). \end{cases} \quad (8.27)$$

Grâce à ces trois estimations, il est possible de prédire la transformation estimée \tilde{T}_t entre u_t et v_t par :

$$\tilde{T}_t = A_{t-1}^{-1} T_{t-1} B_{t-1} \quad (8.28)$$

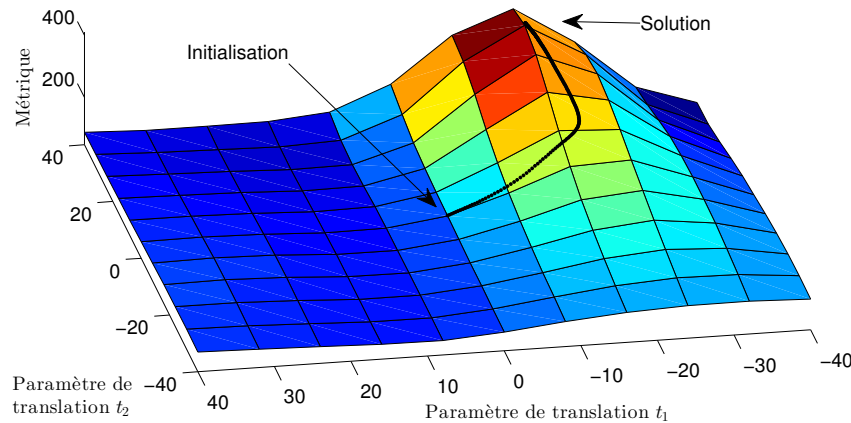


FIGURE 8.2 – Carte du critère de similarité pour une grille de paramètres de translations, à zoom fixé, et évolution des paramètres estimés à chaque itération de la montée de gradient. L’algorithme d’optimisation permet effectivement de maximiser la métrique, et converge vers le maximum.

Cette estimation \tilde{T}_t peut alors être utilisée comme initialisation raisonnable pour le recalage à l’instant t . Cela permet d’assurer que la montée de gradient converge en peu d’itérations, et la bonne initialisation permet d’atteindre un maximum pertinent.

De plus, cette procédure peut également se montrer utile pour contrôler l’énergie et éviter des phénomènes de divergence dans le processus de montée de gradient. En effet, l’algorithme est sujet à des maxima locaux et peut se révéler sensible aux pas de montée utilisés pour la mise à jour de chaque paramètre : si l’optimum est proche et les pas de montée trop grands, la mise à jour des paramètres risque d’être trop importante et de dépasser le maximum, ce qui peut conduire à l’oscillation du critère autour du maximum sans jamais l’atteindre, ou à la divergence vers un autre maximum local. L’hypothèse de régularité temporelle du mouvement permet de réduire cette sensibilité : si l’énergie à la fin de la montée de gradient s’avère plus faible qu’à l’initialisation, cela signifie que l’algorithme a vraisemblablement convergé vers un maximum local, et l’on peut alors choisir de préférer la prédiction \tilde{T}_t .

8.4 Analyse du schéma proposé

8.4.1 Étude du schéma d’optimisation

Les études théoriques et expérimentales décrites ci-dessus ont permis de valider la métrique en termes de performance, de robustesse et de pertinence. La maximisation du critère permet bien de recouvrir les paramètres optimaux de transformation, en théorie comme en pratique. La robustesse au bruit et le comportement de la fonctionnelle ont également été évalués et ont permis de démontrer les performances du critère dans le cadre d’une recherche exhaustive sur l’ensemble de l’espace des

paramètres.

Reste à valider le schéma d'optimisation. En effet, bien que la fonctionnelle ne soit pas concave, la capacité de la montée de gradient à atteindre le maximum global, pourvu que l'initialisation soit suffisamment proche, peut être évaluée.

Sur la figure 8.2 est représentée une carte de l'évolution de la métrique, à niveau de zoom fixé. La ligne noire indique le chemin parcouru par les paramètres de translation t_1 et t_2 au fur et à mesure des itérations de la montée de gradient. Chaque étape permet effectivement de maximiser la métrique, et l'algorithme converge vers l'optimum. De plus, bien que le critère ne soit pas strictement concave, cette carte montre le comportement lisse de la fonctionnelle autour du maximum, ce qui garantit qu'une initialisation "raisonnable" peut permettre d'atteindre le maximum global.

Une telle initialisation "raisonnable" est elle-même assurée par la continuité fournie par la séquence d'images, et par l'utilisation du recalage effectué aux trames antérieures, conformément à ce qui a été présenté dans la section 8.3. L'expérience suivante illustre ce point : toujours à partir des images synthétiques, un mouvement évoluant dans le temps a été simulé sur l'image IR, avec des paramètres de translation t_1 et t_2 compris entre -40 et +40 pixels, et un paramètre de zoom z compris entre 0 et 0.4. La première trame est recalée à l'aide d'une recherche exhaustive grossière, puis pour chaque trame suivante, l'image IL est recalée à l'aide de la montée de gradient, et l'initialisation est basée sur le recalage effectué à l'instant précédent. Cela garantit que l'initialisation n'est pas trop loin de l'optimum, et permet de converger en peu d'itérations. Sur la figure 8.3 est représentée l'évolution des paramètres de transformation simulés (en bleu), ainsi que l'estimation obtenue (étoiles rouges), pour chaque trame. L'algorithme est performant tout au long de la séquence, avec une erreur d'estimation moyenne inférieure à 1 pixels pour le vecteur de translation et inférieure à 0.003 pour le coefficient de zoom.

Ces deux expériences montrent la pertinence de la montée de gradient pour le recalage des deux séquences d'images.

8.4.2 Étude du schéma temporel

Si la position relative entre les deux caméras évolue soudainement, les paramètres de translation vont varier de façon abrupte, et la montée de gradient peut ne plus suivre un tel changement en raison d'une initialisation non adaptée. Le schéma de validation temporelle permet de remédier à cela. Grâce à la détection de posture et aux capteurs inertiels, le mouvement des deux caméras peut être évalué indépendamment afin de propager le mouvement et prédire plus finement le recalage aux trames suivantes. Si le mouvement est global et assimilé à un mouvement de la scène, par exemple dans le cas d'une rafale de vent, alors la position relative des deux caméras est inchangée et la performance du recalage intouchée. Si le changement brusque n'affecte que l'une des deux caméras, par exemple dans le cas d'un mouvement soudain de la tête du pilote, les paramètres de recalage peuvent changer de façon drastique, au risque que la seule montée de gradient ne permette pas de suivre

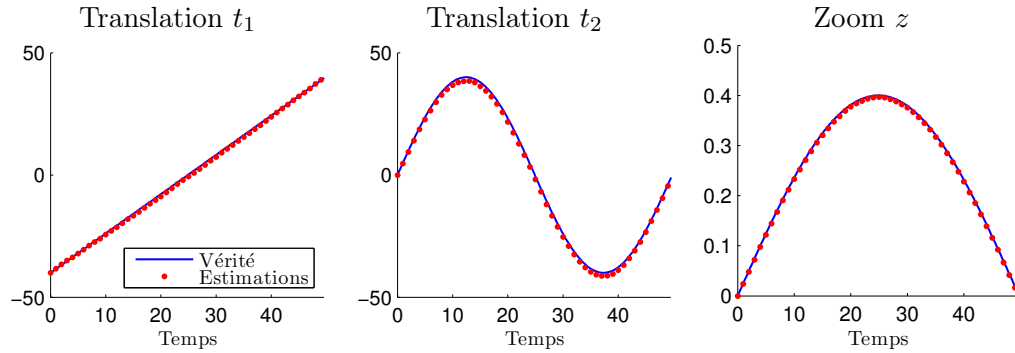


FIGURE 8.3 – Évolution de chacun des paramètres (t_1, t_2, z) de la transformation au cours de la séquences et estimation (étoiles rouges) obtenue à l'aide de la montée de gradient. Certaines erreurs sont liées au fait que l'évolution des paramètres est continue, donc les valeurs ne sont pas entières, ce qui impose des approximations au sein de la transformation.

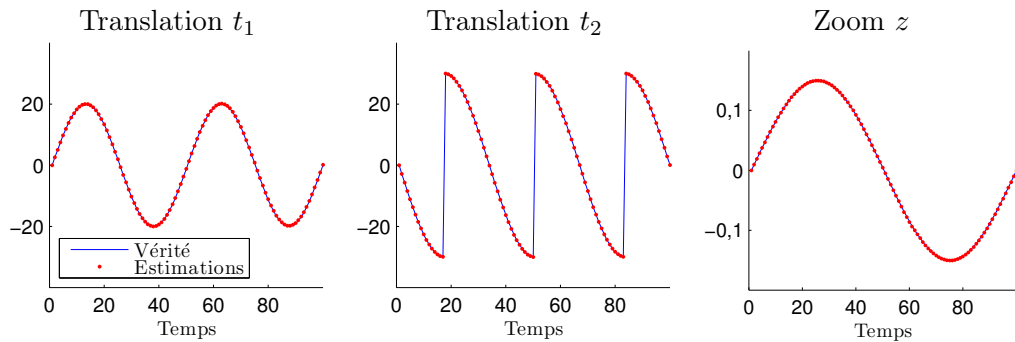


FIGURE 8.4 – Évolution de chacun des paramètres (t_1, t_2, z) de la transformation au cours de la séquences et estimation (étoiles rouges) obtenue à l'aide de la montée de gradient. Les discontinuités observées pour le paramètre de translation horizontale t_2 simulent un mouvement soudain de la tête, géré grâce au schéma de validation temporelle.

une telle variation. D'où l'intérêt du schéma temporel. La figure 8.4 illustre l'évolution (en bleu) et l'estimation (étoiles rouges) de chacun des paramètres simulés, lorsque le paramètre de translation horizontale souffre de discontinuités qui simulent un mouvement rapide de la tête du pilote. Cela montre que même confronté à des sauts, l'algorithme est capable de recalcr avec précision les deux caméras, grâce à la prédiction qui s'appuie sur le mouvement unimodal.

En fait, les deux estimations unimodales et multimodales, à savoir l'estimation multimodale T_t entre les deux caméras d'une part et les estimations unimodales A_t et B_t du mouvement de chaque caméra indépendamment d'autre part (voir figure 8.1), se complètent. En effet, dans le schéma temporel le recalcr unimodal a été introduit en tant que garde-fou pour s'assurer que la montée de gradient fonctionne à souhait, mais le contraire s'applique également : l'accumulation des estimations

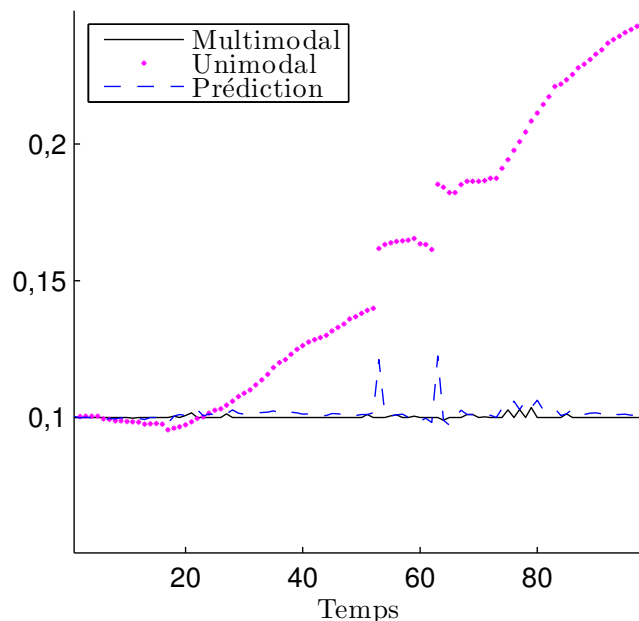


FIGURE 8.5 – Évolution de l'estimation du paramètre de zoom pour une séquences d'images réelles à l'aide de la méthode multimodale (ligne noire), par propagation de l'information de recalage multimodal (étoiles pourpres), et état de l'initialisation avant la montée de gradient (tirets bleu).

unimodales consécutives introduit une dérive dans l'estimation des paramètres, ce qui rend l'étape de recalage multimodale nécessaire pour ajuster l'estimation. La figure 8.5 illustre ce phénomène, à travers l'évolution du coefficient de zoom sur une séquence d'images réelles, dont la vraie valeur est constante égale à 0.1 (cela correspond à la différence intrinsèque de résolution des deux caméras). La ligne noire correspond à l'estimation du paramètre par la méthode multimodale, les étoiles pourpres représentent l'estimation obtenue à l'aide de la seule information unimodale propagée depuis l'initialisation, et les tirets bleu donnent l'initialisation \tilde{T}_t à chaque étape avant la montée de gradient. Ces courbes révèlent deux aspects : d'abord, que le recalage unimodal seul introduit une dérive. Cela permet d'observer également que (presque) chaque étape, le recalage multimodal apporte une correction à la prédiction \tilde{T}_t , illustrés par la différence entre la ligne noire et les tirets bleu.

Cette étude permet de justifier l'association du recalage unimodal et multimodal pour garantir précision et fiabilité.

8.5 Résultats

8.5.1 Implémentation et temps de calcul

La méthode proposée a dans un premier temps été implémentée sous Matlab, puis en C^{++} et enfin optimisée sur carte graphique GPU (*Graphics Processing Unit*)

pour accélérer les temps de calcul. Les images IL initiales ont pour taille 1600×1200 pixels et les images IR 768×576 . Les images sont d'abord ramenées à la même dimension à l'aide d'une interpolation bicubique, puis elles sont sous-échantillonnées afin de réduire la taille des images et donc les temps de calcul.

Au cours de la résolution par recherche exhaustive, la métrique doit être évaluée pour chaque jeu de paramètres testés. Cela implique d'appliquer la transformation testée à l'image courante, puis calculer la métrique, ce qui nécessite le calcul du gradient de chacune des images. Cette étape est répétée pour chaque jeu de paramètres de l'espace de recherche, qui peut contenir au moins 10000 possibilités (dans le cas non projectif!).

Pour la résolution par montée de gradient, à chaque itération l'image recalée doit être calculée afin d'évaluer le gradient de la métrique, ce qui nécessite également le calcul des gradients. Les expériences ont montré que 100 itérations permettent à la montée de gradient de converger, et ce nombre d'itérations peut être d'avantage réduit dans le cadre du schéma temporel décrit dans la section 8.3.

L'efficacité de la montée de gradient par rapport à la résolution par recherche exhaustive est d'autant plus intéressante que le nombre de paramètres à estimer croît, comme par exemple dans le cas projectif. Afin d'illustrer la complexité associée à chacune des méthodes, le tableau 8.1 fournit les temps de calcul en fonction de la taille des images (lié au coefficient de sous-échantillonnage appliqué) pour une étape du calcul de la solution : une itération dans le cas de la montée de gradient ou le calcul de la métrique pour un lot de paramètres dans le cas de la recherche exhaustive. La prise en compte du nombre moyen d'étapes nécessaires, (nombre d'itérations pour la montée de gradient ou taille de l'espace de recherche des paramètres pour la recherche exhaustive), fournit ainsi une indication du temps nécessaire au recalage d'un couple de trames, en fonction de la taille. Ce temps de calcul peut alors être exprimé en termes de fréquence de traitement, c'est-à-dire le nombre de trames qui peuvent être traitées en une seconde, qui s'avère aller jusqu'à 8 trames par secondes.

8.5.2 Validation expérimentale sur données réelles

Afin d'évaluer les différentes métriques sur les données réelles sans connaître les paramètres optimaux de transformation, une méthode basée sur la sélection manuelle de points contrôle a été développée. Sur l'image de référence IR et l'image à recalée IR sont sélectionnées dix paires de points caractéristiques (ce qui peut être une tâche complexe), et la distance (en pixels) entre les coordonnées de chaque paire de points entre les deux images est mesurée. Les mêmes mesures sont effectuées entre l'image de référence et l'image recalée. L'évaluation de la distance est répétée sur des séquences de dix images pour chaque série, puis moyennée pour fournir une erreur de recalage (en pixels) moyenne. Un exemple de sélection de points de contrôle sur l'image de référence et sur l'image recalée associée est représenté sur la figure 8.6. Cinq séries d'images, dont un exemple est reproduit sur la figure 8.7, ont été sélectionnées afin de représenter différentes caractéristiques de la vision de nuit. Concernant les images IL, les trois premières séries sont relativement peu

TABLE 8.1 – Temps de calcul nécessaires au recalage effectué par recherche exhaustive, montée de gradient dans le cas standard zoom/translation, et montée de gradient dans le cas projectif, en fonction de la taille des images et du type d'implémentation. L'implémentation GPU de la montée de gradient permet d'effectuer le recalage en moins d'une seconde.

	Recherche exhaustive		Montée de gradient (Zoom/Translation)		Montée de gradient (Projective)				
	Matlab	C++ GPU	Matlab	C++ GPU	Matlab	C++ GPU			
	Durée d'une itération (ms)								
1600 × 1200	756.37	153.15	7.46	1455.71	195.01	9.38	1657.97	255.23	12.59
800 × 600	175.53	37.26	2.99	328.52	44.56	3.28	394.28	56.36	4.91
400 × 300	46.5	9.19	0.998	88.36	11.38	1.22	99.05	15.09	1.66
	Temps total de recalage (s)								
	≈ 10000 itérations		≈ 100 itérations		≈ 100 itérations				
1600 × 1200	7563.7	1531.5	74.6	145.571	19.501	0.938	165.797	25.523	1.259
800 × 600	1755.3	372.6	29.9	32.852	4.456	0.328	39.428	5.636	0.491
400 × 300	465	91.9	9.98	8.836	1.138	0.122	9.905	1.509	0.166
	Fréquence (Hz)								
1600 × 1200	0.0001	0.0007	0.0134	0.0069	0.0513	1.0661	0.0060	0.0392	0.7943
800 × 600	0.0006	0.0027	0.0334	0.0304	0.2244	3.0488	0.0254	0.1774	2.0367
400 × 300	0.0022	0.0109	0.1002	0.1132	0.8787	8.1967	0.1010	0.6627	6.0241

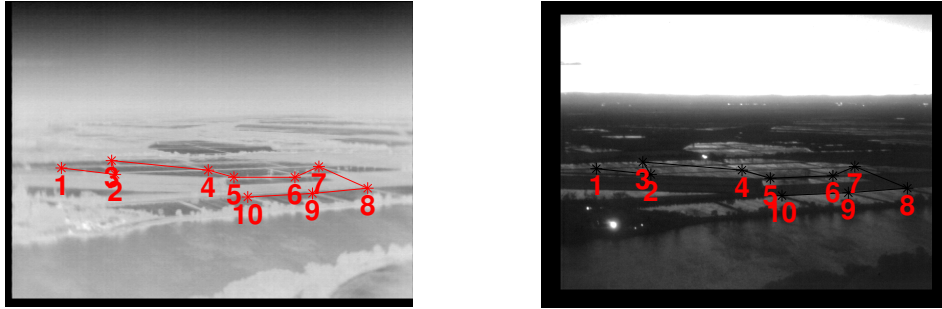


FIGURE 8.6 – Exemple de points de contrôle sélectionnés sur l'image de référence IR (à gauche), et correspondance sur l'image recalée (à droite).

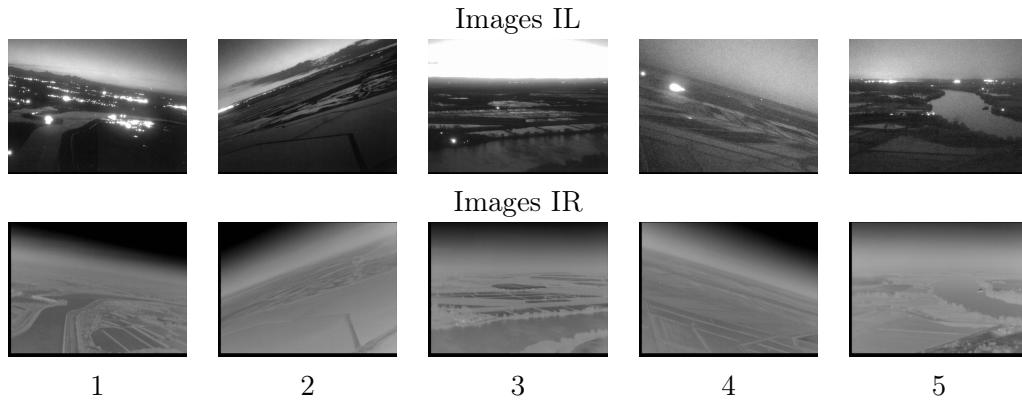


FIGURE 8.7 – Extraits des séquences d'images utilisées pour les tests de contrôle. En fonction des conditions et du niveau de lumière, les images présentent différentes caractéristiques et dégradations.

corrompues par du bruit et offrent un contraste satisfaisant, tandis que les séries 4 et 5 sont affectées par un bruit de Poisson fort et souffrent d'un contraste plus faible. Les séries 1 et 4 présentent également des phénomènes de saturation autour des sources lumineuses qui perturbent la dynamique de la scène et masquent une partie de l'information. D'autre part les images IR sont peu contrastées et apportent parfois très peu d'information, comme c'est le cas pour les séries 2 et 4. L'expérience des points de contrôle a été appliquée pour des images recalées à l'aide de la corrélation croisée (7.1), l'information mutuelle (7.4), l'information mutuelle combinée gradient (7.10), les descripteurs MIND (7.11), le cosinus pondéré (7.5), NGF (7.8) et la métrique proposée. L'erreur moyenne en termes de distance entre les pixels de l'image de référence et de l'image recalée figure dans le tableau 8.2. La méthode proposée fournit un meilleur recalage dans la plupart des cas. Les méthodes basées sur le cosinus pondéré (7.5) ou les descripteurs MIND (7.11) offrent également de bonnes performances de recalage, mais elles semblent plus sensibles aux faibles contrastes et à la présence de bruit ou d'artefacts.

TABLE 8.2 – Distance moyenne entre les coordonnées en pixels des points de contrôle sur l'image de référence et ceux des images à recalcr puis des images recalées à l'aide de la corrélation croisée (CC), de l'information mutuelle (MI), de l'information mutuelle combinée gradient (MI-G), des descripteurs MIND, du cosinus pondéré, de NGF, et de la métrique proposée.

Méthode	1	2	3	4	5
Original	31.53	31.23	19.79	25.57	29.33
CC	33.36	40.76	28.64	31.99	35.02
MI	13.34	11.80	7.97	23.93	10.69
MI-G	17.11	13.16	5.86	10.23	9.01
MIND	11.75	11.81	4.25	8.62	9.40
Cosinus pondéré	6.69	11.96	4.04	8.85	8.10
NGF	40.42	11.81	12.13	13.13	9.44
Proposé	6.65	11.26	4.42	8.58	7.80

8.5.3 Recalage sur données réelles

La figure 8.8 illustre les performances de recalage sur données réelles. Pour chaque colonne, l'image infrarouge, l'image optique à recalcr et l'image IR recalée dans le plan de l'image IR sont représentées. Les images ont été ramenées à la résolution commune de 1024×768 pixels, et sous-échantillonnées d'un facteur deux. Les paramètres de la montée de gradient ont été initialisés à l'aide d'une recherche exhaustive, puis les pas de mise à jour ont été réglés à 10^{-4} pour les paramètres de translation et $1e-8$ pour le coefficient de zoom. Ces pas ont été optimisés à la main, mais sont fixes pour toutes les données testées. De plus, l'estimation du déplacement temporel de chaque caméra fourni normalement par les capteurs de posture et les centrales inertielles, a été simulée à l'aide de Motion2D [Odobez 1995], une méthode rapide et simple de recalage unimodal temporel.

La précision du recalage obtenu peut s'observer à l'aide de la superposition des contours. Un détecteur de contour a été appliqué à l'image IR, et cette information est superposée aux images optiques afin de vérifier l'alignement correct des contours. Les images mosaïques qui alternent des bandes de l'image de référence avec des bandes de l'image optique avant ou après recalage permettent également d'appréhender la qualité du recalage, en observant la continuité des contours d'une modalité à l'autre. Les séquences complètes au format vidéo sont disponibles au téléchargement ¹.

8.5.4 Résultats sur simulations projectives

Faute de données réelles adaptées, des transformations affines ou projectives ont été simulées à partir d'images réelles préalablement alignées. La géométrie projective inclut les rotations et les changements de perspective. De telles transformations ont

1. <http://image.math.u-bordeaux1.fr/Registration>

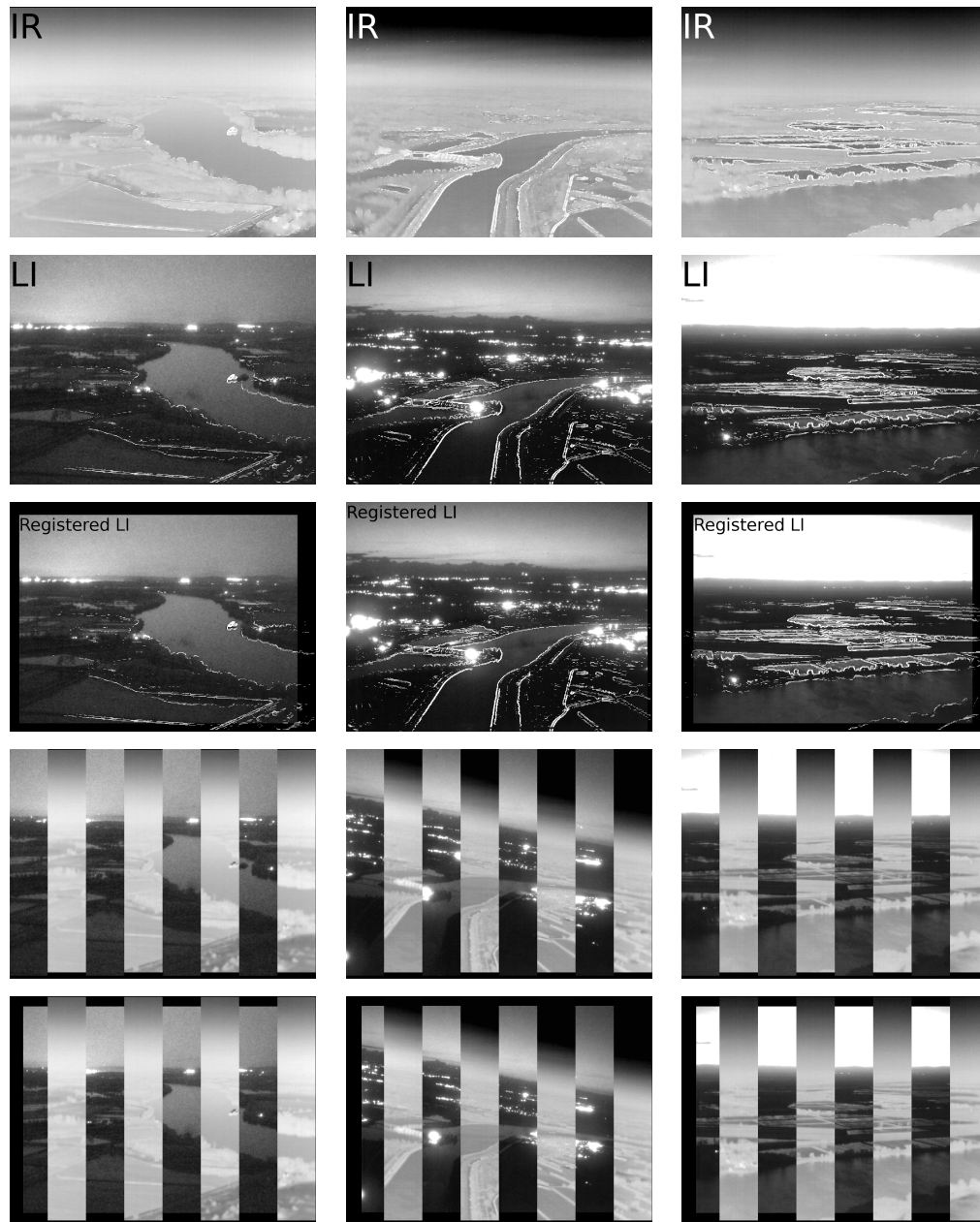


FIGURE 8.8 – Résultats de recalage obtenus sur données réelles à l'aide de la métrique proposée. L'image de référence IR est représentée sur la première ligne, l'image courante IL (à recaler) sur la deuxième, et l'image recalée sur la troisième. Les contours extraits de l'image IR sont superposés sur chaque image pour appréhender visuellement la qualité du recalage. Sur les quatrième et cinquième lignes sont représentées des images mosaïques avant et après recalage qui alternent des bandes de l'image IR et de l'image IL à recaler, pour mieux constater les différences d'alignement. Les séquences complètes peuvent être téléchargées sur <http://image.math.u-bordeaux1.fr/Registration>.

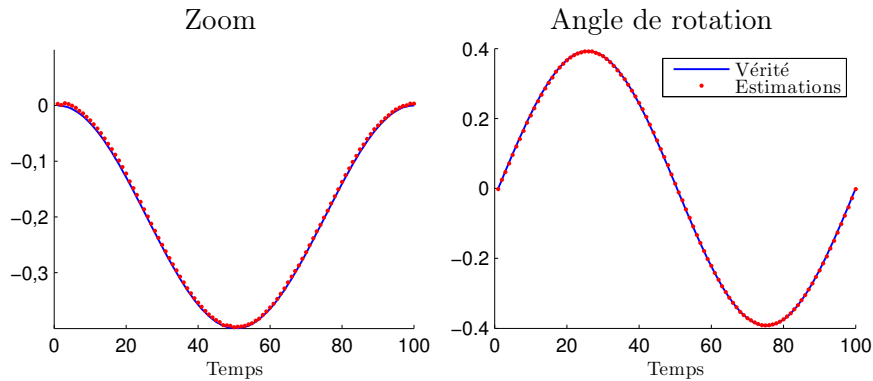


FIGURE 8.9 – Simulation d'un mouvement de rotation de l'image IL, avec un coefficient de zoom et un angle de rotation évoluant dans le temps (en bleu), et estimation des paramètres (étoiles rouges).

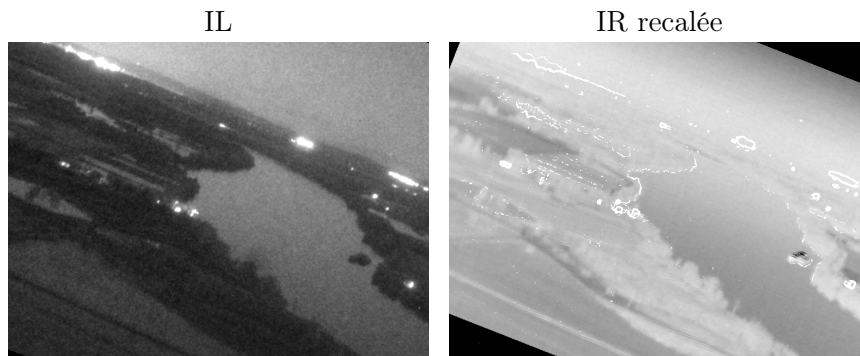


FIGURE 8.10 – Exemple d'une image IL ayant subi un zoom et une rotation, et image IR associée, recalée. Les contours de l'image IL sont superposés à l'image IR pour mieux appréhender la précision du recalage.

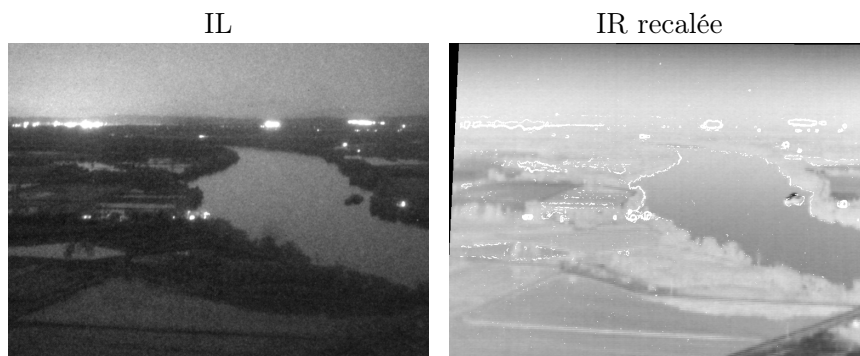


FIGURE 8.11 – Exemple d'une image IL ayant subi un zoom, une translation dans les deux directions et une projection horizontale, et image IR associée, recalée. Les contours de l'image IL sont superposés à l'image IR pour mieux appréhender la précision du recalage.

été appliquées aux séquences d'images optiques, puis l'image infrarouge est recalée en fonction. La connaissance des vrais paramètres de transformation permet de s'assurer que l'estimation est correcte.

La figure 8.9 montre l'évolution des paramètres de zoom et de rotation simulés et estimés au cours de la séquence. Les figures 8.10 et 8.11 donnent un exemple de transformation appliquée : l'image IL est soumise à un zoom et à une rotation pour la figure 8.10, et à un zoom, une translation dans les deux directions et une projection horizontale pour la figure 8.11. Les images IR recalées y sont représentées également. Ces figures illustrent la capacité du modèle projectif à représenter nombre de transformations susceptibles d'être rencontrées en vol.

Cependant, l'approche présentée ici suppose que la scène est plane, pour l'estimation d'une transformation paramétrique. Si l'hélicoptère vole à très faible altitude par exemple, la présence de bâtiments ou d'arbres pourrait contredire cette hypothèse. Dans ce cas, la maximisation de la fonctionnelle correspondrait à la mise en correspondance du plus grand nombre possible de contours, ce qui reviendrait à aligner le plan principal de la scène en considérant le reste comme des outliers. De telles configurations seraient à étudier plus finement.

8.6 Conclusion

Une technique de recalage multimodal basée sur l'alignement des contours a été développée dans cette partie. Celle-ci consiste à aligner les contours présents dans les deux modalités, en se basant sur leur amplitude et leur orientation. La résolution dans le cadre d'une implémentation embarquée pour le dispositif BNL s'effectue à l'aide d'une montée de gradient qui offre un algorithme rapide. Couplée à un schéma temporel qui permet de contrôler les erreurs et garantit ainsi la fiabilité du processus, la méthode proposée s'est révélée robuste et rapide, et l'algorithme développé peut procéder jusqu'à 8 trames par secondes. Des études théoriques et expérimentales ont montré que le critère est adapté, et les résultats sur données réelles valident l'application au contexte de vision de nuit.

De plus, la complexité temporelle pourrait être encore davantage améliorée en exploitant le schéma temporel : en étudiant le comportement du recalage unimodal, il pourrait être possible de n'effectuer le recalage temporel que toutes les n trames (en fonction de la dérive observée), en se limitant aux corrections unimodales entre temps. Cela offrirait un gain de temps non négligeable pour l'application embarquée.

Conclusion et perspectives

Conclusion

Cette thèse s'est consacrée à la problématique de la vision de nuit pour les pilotes d'hélicoptère. Elle a étudié la faisabilité d'un dispositif entièrement numérique de vision bas niveau de lumière (BNL) doté d'un tube intensificateur de lumière (IL) couplé à une caméra CCD d'une part, et d'une caméra infrarouge (IR) d'autre part. L'objectif était d'améliorer la qualité du dispositif de vision de nuit en rehaussant les images BNL. Nous avons pour cela développé une méthode de débruitage des séquences d'images IL, qui s'appuie sur l'estimation automatique du bruit dans les images et un algorithme de débruitage robuste et adaptatif, ainsi qu'une méthode de recalage des images IL et IR dans le même référentiel afin de permettre la fusion des informations fournies par les deux modalités.

Estimation automatique du bruit. La première partie de cette thèse s'est consacrée à l'analyse du bruit présent dans les images IL. Nous avons réalisé dans un premier temps une étude du dispositif d'acquisition des images issues du couplage IL-CCD, afin d'établir un modèle pour le bruit en fonction de la luminosité. Cette étude a ensuite été complétée par la mise en place d'une méthode d'estimation automatique du niveau de bruit dans une image. Cette estimation repose sur les statistiques extraites des régions homogènes de l'image. Nous avons pour cela proposé un test non paramétrique de détection des zones homogènes, basé sur la mesure de corrélation de Kendall. Le test consiste à évaluer la corrélation de rang entre des séquences de pixels voisins, afin de détecter la présence de structure au sein d'un bloc de l'image. Les performances de détection, caractérisées par les probabilités de fausse alarme et de fausse détection, sont contrôlées par le rapport signal à bruit et le nombre d'échantillons disponibles, et elles requièrent peu d'hypothèses sur la distribution du bruit, hormis le fait qu'il doit être non stationnaire et décorréolé spatialement, ce qui le rend applicable dans de nombreux problèmes pour lesquels la nature du bruit est inconnue. Une fois les zones homogènes détectées, la fonction de niveau de bruit, c'est-à-dire la fonction qui relie la variance du bruit à l'intensité du signal, est assimilée à une fonction polynomiale d'ordre deux et ses paramètres sont estimés à partir des statistiques extraites des régions homogènes à l'aide de l'estimateur robuste des moindres déviations. Grâce à cette méthode d'estimation, un modèle de bruit a pu être associé aux images BNL, et l'estimation automatique des paramètres de la fonction de niveau de bruit permet son utilisation dans un algorithme embarqué.

Débruitage adaptatif des séquences d'images. Grâce à l'estimation du bruit présent dans les images IL, nous avons pu entreprendre de les débruiter. Pour cela, nous nous sommes appuyés sur deux méthodes classiques de débruitage, les méthodes variationnelles et les moyennes non locales. Ces deux méthodes bénéficient d'une implémentation simple et sont adaptables à de nombreux types de bruits, ce qui les rend appropriées pour le débruitage des images BNL. En associant les deux méthodes, nous avons pu réduire leurs défauts respectifs. Nous avons tout d'abord proposé de réduire l'effet de *jittering* lié à la sélection de candidats non pertinents dans l'algorithme des moyennes non locales. Pour cela, nous avons défini un indice de *jittering* qui s'appuie sur la variance non locale des candidats sélectionnés et évalue localement la qualité du débruitage. Cela permet de réintroduire de la donnée bruitée lorsque la réduction de bruit est jugée non pertinente, et donc de fournir une solution débiaisée pour laquelle la réduction de bruit est fiable. L'étape de *dejittering* fournit ainsi un indicateur de la qualité du débruitage, utilisé pour mettre en place une régularisation adaptative des moyennes non locales. Cette régularisation vise à réduire l'effet de patch rare lié à la présence de bruit résiduel, lorsque les moyennes non locales ne parviennent pas à trouver suffisamment de candidats pour réduire significativement le bruit, autour des structures singulières par exemple. Pour cela, nous avons défini une attache aux données non locale pondérée par l'indice de confiance des NL-means, afin d'appliquer une régularisation TV localement adaptative, là où le débruitage effectué par les NL-means est jugé insuffisant. Cela permet ainsi de réduire l'effet de patch rare, mais sans introduire d'effet de crénelage associé à la minimisation TV, grâce à l'adaptativité de la méthode. Ce modèle a pu être étendu à de nombreux types de bruits non gaussiens, et nous avons en particulier proposé une adaptation du modèle au cas de bruit additif à variance polynomiale, afin d'exploiter l'estimation de bruit réalisée en première partie. Enfin, afin de bénéficier de la redondance d'information apportée par le flux vidéo des images BNL, nous avons étendu la régularisation adaptative des moyennes non locales au débruitage de séquences d'images. Nous avons utilisé l'extension des NL-means au débruitage de vidéos qui utilise une fenêtre de recherche spatio-temporelle, que nous avons associée à l'utilisation de patches tridimensionnels afin d'augmenter la stabilité temporelle. L'effet de patch rare résultant, accru en raison de la plus grande sélectivité des patches 3D, nous a conduit à mettre au point une régularisation TV spatio-temporelle adaptative qui réduit localement le bruit résiduel et garantit la stabilité temporelle de la solution. La mise en place d'un algorithme adaptatif de débruitage spatio-temporel de séquences d'images, ainsi que son adaptation au bruit à variance polynomiale à partir de l'estimation réalisée en première partie a ainsi permis de débruiter les images IL, afin de fournir des images exploitables pour d'autres applications de traitement d'images.

Recalage multimodal des images IL et IR. Enfin, la dernière partie de cette étude a proposé une méthode de recalage multimodal des images optiques issues de l'intensificateur de lumière et des images infrarouge, en vue d'une fusion

ultérieure. En effet, les différentes informations apportées par les deux capteurs permettraient de rehausser la qualité des images en les associant. Cependant, en raison des différences de perspective entre les deux caméras, la fusion des deux modalités nécessite un recalage préalable des images dans le même référentiel. Nous avons donc développé un critère de recalage adapté à la multimodalité du problème, qui propose d'aligner les contours des deux images à recaler. En se basant sur l'amplitude et la direction des gradients des deux images, le critère vise à aligner les contours présents dans les deux modalités, sans tenir compte de l'orientation de ces derniers, ce qui le rend particulièrement adapté au problème de recalage multimodal des images BNL. Afin de permettre une résolution rapide pour une implémentation embarquée, nous avons ensuite mis au point une méthode d'optimisation basée sur une montée de gradient couplée à un schéma temporel qui garantit que la méthode est robuste. Ce schéma temporel s'appuie sur la redondance temporelle apportée par le flux vidéo et fournit une prédiction fine du recalage pour un couple d'images à l'aide du recalage effectué aux trames précédentes et des informations de déplacement des caméras fournies par les capteurs de position embarqués. Ce schéma d'optimisation permet de recaler des images soumises à des transformations projectives avec une fréquence de six trames par seconde, et la méthode s'est montrée fiable sur les jeux de données réelles fournis par Thales.

L'étude menée ici a donc exploré deux voies d'amélioration des images IL, et de façon plus générale de la vision de nuit des pilotes d'hélicoptère. Ces améliorations reposent d'une part sur l'étude et la réduction du bruit dont souffrent les images issues du dispositif IL, et d'autre part sur le recalage des images IL et IR en vue d'offrir une image rehaussée par la fusion des deux modalités.

Perspectives

Les perspectives liées à ces travaux se présentent sous deux axes. Un axe pratique lié à la problématique de vision bas niveau lumière, et un axe plus académique visant à étendre les modèles proposés ci-dessus à des problèmes plus généraux.

Exploitation pratique pour la vision nocturne

Implémentation temps réel. L'une des contraintes majeures pour les méthodes de rehaussement des images BNL réside dans la nécessité de fournir des algorithmes embarqués, et temps réel. Si nous avons toujours gardé à l'esprit ces contraintes dans la sélection de méthodes ou dans certains choix d'implémentation, le développement des algorithmes répondant à ces critères est à étudier.

L'estimation automatique du bruit n'est pas nécessairement à appliquer à chaque trame ; il s'agit plutôt d'une étape d'initialisation de l'algorithme de débruitage qui s'effectuerait sur les premières trames de la séquence, avec une éventuelle mise à jour pour tenir compte des changements de luminosité de la scène. Les performances

actuelles de l'implémentation, qui détecte les zones homogènes et estime les paramètres de la NLF en quelques secondes, devraient donc suffire. En revanche la complexité calculatoire et la mémoire requises par l'algorithme de débruitage sont à optimiser. Un stage de master co-encadré par Thales et nous-mêmes a été défini dans cette optique. En donnant suite aux travaux fournis par Anne-Sophie Auger au cours de ce stage, une étude des choix d'implémentation optimaux (patches 2D ou 3D, régularisation adaptative ou non, recalage préalable des trames...) en termes de compromis efficacité/complexité, couplée à une parallélisation de l'algorithme, devrait permettre d'optimiser les temps de calcul.

Concernant la méthode de recalage proposée, les études des temps de calcul réalisées offrent des performances satisfaisantes. On pourrait néanmoins envisager d'améliorer encore ces dernières en optimisant l'apport du recalage unimodal et la prédiction de mouvement.

Cependant, la mise en place du dispositif étudié serait envisagée pour l'horizon 2020-2030, donc les systèmes d'acquisition ainsi que les processus embarqué auront évolué. Les performances en terme de qualité des données mais aussi de capacité calculatoire devraient ainsi permettre plus de flexibilité.

Fusion des données. Suite naturelle au recalage des données optiques et infrarouge, la fusion des deux capteurs devraient permettre d'une part de valider l'intérêt et la performance du recalage effectué, mais surtout de rehausser visiblement la qualité des images. Des techniques de fusion [Hall 1997] ont été étudiées au sein de Thales Avionics. Il serait donc intéressant de combiner l'algorithme de fusion à celui de recalage afin de fournir des images hybrides de qualité rehaussée [Li 2013]. La figure 9.1 illustre l'apport de la fusion d'une image infrarouge et d'une image IL. Cette dernière souffre d'un manque de contraste en raison de la saturation des sources lumineuses. La fusion avec l'image infrarouge rehausse les informations de contour, ce qui permet de distinguer les routes et la rivière plus finement.

La mise en place d'autres dispositifs de vision bas niveau de lumière pourraient également contribuer à l'amélioration de la vision de nuit des pilotes. Comme illustré sur la figure 9.2, les nouvelles images SWIR (*Short Wave InfraRed*) fournies par Thales présentent une bonne résolution et un niveau de détail supérieur aux images infrarouge standards. Les caméras SWIR ne sont pas mobiles, mais combinées avec les images IL elles offriraient des perspectives de rehaussement intéressantes.

Au delà de la fusion, l'incrustation de données de natures différentes [Zendjebil 2008] comme des bases de données cartographiques de type modèles numériques de terrain ou des informations capteurs comme la détection d'une cible pourrait venir s'ajouter aux images BNL.

Filtrage guidé. Le rehaussement potentiel apporté par la fusion des capteurs optique et infrarouge ouvre la voie à des techniques de filtrage guidé qui utiliseraient conjointement les deux modalités dans les algorithmes de traitement. Le filtrage guidé a été introduit dans [He 2013], et propose d'effectuer une moyenne non locale

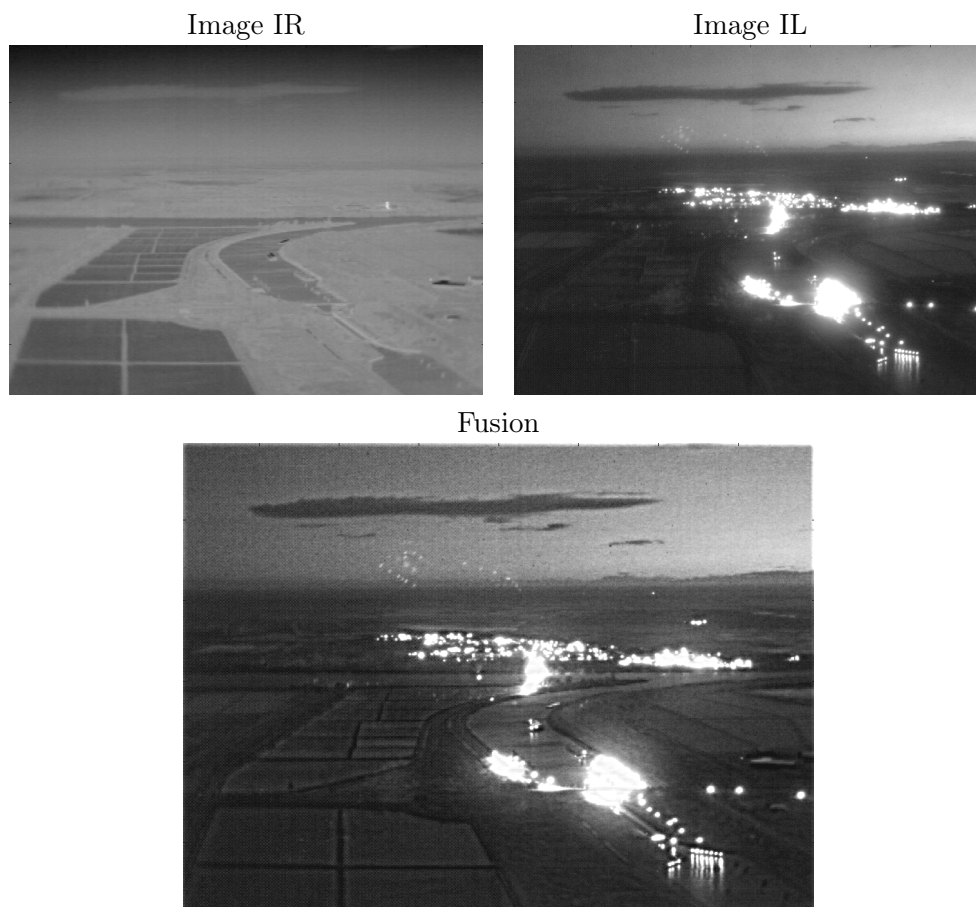


FIGURE 9.1 – Rehaussement des images BNL par la fusion d’une image infrarouge et d’une image IL.

dont les poids sont calculés à partir d’une image guide. Cela permet d’exploiter par exemple l’information des contours de l’image guide. L’idée du filtrage guidé a par la suite été exploitée dans un contexte multimodal. Dans le domaine de l’imagerie médicale, les auteurs de [Yan 2013] ont proposé de restaurer des images PET en utilisant l’information issue de données IRM, tandis qu’en télédétection des images optiques sont utilisées pour supprimer le bruit de *speckle* des images SAR dans [Verdolina 2014]. L’utilisation des images infrarouge pour guider le rehaussement des images optiques pourrait donc être envisagé dans le contexte de vision nocturne.

Correction des défauts. L’amélioration des images BNL peut s’accompagner également de rehaussement de contraste, afin d’équilibrer la dynamique des zones sombres et claires de l’image. Ces opérations peuvent s’appuyer sur l’égalisation globale ou locale de l’histogramme de l’image [Zimmerman 1988, Stark 2000], ou sur un traitement fréquentiel [FitzHenry 1991, Dippel 2002].

D’autres défauts plus locaux sont également à corriger, comme les halos autour des sources lumineuses et les reflets des sources lumineuses sur la vitre du cockpit



FIGURE 9.2 – Exemple d’une image SWIR de haute résolution qui pourrait être incluse dans le dispositif de vision de nuit.

[Fainstain 2010]. On note aussi la présence de défauts globaux systématiques comme la présence d’un maillage ou la différence de luminosité entre les parties gauche et droite de l’image. Cependant, ces derniers sont liés au système de couplage IL-CCD et sont susceptibles d’évoluer avec les futurs dispositifs d’acquisition, tandis que d’autres défauts pourront voir le jour.

L’étude du type d’implémentation souhaitée reste à envisager également. L’ensemble de ces processus de rehaussement, y compris le débruitage et la fusion, peuvent être envisagés de façon systématique si l’implémentation temps réel le permet, ou bien sur demande du pilote dans des conditions définies.

Extensions des modèles théoriques

Si l’étude réalisée au cours de cette thèse s’est appuyée sur une application pratique, les solutions proposées ont été développées dans un contexte plus théorique. Elles peuvent s’accompagner d’améliorations ou d’extensions vouées à des problèmes plus académiques, au moins dans un premier temps.

Estimation non paramétrique du bruit. Une première extension assez directe de la méthode d’estimation du bruit proposée repose sur la prise en compte de bruit non stationnaire. En effet dans beaucoup de systèmes d’imagerie la réponse du capteur n’est pas uniforme, donc la nature du bruit peut varier en fonction de la distance au centre de l’image par exemple. On pourrait donc envisager que la fonction de niveau de bruit ne soit pas décrite par la même paramétrisation en tout point de l’image, ce qui nécessiterait de prendre en compte l’information spatiale et définir cette fonction de niveau de bruit en fonction de paramètres d’intensité et de localisation.

De plus, bien qu’appliqué aux images BNL en sous-échantillonnant ces dernières

afin de réduire la corrélation spatiale, le test de Kendall que nous avons utilisé pour la détection des zones homogènes d'une image n'est pas adapté pour gérer du bruit spatialement corrélé. En effet, la corrélation spatiale se traduit par de la dépendance entre voisins, donc le test de Kendall détecte de la corrélation, interprétée alors comme la présence de structure au sein d'une zone. La détection du bruit corrélé nécessiterait donc d'adapter le test de Kendall.

Enfin, outre la présence de bruit, les images naturelles sont souvent corrompues par d'autres dégradations comme le flou ou des artefacts de compression. Ces derniers peuvent se traduire par de la corrélation spatiale également, et il serait intéressant de les détecter et de les quantifier, afin de pouvoir les corriger de façon automatique.

Régularisation adaptative des moyennes non locales. La régularisation adaptative a été étudiée dans le cadre du débruitage d'images ou de vidéos, en proposant de corriger localement la solution des NL-means. Nous avons mentionné dans la discussion du chapitre 5 des approches qui utilisent une attache aux données non locale pondérée par les poids des NL-means dans un contexte de super-résolution [Protter 2009, d'Angelo 2011] ou d'inpainting [Mignotte 2008]. L'attache aux données localement adaptative que nous proposons pourrait elle aussi s'adapter à des problèmes plus larges de super-résolution, d'inpainting ou de déconvolution.

De plus, l'étude d'autres régularisations que la variation totale, par exemple l'utilisation de dictionnaires de patchs dans l'approche adaptative pourrait être une extension intéressante à poursuivre.

Une autre approche à envisager serait de développer une méthode de débruitage non paramétrique. De façon analogue à ce qui a été proposé pour l'estimation de bruit, il serait intéressant pour cela de définir une mesure de similarité entre patchs indépendante de la distribution du bruit et des intensités des pixels, en se basant plutôt sur l'information géométrique.

Recalage multimodal. La métrique de recalage proposée se base sur l'alignement des contours présents dans les deux modalités. Il serait donc intéressant d'améliorer l'étape de détection des contours [Cao 2008] pour rendre l'algorithme plus robuste, ou plus léger à optimiser. On pourrait également envisager une approche multi-résolution qui permettrait de combiner efficacité d'implémentation et précision du recalage.

Si l'approche proposée a été validée dans le cadre du recalage multimodal des images optiques et infrarouge, l'étude de la faisabilité du modèle dans un contexte plus général pour l'adapter au domaine d'imagerie médicale (où les données sont éventuellement des volumes 3D) ou de télédétection, ou pour recalage des données de nature différente (par exemple des nuages de points) serait à envisager.

Liste des publications

Papiers de journaux

1. C. Sutour, C.-A. Deledalle et J.-F. Aujol. Adaptive regularization of the NL-means: Application to image and video denoising. *IEEE Transactions on image processing*, 23(8):3506-3521, 2014.
2. C. Sutour, J.-F. Aujol, C.-A. Deledalle et B.D. De-Senneville. Edge-based multi-modal registration and application for night vision devices. *Journal of Mathematical Imaging and Vision*, pages 1–20, 2015.
3. C. Sutour, C.-A. Deledalle et J.-F. Aujol. Estimation of the noise level function based on a non-parametric detection of homogeneous image regions. *Soumis à Siam Journal on Imaging Sciences*, 2015.

Papiers de conférence

1. C. Sutour, J.-F. Aujol et C.-A. Deledalle. TV-NL : Une coopération entre les NL-means et les méthodes variationnelles. *GRETSI*, 2013.
2. C. Sutour, J.-F. Aujol, C.-A. Deledalle et J.-P. Domenger. Adaptive regularization of the NL-means for video denoising. *International Conference on Image Processing (ICIP)*, pages 2704–2708, 2014.
3. C. Sutour, C.-A. Deledalle et J.-F. Aujol. Estimation of the noise level function based on a non-parametric detection of homogeneous image regions. *Soumis à International Conference on Image Processing (ICIP)*, 2015.
4. C. Sutour, C.-A. Deledalle et J.-F. Aujol. Estimation du niveau de bruit par la détection non paramétrique de zones homogènes. *Soumis au GRETSI*, 2015.

Bibliographie

- [Allaire 2008] S. Allaire, J. Kim, S. Breen, D. Jaffray et V. Pekar. Full orientation invariance and improved feature selectivity of 3d sift with application to medical image analysis. *In IEEE Comput. Vis. and Pattern Recognition (CVPR)*, pages 1–8, 2008. (Cité en page 189.)
- [Alliney 1992] S. Alliney. Digital filters as absolute norm regularizers. *IEEE Trans. Signal Process.*, 40(6):1548–1562, 1992. (Cité en page 117.)
- [Anscombe 1948] F. Anscombe. The transformation of poisson, binomial and negative-binomial data. *Biometrika*, pages 246–254, 1948. (Cité en pages 54 et 64.)
- [Anthoine 2012] S. Anthoine, J.F. Aujol, Y. Boursier et C. Mélot. Some proximal methods for poisson intensity CBCT and PET. *Inverse Problems and Imaging*, 6(4):565–598, 2012. (Cité en pages 14, 17, 42 et 123.)
- [Argenti 2002] F. Argenti et L. Alparone. Speckle removal from sar images in the undecimated wavelet domain. *IEEE Trans. Geosci. Remote Sens.*, 40(1):2363–2374, 2002. (Cité en page 121.)
- [Argenti 2006] F. Argenti, T. Bianchi et L. Alparone. Multiresolution map despeckling of sar images based on locally adaptive generalized gaussian pdf modeling. *IEEE Trans. Image Process.*, 15(11):3385–3399, 2006. (Cité en page 121.)
- [Arias 2012] P. Arias, V. Caselles et G. Facciolo. Analysis of a variational framework for exemplar-based image inpainting. *SIAM Journal Multiscale Model. Simul.*, 10(2):473–514, 2012. (Cité en page 152.)
- [Attouch 2013] H. Attouch, J. Bolte et B. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. *Mathematical Programming*, 137(1–2):91–129, 2013. (Cité en page 124.)
- [Aubert 2006] G. Aubert et P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations*, 147. Springer, 2006. (Cité en page 112.)
- [Aubert 2008] G. Aubert et J.-F. Aujol. A variational approach to removing multiplicative noise. *SIAM Journal Applied Math.*, 68(4):925–946, 2008. (Cité en pages 14, 42, 123 et 124.)
- [Aujol 2003] J.-F. Aujol, G. Aubert et L. Blanc-Féraud. Wavelet-based level set evolution for classification of textured images. *IEEE Trans. Image Process.*, 12(12):1634–1641, 2003. (Cité en page 196.)
- [Aujol 2005] J.-F. Aujol, G. Aubert, L. Blanc-Féraud et A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1):71–88, 2005. (Cité en page 124.)

- [Aujol 2009] J.-F. Aujol. Some first-order algorithms for total variation based image restoration. *Journal of Mathematical Imaging and Vision*, 34(3):307–327, 2009. (Cit  en page 124.)
- [Awate 2006] S. Awate et R. Whitaker. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):364–376, 2006. (Cit  en page 152.)
- [Azzabou 2007] N. Azzabou, N. Paragios et F. Guichard. Image denoising based on adapted dictionary computation. In *IEEE Int. Conf. Image Process. (ICIP)*, 3:109–112, 2007. (Cit  en page 134.)
- [Azzari 2014a] L. Azzari et A. Foi. Gaussian-cauchy mixture modeling for robust signal-dependent noise estimation. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5357–5361, 2014. (Cit  en page 64.)
- [Azzari 2014b] L. Azzari et A. Foi. Indirect estimation of signal-dependent noise with nonadaptive heterogeneous samples. *IEEE Trans. Image Process.*, 23(8):34–59, 2014. (Cit  en page 64.)
- [Bardsley 2009] J. Bardsley et A. Luttmann. Total variation-penalized poisson likelihood estimation for ill-posed problems. *Advances in Computational Mathematics*, 31(1–3):35–59, 2009. (Cit  en page 121.)
- [Barnes 2009] C. Barnes, E. Shechtman, A. Finkelstein et D. Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graphics*, 28(3):24–33, 2009. (Cit  en page 134.)
- [Bay 2008] H. Bay, A. Ess, T. Tuytelaars et L. Van Gool. SURF: Speeded up robust features. *Computer vision and image understanding*, 110(3):346–359, 2008. (Cit  en page 189.)
- [Beaurepaire 1997] L. Beaurepaire, K. Chehdi et B. Vozel. Identification of the nature of noise and estimation of its statistical parameters by analysis of local histograms. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 4:2805–2808, 1997. (Cit  en page 64.)
- [Beck 2009a] A. Beck et M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, 18(11):2419–2434, 2009. (Cit  en page 120.)
- [Beck 2009b] A. Beck et M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal Imaging Sci.*, 2(1):183–202, 2009. (Cit  en page 120.)
- [Beck 2010] A. Beck et M. Teboulle. Gradient-based algorithms with applications to signal recovery problems. In *Convex Optimization in Signal Processing and Communications*, pages 33–88. D. Palomar and Y. Eldar Eds., 2010. (Cit  en pages 119, 120, 141 et 160.)
- [Bougleux 2007] S. Bougleux, A. Elmoataz et M. Melkemi. Discrete regularization on weighted graphs for image and mesh filtering. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 128–139. Springer, 2007. (Cit  en page 152.)

- [Boulanger 2007] J. Boulanger, C. Kervrann et P. Bouthemy. Space-time adaptation for patch-based image sequence restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):1096–1102, 2007. (Cité en pages 171, 172 et 178.)
- [Boulanger 2010] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita et J. Salamero. Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Trans. Med. Imag.*, 29(2):442–454, 2010. (Cité en pages 55, 64 et 172.)
- [Brown 1992] L.G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992. (Cité en page 187.)
- [Brox 2007] T. Brox et D. Cremers. Iterated nonlocal means for texture restoration. *In Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 13–24. Springer, 2007. (Cité en pages 128 et 152.)
- [Brox 2008] T. Brox, O. Kleinschmidt et D. Cremers. Efficient nonlocal means for denoising of textural patterns. *IEEE Trans. Image Process.*, 17(7):1083–1092, 2008. (Cité en page 134.)
- [Brox 2011] T. Brox et J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(3):500–513, 2011. (Cité en page 189.)
- [Buades 2005a] A. Buades, B. Coll et J.-M. Morel. Denoising image sequences does not require motion compensation. *IEEE Int. Conf. Advanced Video and Signal-Based Surveillance (AVSS)*, pages 70–74, 2005. (Cité en pages 14, 42, 171 et 178.)
- [Buades 2005b] A. Buades, B. Coll et J.-M. Morel. A review of image denoising algorithms, with a new one. *SIAM Journal Multiscale Model. Simul.*, 4(2):490–53, 2005. (Cité en pages 14, 42, 63, 65, 84, 126, 128, 136 et 192.)
- [Buades 2006] A. Buades, B. Coll et J.-M. Morel. Image enhancement by non-local reverse heat equation. *Technical report 22, CMLA, ENS-Cachan, France*, 2006. (Cité en page 152.)
- [Buades 2008] A. Buades, B. Coll et J.-M. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, 2008. (Cité en pages 18, 111, 171 et 176.)
- [Candes 1999] E. Candes et D.L. Donoho. Curvelets: A surprisingly effective nonadaptive representation for objects with edges. DTIC Document, 1999. (Cité en page 115.)
- [Cao 2008] F. Cao, J. Lisani, J.-M. Morel, P. Musé et F. Sur. A theory of shape identification. Springer, 2008. (Cité en page 235.)
- [Chambolle 1997] A. Chambolle et P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997. (Cité en page 149.)
- [Chambolle 2004] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1–2):89–97, 2004. (Cité en pages 32, 117, 118, 124 et 139.)

- [Chambolle 2011] A. Chambolle et T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. (Cit  en pages 12, 17, 80, 81, 120, 121, 143 et 157.)
- [Chan 1998] T. Chan et C. Wong. Total variation blind deconvolution. *IEEE Trans. Image Process.*, 7(3):370–375, 1998. (Cit  en page 124.)
- [Chan 2000] T. Chan, A. Marquina et P. Mulet. High-order total variation-based image restoration. *SIAM Journal Sci. Comput.*, 22(2):503–516, 2000. (Cit  en page 124.)
- [Chan 2005a] T. Chan et S. Esedoglu. Aspects of total variation regularized l1 function approximation. *SIAM Journal Applied Math.*, 65(5):1817–1837, 2005. (Cit  en page 117.)
- [Chan 2005b] T. Chan, A. Yip et F. Park. Simultaneous total variation image inpainting and blind deconvolution. *Int. Journal Imaging Syst. Technol.*, 15(1):92–102, 2005. (Cit  en page 124.)
- [Chatterjee 2012] P. Chatterjee et P. Milanfar. Patch-based near-optimal image denoising. *IEEE Trans. Image Process.*, 21(4):1635–1649, 2012. (Cit  en page 135.)
- [Chehdi 1992] K. Chehdi et M. Sabri. A new approach to identify the nature of the noise affecting an image. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 3:285–288, 1992. (Cit  en page 64.)
- [Chen 1998] S. Chen, D.L. Donoho et M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal Sci. Comput.*, 20(1):33–61, 1998. (Cit  en page 116.)
- [Chen 2007] Y. Chen et M. Das. An automated technique for image noise identification using a simple pattern classification approach. In *IEEE Int. Midwest Symp. Circuits and Systems (MWSCAS)*, pages 819–822, 2007. (Cit  en page 64.)
- [Cho 2008] E. Cho et M.J. Cho. Variance of sample variance. In *Joint Statistical Meetings, Section on Survey Research Methods*, pages 1291–1293, 2008. (Cit  en page 76.)
- [Collins 2001] M. Collins, S. Dasgupta et R. Schapire. A generalization of principal components analysis to the exponential family. In *Advances in Neural Inf. Process. Syst. (NIPS)*, pages 617–624, 2001. (Cit  en page 155.)
- [Colom 2013] M. Colom et A. Buades. Analysis and extension of the percentile method, estimating a noise curve from a single image. *Image Process. On Line*, 3:332–359, 2013. (Cit  en pages 85, 86 et 87.)
- [Colom 2014a] M. Colom et A. Buades. Analysis and extension of the pca method, estimating a noise curve from a single image. *Image Process. On Line*, 2014. (Cit  en pages 64, 85, 86 et 87.)
- [Colom 2014b] M. Colom, M. Lebrun, A. Buades et J.-M. Morel. A non-parametric approach for the estimation of intensity-frequency dependent noise. In *IEEE*

- Int. Conf. Image Process. (ICIP)*, pages 4261–4265. IEEE, 2014. (Cité en pages 64, 85 et 86.)
- [Combettes 2005] P. Combettes et V. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Journal Multiscale Model. Simul.*, 4(4):1168–1200, 2005. (Cité en pages 17, 119, 140 et 160.)
- [Combettes 2010] P. Combettes et J.-C. Pesquet. Proximal splitting methods in signal processing, 2010. (Cité en page 141.)
- [Cornillon 2013] P.-A. Cornillon, N. Hengartner, N. Jegou et E. Matzner-Løber. Iterative bias reduction: a comparative study. *Statistics and Computing*, 23(6):777–791, 2013. (Cité en page 146.)
- [Coupé 2006] P. Coupé, P. Yger et C. Barillot. Fast non local means denoising for 3D MR images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 33–40. Springer, 2006. (Cité en page 134.)
- [Coupé 2008] P. Coupé, P. Hellier, C. Kervrann et C. Barillot. Bayesian non local means-based speckle filtering. In *IEEE Int. Symp. Biomedical Imaging (ISBI)*, pages 1291–1294, 2008. (Cité en page 133.)
- [Dabov 2007a] K. Dabov, A. Foi et K. Egiazarian. Video denoising by sparse 3D transform-domain collaborative filtering. In *European Signal Processing Conf. (EUSIPCO)*, 1(2):145–149, 2007. (Cité en pages 174 et 176.)
- [Dabov 2007b] K. Dabov, A. Foi, V. Katkovich et K. Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, 2007. (Cité en pages 131, 135 et 163.)
- [d’Angelo 2011] E. d’Angelo et P. Vanderghenst. Fully non-local super-resolution via spectral hashing. *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1137–1140, 2011. (Cité en pages 145, 146, 154, 155, 163 et 235.)
- [Darbon 2005] J. Darbon et M. Sigelle. A fast and exact algorithm for total variation minimization. In *Pattern recognition and image analysis*, pages 351–359. Springer, 2005. (Cité en page 124.)
- [Darbon 2008] J. Darbon, A. Cunha, T.F. Chan, S. Osher et G.J. Jensen. Fast non-local filtering applied to electron cryomicroscopy. *IEEE Int. Symp. Biomedical Imaging (ISBI)*, pages 1331–1334, 2008. (Cité en pages 17, 134, 156, 177 et 178.)
- [Daubechies 1992] I. Daubechies *et al.* Ten lectures on wavelets, 61. SIAM, 1992. (Cité en page 115.)
- [Dauwe 2008] A. Dauwe, B. Goossens, H. Luong et W. Philips. A fast non-local image denoising algorithm. In *SPIE-IS&T Electronic Imaging*, 6812(10):1–8. International Society for Optics and Photonics, 2008. (Cité en page 134.)
- [De Stefano 2004] A. De Stefano, P.R. White et W.B. Collis. Training methods for image noise level estimation on wavelet components. *EURASIP Journal on Applied Signal Processing*, pages 2400–2407, 2004. (Cité en page 63.)

- [Deledalle 2009] C.-A. Deledalle, L. Denis et F. Tupin. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Trans. Image Process.*, 18(12):2661–2672, 2009. (Cité en pages 132, 133 et 134.)
- [Deledalle 2010] C.-A. Deledalle, F. Tupin et L. Denis. Poisson NL-means: Unsupervised non local means for poisson noise. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 801–804, 2010. (Cité en pages 14 et 42.)
- [Deledalle 2011] C.-A. Deledalle, V. Duval et J. Salmon. Non-local methods with shape-adaptive patches (NLM-SAP). *Journal of Mathematical Imaging and Vision*, pages 1–18, 2011. (Cité en pages 130, 134 et 136.)
- [Deledalle 2012] C.-A. Deledalle, L. Denis et F. Tupin. How to compare noisy patches? Patch similarity beyond Gaussian noise. *International Journal of Computer Vision*, 99(1):86–102, 2012. (Cité en pages 128, 132 et 133.)
- [Deledalle 2014] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber et M. Jäger. NL-SAR: a unified non-local framework for resolution-preserving (pol)(in) SAR denoising. *IEEE Trans. Geosci. Remote Sens.*, 53(4):2021–2038, 2014. (Cité en pages 14, 42, 133 et 147.)
- [Deledalle 2015] C.-A. Deledalle, N. Papadakis et J. Salmon. On debiasing restoration algorithms: applications to total-variation and nonlocal-means. *arXiv preprint*, 2015. (Cité en page 132.)
- [Dellinger 2012] F. Dellinger, J. Delon, Y. Gousseau, J. Michel et F. Tupin. SAR-SIFT: A SIFT-like algorithm for applications on SAR images. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 3478–3481. IEEE, 2012. (Cité en page 189.)
- [Delon 2012] J. Delon et A. Desolneux. A patch-based approach for removing mixed gaussian-impulse noise. *SIAM Journal Imaging Sci.*, 6(2):1140–1174, 2012. (Cité en pages 14, 42, 128, 132 et 133.)
- [Dippel 2002] S. Dippel, M. Stahl, R. Wiemker et T. Blaffert. Multiscale contrast enhancement for radiographies: Laplacian pyramid versus fast wavelet transform. *IEEE Trans. Med. Imag.*, 21(4):343–353, 2002. (Cité en pages 40 et 233.)
- [Do 2003] M. Do et M. Vetterli. Framing pyramids. *IEEE Trans. Signal Process.*, 51(9):2329–2342, 2003. (Cité en page 115.)
- [Dobson 1996] D. Dobson et F. Santosa. Recovery of blocky images from noisy and blurred data. *SIAM Journal Applied Math.*, 56(4):1181–1198, 1996. (Cité en page 125.)
- [Donoho 1994] D.L. Donoho et J.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. (Cité en pages 115 et 149.)
- [Donoho 1995] D.L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory*, 41(3):613–627, 1995. (Cité en pages 63, 85, 86, 87 et 115.)
- [Donoho 1999] D.L. Donoho *et al.* Wedgelets: Nearly minimax estimation of edges. *The Annals of Statistics*, 27(3):859–897, 1999. (Cité en page 115.)

- [Duval 2009] V. Duval, J.-F. Aujol et Y. Gousseau. The TVL1 model: a geometric point of view. *SIAM Journal Multiscale Model. Simul.*, 8(1):154–189, 2009. (Cité en page 117.)
- [Duval 2011] V. Duval, J.-F. Aujol et Y. Gousseau. A biais-variance approach for the non-local means. *SIAM Journal Imaging Sci.*, 4(2):760–788, 2011. (Cité en pages 129, 133 et 134.)
- [Elad 2006] M. Elad et M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745, 2006. (Cité en page 116.)
- [Fainstain 2010] E. Fainstain et Y. Lavi. Smear correction in a digital camera, 2010. US Patent 7,693,347. (Cité en pages 40 et 234.)
- [Figueiredo 2009] M. Figueiredo et J. Bioucas-Dias. Deconvolution of poissonian images using variable splitting and augmented lagrangian optimization. In *IEEE Statistical Signal Processing (SSP)*, pages 733–736, 2009. (Cité en page 121.)
- [Fischler 1981] M.A. Fischler et R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. (Cité en page 189.)
- [FitzHenry 1991] P. FitzHenry et L. Scanlan. Image contrast enhancement technique, 1991. US Patent 5,063,607. (Cité en pages 40 et 233.)
- [Foi 2008] A. Foi, M. Trimeche, V. Katkovnik et K. Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans. Image Process.*, 17(10):1737–1754, 2008. (Cité en pages 64, 85, 86 et 87.)
- [Freeman 2002] W. Freeman, T. Jones et E. Pasztor. Example-based super-resolution. *IEEE Comput. Graph. Appl.*, 22(2):56–65, 2002. (Cité en page 63.)
- [Gasser 1986] T. Gasser, L. Sroka et C. Jennen-Steinmetz. Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73(3):625–633, 1986. (Cité en page 64.)
- [Genin 2012] L. Genin, F. Champagnat et G. Le Besnerais. Background first-and second-order modeling for point target detection. *Journal of the Optical Society of America*, 51(31):7701–7713, 2012. (Cité en page 49.)
- [Gilboa 2006] G. Gilboa, N. Sochen et Y. Zeevi. Variational denoising of partly textured images by spatially varying constraints. *IEEE Trans. Image Process.*, 15(8):2281–2289, 2006. (Cité en pages 117, 125, 136 et 146.)
- [Gilboa 2007] G. Gilboa et S. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Journal Multiscale Model. Simul.*, 6(2):595–630, 2007. (Cité en page 154.)

- [Gilboa 2008] G. Gilboa et S. Osher. Nonlocal operators with applications to image processing. *SIAM Journal Multiscale Model. Simul.*, 7(3):1005–1028, 2008. (Cité en pages 145 et 152.)
- [Goossens 2008] B. Goossens, H. Luong, A. Pizurica et W. Philips. An improved non-local denoising algorithm. *Local and Non-Local Approximation (LNLA)*, pages 143–156, 2008. (Cité en pages 128 et 131.)
- [Gousseau 2001] Y. Gousseau et J.-M. Morel. Are natural images of bounded variation? *SIAM Journal Mathematical Anal.*, 33(3):634–648, 2001. (Cité en page 125.)
- [Guichard 1998] F. Guichard et F. Malgouyres. Total variation based interpolation. *In European Signal Processing Conf. (EUSIPCO)*, 3:1741–1744, 1998. (Cité en page 124.)
- [Haber 2006] E. Haber et J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 46:292–299, 2006. (Cité en pages 21, 191, 204 et 205.)
- [Hall 1997] D. Hall et J. Llinas. An introduction to multisensor data fusion. *Proc. IEEE*, 85(1):6–23, 1997. (Cité en pages 40 et 232.)
- [Hansen 1993] P. Hansen et D. O’Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal Sci. Comput.*, 14(6):1487–1503, 1993. (Cité en pages 117 et 125.)
- [Harmany 2009] Z. Harmany, R. Marcia et R. Willett. Sparse Poisson intensity reconstruction algorithms. *In IEEE Statistical Signal Processing (SSP)*, pages 634–637, 2009. (Cité en page 121.)
- [Hartley 2003] R. Hartley et A. Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003. (Cité en pages 22 et 206.)
- [He 2013] K. He, J. Sun et X. Tang. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(6):1397–1409, 2013. (Cité en page 232.)
- [Healey 1994] G. Healey et R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(3):267–276, 1994. (Cité en page 63.)
- [Heinrich 2012] M.P. Heinrich, M. Jenkinson, M. Bhushan, T. Martin, F.V. Gleeson, S.M. Brady et J.A. Schnabel. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423:1435, 2012. (Cité en pages 21, 192, 204 et 205.)
- [Hensel 2007] M. Hensel, T. Pralow et R. Grigat. Modeling and real-time estimation of signal-dependent noise in quantum-limited imaging. *In WSEAS Int. Conf. Signal Processing, Robotics and Automation (ISPRA)*, pages 183–191, 2007. (Cité en page 64.)
- [Hermosillo 2002] G. Hermosillo, C. Chefd’Hotel et O. Faugeras. Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50(3):329–343, 2002. (Cité en page 190.)

- [Horn 1981] B. Horn et B. Schunck. Determining optical flow. *In Technical symposium east*, pages 319–331. International Society for Optics and Photonics, 1981. (Cité en page 188.)
- [Huber 1964] P.J. Huber *et al.* Robust estimation of a location parameter. *The Annals of Statistics*, 35(1):73–101, 1964. (Cité en page 81.)
- [Irani 1998] M. Irani et P Anandan. Robust multi-sensor image alignment. *In Int. Conf. Computer Vision (ICCV)*, pages 959–966, 1998. (Cité en pages 21 et 191.)
- [Johnson 1995] C. Johnson et L. Owen. Image tube intensified electronic imaging. *Handbook of Optics*, 1(21):1–32, 1995. (Cité en page 55.)
- [Kendall 1938] M.G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1–2):81–93, 1938. (Cité en pages 9, 41, 65 et 68.)
- [Kendall 1945] M.G. Kendall. The treatment of ties in ranking problems. *Biometrika*, pages 239–251, 1945. (Cité en pages 9, 69 et 70.)
- [Kendall 1948] M.G. Kendall. Rank correlation methods. Griffin, 1948. (Cité en pages 10 et 70.)
- [Kervrann 2006] C. Kervrann et J. Boulanger. Optimal spatial adaptation for patch-based image denoising. *IEEE Trans. Image Process.*, 15(10):2866–2878, 2006. (Cité en pages 129, 130 et 136.)
- [Kervrann 2007] C. Kervrann, J. Boulanger et P. Coupé. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. *In Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 520–532. Springer, 2007. (Cité en pages 14, 15, 128, 133 et 134.)
- [Kervrann 2008] C. Kervrann et J. Boulanger. Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *International Journal of Computer Vision*, 79(1):45–69, 2008. (Cité en pages 134, 136, 146 et 163.)
- [Kindermann 2005] S. Kindermann, S. Osher et P. Jones. Deblurring and denoising of images by nonlocal functionals. *SIAM Journal Multiscale Model. Simul.*, 4(4):1091–1115, 2005. (Cité en pages 145 et 152.)
- [Knight 1966] W. Knight. A computer method for calculating kendall’s tau with ungrouped data. *Journal of the American Statistical Association*, 61(314):436–439, 1966. (Cité en page 75.)
- [Koenker 2005] R. Koenker. Quantile regression. Cambridge university press, 2005. (Cité en page 81.)
- [Kuan 1985] D. Kuan, A. Sawchuk, T. Strand et P. Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 2:165–177, 1985. (Cité en page 147.)
- [Le Pennec 2005] E. Le Pennec et S. Mallat. Sparse geometric image representations with bandelets. *IEEE Trans. Image Process.*, 14(4):423–438, 2005. (Cité en page 115.)

- [Le 2007] T. Le, R. Chartrand et T. Asaki. A variational approach to reconstructing images corrupted by poisson noise. *Journal of Mathematical Imaging and Vision*, 27(3):257–263, 2007. (Cité en page 121.)
- [Lebrun 2013] M. Lebrun, A. Buades et J.-M. Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal Imaging Sci.*, 6(3):1665–1688, 2013. (Cité en pages 85, 128 et 135.)
- [Lebrun 2014] M. Lebrun, M. Colom et J.-M. Morel. The noise-clinic: a universal blind denoising algorithm. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 2674–2678, 2014. (Cité en pages 85 et 86.)
- [Lebrun 2015] M. Lebrun, M. Colom et J.-M. Morel. The noise-clinic: a blind image denoising algorithm. *Image Process. On Line*, 5:1–54, 2015. (Cité en pages 64, 85, 86 et 87.)
- [Lee 1981] J.-S. Lee. Refined filtering of image noise using local statistics. *Computer Graphics and Image Processing*, 15(4):380–389, 1981. (Cité en page 147.)
- [Lefébure 2001] M. Lefébure et L. Cohen. Image registration, optical flow and local rigidity. *Journal of Mathematical Imaging and Vision*, 14(2):131–147, 2001. (Cité en page 188.)
- [Li 2013] S. Li, X. Kang et J. Hu. Image fusion with guided filtering. *IEEE Trans. Image Process.*, 22(7):2864–2875, 2013. (Cité en page 232.)
- [Liu 2006] C. Liu, W. Freeman, R. Szeliski et S. Kang. Noise estimation from a single image. In *IEEE Comput. Vis. and Pattern Recognition (CVPR)*, 1:901–908, 2006. (Cité en page 65.)
- [Liu 2010] C. Liu et W. Freeman. A high-quality video denoising algorithm based on reliable motion estimation. *European Conf. Computer Vision (ECCV)*, pages 706–719, 2010. (Cité en pages 20 et 178.)
- [Liu 2013] X. Liu, M. Tanaka et M. Okutomi. Single-image noise level estimation for blind denoising. *IEEE Trans. Image Process.*, 22(12):5226–5237, 2013. (Cité en page 63.)
- [Louchet 2011] C. Louchet et L. Moisan. Total variation as a local filter. *SIAM Journal Imaging Sci.*, 4(2):651–694, 2011. (Cité en pages 136, 145 et 154.)
- [Lowe 2004] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. (Cité en pages 21, 63 et 189.)
- [Lu 1997] J. Lu et M. Liou. A simple and efficient search algorithm for block-matching motion estimation. *IEEE Trans. Circuits Syst. Video Technol.*, 7(2):429–433, 1997. (Cité en page 40.)
- [M. 2006] Charest M., M. Elad et Peyman. Milanfar. A general iterative regularization framework for image denoising. In *Conf. Information Sciences and Systems*, pages 452–457, 2006. (Cité en page 146.)
- [Maes 1997] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal et P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.*, 16(2):187–198, 1997. (Cité en pages 21 et 190.)

- [Maggioni 2012] M. Maggioni, G. Boracchi, A. Foi et K. Egiazarian. Video denoising, deblocking and enhancement through separable 4-D nonlocal spatio-temporal transforms. *IEEE Trans. Image Process.*, 21(9):3952–3966, 2012. (Cité en pages 174 et 176.)
- [Mahmoudi 2005] M. Mahmoudi et G. Sapiro. Fast image and video denoising via non-local means of similar neighborhoods. *IEEE Signal Process. Lett.*, 12(12), 2005. (Cité en pages 134, 136 et 171.)
- [Mairal 2009] J. Mairal, F. Bach, J. Ponce, G. Sapiro et A. Zisserman. Non-local sparse models for image restoration. In *Int. Conf. Computer Vision (ICCV)*, pages 2272–2279, 2009. (Cité en page 135.)
- [Mäkitalo 2010] M. Mäkitalo, A. Foi, D. Fevralev et V. Lukin. Denoising of single-look SAR images based on variance stabilization and nonlocal filters. In *Int. Conf. Mathematical Methods in Electromagnetic Theory (MMET)*, pages 1–4, 2010. (Cité en page 163.)
- [Mäkitalo 2011] M. Mäkitalo et A. Foi. Optimal inversion of the anscombe transformation in low-count poisson image denoising. *IEEE Trans. Image Process.*, 20(1):99–109, 2011. (Cité en page 55.)
- [Mäkitalo 2012] M. Mäkitalo et A. Foi. Poisson-gaussian denoising using the exact unbiased inverse of the generalized anscombe transformation. *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1081–1084, 2012. (Cité en page 163.)
- [Mallat 1993] S. G Mallat et Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993. (Cité en page 116.)
- [Mallat 1999] S. Mallat. A wavelet tour of signal processing. Academic press, 1999. (Cité en page 115.)
- [Mallat 2008] S. Mallat. A wavelet tour of signal processing: the sparse way. Academic press, 2008. (Cité en page 115.)
- [Mann 1997] S. Mann et R. Picard. Video orbits of the projective group a simple approach to featureless estimation of parameters. *IEEE Trans. Image Process.*, 6(9):1281–1295, 1997. (Cité en page 215.)
- [Mignotte 2008] M. Mignotte. A non-local regularization strategy for image deconvolution. *Pattern Recognition Lett.*, 29:2206–2212, 2008. (Cité en pages 145, 152 et 235.)
- [Mikolajczyk 2005] K. Mikolajczyk et C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630, 2005. (Cité en page 189.)
- [Milanfar 2013] P. Milanfar. A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE Signal Process. Mag.*, 30(1):106–128, 2013. (Cité en page 146.)

- [Mittal 2012] A. Mittal, A. Moorthy et A. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.*, 21(12):4695–4708, 2012. (Cité en page 111.)
- [Moorthy 2011] A. Moorthy et A. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Trans. Image Process.*, 20(12):3350–3364, 2011. (Cité en page 111.)
- [Muresan 2003] D. Muresan et T. Parks. Adaptive principal components and image denoising. *In IEEE Int. Conf. Image Process. (ICIP)*, 2003. (Cité en page 115.)
- [Murtagh 1995] F. Murtagh, J.-L. Starck et A. Bijaoui. Image restoration with noise suppression using a multiresolution support. *Astronomy and Astrophysics Suppl. Ser.*, 112:179–189, 1995. (Cité en page 55.)
- [Odobez 1995] J.-M. Odobez et P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, 1995. (Cité en pages 179 et 224.)
- [Olsen 1993] S. Olsen. Estimation of noise in images: an evaluation. *Graphical Models and Image Processing (CVGIP)*, 55(4):319–323, 1993. (Cité en page 63.)
- [Orchard 2008] J. Orchard, M. Ebrahimi et A. Wong. Efficient nonlocal-means denoising using the svd. *In IEEE Int. Conf. Image Process. (ICIP)*, pages 1732–1735, 2008. (Cité en page 134.)
- [Osher 2003] S. Osher, A. Solé et L. Vese. Image decomposition and restoration using total variation minimization and the h1. *SIAM Journal Multiscale Model. Simul.*, 1(3):349–370, 2003. (Cité en page 124.)
- [Paganelli 2012] C. Paganelli, M. Peroni, F. Pennati, G. Baroni, P. Summers, M. Bellomi et M. Riboldi. Scale invariant feature transform as feature tracking method in 4D imaging: A feasibility study. *In IEEE Int. Conf. Engineering in Medicine and Biology Society (EMBS)*, pages 6543–6546, 2012. (Cité en page 189.)
- [Pati 1993] Y. Pati, R. Rezaifar et P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. *In IEEE Asilomar Conf. Signals, Systems and Computers*, pages 40–44, 1993. (Cité en page 116.)
- [Perona 1990] P. Perona et J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990. (Cité en page 113.)
- [Peyré 2011] G. Peyré, S. Bogleux et L. D. Cohen. Non-local regularization of inverse problems. *Inverse Problems and Imaging*, 5(2):511–530, 2011. (Cité en pages 145 et 154.)
- [Pizarro 2010] L. Pizarro, P. Mrázek, S. Didas, S. Grewenig et J. Weickert. Generalised nonlocal image smoothing. *International Journal of Computer Vision*, 90(1):62–87, 2010. (Cité en page 152.)

- [Pluim 2000] J.P.W. Pluim, J.B.A. Maintz et M.A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Med. Imag.*, 19(8):809–814, 2000. (Cit  en pages 192, 204 et 205.)
- [Pluim 2003] J. Pluim, J. Maintz et M. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Trans. Med. Imag.*, 22(8):986–1004, 2003. (Cit  en page 190.)
- [Pollard 1991] D. Pollard. Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(2):186–199, 1991. (Cit  en page 80.)
- [Polzehl 2006] J. Polzehl et V. Spokoiny. Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135(3):335–362, 2006. (Cit  en pages 132 et 155.)
- [Ponomarenko 2010] N. Ponomarenko, V. Lukin, K. Egiazarian et J. Astola. A method for blind estimation of spatially correlated noise characteristics. In *SPIE-IS&T Electronic Imaging*, pages 7532(8):1–12. International Society for Optics and Photonics, 2010. (Cit  en page 85.)
- [Portilla 2003] J. Portilla, V. Strela, M. Wainwright et E. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Image Process.*, 12(11):1338–1351, 2003. (Cit  en page 115.)
- [Portilla 2004] J. Portilla. Blind non-white noise removal in images using gaussian scale mixtures in the wavelet domain. In *Benelux Signal Processing Symp.*, 2004. (Cit  en page 64.)
- [Pratt 1978] W.K. Pratt. Digital image processing. John Wiley and Sons, Inc., NY, 1978. (Cit  en pages 21, 190 et 205.)
- [Protter 2009] M. Protter, M. Elad, H. Takeda et P. Milanfar. Generalizing the non-local means to super-resolution reconstruction. *IEEE Trans. Image Process.*, 18(1):36–51, 2009. (Cit  en pages 145, 146, 154 et 235.)
- [Pyatykh 2013] S. Pyatykh, J.n Hesser et L. Zheng. Image noise level estimation by principal component analysis. *IEEE Trans. Image Process.*, 22(2):687–699, 2013. (Cit  en pages 63 et 85.)
- [Pyatykh 2014] S. Pyatykh et J. Hesser. Image sensor noise parameter estimation by variance stabilization and normality assessment. *IEEE Trans. Image Process.*, 23(9):3990–3998, 2014. (Cit  en page 64.)
- [Raginsky 2010] M. Raginsky, R. Willett, Z. Harmany et R. Marcia. Compressed sensing performance bounds under poisson noise. *IEEE Trans. Signal Process.*, 58(8):3990–4002, 2010. (Cit  en page 121.)
- [Raguet 2013] H. Raguet, J. Fadili et G. Peyr . Generalized forward-backward splitting. *SIAM Journal Imaging Sci.*, 6(3):1199–1226, 2013. (Cit  en pages 120, 124 et 141.)
- [Ramani 2008] S. Ramani, T. Blu et M. Unser. Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms.

- IEEE Trans. Image Process.*, 17(9):1540–1554, 2008. (Cité en pages 117 et 125.)
- [Rank 1999] K. Rank, M. Lendl et R. Unbehauen. Estimation of image noise variance. *Proc. Vision, Image, and Signal Processing*, 146(2):80–84, 1999. (Cité en page 63.)
- [Roche 1998] A. Roche, G. Malandain, X. Pennec et N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. *In Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 1115–1124. 1998. (Cité en page 190.)
- [Rockafellar 1970] R.T. Rockafellar. Convex analysis. Princeton, NJ, 1970. (Cité en page 137.)
- [Roshni 2008] V. Roshni et K. Revathy. Using mutual information and cross-correlation as metrics for registration of images. *Journal of Theoretical and Applied Information Technology*, 4(6), 2008. (Cité en pages 21 et 190.)
- [Roudot 2013] P. Roudot, C. Kervrann, J. Boulanger et F. Waharte. Noise modeling for intensified camera in fluorescence imaging: Application to image denoising. *In IEEE Int. Symp. Biomedical Imaging (ISBI)*, pages 600–603, 2013. (Cité en page 59.)
- [Rudin 1992] L. Rudin, S. Osher et E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1):259–268, 1992. (Cité en pages 14, 42, 116 et 136.)
- [Saad 2012] M. Saad, A. Bovik et C. Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE Trans. Image Process.*, 21(8):3339–3352, 2012. (Cité en page 111.)
- [Salmon 2010a] J. Salmon. On two parameters for denoising with non-local means. *IEEE Signal Process. Lett.*, 17(3):269–272, 2010. (Cité en page 130.)
- [Salmon 2010b] J. Salmon et Y. Strozecski. From patches to pixels in non-local methods: Weighted-average reprojection. *In IEEE Int. Conf. Image Process. (ICIP)*, 26, 2010. (Cité en page 134.)
- [Shin 2005] D. Shin, R. Park, S. Yang et J. Jung. Block-based noise estimation using adaptive gaussian filtering. *IEEE Trans. Consum. Electron.*, 51(1):218–226, 2005. (Cité en page 63.)
- [Smith 1997] S. Smith et J. Brady. SUSAN — a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997. (Cité en page 126.)
- [Spontón 2012] H. Spontón et J. Cardelino. A review of classic edge detectors. *Image Process. On Line*, 2012. (Cité en page 193.)
- [Stark 2000] A. Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.*, 9(5):889–896, 2000. (Cité en pages 49 et 233.)

- [Stein 1981] C. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, pages 1135–1151, 1981. (Cit  en page 134.)
- [Strong 1997] D. Strong, P. Blomgren et T. Chan. Spatially adaptive local-feature-driven total variation minimizing image restoration. *In Optical Science, Engineering and Instrumentation*, pages 222–233. International Society for Optics and Photonics, 1997. (Cit  en page 117.)
- [Strong 2003] D. Strong et T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):165–187, 2003. (Cit  en page 124.)
- [Sun 2004] Y. Sun, M.-P. Jolly et J.M. F. Moura. Integrated registration of dynamic renal perfusion MR images. *In IEEE Int. Conf. Image Process. (ICIP)*, pages 1923–1926, 2004. (Cit  en pages 21, 43, 191, 193, 194 et 205.)
- [Sutour 2013] C. Sutour, J.-F. Aujol et C.-A. Deledalle. TV-NL : Une coop ration entre les NL-means et les m thodes variationnelles. *In GRETSI*, 2013. (Cit  en page 43.)
- [Sutour 2014a] C. Sutour, J.-F. Aujol, C.-A. Deledalle et J.-P. Domenger. Adaptive regularization of the NL-means for video denoising. *In IEEE Int. Conf. Image Process. (ICIP)*, pages 2704–2708, 2014. (Cit  en page 43.)
- [Sutour 2014b] C. Sutour, C.-A. Deledalle et J.-F. Aujol. Adaptive regularization of the NL-means: Application to image and video denoising. *IEEE Trans. Image Process.*, 23(8):3506–3521, 2014. (Cit  en pages 43, 63 et 130.)
- [Sutour 2015a] C. Sutour, J.-F. Aujol, C.-A. Deledalle et B.D. De-Senneville. Edge-based multi-modal registration and application for night vision devices. *Journal of Mathematical Imaging and Vision*, pages 1–20, 2015. (Cit  en page 43.)
- [Sutour 2015b] C. Sutour, C.-A. Deledalle et J.-F. Aujol. Estimation of the noise level function based on a non-parametric detection of homogeneous image regions. *Submitted*, 2015. (Cit  en page 41.)
- [Tai 2008] S. Tai et S. Yang. A fast method for image noise estimation using laplacian operator and adaptive edge detection. *In Int. Symp. Communications, Control and Signal Processing (ISCCSP)*, pages 1077–1081, 2008. (Cit  en page 63.)
- [Talebi 2013] H. Talebi, X. Xhu et P. Milanfar. How to SAIF-ly boost denoising performance. *IEEE Trans. Image Process.*, 22(4):1470–1485, 2013. (Cit  en pages 134, 136, 146 et 163.)
- [Theil 1992] H. Theil. A rank-invariant method of linear and polynomial regression analysis. *In Henri Theil Contributions to Economics and Econometrics*, pages 345–381. Springer, 1992. (Cit  en page 81.)
- [Tibshirani 1996] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996. (Cit  en page 115.)

- [Tikhonov 1977] A. Tikhonov, V. Arsenin et F. John. Solutions of ill-posed problems. Winston Washington, DC, 1977. (Cit  en page 113.)
- [Toews 2013] M. Toews, L. Z lle et W.M. Wells III. Feature-based alignment of volumetric multi-modal images. *In Information Processing in Medical Imaging*, pages 25–36. Springer, 2013. (Cit  en page 189.)
- [Tomasi 1998] C. Tomasi et R. Manduchi. Bilateral filtering for gray and color images. *In Int. Conf. Computer Vision (ICCV)*, pages 839–846, 1998. (Cit  en page 126.)
- [Tukey 1977] J. Tukey. Exploratory data analysis. 1977. (Cit  en page 146.)
- [Van-De-Ville 2009] D. Van-De-Ville et M. Kocher. Sure-based non-local means. *IEEE Signal Process. Lett.*, 16(11):973–976, 2009. (Cit  en page 129.)
- [Van-de Ville 2011] D. Van-de Ville et M. Kocher. Non-local means with dimensionality reduction and SURE-based parameter selection. *IEEE Trans. Image Process.*, 20(9):2683–2690, 2011. (Cit  en pages 134 et 136.)
- [Verdoliva 2014] L. Verdoliva, D. Amitrano, R. Gaetano, G. Ruello et G. Poggi. Sar despeckling guided by an optical image. *In IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 3698–3701, 2014. (Cit  en page 233.)
- [Viola 1997] P. Viola et W.M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997. (Cit  en pages 21, 190 et 205.)
- [Wang 2004] Z. Wang, A. Bovik, H. Sheikh et E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. (Cit  en pages 111 et 162.)
- [Weiss 2009] P. Weiss, L. Blanc-F raud et G. Aubert. Efficient schemes for total variation minimization under constraints in image processing. *SIAM Journal Sci. Comput.*, 31(3):2047–2080, 2009. (Cit  en page 124.)
- [Wexler 2007] Y. Wexler, E. Shechtman et M. Irani. Space-time completion of video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(3):463–476, 2007. (Cit  en pages 18, 171 et 172.)
- [Willett 2010] R. Willett, Z. Harmany et M. Poisson image reconstruction with total variation regularization. *In IEEE Int. Conf. Image Process. (ICIP)*, pages 4177–4180, 2010. (Cit  en page 121.)
- [Williams Jr 1992] G. Williams Jr. High-performance lltv ccd camera for night-time pilotage. *In SPIE-IS&T Electronic Imaging*, pages 14–32. International Society for Optics and Photonics, 1992. (Cit  en page 55.)
- [Wu 1982] F. Wu. The potts model. *Reviews of modern physics*, 54(1):235–268, 1982. (Cit  en page 117.)
- [Xie 2002a] H. Xie, L. Pierce et F. Ulaby. Sar speckle reduction using wavelet denoising and markov random field modeling. *IEEE Trans. Geosci. Remote Sens.*, 40(10):2196–2212, 2002. (Cit  en page 55.)

- [Xie 2002b] H. Xie, L. Pierce et F. Ulaby. Statistical properties of logarithmically transformed speckle. *IEEE Trans. Geosci. Remote Sens.*, 40(3):721-727, 2002. (Cité en page 54.)
- [Yan 2013] J. Yan, J. Chu-Shern et D. Townsend. Mri guided pet image filtering and partial volume correction. In *IEEE Nuclear Science Symp. and Medical Imaging Conf. (NSS/MIC)*, pages 1–5, 2013. (Cité en page 233.)
- [Yaroslavsky 1985] L. Yaroslavsky. Digital picture processing. an introduction. *Springer Series in Information Sciences*, 9, 1985. (Cité en page 126.)
- [Yu 2008] L. Yu, D. Zhang et E.-J. Holden. A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. *Computers & Geosciences*, 34(7):838–848, 2008. (Cité en page 189.)
- [Zendjebil 2008] I. Zendjebil, F. Ababsa, J.-Y. Didier, J. Vairon, L. Frauciel, M. Hachet, P. Guitton et R. Delmont. Outdoor augmented reality: State of the art and issues. In *ACM/IEEE Virtual Reality Int. Conf. (VRIC)*, pages 177–187, 2008. (Cité en page 232.)
- [Zhang 2010a] L. Zhang, W. Dong, D. Zhang et G. Shi. Two-stage image denoising by principal component analysis with local pixel grouping. *Pattern Recognition*, 43(4):1531–1549, 2010. (Cité en page 115.)
- [Zhang 2010b] X. Zhang, M. Burger, X. Bresson et S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal Imaging Sci.*, 3(3):253–276, 2010. (Cité en pages 145 et 154.)
- [Zhong 2011] H. Zhong, Y. Li et L. Jiao. Sar image despeckling using bayesian nonlocal means filter with sigma preselection. *IEEE Geosci. Remote Sens. Lett.*, 8(4):809–813, 2011. (Cité en page 133.)
- [Zhou 2004] D. Zhou et B. Scholkopf. A regularization framework for learning from graph data. *Statistical Relational Learning and its Connections to other fields*, 2004. (Cité en page 152.)
- [Zimmerman 1988] J. Zimmerman, S. Pizer, E. Staab, J Randolph R. Perry, W. McCarty et B. Brenton. An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement. *IEEE Trans. Med. Imag.*, 7(4):304–312, 1988. (Cité en pages 49 et 233.)
- [Zitova 2003] B. Zitova et J. Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977–1000, 2003. (Cité en page 187.)
- [Zoran 2009] D. Zoran et Y. Weiss. Scale invariance and noise in natural images. In *Int. Conf. Computer Vision (ICCV)*, pages 2209–2216, 2009. (Cité en page 63.)

Résumé : La vision de nuit des pilotes d'hélicoptère est artificiellement assistée par un dispositif de vision bas niveau de lumière constitué d'un intensificateur de lumière (IL) couplé à une caméra numérique d'une part, et d'une caméra infrarouge (IR) d'autre part. L'objectif de cette thèse est d'améliorer ce dispositif en ciblant les défauts afin de les corriger.

Une première partie consiste à réduire le bruit dont souffrent les images IL. Cela nécessite d'évaluer la nature du bruit qui corrompt ces images. Pour cela, une méthode d'estimation automatique du bruit est mise en place. L'estimation repose sur la détection non paramétrique de zones homogènes de l'image. Les statistiques du bruit peuvent être alors être estimées à partir de ces régions homogènes à l'aide d'une méthode d'estimation robuste de la fonction de niveau de bruit par minimisation ℓ^1 .

Grâce à l'estimation du bruit, les images IL peuvent alors être débruitées. Nous avons pour cela développé dans la seconde partie un algorithme de débruitage d'images qui associe les moyennes non locales aux méthodes variationnelles en effectuant une régularisation adaptative pondérée par une attache aux données non locale. Une adaptation au débruitage de séquences d'images permet ensuite de tenir compte de la redondance d'information apportée par le flux vidéo, en garantissant stabilité temporelle et préservation des structures fines.

Enfin, dans la troisième partie les informations issues des capteurs optique et infrarouge sont recalées dans un même référentiel. Nous proposons pour cela un critère de recalage multimodal basé sur l'alignement des contours des images. Combiné à une résolution par montée de gradient et à un schéma temporel, l'approche proposée permet de recalibrer de façon robuste les deux modalités, en vue d'une ultérieure fusion.

Mots clés : Vision nocturne, estimation automatique du bruit, débruitage vidéo adaptatif, recalage multimodal

Numerical night vision system: Automatic restoration and multimodal registration of low light level images

Abstract: Night vision for helicopter pilots is artificially enhanced by a night vision system. It consists in a light intensifier (LI) coupled with a numerical camera, and an infrared camera. The goal of this thesis is to improve this device by analyzing the defaults in order to correct them.

The first part consists in reducing the noise level on the LI images. This requires to evaluate the nature of the noise corrupting these images, so an automatic noise estimation method has been developed. The estimation is based on a non parametric detection of homogeneous areas. Then the noise statistics are estimated using these homogeneous regions by performing a robust ℓ^1 estimation of the noise level function.

The LI images can then be denoised using the noise estimation. We have developed in the second part a denoising algorithm that combines the non local means with variational methods by applying an adaptive regularization weighted by a non local data fidelity term. Then this algorithm is adapted to video denoising using the redundancy provided by the sequences, hence guaranteeing temporal stability and preservation of the fine structures.

Finally, in the third part data from the optical and infrared sensors are registered. We propose an edge based multimodal registration metric. Combined with a gradient ascent resolution and a temporal scheme, the proposed method allows robust registration of the two modalities for later fusion.

Keywords: Night vision, automatic noise estimation, adaptive video denoising, multimodal registration
