



**HAL**  
open science

# Détection de petits objets dans une image en utilisant les techniques de super-résolution

Fabrice Humblot

► **To cite this version:**

Fabrice Humblot. Détection de petits objets dans une image en utilisant les techniques de super-résolution. Interface homme-machine [cs.HC]. Université Paris Sud - Paris XI, 2005. Français. NNT : . tel-00011319

**HAL Id: tel-00011319**

**<https://theses.hal.science/tel-00011319v1>**

Submitted on 7 Jan 2006

**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.

ORSAY

N° d'ordre : 8130

# THÈSE

présentée pour obtenir

## LE GRADE DE DOCTEUR EN SCIENCES DE L'UNIVERSITÉ PARIS-SUD XI

Spécialité : **automatique et traitement du signal**

par

**Fabrice HUMBLLOT**

—o—

### DÉTECTION DE PETITS OBJETS DANS UNE IMAGE EN UTILISANT LES TECHNIQUES DE SUPER-RÉSOLUTION

—o—

Soutenue le **6 décembre 2005** devant la commission d'examen composée de :

M. Christophe COLLET	Professeur, Université Strasbourg I	Rapporteur
M. Bertrand COLLIN	Chargé de recherche, DGA	Co-directeur de thèse
M. Xavier DESCOMBES	Chargé de recherche, INRIA	Rapporteur
M. Ali MOHAMMAD-DJAFARI	Directeur de recherche, CNRS	Directeur de thèse
M. Wojciech PIECZYNSKI	Professeur, INT	Examineur
M. Bertrand ZAVIDOVIQUE	Professeur, Université Paris-Sud XI	Président



# Remerciements

Je remercie toutes les personnes qui ont contribué à la réalisation et à l'amélioration de mes travaux de thèse, à la rédaction de mon mémoire et à la préparation de ma soutenance.

Je remercie mes directeur et co-directeur de thèse, Ali et Bertrand, pour la qualité de leur encadrement durant mes trois années de préparation. Leurs conseils d'orientation pertinents, leurs explications pédagogiques, leurs encouragements répétés, et leur bonne humeur communicative ont été déterminants dans ma réussite.

Je remercie mes deux rapporteurs Xavier Descombes et Christophe Collet, ainsi que Hassan Foroosh, pour la relecture attentive de mon mémoire dans un temps très court imposé par l'organisation de la soutenance, et pour leurs suggestions. Ce travail a été réalisé par correspondance pour Hassan Foroosh qui ne pouvait assister à la soutenance, mais a envoyé un avis positif élogieux au jury, joint au procès-verbal de l'examen. Leurs commentaires constructifs m'ont beaucoup aidé dans ma présentation et l'amélioration de mon mémoire.

Je remercie le président du jury Bertrand Zavidovique, ainsi que l'examineur Wojciech Pieczynski d'avoir accepté ces fonctions, et pour leurs interventions critiques conciliantes lors de ma soutenance, pour leurs propositions de références complémentaires et leurs suggestions de prolongements possibles de mes recherches.

Je remercie mes collègues et amis du GPI, GIP et de l'ESIEE pour le climat amical et fructueux de mes trois années d'étude. Leurs conseils scientifiques et leurs suggestions m'ont été précieux. J'ai aussi apprécié les agréables moments de détente (et parfois de vacances...) passés en leur compagnie durant ces trois années. Merci Adel, Antoine, Aurélien, Cédric, Christian, Daniel, Denis, Francis, Franck, Françoise, Geneviève, Gildas, Guy, Jacques, Jean-François, Jean-Paul, Jean-Philippe, Jérôme, Laurent, Magali, Mehdi, Nadia, Olivier, Patrice, Stéphane, Thomas, Tristan, Zouaoui, et tous ceux que j'ai peut-être oubliés.

Enfin, je remercie mon père pour ses relectures attentives d'un mémoire parfois « austère », et pour ses suggestions d'améliorations de forme susceptibles d'en faciliter la lecture. Je le remercie aussi d'une manière beaucoup plus générale, ainsi que ma mère, Jean-Philippe, Gaëlle, Delon, Marguerite, Hubert, Nathalie, Bidou et tous les amis très proches Bont, Charlotte, Christo, Git, Henri, Isa, Jérémie, Johan, Laeti, Lo, Mégale, Nath et Tawa pour leur soutien affectueux tout au long de cette aventure.



# Table des matières

<b>1</b>	<b>Introduction et position du travail</b>	<b>7</b>
1.1	Position du contexte applicatif . . . . .	9
1.2	Problématiques rencontrées . . . . .	10
1.3	Quelques définitions et notations . . . . .	10
1.3.1	Résolutions spatiale et en niveaux de gris . . . . .	10
1.3.2	Notations et conventions utilisées . . . . .	11
1.4	Plan de la thèse . . . . .	13
<b>2</b>	<b>Évaluation d'une méthode de détection de petits objets</b>	<b>15</b>
2.1	Fonctionnement de l'algorithme . . . . .	16
2.1.1	Modèle direct . . . . .	16
2.1.2	Modèle <i>a priori</i> . . . . .	16
2.1.3	Critère robuste . . . . .	18
2.1.4	Étude locale . . . . .	19
2.1.5	Interprétation semi-quadratique . . . . .	20
2.2	Simulations . . . . .	21
2.2.1	Protocole expérimental d'évaluation . . . . .	21
2.2.2	Résultats . . . . .	23
2.3	Interprétations et conclusions . . . . .	27
<b>3</b>	<b>Recalage d'images par méthodes de corrélation de phase</b>	<b>29</b>
3.1	Introduction . . . . .	31
3.2	État de l'art sur les méthodes de recalage . . . . .	32
3.2.1	Techniques basées sur le gradient . . . . .	32
3.2.2	Méthodes pixel-récurrentes . . . . .	34
3.2.3	Techniques par correspondance de blocs . . . . .	35
3.2.4	Méthodes du domaine fréquentiel . . . . .	36
3.2.5	Techniques à base d'ensemble de niveaux . . . . .	36
3.3	Recalage d'images par corrélation de phase . . . . .	38
3.3.1	Principe de fonctionnement . . . . .	38
3.3.2	Extension au recalage sous-pixellique . . . . .	39
3.3.3	Extension avec utilisation du gradient . . . . .	43
3.3.4	Équation de recalage considérant les bords . . . . .	44
3.4	Exemple d'applications, résultats . . . . .	46
3.4.1	Cas d'un déplacement précis réalisé à partir d'une fonction numérique . . . . .	46

3.4.2	Cas d'un déplacement obtenu par simulation . . . . .	50
3.4.3	Cas d'un déplacement réel sur images de nuages . . . . .	51
3.4.4	Cas d'un grand déplacement réel vertical . . . . .	55
3.5	Extension de ce modèle à la rotation et à l'homothétie . . . . .	57
3.6	Conclusions . . . . .	58
<b>4</b>	<b>Super-résolution (SR)</b>	<b>59</b>
4.1	Définition . . . . .	61
4.2	Modélisation du lien entre les images BR et l'image HR . . . . .	63
4.3	Problème inverse mal posé . . . . .	65
4.4	État de l'art . . . . .	67
4.4.1	Les méthodes du domaine fréquentiel . . . . .	68
4.4.2	Les méthodes du domaine spatial . . . . .	76
4.4.3	Comparaisons et conclusions . . . . .	103
4.5	Les limites . . . . .	105
<b>5</b>	<b>Approche bayésienne en super-résolution</b>	<b>107</b>
5.1	Cadre général . . . . .	109
5.2	Modèle gaussien et lien avec la régularisation . . . . .	110
5.3	Modélisation markovienne . . . . .	111
5.3.1	Modélisation d'une image par champ de Markov . . . . .	111
5.3.2	Modèle de Gauss–Markov . . . . .	115
5.3.3	Modèle markovien avec variable de contours . . . . .	116
5.3.4	Modèle markovien avec variable de régions . . . . .	117
5.4	Mise en œuvre du calcul de la solution . . . . .	121
5.4.1	L'échantillonneur de Gibbs . . . . .	121
5.4.2	Algorithmes proposés . . . . .	122
5.5	Liens avec les méthodes existantes . . . . .	126
5.6	Méthodes d'évaluation . . . . .	127
5.7	Évaluation des performances . . . . .	128
5.7.1	Effets du bruit . . . . .	129
5.7.2	Effets des erreurs dans les estimées du mouvement . . . . .	134
5.7.3	Importance du nombre d'images BR utilisées vs. répartition déplacements . . . . .	136
5.7.4	Résultats obtenus à partir de données réelles . . . . .	138
5.7.5	Bilan de l'évaluation et interface Matlab de simulation . . . . .	139
<b>6</b>	<b>Conclusions et perspectives</b>	<b>141</b>
	<b>Annexes</b>	<b>144</b>
	<b>A Calculs de lois marginales <i>a posteriori</i></b>	<b>145</b>
	<b>B Acronymes et glossaire</b>	<b>147</b>
	<b>Bibliographie</b>	<b>149</b>

# Chapitre 1

## Introduction et position du travail

### Sommaire

---

<b>1.1</b>	<b>Position du contexte applicatif</b>	<b>9</b>
<b>1.2</b>	<b>Problématiques rencontrées</b>	<b>10</b>
<b>1.3</b>	<b>Quelques définitions et notations</b>	<b>10</b>
1.3.1	Résolutions spatiale et en niveaux de gris	10
1.3.2	Notations et conventions utilisées	11
<b>1.4</b>	<b>Plan de la thèse</b>	<b>13</b>

---

### Table des figures

---

1.1	Comparaison de deux approches pour la détection.	10
1.2	Convention utilisée pour représenter les coordonnées des images numériques.	12

---



Le travail de recherche présenté dans ce mémoire s'intéresse à la détection de petits objets dans une image en utilisant les techniques de super-résolution (SR). Il considère que la reconstruction (ou restauration) d'une image par une méthode de SR consiste à produire une image de haute résolution (HR) à partir de plusieurs images de résolution plus faible qualifiées d'images de basse résolution (BR).

Les enjeux de la maîtrise d'une bonne méthode de SR sont considérables. Une « bonne » méthode de SR suppose qu'elle soit rapide en temps de calcul pour des applications destinées à un traitement en temps réel, et qu'elle augmente sensiblement la résolution des images BR qui sont représentatives d'une scène observée. Plus précisément, elle doit présenter les caractères suivants :

- le contraste des images obtenues est plus élevé que dans les images BR initiales,
- la taille des images HR (en nombre de pixels<sup>1</sup>) est plus grande que celle des images BR,
- les contours des objets présents dans la scène sont plus nets,
- elle permet de distinguer et/ou détecter des objets dans les images HR qui n'étaient pas discernables dans les images BR.

D'autre part, il existe de nombreuses situations où les dégradations de la scène originale sont trop importantes pour l'application visée. Lorsque la résolution ne peut pas être améliorée en remplaçant les capteurs d'acquisition, soit à cause de leur prix, soit en raison des limites physiques, il est possible d'utiliser des algorithmes de SR. Dans tous les cas, même celui où des équipements de haute qualité sont disponibles, ces algorithmes proposent une alternative à faible coût. Ils permettent non seulement d'améliorer la qualité visuelle des images, mais aussi d'augmenter les performances des traitements postérieurs tels que la détection de formes ou d'objets, la reconnaissance de visage de personnes, ou l'identification de type d'objets.

Ces dernières années, plusieurs thèses en France et à l'étranger ont été consacrées à la SR. On peut citer par exemple [Nguyen 00, Dekeyser 01, Roullot 01, Borman 04, Rochefort 05] pour des textes en français ou en anglais. Cet intérêt croissant pour ce type de techniques s'explique par le nombre considérable des applications dédiées à ces méthodes. On peut relever les domaines d'application tels que :

- l'imagerie satellitaire [Tsai 84, Omrane 04],
- l'imagerie astronomique [Sheppard 98],
- l'imagerie infra-rouge [Hardie 97a, Hardie 98, Alam 00, Farsiu 04],
- l'amélioration d'écriture [Capel 00, Cortijo 03],
- l'amélioration et la reconnaissance de visage [Baker 02, Gunturk 03],
- l'amélioration vidéo et la restauration [Schultz 96, Alvarez 04, Shechtman 05],
- la conversion des standards vidéo [Patti 95],
- l'imagerie SAR (*Synthetic Aperture Radar* ou radar à ouverture synthétique) et ISAR (*Inverse SAR*) [Corral 02, Xu 05],
- la création de mosaïque numérique [Zomet 00],
- l'imagerie médicale [Greenspan 02],
- la restauration de flux vidéo compressés [Chen 98, Patti 99, Gunturk 02, Gunturk 04, Segall 04].

Cette liste est loin d'être exhaustive car de nouvelles applications sont sans cesse développées.

---

<sup>1</sup>Point représentant le plus petit élément de teinte homogène d'une image.

On distingue les **techniques de restauration d'image**, qui sont des traitements s'appliquant à une seule image à la fois, des **techniques de SR**, qui utilisent plusieurs images en même temps pour en obtenir une seule qualifiée de **super-résolue**. La SR est parfois appelée « **restauration multi-images** ».

La position du **contexte applicatif** est précisée dans la première partie de ce chapitre d'introduction, et les différentes **problématiques rencontrées** au cours de cette étude sont énumérées dans la section **1.2**.

Des **définitions** et les **notations** utilisées dans ce travail sont données dans la partie **1.3**.

Enfin, le **plan** suivi par ce mémoire est précisé dans la section **1.4**.

## 1.1 Position du contexte applicatif

Le contexte applicatif de l'étude est celui de la veille et de la surveillance. On se place dans un cadre où l'on observe des positions éloignées du capteur. Ce contexte suppose que le traitement de SR soit réalisé à partir d'images obtenues par un seul capteur numérique d'acquisition vidéo, que l'on appellera également caméra par la suite. On définit une vidéo numérique comme étant une séquence d'images prises à une certaine vitesse d'acquisition, 25 images par seconde avec les appareils utilisés dans les expérimentations. La caméra est mobile et ses déplacements, entre deux images consécutives, sont faibles. Il en est de même pour ses déplacements angulaires dans l'espace. Dans un repère commun, le déplacement d'un pixel entre 2 images consécutives de la position  $\begin{pmatrix} x & y \end{pmatrix}^t$  à la position  $\begin{pmatrix} x' & y' \end{pmatrix}^t$  peut être modélisé par l'équation affine suivante :

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Cette équation permet de représenter les mouvements suivants : translation par le choix des paramètres  $x_0$  et  $y_0$ , homothétie et rotation par le choix de  $a$ ,  $b$ ,  $c$  et  $d$ .

Le contexte applicatif de ce travail permet de limiter le champ des transformations possibles entre deux images au mouvement de translation global. En effet, il semble réaliste d'utiliser une caméra fixée sur un mobile contrôlé et stabilisé pour l'obtention de la séquence vidéo BR initiale. On peut alors supposer qu'il n'y a ni de rotation de l'axe de la caméra puisque celui-ci est stabilisé, ni d'effets de zoom (équivalent à une transformation par homothétie) car l'opérateur humain ou automatique utilise déjà un zoom maximum.

D'autre part, la petite taille des cibles pixelliques susceptibles d'être observées permet de les supposer lointaines. Dans ce cas, même si elles sont en mouvement, elles peuvent être considérées comme fixes dans la scène d'observation. Cette remarque permet de justifier que l'on s'intéresse à un mouvement global dans la scène, et non pas local. Ce domaine d'application permet de limiter les transformations possibles entre deux images BR, dues au mouvement de la caméra pendant l'acquisition, à l'unique **translation globale de la scène**. Ce choix a été validé par les résultats expérimentaux obtenus à partir de séquences réelles.

## 1.2 Problématiques rencontrées

Cette étude a conduit à poser plusieurs questions auxquelles ce mémoire apporte des réponses.

- Tout d’abord, **comment caractériser une méthode** de détection de petits objets ?
- Peut-on trouver un **critère quantitatif** permettant de dire à quelle condition un pixel sera détecté par l’algorithme choisi ?
- Il a fallu également se poser la question du choix d’une **méthode de recalage**, et en évaluer correctement **la précision**. Fallait-il se concentrer sur une estimation globale du mouvement inter-images, ou sur des estimations locales du mouvement dans l’image ?
- Une autre question à laquelle cette thèse souhaitait apporter des réponses est de savoir si on peut obtenir des résultats équivalents de détection sur des images, avec une caméra qualifiée de « haute résolution », et à l’aide d’une caméra de moins bonne résolution que la première, mais dont les images sont retraitées par SR. Cette problématique est illustrée par la FIG. 1.1. L’ensemble des attributs permettant de décrire les différents objets détectés sont par exemple l’aire (nombre de pixels et leur position), la forme, la luminance et/ou l’emplacement précis du centre de gravité de tous les objets détectés.
- L’**évaluation** des méthodes de SR est également un problème en soi. Comment comparer entre eux les résultats obtenus avec différentes méthodes ? Comment qualifier une méthode de SR comme **meilleure** par rapport à une autre ?
- Enfin, il a fallu proposer une **méthode de SR** susceptible de faciliter un traitement postérieur, automatisé ou humain, de détection.

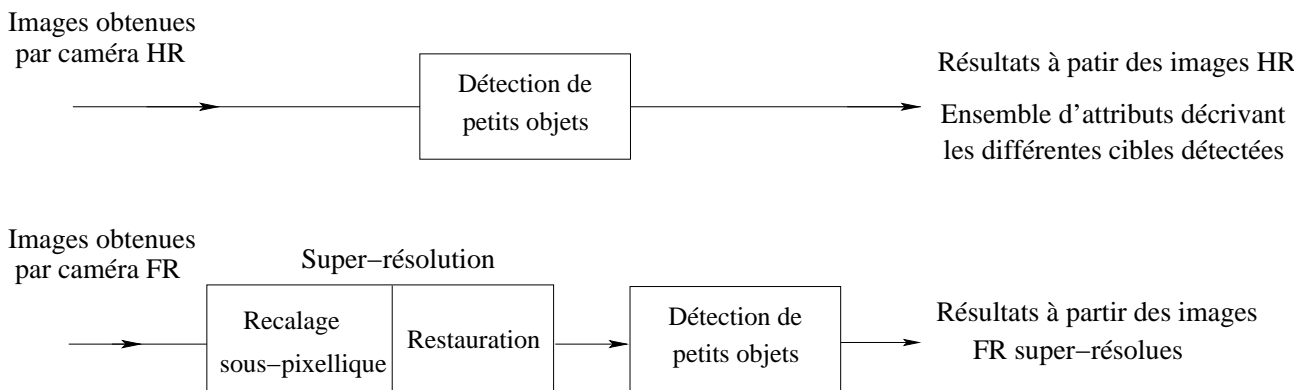


FIG. 1.1: Comparaison de deux approches pour la détection.  
À partir d’images HR et à partir d’images BR super-résolues.

## 1.3 Quelques définitions et notations

### 1.3.1 Résolutions spatiale et en niveaux de gris

Le type de capteurs numériques présents dans les caméras employées pour l’acquisition de séquences vidéo est une matrice 2-D de capteurs carrés sensibles à la lumière, communément appelée matrice CCD (*charge-coupled device* ou composant à couplage de charge). La réponse de chaque capteur est proportionnelle à l’intégrale de l’énergie de la lumière projetée à sa surface. En général les caméras numériques

du commerce, qui incluent les webcams, utilisent ce type de capteurs. Pour avoir un ordre de grandeur, un composant de résolution moyenne peut avoir 307 200 capteurs sur une matrice CCD de taille  $5 \times 5 \text{ mm}^2$  (ceci équivaut à une matrice de  $640 \times 480$  capteurs).

Afin de créer une image numérique, il faut convertir sous forme numérique les données continues captées. Ceci s'effectue à l'aide de deux processus : l'échantillonnage et la quantification. Soit une image continue selon ces deux coordonnées  $x$  et  $y$ , et également en amplitude. L'**échantillonnage** revient à **numériser** la valeur des coordonnées ; c'est-à-dire à prendre des échantillons uniformément espacés dans l'image continue. La **quantification** revient à numériser la valeur des amplitudes ; elle est réalisée en convertissant ces valeurs en quantités discrètes en allant du noir (0 par convention) au blanc. Dans le cas d'une acquisition avec une matrice CCD, c'est le nombre de capteurs dans la matrice qui détermine l'échantillonnage selon les deux axes. La qualité d'une image numérique est déterminée en grande partie par le nombre d'échantillons et le nombre de niveaux de gris discrets utilisés pour l'échantillonnage et la quantification.

L'échantillonnage est le principal facteur qui détermine la **résolution spatiale d'une image**. La résolution spatiale représente le nombre de pixels utilisé pour représenter l'image.

De la même manière, la **résolution en niveaux de gris** est le plus petit changement qui soit discernable dans un niveau de gris. C'est aussi le nombre de niveaux de gris utilisés pour la représentation de l'amplitude des pixels de l'image. Cependant, on peut signaler que mesurer un changement discernable dans un niveau de gris est un processus très subjectif. Pour des considérations matérielles (*hardware*) liées au traitement, au stockage et à l'échantillonnage, le nombre de niveaux de gris est habituellement une puissance de 2. La valeur souvent utilisée, et également dans cette thèse, consiste à coder les niveaux de gris sur 8 bits (*binary digit* ou nombre binaire). Ceci donne une représentation des images sur  $N = 256$  niveaux de gris, généralement exprimés par des valeurs entières comprises entre 0 et 255.

Ainsi, on dit qu'une image numérique codée sur  $N$  niveaux et de taille  $H \times L$ , a une résolution spatiale de  $H \times L$  pixels et une résolution en niveaux de gris de  $N$  niveaux. Elle peut ainsi être vue comme une matrice de dimension finie, avec  $H$  représentant le nombre de pixels présents sur une colonne de l'image, et  $L$  le nombre de pixels présents sur une ligne.  $H$  et  $L$  peuvent également être vues comme respectivement la hauteur et la largeur de l'image. C'est la terminologie utilisée dans ce document.

Le lecteur pourra se rapporter au chapitre 2 de [Gonzalez 03] pour obtenir plus d'informations concernant les bases de l'imagerie numérique.

### 1.3.2 Notations et conventions utilisées

On note les images par des fonctions à deux dimensions de la forme  $f(x, y)$  dans le cas continu, et  $f[x, y]$  dans le cas discret. La valeur, ou amplitude, de  $f$  aux coordonnées spatiales  $(x, y)$  est une quantité scalaire positive dont la signification physique est déterminée par la source d'illumination dans l'image. Les images auxquelles s'intéresse le présent document sont **monochromatiques**, et leurs valeurs s'évaluent sur l'échelle de gris. Lorsque les images sont générées à partir d'un processus physique, ces

valeurs sont proportionnelles à l'énergie rayonnée par une source physique (par exemples des ondes électromagnétiques). En conséquence,  $f(x, y)$  doit être non nul et fini :

$$0 < f(x, y) < \infty.$$

On appelle **intensité** d'une image monochrome, à un point de coordonnées  $(x_0, y_0)$ , le **niveau de gris**  $n$  de l'image en ce point :  $n = f(x_0, y_0)$ .

On pose également que  $N_{\min} \leq n \leq N_{\max}$ , ce qui permet de définir l'échelle de gris par l'intervalle  $[N_{\min}, N_{\max}]$ . En pratique, on décale numériquement cet intervalle de manière à obtenir l'intervalle  $[0, N - 1]$ , où  $n = 0$  est considéré noir, et  $n = N - 1$  est considéré blanc sur l'échelle de gris. Toutes les valeurs intermédiaires sont des nuances de gris variant du noir au blanc. Ces niveaux discrets sont espacés de manière égale et sont des nombres entiers compris dans l'intervalle  $[0, N - 1]$ .

L'intervalle de valeurs réparties sur l'échelle de gris est appelé **intervalle dynamique** d'une image. Pour les images dont les niveaux de gris s'étalent sur une portion significative de l'échelle de gris on parle d'**intervalle dynamique élevé**, ou on dit que l'image a une **dynamique élevée**. Quand une proportion importante de pixels montre cette propriété, l'image a un **contraste élevé**. À l'inverse, une image avec un **intervalle dynamique faible** a un rendu terne et **faiblement contrasté**.

Il existe plusieurs manières pour représenter les images numériques. On suppose qu'une image  $f(x, y)$  est échantillonnée de manière à ce que l'image numérique résultante  $f[x, y]$  ait  $H$  lignes et  $L$  colonnes. Ainsi, les valeurs aux coordonnées  $(x, y)$  deviennent des quantités discrètes. Dans un but de clarté et de commodité, les formules suivantes utilisent des valeurs entières pour ces coordonnées discrètes. Les valeurs des coordonnées à l'origine sont  $(x, y) = (0, 0)$ . Les valeurs de coordonnées suivantes, selon la première ligne de l'image, sont représentées comme  $(x, y) = (0, 1)$ . **Il faut garder à l'esprit** que  $(0, 1)$  est utilisé pour signifier le second échantillon le long de la première ligne. La FIG. 1.2 montre la convention utilisée pour la représentation des coordonnées d'une image.

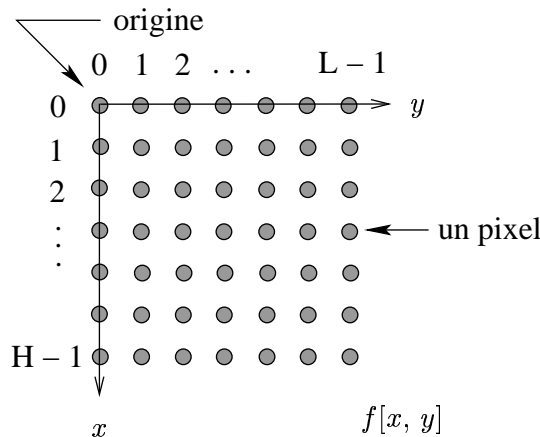


FIG. 1.2: Convention utilisée pour représenter les coordonnées des images numériques.

La notation proposée permet d'écrire l'image numérique  $H \times L$  complète en utilisant la forme matricielle compacte suivante :

$$f[x, y] = \begin{bmatrix} f[0, 0] & f[0, 1] & \dots & f[0, L-1] \\ f[1, 0] & f[1, 1] & \dots & f[1, L-1] \\ \vdots & \vdots & & \vdots \\ f[H-1, 0] & f[H-1, 1] & \dots & f[H-1, L-1] \end{bmatrix}. \quad (1.1)$$

Le côté droit de cette égalité est, par définition, une image numérique. Chaque élément de cette matrice est appelé **élément d'image**, **pixel** ou **pel**. Ces deux derniers termes proviennent de l'anglais *picture element*. Un pixel peut être défini comme un point représentant le plus petit élément de teinte homogène d'une image, et pouvant être transmis sous forme numérique. Dans la suite du texte, les termes image et pixel sont utilisés pour désigner une image numérique et ses éléments.

Une autre écriture plus compacte consiste aussi à noter une image  $f(\mathbf{r})$ , où  $\mathbf{r} = (x, y) \in \mathbb{N}^2$  représente la position d'un pixel. De cette manière, une image est constituée d'un ensemble fini  $\mathcal{R}$  de sites  $\mathbf{r}$  correspondants aux pixels :

$$\mathcal{R} = \{ \mathbf{r} = (x, y) ; 0 \leq x \leq H-1, 0 \leq y \leq L-1 \}.$$

Enfin, parfois, il est préférable d'utiliser une notation matricielle plus traditionnelle pour symboliser une image numérique et ses éléments :

$$\mathbf{F} = \begin{bmatrix} f_{1,1} & f_{1,2} & \dots & f_{1,L} \\ f_{2,1} & f_{2,2} & \dots & f_{2,L} \\ \vdots & \vdots & & \vdots \\ f_{H,1} & f_{H,2} & \dots & f_{H,L} \end{bmatrix}. \quad (1.2)$$

On peut écrire  $f_{i,j} = f[x = i-1, y = j-1] = f[i-1, j-1]$ ,  $i = 1, \dots, H$ ,  $j = 1, \dots, L$ . Ainsi, les matrices (1.1) et (1.2) sont identiques.

Pour rappel, la matrice  $\mathbf{F}$  transposée a une taille  $L \times H$  et s'écrit :

$$\mathbf{F}^t = \begin{bmatrix} f_{1,1} & f_{2,1} & \dots & f_{H,1} \\ f_{1,2} & f_{2,2} & \dots & f_{H,2} \\ \vdots & \vdots & & \vdots \\ f_{1,L} & f_{2,L} & \dots & f_{H,L} \end{bmatrix}. \quad (1.3)$$

## 1.4 Plan de la thèse

Les travaux de recherche présentés dans ce document concernent la détection de petits objets dans une image en utilisant les techniques de super-résolution (SR).

Ce sujet repose sur trois problématiques distinctes : la **détection d'objets dans une image**, le **recalage des images** et les **méthodes de SR**.

Le sujet de cette thèse est décrit dans le **Chapitre 1**, p. 7 ; des concepts et les notations utilisées dans ce document y sont également définis.

La **détection de petits objets** se réalise dans un repère lié à une image de référence. Dans cette étude, la détection consiste en la recherche des pixels ayant une intensité assez différente de l'intensité de leurs voisins les plus proches. C'est dans ce contexte que l'on définit le « contraste » d'un pixel cible comme son intensité (ou son niveau de gris en imagerie mono-spectrale) diminuée de la moyenne des intensités de ses voisins les plus proches.

Le **Chapitre 2**, p. 15, traite de **l'évaluation d'un algorithme de détection de petits objets**. Il s'agit de l'algorithme présenté dans la thèse de [Samson 02]. J'ai présenté ce travail lors de la réunion plénière du thème E du GDR ISIS, le 4 décembre 2003.

Le **recalage des images** entre elles est une phase essentielle dans le processus de SR. Cette étape consiste à faire correspondre dans un repère commun, deux ou plusieurs images représentatives d'une même scène, à des instants distincts, depuis des points de vue différents, et/ou à partir de capteurs différents. Pour avoir une bonne reconstruction par SR, il est fondamental de connaître précisément la transformation permettant de passer d'une image à une autre.

Le **Chapitre 3**, p. 29, présente et compare **deux méthodes d'estimation du mouvement utilisant le principe de corrélation de phase**. Ces méthodes permettent d'obtenir une précision sous-pixelique, c'est-à-dire inférieure à la taille d'un pixel. Après la présentation d'un rapide **état de l'art** dans la section **3.2** de ce chapitre, cette **précision est rigoureusement évaluée**. J'ai présenté ce travail dans la conférence internationale PSIP 2005 en janvier 2005 [Humblot 05a].

Dans cette étude, par hypothèse, la **restauration d'une image par une méthode de SR** consiste à produire une image HR à partir de plusieurs images BR de résolution plus faible. Ces images BR peuvent être extraites d'une séquence vidéo obtenue à partir d'une caméra fixe montrant une scène en mouvement, d'une caméra en mouvement observant une scène fixe, ou encore de plusieurs caméras fixes observant une même scène : **le mouvement et l'information non redondante rendent le processus de SR possible**.

Ainsi, l'image HR obtenue contient plus d'information utile que dans une seule image BR prise dans la séquence vidéo initiale.

Un **état de l'art** des méthodes de SR est présenté dans le **Chapitre 4**, p. 59.

Le chapitre **Chapitre 5**, p. 107, est consacré à une méthode de SR associée à un traitement joint de segmentation de l'image HR en régions homogènes. Cette méthode a été développée dans le cadre de l'estimation bayésienne. Celui-ci permet d'introduire une information issue de notre connaissance *a priori* de la scène originale. Cette approche a été inspirée et adaptée d'un processus de fusion avec segmentation jointe, objet de l'article [Féron 05].

Ce travail a été accepté pour publication dans un numéro spécial de la revue internationale EURASIP *Journal on Applied Signal Processing* consacré à la SR, en juin 2005 [Humblot 05c].

La méthode proposée a fait l'objet d'une extension présentée en août 2005 au cours du 25<sup>e</sup> *workshop* international consacré aux méthodes d'inférence bayésienne et de maximum d'entropie [Humblot 05b].

Enfin, les conclusions de cette étude sont présentées au **Chapitre 6**, p. 141.

## Chapitre 2

# Évaluation d'une méthode de détection de petits objets

### Sommaire

---

<b>2.1</b>	<b>Fonctionnement de l'algorithme</b> . . . . .	<b>16</b>
2.1.1	Modèle direct . . . . .	16
2.1.2	Modèle <i>a priori</i> . . . . .	16
2.1.3	Critère robuste . . . . .	18
2.1.4	Étude locale . . . . .	19
2.1.5	Interprétation semi-quadratique . . . . .	20
<b>2.2</b>	<b>Simulations</b> . . . . .	<b>21</b>
2.2.1	Protocole expérimental d'évaluation . . . . .	21
2.2.2	Résultats . . . . .	23
<b>2.3</b>	<b>Interprétations et conclusions</b> . . . . .	<b>27</b>

---

### Table des figures

---

2.1	Pixels tests et leurs 4 voisins (dans le même ordre que dans TAB. 2.1). . . . .	22
2.2	Localisation des 5 pixels dans l'image IR utilisée pour les simulations. . . . .	24
2.3	Comparaison résultats cas entier/réel - Pixel (128, 54), zone chaude. . . . .	24
2.4	Comparaison résultats cas entier/réel - Pixel (218, 292), 3 voisins chauds. . . . .	25
2.5	Comparaison résultats cas entier/réel - Pixel (30, 260), zone froide. . . . .	25
2.6	Superposition des 5 cas entiers. . . . .	26
2.7	Superposition des 5 cas réels. . . . .	26

---

### Liste des tableaux

---

2.1	Niveaux de gris des pixels voisins aux pixels tests et leur moyenne. . . . .	22
2.2	Temps de calcul pour 256 réalisations de l'algorithme (cas entier et réel). . . . .	23

---



Ce chapitre est consacré à l'étude d'une méthode de **détection** de petits objets de la taille d'environ **un pixel** dans une image. Le but est d'obtenir un **critère permettant de déterminer à quelle condition un objet sera détecté par une méthode donnée**, et susceptible d'être utilisée pour des comparaisons dans des études ultérieures.

Le cadre bayésien d'estimation permet d'obtenir de meilleures performances que les approches classiques<sup>1</sup>. En particulier, la méthode bayésienne utilisée par Samson dans sa thèse [Samson 02] donne de bons résultats de détection avec un faible taux de fausses alarmes. Cette méthode est prise comme base de départ pour la présente étude. Les aspects « taux de fausses alarmes » ne sont pas traités ici.

Le **modèle de détection** choisi est présenté dans la première partie de ce chapitre. Les résultats de **simulation** font l'objet de la partie **2.2**. Des **conclusions** sont données dans la partie **2.3**.

## 2.1 Fonctionnement de l'algorithme

### 2.1.1 Modèle direct

Dans sa thèse, Samson aborde le problème de la détection par une phase d'estimation ou de transformation des données préalable à toute décision. Samson a cherché à définir non pas une règle de décision mais seulement une statistique, semblable aux résidus de prédiction linéaire, qu'il se contente ensuite de « seuiller » afin d'obtenir la carte des plots de détection.

L'image globale des données est modélisée comme la somme de l'image du fond et d'une image des cibles potentielles. L'objectif est d'estimer cette image inconnue des cibles, l'image estimée constituant précisément la collection des valeurs de la statistique recherchée en chaque pixel. Nous notons  $\mathbf{g} = g[x, y]$  avec  $x = 1, \dots, H$  et  $y = 1, \dots, L$  l'image observée,  $\mathbf{f}$  l'image du fond et  $\mathbf{t}$  l'image des cibles. Le modèle global d'observation s'écrit sous la forme suivante :

$$\mathbf{g} = \mathbf{f} + \mathbf{t} \quad (2.1)$$

### 2.1.2 Modèle *a priori*

En ce qui concerne le fond  $\mathbf{f}$ , un modèle de type Gibbs-Markov est choisi, d'usage courant en restauration d'images (cf. **5.3** pour avoir des précisions sur la modélisation markovienne et les notations). Le caractère corrélé est introduit par des termes d'interaction entre pixels voisins. Plus précisément, le modèle retenu est un modèle de Gibbs stationnaire non gaussien et isotrope, pénalisant les différences entre pixels voisins par l'intermédiaire d'une fonction  $\phi$  non quadratique :

$$p(\mathbf{F} = \mathbf{f}) \propto \exp(-\mathcal{U}(\mathbf{f})) \quad \text{avec} \quad \mathcal{U}(\mathbf{f}) \propto \sum_{\mathbf{c} \in \mathcal{C}} \phi(\mathbf{v}_{\mathbf{c}}^t \mathbf{f}).$$

---

<sup>1</sup>toutefois aucun état de l'art du domaine, qui aurait permis de renforcer cette affirmation, n'a été mené au cours de cette thèse.

Ce type de modèle avec  $\phi$ -fonction a été préconisé afin de préserver les fortes différences aux contours des objets dans une image. En fonction du choix de  $\phi$ , un tel modèle isotrope peut être **non paramétrique**. Toutefois, le choix de la fonction de pénalisation joue un rôle important, et son influence fait souvent intervenir au moins un « hyper-paramètre » qu'il faut régler empiriquement. Quant au système de voisinage, nous considérons le modèle le plus simple faisant intervenir les différences premières avec les quatre plus proches voisins.

Par la suite, la **loi de probabilité jointe** d'une image sera notée indifféremment  $p(\mathbf{F} = \mathbf{f})$  ou  $p(\mathbf{f})$  : la première lorsque l'image est considérée comme un **vecteur aléatoire discret**, et la deuxième lorsque l'image est considérée comme un **vecteur aléatoire continu**,  $p(\mathbf{f})$  représente alors sa **densité de probabilité**. De manière équivalente encore, ces lois peuvent s'écrire  $p_{\mathbf{F}}(\mathbf{F} = \mathbf{f})$  et  $p_{\mathbf{F}}(\mathbf{f})$ .

Pour l'image des cibles  $\mathbf{t}$ , on choisit un modèle séparable non quadratique, c'est-à-dire en termes probabilistes, un ensemble de variables aléatoires (v.a.) indépendantes pour chaque pixel  $\mathbf{r} \in \mathcal{R}$ , identiquement distribuées (i.i.d.) et **non gaussiennes** :

$$p(\mathbf{t}) = \prod_{\mathbf{r} \in \mathcal{R}} p(t_{\mathbf{r}}) \quad \text{avec} \quad -\log p(t_{\mathbf{r}}) \propto \chi(t_{\mathbf{r}}) + \text{constante.}$$

L'intensité de chaque pixel est pénalisée par une même fonction continue  $\chi(\cdot)$ , qui peut dépendre également d'un paramètre. Cette distribution rend possible la présence de plusieurs cibles d'amplitudes variables dans une image. L'information apportée par une telle modélisation des cibles, de même que celle apportée par l'utilisation d'un modèle markovien sur l'image du fond, permet de structurer la solution sans pour autant être en parfaite adéquation avec les images recherchées. Un modèle « ponctuel » faisant intervenir explicitement la rareté des pixels non nuls et la position de chaque cible éventuelle est utilisé pour l'image des cibles (type bernouilli-gaussien [Champagnat 96] ou poisson-gaussien [Dublanche 97] si l'on veut modéliser des cibles sous-pixelliques).

En ce qui concerne les fonctions de potentiel  $\phi(u)$  et  $\chi(u)$ , Samson choisit de se limiter à des pénalisations convexes (c'est-à-dire à des lois *a priori* log-concaves). Il s'intéresse plus particulièrement aux potentiels convexes non quadratiques, et à la famille de fonctions «  $L_2 - L_1$  » quadratiques au voisinage de zéro, et linéaires à l'infini. Son choix se porte sur la branche d'hyperbole :  $u \mapsto \sqrt{u^2 + T^2} - T$ . Cette fonction fait intervenir un paramètre d'échelle  $T$  qui règle le passage du comportement quadratique au comportement linéaire.

### 2.1.3 Critère robuste

Le contexte de l'estimation bayésienne est adopté. Les deux composantes  $\mathbf{f}$  et  $\mathbf{t}$  sont considérées comme les réalisations des champs aléatoires **indépendants** dont les densités de probabilité s'écrivent respectivement :

$$-\log p(\mathbf{f}) \propto \sum_{\mathbf{r}' \sim \mathbf{r}} \phi(f_{\mathbf{r}'} - f_{\mathbf{r}}) + \text{constante},$$

où «  $\mathbf{r}' \sim \mathbf{r}$  » désigne une clique  $\mathbf{c}_2$  constituée de deux plus proches voisins, et

$$-\log p(\mathbf{t}) \propto \sum_{\mathbf{r} \in \mathcal{R}} \chi(t_{\mathbf{r}}) + \text{constante}.$$

L'approche bayésienne offre différents choix d'estimateurs suivant la règle de décision adoptée : MAP (maximum *a posteriori*), MAPM (MAP marginal), MP (moyenne *a posteriori*). L'estimateur le plus couramment utilisé en traitement d'images est sans aucun doute le MAP, choisi ici. Pour une distribution *a posteriori* identique et non gaussienne, l'estimateur MAP a des propriétés différentes de l'estimateur MP. L'estimateur MAP pourra être égal à la donnée si on considère un terme d'attache aux données linéaires [Nikolova 01], alors que ce n'est pas le cas pour MP.

Ainsi, l'image  $\mathbf{t}$  est estimée par l'argument  $\hat{\mathbf{t}}$  qui rend maximum la distribution *a posteriori* connaissant l'image observée  $\mathbf{g}$  :

$$\hat{\mathbf{t}} = \arg \max_{\mathbf{t} \in \mathbb{R}^{H \times L}} [p(\mathbf{t} | \mathbf{g})] \quad \text{avec} \quad \mathbf{g} = \mathbf{f} + \mathbf{t}. \quad (2.2)$$

On obtient de la règle de Bayes :

$$p(\mathbf{t} | \mathbf{g}) = \frac{p(\mathbf{t}, \mathbf{g})}{p(\mathbf{g})} = \frac{p(\mathbf{g} | \mathbf{t}) p(\mathbf{t})}{p(\mathbf{g})}. \quad (2.3)$$

Du point de vue de la maximisation en  $\mathbf{t}$  de la distribution de probabilité *a posteriori*, la probabilité d'observation de l'image  $p(\mathbf{g})$  est constante. Ainsi,

$$p_{\mathbf{T} | \mathbf{G}}(\mathbf{t} | \mathbf{g}) \propto p_{\mathbf{G} | \mathbf{T}}(\mathbf{g} | \mathbf{t}) p_{\mathbf{T}}(\mathbf{t}) \quad (2.4)$$

Qui peut être réécrit de manière tout à fait équivalente :

$$p_{\mathbf{T} | \mathbf{G}}(\mathbf{t} | \mathbf{g}) \propto p_{\mathbf{F}}(\mathbf{g} - \mathbf{t}) p_{\mathbf{T}}(\mathbf{t}). \quad (2.5)$$

Par un changement de variable, on montre que cela revient à définir  $\hat{\mathbf{t}}$  comme l'image des résidus :

$$\hat{\mathbf{t}} = \mathbf{g} - \hat{\mathbf{f}} \quad \text{avec} \quad \hat{\mathbf{f}} = \arg \max_{\mathbf{f} \in \mathbb{R}^{H \times L}} [p(\mathbf{f} | \mathbf{g})].$$

$\mathbf{f}$  correspond à l'estimée du fond au sens du MAP et minimise donc un critère  $\mathbf{J}(\mathbf{f})$  de la forme :

$$\mathbf{J}(\mathbf{f}) = \sum_{\mathbf{r} \in \mathcal{R}} \chi(g_{\mathbf{r}} - f_{\mathbf{r}}) + \lambda \sum_{\mathbf{r}' \sim \mathbf{r}} \phi(f_{\mathbf{r}'} - f_{\mathbf{r}}) \quad (2.6)$$

où  $\lambda$  est un paramètre de régularisation. Le critère (2.6) à minimiser en  $\mathbf{f}$  ressemble à celui obtenu par une technique de débruitage impulsif préservant les contours. Dans le cas présent, c'est l'image des cibles  $\mathbf{t}$  qui constitue le signal d'intérêt. Et par analogie, les cibles peuvent donc s'assimiler à un bruit impulsif sur le fond  $\mathbf{f}$ . Les deux potentiels robustes  $\chi$  et  $\phi$  pénalisent respectivement les écarts aux données, et les différences entre pixels voisins du fond.

On peut montrer que le paramètre  $\lambda$  doit être plus grand qu'une certaine valeur  $\lambda^{min}$  dépendant de  $\chi$  et  $\phi$ , pour que les valeurs aberrantes de  $\mathbf{g}_r$  soient écartées. Les dérivées  $\chi'$  et  $\phi'$  étant bornées, on définit :

$$\lambda^{min} = \frac{\sup |\chi'|}{k \sup |\phi'|}$$

avec  $k$  le nombre de voisins pris en considération dans le modèle. Un modèle markovien d'ordre 1 a été choisi :  $k = 4$  (voir FIG. 5.1).

D'autre part,  $\chi'(u) = \phi'(u) = \frac{u}{\sqrt{u^2 + T^2}}$ , et ces fonctions dérivées sont donc majorées par 1.

Ainsi,  $\lambda^{min} = \frac{1}{4}$ .

Dans les simulations qui suivent, la valeur 0,3 a été choisie comme suggéré dans [Samson 02].

On peut montrer qu'il convient de choisir  $\lambda$  assez grand si l'on souhaite séparer les objets ponctuels du fond, mais le compromis veut que l'on ne prenne pas  $\lambda$  trop élevé pour ne pas trop lisser le fond et créer trop de fausses alarmes (c'est-à-dire trop de détections erronées). Samson conclut que  $\lambda^{min}$  fixe un ordre de grandeur très utile en pratique. Les tests qu'il a réalisés attestent que le choix d'un  $\lambda$  proche de  $\lambda^{min}$  permet d'obtenir une bonne séparation entre le fond et les cibles (voir III.3 de [Samson 02] pour plus de détails).

Le critère défini par l'équation (2.6) peut également être introduit dans un cadre purement déterministe, en le formulant en termes de fonctions de pénalisation à minimiser sous la contrainte exprimée par le modèle direct (2.1) :

$$\min_{\mathbf{f}, \mathbf{t}} \sum_{\mathbf{r} \in \mathcal{R}} \chi(t_{\mathbf{r}}) + \lambda \sum_{\mathbf{r}' \sim \mathbf{r}} \phi(f_{\mathbf{r}'} - f_{\mathbf{r}}) \quad \text{sous la contrainte} \quad \mathbf{f} + \mathbf{t} = \mathbf{g}. \quad (2.7)$$

#### 2.1.4 Étude locale

En marge de l'aspect algorithmique, on peut s'intéresser au comportement des estimateurs en un site  $\mathbf{r}$  fixé, où l'on observe la donnée  $\mathbf{g}_r$ .

On suppose que les valeurs des variables  $f_{\mathbf{r}'}$  et  $t_{\mathbf{r}'}$  sont connues pour  $\mathbf{r}' \in \mathcal{R}$ ,  $\mathbf{r}' \neq \mathbf{r}$  (en fait seulement pour les pixels voisins  $\mathbf{r}'$  appartenant au voisinage de  $\mathbf{r}$  :  $\mathcal{V}(\mathbf{r})$ , étant donné le caractère markovien du modèle), et on cherche à estimer  $f_{\mathbf{r}}$  et  $t_{\mathbf{r}}$  au pixel  $\mathbf{r}$ . À chaque itération d'un algorithme de remise à jour « pixel par pixel » de type Gauss-Seidel, qui est un cas particulier d'une **descente coordonnée par coordonnée** (« *coordinate descent* ») [Bertsekas 95] dans laquelle le critère est minimisé suivant une direction de coordonnée à chaque étape, on a :

$$\hat{f}_{\mathbf{r}} = \arg \max_{f_{\mathbf{r}} \in \mathbb{R}} [p(f_{\mathbf{r}} | \mathbf{g}, f_{\mathbf{r}'}, \mathbf{r}' \neq \mathbf{r})],$$

$$\widehat{f}_{\mathbf{r}} = \arg \min_{f_{\mathbf{r}} \in \mathbb{R}} \left[ \chi(g_{\mathbf{r}} - f_{\mathbf{r}}) + \lambda \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \phi(f_{\mathbf{r}'} - f_{\mathbf{r}}) \right]. \quad (2.8)$$

En initialisant  $\mathbf{f}$  par l'image des données  $\mathbf{g}$ , l'estimée  $\widehat{f}_{\mathbf{r}}$  après la première itération peut être considérée comme la sortie d'un filtrage local robuste faisant intervenir la donnée  $g_{\mathbf{r}}$  au pixel central et les données  $f_{\mathbf{r}'} = g_{\mathbf{r}'}$  aux pixels voisins.

Le point de vue adopté est de supposer que l'on se trouve au minimum du critère et que l'on cherche les relations vérifiées par l'estimateur global  $\widehat{\mathbf{f}}$  qui minimise  $\mathbf{J}(\mathbf{f})$  dans toutes les « directions », et en particulier dans la direction du site  $\mathbf{r}$ . L'estimateur  $\widehat{\mathbf{f}}$  annule donc les dérivées partielles, soit  $\frac{\delta \mathbf{J}(\mathbf{f})}{\delta f_{\mathbf{r}}} = 0$  pour tout  $\mathbf{r}$ , et à partir de l'équation (2.6) ou (2.8) on vérifie :

$$-\chi'(g_{\mathbf{r}} - \widehat{f}_{\mathbf{r}}) - \lambda \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \phi'(\widehat{f}_{\mathbf{r}'} - \widehat{f}_{\mathbf{r}}) = 0$$

soit,

$$\chi'(g_{\mathbf{r}} - \widehat{f}_{\mathbf{r}}) = -\lambda \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \phi'(\widehat{f}_{\mathbf{r}'} - \widehat{f}_{\mathbf{r}}) \quad (2.9)$$

### 2.1.5 Interprétation semi-quadrique

L'écriture des équations de remise à jour locale permet d'étudier le comportement des estimées suivant le type de fonctions de potentiel choisies, et les paramètres associés. Cependant il n'est pas possible d'obtenir une expression analytique locale de l'argument minimum du critère dans le cas général. Une alternative est de recourir à une interprétation semi-quadrique en introduisant des variables auxiliaires de type GR [Geman 92] ou GY [Geman 95].

Dans le cas des variables de lignes GR (utilisées dans l'implémentation de l'algorithme), et ce pour les deux termes du critère  $\mathbf{J}(\mathbf{f})$  (2.6), sous certaines conditions non détaillées ici mais vérifiées, on a les relations de dualité suivantes :

$$\chi(g_{\mathbf{r}} - f_{\mathbf{r}}) = \inf_{l_{\mathbf{r}} \in \mathbb{R}} \left\{ l_{\mathbf{r}} (g_{\mathbf{r}} - f_{\mathbf{r}})^2 + \widetilde{\chi}(l_{\mathbf{r}}) \right\}, \quad (2.10)$$

$$\phi(f_{\mathbf{r}'} - f_{\mathbf{r}}) = \inf_{b_{\mathbf{r}'\mathbf{r}} \in \mathbb{R}} \left\{ b_{\mathbf{r}'\mathbf{r}} (f_{\mathbf{r}'} - f_{\mathbf{r}})^2 + \widetilde{\phi}(b_{\mathbf{r}'\mathbf{r}}) \right\} \quad (2.11)$$

En dérivant l'expression (2.10) par rapport à  $f_{\mathbf{r}}$ , on obtient :

$$-\chi'(g_{\mathbf{r}} - f_{\mathbf{r}}) = -2 \widehat{l}_{\mathbf{r}} (g_{\mathbf{r}} - f_{\mathbf{r}})$$

soit,

$$\widehat{l}_{\mathbf{r}} = \frac{\chi'(g_{\mathbf{r}} - f_{\mathbf{r}})}{2(g_{\mathbf{r}} - f_{\mathbf{r}})} \quad (2.12)$$

De même, en dérivant l'expression (2.11) par rapport à  $f_{\mathbf{r}}$  également, on obtient :

$$-\phi'(f_{\mathbf{r}'} - f_{\mathbf{r}}) = -2\widehat{b}_{\mathbf{r}'\mathbf{r}}(f_{\mathbf{r}'} - f_{\mathbf{r}})$$

soit,

$$\widehat{b}_{\mathbf{r}'\mathbf{r}} = \frac{\phi'(f_{\mathbf{r}'} - f_{\mathbf{r}})}{2(f_{\mathbf{r}'} - f_{\mathbf{r}})} \quad (2.13)$$

L'équation (2.9) permet d'écrire :

$$2\widehat{l}_{\mathbf{r}}(g_{\mathbf{r}} - \widehat{f}_{\mathbf{r}}) = -\lambda \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} 2\widehat{b}_{\mathbf{r}'\mathbf{r}}(\widehat{f}_{\mathbf{r}'} - \widehat{f}_{\mathbf{r}})$$

Après mise en facteur on a :

$$\widehat{f}_{\mathbf{r}} \left( \widehat{l}_{\mathbf{r}} + \lambda \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \widehat{b}_{\mathbf{r}'\mathbf{r}} \right) = \widehat{l}_{\mathbf{r}} g_{\mathbf{r}} + \lambda \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \widehat{b}_{\mathbf{r}'\mathbf{r}} \widehat{f}_{\mathbf{r}'}$$

Soit finalement l'équation de mise à jour des pixels utilisée dans l'algorithme :

$$\widehat{f}_{\mathbf{r}} = \frac{\widehat{l}_{\mathbf{r}} g_{\mathbf{r}} + \lambda \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \widehat{b}_{\mathbf{r}'\mathbf{r}} \widehat{f}_{\mathbf{r}'}}{\widehat{l}_{\mathbf{r}} + \lambda \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \widehat{b}_{\mathbf{r}'\mathbf{r}}} \quad (2.14)$$

En conclusion, l'algorithme étudié consiste à initialiser  $\mathbf{f}$  avec l'image observée  $\mathbf{g}$ , à mettre à jour les pixels de l'image un par un à partir des équations (2.12), (2.13) et (2.14), puis de calculer la valeur du critère  $\mathbf{J}(\mathbf{f})^{(n)}$  à l'itération  $n$ , à l'aide de l'équation (2.6). On arrête les itérations lorsque  $\mathbf{J}(\mathbf{f}^{(n-1)}) - \mathbf{J}(\mathbf{f}^{(n)})$  est inférieur à une valeur faible (choisie égale à 1 dans l'implémentation de cet algorithme). Il suffit ensuite de soustraire – pixel par pixel – l'image  $\mathbf{f}$  à l'image  $\mathbf{g}$  pour obtenir l'image  $\mathbf{t}$ . On peut alors appliquer une détection par seuil afin de détecter les cibles. La méthode proposée permet d'obtenir un minimum local, et non pas le MAP.

## 2.2 Simulations

### 2.2.1 Protocole expérimental d'évaluation

L'image sur laquelle cet algorithme a été testé est tirée d'une séquence vidéo, obtenue à partir d'une caméra infrarouge (IR) disponible au département GIP. La taille des images est de  $220 \times 320$  pixels. Cette image initialement codée sur 16384 niveaux de gris (14 bits), mais dont la dynamique

s'étendait seulement de 6 524,77 à 7 704,80 (soit une quantification possible sur 1 180 valeurs entières), a été ramenée à une dynamique de  $N = 256$  niveaux de gris : de 0 à 255. L'image de test apparaît sur FIG. 2.2.

Lors du traitement d'une image sous GAMMAG<sup>2</sup>, le pixel en haut à gauche est référencé par ses coordonnées  $x = 0$  et  $y = 0$ ; et le pixel en bas à droite par  $x = 219$  et  $y = 319$ .

Deux versions de cet algorithme ont été développées. La première utilisant le type IMAGE<sup>3</sup> : le niveau de gris des images de ce type est codé par des « **char** », donc sur 8 bits, soit une valeur entière comprise entre 0 et 255. La seconde utilisant le type DIMAGE<sup>3</sup> : le niveau de gris de ces images est codé par des données du type « **double** », ce qui permet une certaine précision dans le développement des calculs.

Pour l'évaluation de l'algorithme, 5 pixels particuliers ont été choisis – les raisons de ce choix sont exposées plus tard. L'emplacement de ces pixels est montré sur l'image FIG. 2.2.

Le protocole expérimental d'évaluation de cet algorithme est le suivant : pour chacun de ces 5 pixels pris individuellement ( $g_r$ ), le niveau de gris varie de 0 à 255, par pas de 1. L'algorithme étant basé sur un modèle de Markov de connexité 4, et donc sur les différences de niveaux de gris entre le pixel de test et ses 4 voisins, il a été choisi d'étudier **la valeur du pixel  $t_r$  en sortie de l'algorithme en fonction de sa valeur  $g_r$  d'entrée, moins la moyenne des 4 pixels voisins.**

Le tableau TAB. 2.1 récapitule ces différentes valeurs classées par valeurs d'abscisses  $x$  croissantes.

	Pix. haut	Pix. gauche	Pix. droit	Pix. bas	Moyenne
( 30, 260)	15	16	15	15	15,25
( 63, 64)	116	121	119	<b>177</b>	133,25
( 90, 66)	140	<b>209</b>	149	<b>223</b>	180,25
(128, 54)	<b>240</b>	<b>242</b>	<b>241</b>	<b>244</b>	241,75
(218, 292)	116	<b>186</b>	<b>188</b>	<b>181</b>	167,75

TAB. 2.1: Niveaux de gris des pixels voisins aux pixels tests et leur moyenne.

La figure FIG. 2.1 donne un aperçu visuel de ces niveaux de gris.

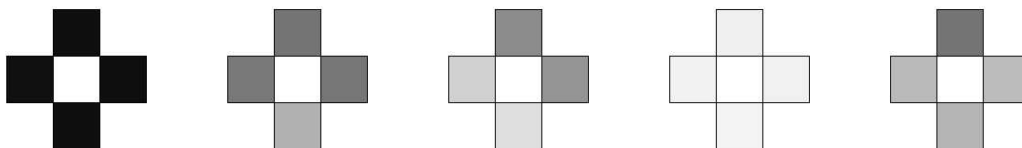


FIG. 2.1: Pixels tests et leurs 4 voisins (dans le même ordre que dans TAB. 2.1).

Sur ce tableau, les pixels voisins considérés comme « **chauds** » sont en gras, les autres étant réputés « froids ». Un pixel **chaud** est défini comme un pixel d'intensité « élevée », tandis qu'un pixel froid

<sup>2</sup>Bibliothèque de traitement d'image en langage C développée au laboratoire GIP.

<sup>3</sup>Types propres à GAMMAG<sup>1</sup>.

comme un pixel d'intensité « faible ». En pratique il s'agit d'une différence d'au moins 50 niveaux de gris entre les pixels considérés comme froids et ceux considérés comme **chauds**.

Le choix de ces 5 pixels de tests répond à la logique suivante : ils ont tous une configuration spécifique de « voisins ». Le pixel (128, 54) se trouve dans une région chaude avec 4 voisins **chauds**, (218, 292) a 3 voisins **chauds**, (90, 66) a 2 voisins **chauds**, (63, 64) a un voisin **chaud**, et (30, 260) est dans une région froide avec 4 voisins froids.

Pour cette implémentation de l'algorithme, en ce qui concerne le choix des paramètres des fonctions de potentiel, les préconisations de [Samson 02] ont été suivies :

$$T = T_\phi = T_\chi = 10,$$

(pour rappel :  $\phi(u) = \chi(u) = \sqrt{u^2 + T^2} - T$ ) et  $\lambda = 0,3$ .

### 2.2.2 Résultats

Pour chacun des 5 pixels tests, les deux versions de l'algorithme ont été successivement appliquées. Le temps de calcul est récapitulé dans le tableau TAB. 2.2.

On constate tout d'abord une différence de 8 minutes entre ces 2 versions, soit entre 20 et 25 % de temps en moins pour le cas entier par rapport au cas réel. Ce gain se fait toutefois au détriment de la précision des calculs (ceci peut être constaté sur les figures qui suivent).

D'autre part, pour les deux cas où les pixels sont dans une zone homogène, c'est-à-dire soit intégralement chaude (les 4 voisins sont chauds, (128, 54)), soit totalement froide (les 4 voisins sont froids, (30, 260)), le temps de calcul est de 20 à 25 % plus long que dans les cas non homogènes.

Rappel : ces résultats ont été obtenus dans le cas d'une seule cible présente dans l'image originale.

	Cas entier	Cas réel	Différence
( 30, 260)	34 min 11 sec	42 min 19 sec	8 min 08 sec
( 63, 64)	25 min 01 sec	33 min 04 sec	8 min 03 sec
( 90, 66)	26 min 28 sec	34 min 14 sec	7 min 46 sec
( 128, 54)	34 min 10 sec	42 min 34 sec	8 min 24 sec
(218, 292)	25 min 35 sec	33 min 26 sec	7 min 51 sec

TAB. 2.2: Temps de calcul pour 256 réalisations de l'algorithme (cas entier et réel).





FIG. 2.2: Localisation des 5 pixels dans l'image IR utilisée pour les simulations.

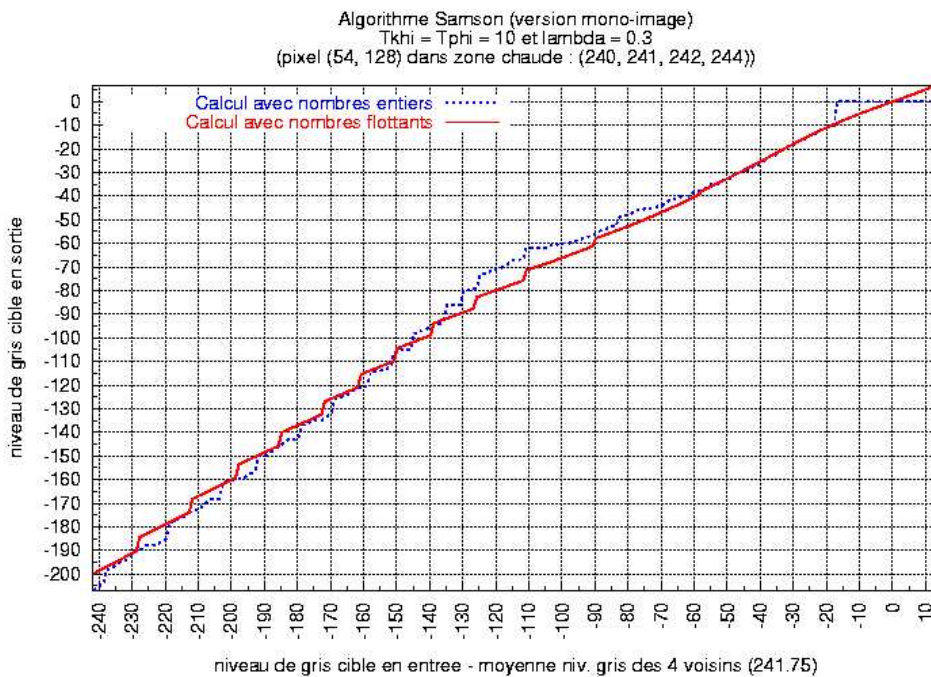


FIG. 2.3: Comparaison résultats cas entier/réel - Pixel (128, 54), zone chaude.

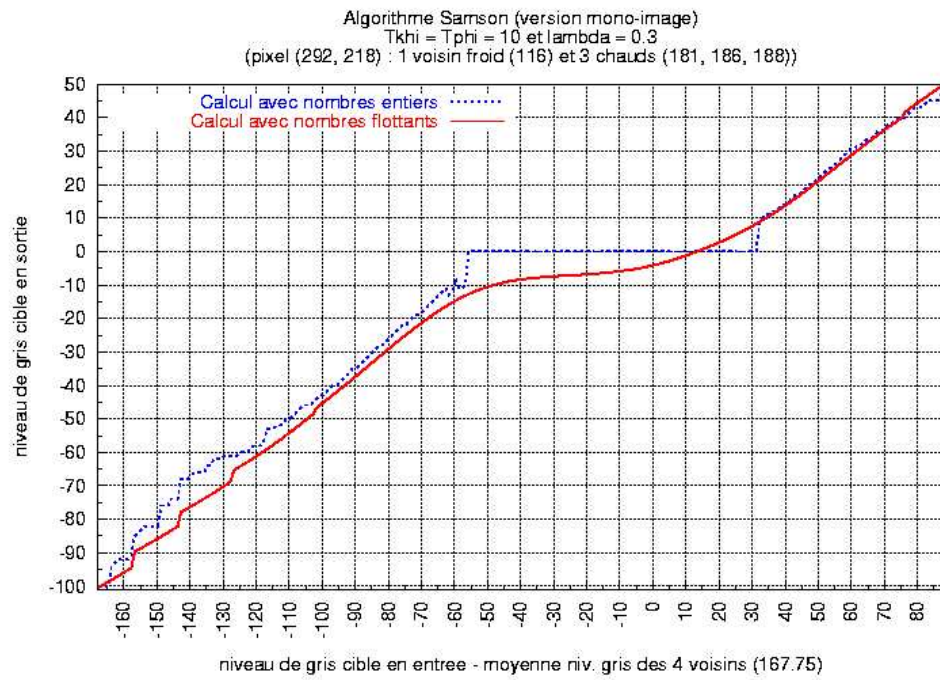


FIG. 2.4: Comparaison résultats cas entier/réel - Pixel (218, 292), 3 voisins chauds.

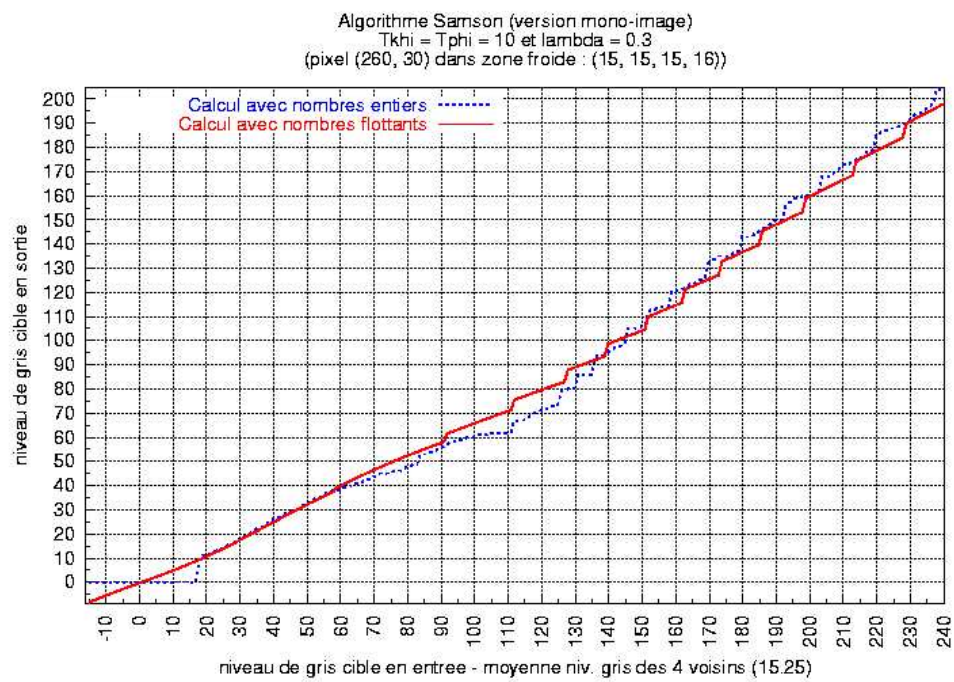


FIG. 2.5: Comparaison résultats cas entier/réel - Pixel (30, 260), zone froide.

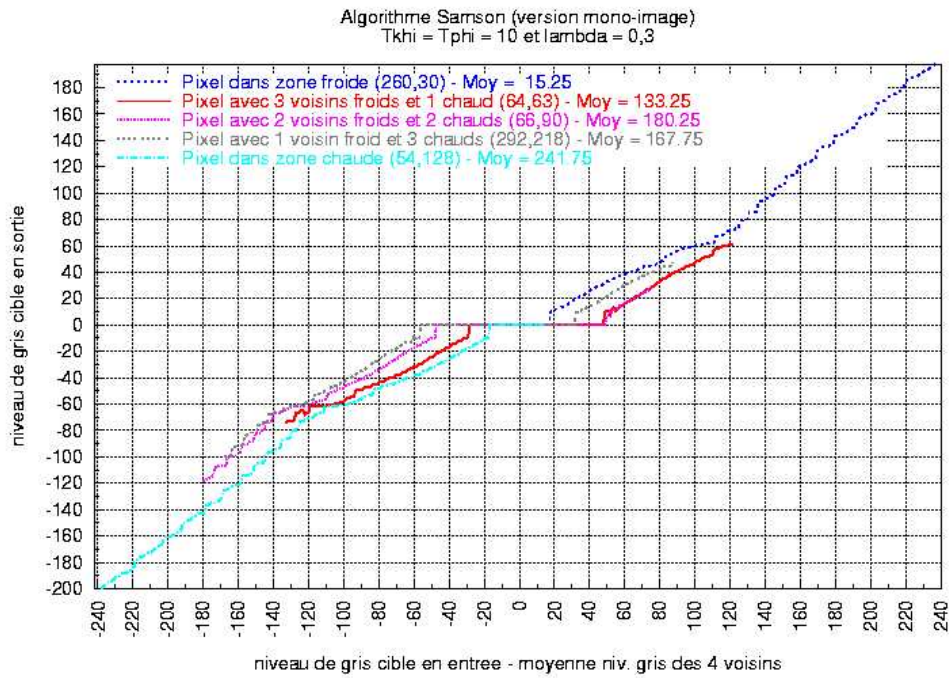


FIG. 2.6: Superposition des 5 cas entiers.

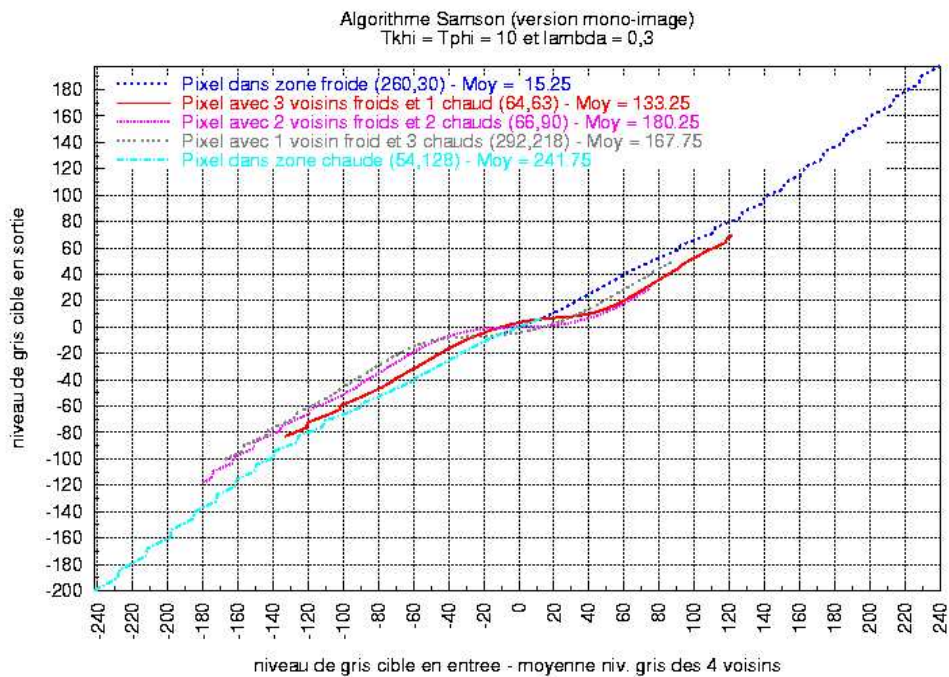


FIG. 2.7: Superposition des 5 cas réels.

## 2.3 Interprétations et conclusions

Ces résultats montrent une certaine robustesse de l'algorithme de Samson dans le cas mono-image. Pour une abscisse de zéro (c'est-à-dire pour un niveau d'entrée du pixel test égal à la moyenne de ses 4 voisins), le niveau de sortie de ce pixel est égal à zéro, plus ou moins 5 selon les cas. Ce résultat est prévisible : si le pixel cible a un niveau de gris similaire à celui de ses voisins, il n'y a pas de raison qu'il soit détecté comme étant une « cible ».

La deuxième remarque importante à propos de ces simulations concerne le niveau de gris détecté pour les 4 pixels voisins du pixel cible. Cette information n'est pas disponible à partir des figures FIG. 2.3, FIG. 2.4 et FIG. 2.5 qui montrent une comparaison des performances des deux modes de codage des données, mais c'est un aspect étudié lors des simulations. Pour les 5 cas traités (cf. TAB. 2.1), dans le cas d'un codage réel des données, le niveau de gris des 4 pixels voisins n'a jamais dépassé l'intervalle  $[-9; 9]$  (pour rappel, on a  $t = g - f$ , et ces niveaux sont observés sur l'image  $t$ ). Les niveaux extrêmes sont observés lorsque le pixel cible a un seul voisin chaud (respectivement froid) et trois voisins froids (respectivement chauds). On observe un niveau relativement élevé (7) dans le cas de deux voisins chauds et deux voisins froids. Dans les cas d'une cible dans une zone homogène, les pixels voisins ont un niveau compris généralement dans  $[-3; 3]$ , quel que soit le niveau de gris initial de la cible.

Pour le cas d'un codage entier des données, le niveau de gris des voisins reste en général égal à 0, ou vaut très occasionnellement 1 (afin de rester dans l'ensemble  $[0; 255]$ , puisqu'on utilise un codage des valeurs par type **char**).

À partir de la remarque précédente, et en se référant à FIG. 2.7, un seuil de détection égal à 10 en valeur absolue permettra de détecter tous les pixels ayant un contraste inférieur à -50 ou supérieur à 55. Il est entendu implicitement par

« **contraste** » : le niveau de gris du pixel cible moins la moyenne de ses 4 voisins.

Un choix de seuil plus élevé, égal à 20 en valeur absolue par exemple, permettra de détecter les pixels ayant un contraste inférieur à -70 ou supérieur à 70 (ou 27,3 % de la dynamique de l'image observée). Pour le cas entier, FIG. 2.6 montre que la valeur du pixel en sortie de l'algorithme est égale à zéro lorsque son contraste en entrée est supérieur à -55 ou inférieur à 55 (ou 21,5 % de la dynamique de l'image observée).

Un aspect non étudié ici, mais à approfondir pour la suite de l'étude, concerne le taux de fausses alarmes à un niveau de seuil fixé. Cet aspect est essentiel car si on ne prend pas en compte le taux de fausses alarmes, on peut avoir un taux de détection de 100 %.

**En conclusion de ces expérimentations, une cible sera bien détectée lorsque son contraste sera d'au moins 21,5 % de la dynamique de l'image en valeur absolue.**



# Chapitre 3

## Recalage d'images par méthodes de corrélation de phase

### Sommaire

---

<b>3.1</b>	<b>Introduction</b>	<b>31</b>
<b>3.2</b>	<b>État de l'art sur les méthodes de recalage</b>	<b>32</b>
3.2.1	Techniques basées sur le gradient	32
3.2.2	Méthodes pixel-récurrentes	34
3.2.3	Techniques par correspondance de blocs	35
3.2.4	Méthodes du domaine fréquentiel	36
3.2.5	Techniques à base d'ensemble de niveaux	36
<b>3.3</b>	<b>Recalage d'images par corrélation de phase</b>	<b>38</b>
3.3.1	Principe de fonctionnement	38
3.3.2	Extension au recalage sous-pixelique	39
3.3.3	Extension avec utilisation du gradient	43
3.3.4	Équation de recalage considérant les bords	44
<b>3.4</b>	<b>Exemple d'applications, résultats</b>	<b>46</b>
3.4.1	Cas d'un déplacement précis réalisé à partir d'une fonction numérique	46
3.4.2	Cas d'un déplacement obtenu par simulation	50
3.4.3	Cas d'un déplacement réel sur images de nuages	51
3.4.4	Cas d'un grand déplacement réel vertical	55
<b>3.5</b>	<b>Extension de ce modèle à la rotation et à l'homothétie</b>	<b>57</b>
<b>3.6</b>	<b>Conclusions</b>	<b>58</b>

---

**Table des figures**

3.1	Exemple d'un déplacement sous-pixelique sur une image de 9 pixels. . . . .	45
3.2	Image numérique utilisée pour l'évaluation des méthodes de recalage. . . . .	48
3.3	Différence entre les images obtenues en prenant $q = 1$ et $q = 11$ . . . . .	48
3.4	Résultats d'estimation des déplacements sous la forme de constellation. . . . .	49
3.5	Effet du recalage observé à partir d'images de différence. . . . .	49
3.6	Exemple de déplacement sous-pixelique connu. . . . .	52
3.7	Image translatée corrigée par une translation de $(-10,17, -12,70)$ . . . . .	52
3.8	(a) Différences entre images réelle et translatée, (b) et entre images réelle et recalée. . .	52
3.9	Deux images IR réelles montrant un déplacement de nuages. . . . .	53
3.10	Image recalée et normes du gradient des deux images de la FIG. 3.9. . . . .	53
3.11	(a) Différences entre 1 <sup>ère</sup> et 350 <sup>e</sup> image, (b) et entre 1 <sup>ère</sup> et 350 <sup>e</sup> image recalée. . . . .	53
3.12	Zoom sur le point de recalage et valeur des niveaux de gris des pixels voisins. . . . .	54
3.13	1 <sup>ère</sup> et 21 <sup>e</sup> image d'une séquence montrant un mouvement vertical. . . . .	56
3.14	Image recalée et normes du gradient des deux images de la FIG. 3.13. . . . .	56
3.15	(a) Différences entre 1 <sup>ère</sup> et 21 <sup>e</sup> image, (b) et entre 1 <sup>ère</sup> et 21 <sup>e</sup> image recalée. . . . .	56

**Liste des tableaux**

3.1	Moyenne et maximum des erreurs (réalisés à partir de 400 estimations, exprimés en %). 48
-----	--

Ce chapitre présente les méthodes fréquentielles de recalage par corrélation de phase (CP), et l'évaluation de leur précision. La méthode principale, et son extension qui utilise les gradients des images à la place des images elles-mêmes, sont bien expliquées dans [Frooosh 02, Argyriou 03]. Ces méthodes sont intéressantes à plusieurs points de vue : pour leur bonne précision sous-pixellique, c'est-à-dire inférieure à la taille d'un pixel, la vitesse de leur exécution sur un ordinateur, et la facilité de leur mise en pratique. L'évaluation de ces deux méthodes que je reprends ici a fait l'objet de la communication [Humbly 05a].

Après une première partie consacrée à l'introduction du recalage des images, un **état de l'art** sur les principales familles de méthodes de recalage d'images est exposé dans la section **3.2**.

La **méthode de recalage par CP** et son extension sont présentées dans la partie **3.3**. Cette section est également l'occasion d'exposer l'**équation de recalage** utilisée dans le cas d'une approximation bilinéaire.

Des résultats de **simulation** sont ensuite proposés dans la partie **3.4**.

Quelques éléments permettant d'étendre cette méthode au traitement de **modèles de mouvement plus généraux** sont présentés dans la section **3.5**.

Enfin, des **conclusions** sur l'étude menée dans ce chapitre sont exposées dans la partie **3.6**.

## 3.1 Introduction

Un des problèmes fondamentaux et connexes à la problématique de la super-résolution (SR) est celui du recalage des images (*image registration*). Le recalage est défini comme la manière de faire concorder deux ou plusieurs images montrant une même scène à des instants distincts, depuis des points de vue différents, ou à partir de capteurs différents. Ainsi, les images observées doivent être utilisées dans un repère commun si on veut pouvoir utiliser les informations non redondantes contenues dans chacune d'elles. Pour avoir une bonne reconstruction par SR, il est indispensable de connaître précisément la transformation qui permet d'aller d'une image BR à une autre.

Les transformations possibles entre deux images peuvent être de type global et ainsi une seule transformation s'applique à tous les pixels de l'image, ou de type local. Dans ce dernier cas, des transformations différentes peuvent être appliquées à différentes régions homogènes de la scène observée. Le champ des transformations possibles entre deux images montrant une même scène consiste en la translation, la rotation de l'axe de l'appareil d'acquisition, et l'homothétie si les prises de vue ont été faites à différents niveaux de résolution (en utilisant différents niveaux de zoom lors de l'acquisition).

Le contexte d'application choisi (cf. **1.1**) permet de limiter le champ des transformations possibles entre deux images de la séquence d'observation à la simple translation globale. C'est pourquoi la **méthode de recalage d'images par CP** a été utilisée pour évaluer le mouvement global de translation dans une scène. Il s'agit de la méthode la plus simple d'un point de vue mise en place pratique, tout en restant robuste au bruit. Elle utilise une propriété de la transformée de Fourier (TF). Il est intéressant de noter que la TF peut être réalisée efficacement en hardware, et en utilisant la TF rapide (TFR).



## 3.2 État de l'art sur les méthodes de recalage

Plusieurs états de l'art très complets ont été publiés [Brown 92, Dufaux 95, Zitová 03]. Celui présenté dans cette partie sera sommaire. Il a pour objectif de décrire les principales méthodes utilisées pour le recalage des images, sans détailler toutes les extensions existantes. Le lecteur peut se reporter aux articles mentionnés ci-dessus pour obtenir des précisions.

Le recalage des images consiste d'abord à estimer le mouvement présent entre deux images, puis à utiliser une fonction de transformation permettant de faire concorder au mieux les deux images dans un repère commun. La partie présentée ici traitera uniquement l'aspect estimation du mouvement.

Les techniques d'estimation du mouvement peuvent être divisées en cinq groupes principaux :

- Les techniques basées sur le gradient se fondent sur l'hypothèse que l'intensité lumineuse (ou luminance) des images est invariante au cours du mouvement. Elles résolvent une équation de contrainte de flux optique qui mène à un champ de mouvement dense [Barron 94].
- Les techniques pixel-récurrentes peuvent être considérées comme un sous ensemble des techniques à base de gradients. Elles minimisent récursivement la contrainte spatio-temporelle [Musmann 85].
- Les techniques par correspondance de blocs (*block matching*) se basent sur une recherche de correspondance de blocs entre deux images, dans le but de minimiser une mesure de disparité [Jain 81].
- Les techniques du domaine fréquentiel utilisent les relations existantes entre les coefficients des différentes images dans un domaine transformé (de Gabor ou Fourier) [Fleet 90].
- Enfin, il existe des méthodes basées sur les ensembles de niveaux. Un ensemble de niveau représente un groupe de pixels ayant la même intensité dans une image. Le but de ces méthodes est de faire correspondre des sections de deux images en utilisant des mesures de surface et de moments d'inertie [Monasse 00b].

Dans la suite, l'intensité au pixel  $\mathbf{r} = (x, y)$  et à l'instant  $t$  d'une image tirée d'une séquence vidéo s'écrit  $I(\mathbf{r}, t)$ , et  $\mathbf{s} = (d_x, d_y)$  est le déplacement pendant l'intervalle  $\Delta t$ .

Les techniques présentées reposent parfois sur l'hypothèse qu'un changement dans l'intensité de  $I(\mathbf{r}, t)$  est seulement dû au déplacement  $\mathbf{s}$ . Celle-ci s'écrit :  $I(\mathbf{r}, t) = I(\mathbf{r} - \mathbf{s}, t - \Delta t)$ .

La différence entre deux images espacées de  $\Delta t$  dans le temps s'écrit :

$$\Delta_I(\mathbf{r}, \mathbf{s}, t) = I(\mathbf{r}, t) - I(\mathbf{r} - \mathbf{s}, t - \Delta t). \quad (3.1)$$

### 3.2.1 Techniques basées sur le gradient

Les méthodes à base de gradient reposent sur l'hypothèse que la luminance est invariante au cours du mouvement. Le développement limité en série de Taylor de  $I(\mathbf{r} - \mathbf{s}, t - \Delta t)$  donne :

$$I(\mathbf{r} - \mathbf{s}, t - \Delta t) = I(\mathbf{r}, t) - \mathbf{s} \cdot \nabla I(\mathbf{r}, t) - \Delta t \frac{\delta I(\mathbf{r}, t)}{\delta t} + \text{termes d'ordre plus élevés} \quad (3.2)$$

où  $\nabla \cdot = \left[ \frac{\delta \cdot}{\delta x}, \frac{\delta \cdot}{\delta y} \right]$  est l'opérateur gradient.

Dans le cas d'une approximation du premier ordre, en supposant que la limite  $\Delta t \rightarrow 0$ , et en définissant le vecteur de mouvement  $\mathbf{v} = (v_x, v_y) = \frac{\mathbf{s}}{\Delta t}$ , on obtient :

$$\mathbf{v} \cdot \nabla I(\mathbf{r}, t) + \frac{\delta I(\mathbf{r}, t)}{\delta t} = 0. \quad (3.3)$$

L'équation (3.3) est connue comme l'**équation de contrainte spatio-temporelle** ou l'**équation de contrainte du flux optique** [Horn 81].

Le changement d'intensité dû au mouvement dans l'image en un point ne donnant qu'une contrainte (3.3), et le vecteur mouvement au même point ayant deux composantes, le champ de mouvement (ou **flux optique**) ne peut pas être calculé sans une contrainte supplémentaire.

En réalité, l'équation (3.3) ne peut que déterminer la projection de  $\mathbf{v}$  sur  $\nabla I$ , c'est-à-dire la composante de  $\mathbf{v}$  parallèle au gradient d'intensité. Ainsi, une contrainte additionnelle doit être introduite pour régulariser le problème mal-posé, et résoudre le flux optique.

Dans [Horn 81], Horn et Schunck introduisent une contrainte de douceur. Celle-ci consiste à minimiser la norme au carré du gradient du flux optique :

$$\left(\frac{\delta v_x}{\delta x}\right)^2 + \left(\frac{\delta v_x}{\delta y}\right)^2 \quad \text{et} \quad \left(\frac{\delta v_y}{\delta x}\right)^2 + \left(\frac{\delta v_y}{\delta y}\right)^2.$$

De cette façon, le flux optique est obtenu en minimisant le terme d'erreur suivant :

$$\int \int \left( \mathbf{v} \cdot \nabla I + \frac{\delta I}{\delta t} \right)^2 + \lambda^2 \left( \left(\frac{\delta v_x}{\delta x}\right)^2 + \left(\frac{\delta v_x}{\delta y}\right)^2 + \left(\frac{\delta v_y}{\delta x}\right)^2 + \left(\frac{\delta v_y}{\delta y}\right)^2 \right) dx dy$$

où  $\lambda^2$  est un facteur de pondération. Ce problème de minimisation est résolu à l'aide du calcul variationnel, et une procédure itérative de Gauss-Seidel.

De nombreuses variations de cet algorithme ont été proposées. La contrainte de douceur a été remplacée par une hypothèse d'uniformité locale par Lucas et Kanade dans [Lucas 85].

Dans [Nagel 83], Nagel développe la série de Taylor de (3.2) jusqu'aux termes de deuxième ordre.

L'approche présentée traite des images à une résolution donnée, observées à la même échelle. Par extension, des schémas hiérarchiques basés sur l'équation de contrainte spatio-temporelle ont été développés dans [Glazer 84, Enkelmann 88].

Des discussions détaillées traitant des variations de la méthode de Horn et Schunck peuvent être trouvées dans [Nagel 86, Aggarwal 88, Barron 94].

Toutes ces techniques résultent en un **champ de mouvement dense** (c'est-à-dire en un **vecteur mouvement par pixel**). C'est un résultat intéressant lorsqu'on s'intéresse à une **estimation locale** du mouvement dans une image.

### 3.2.2 Méthodes pixel-récurrentes

Les techniques pixel-récurrentes minimisent l'erreur de prédiction, ou l'image de différence  $\Delta_I$  de (3.1). La récurrence est habituellement réalisée sur une base pixel par pixel, produisant un champ vectoriel de mouvement dense. Ces méthodes ont été conçues avec une contrainte de très faible complexité calculatoire.

Elles peuvent être considérées comme un sous ensemble des techniques basées sur le gradient, et dans lequel la contrainte spatio-temporelle est minimisée récursivement.

Le premier algorithme pixel-récurrent a été proposé par Netravali et Robbins dans [Netravali 79]. Dans cette méthode,  $\Delta_I^2$  est minimisée de manière récursive par une technique de descente à pas fixe :

$$\mathbf{s}^{(n+1)} = \mathbf{s}^{(n)} - \frac{\lambda}{2} \nabla_{\mathbf{s}} \Delta_I^2(\mathbf{r}, \mathbf{s}^{(n)}, t) \quad (3.4)$$

avec un gain constant  $\lambda > 0$ , et où  $n$  représente l'index d'itération. À partir de la définition de  $\Delta_I$  dans (3.1), on a :

$$\nabla_{\mathbf{s}} \Delta_I^2(\mathbf{r}, \mathbf{s}, t) = 2 \Delta_I(\mathbf{r}, \mathbf{s}, t) \cdot \nabla_{\mathbf{r}} I(\mathbf{r} - \mathbf{s}, t - \Delta t). \quad (3.5)$$

En substituant (3.5) dans (3.4), la mise à jour du vecteur déplacement devient :

$$\mathbf{s}^{(n+1)} = \mathbf{s}^{(n)} - \lambda \Delta_I(\mathbf{r}, \mathbf{s}^{(n)}, t) \cdot \nabla_{\mathbf{r}} I(\mathbf{r} - \mathbf{s}^{(n)}, t - \Delta t). \quad (3.6)$$

Les performances des techniques pixel-récurrentes dépendent fortement de la manière dont on calcule le terme de mise à jour dans (3.6).

Dans [Netravali 79], l'itération de  $n$  à  $n + 1$  est réalisée soit à l'emplacement d'un pixel  $\mathbf{r}$ , soit d'un pixel à son voisin. Pour faire disparaître l'effet du bruit, l'algorithme peut être étendu en calculant une moyenne du terme estimé sur plusieurs pixels.

Des améliorations de cette méthode ont été proposées, par exemple, dans [Cafforio 83, Walker 84]. En comparaison à [Netravali 79], le gain  $\lambda$  est substitué par des variables qui s'adaptent aux statistiques locales de l'image et par conséquent, les algorithmes convergent mieux.

Une description détaillée des diverses méthodes pixel-récurrentes et des résultats comparatifs sont proposés dans [Musmann 85].

Les algorithmes pixel-récurrents ont deux inconvénients.

D'abord, en général, la fonction d'erreur à minimiser contient plusieurs minimums locaux, et la procédure itérative peut converger vers l'un d'entre eux au lieu du minimum global. Et puis, ces méthodes sont très sensibles au bruit.

Enfin, des grands déplacements et des discontinuités dans le champ de mouvement ne peuvent pas être facilement estimés.

### 3.2.3 Techniques par correspondance de blocs

Les techniques par correspondance de blocs sont basées sur une recherche de correspondance entre les blocs de deux images, en minimisant une mesure de disparité [Jain 81, Musmann 85].

À l'origine, ces méthodes ont été développées spécifiquement pour le codage et la compression des séquences d'images, et elles sont largement utilisées dans ce domaine.

Dans ces techniques d'estimation du mouvement, l'image est décomposée en blocs et le même vecteur déplacement est attribué à tous les pixels d'un bloc. Le modèle de mouvement fait habituellement l'hypothèse qu'une image est composée d'objets rigides qui suivent un mouvement de translation.

Bien que ce modèle soit restrictif, il est justifié par le fait qu'un mouvement complexe peut être décomposé en une somme de mouvements de translation.

Pour chaque bloc, le vecteur déplacement est évalué en faisant correspondre le contenu informationnel d'une fenêtre de mesure  $\mathcal{W}$  avec celui d'une fenêtre de mesure correspondante dans une région de recherche  $\mathcal{S}$ , placée dans la précédente image. Il faut alors chercher l'endroit spatial qui minimise le critère de concordance :

$$\mathbf{s} = \arg \min_{\mathbf{s} \in \mathcal{S}} \left[ \sum_{\mathbf{r} \in \mathcal{W}} \phi(I(\mathbf{r}, t) - I(\mathbf{r} - \mathbf{s}, t - \Delta t)) \right]$$

où les mesures de distance  $\phi$  les plus largement utilisées sont :

- la norme quadratique  $\phi(x) = x^2$ ,
- et la valeur absolue  $\phi(x) = |x|$ .

Trouver un minimum absolu pour le critère de correspondance peut uniquement être garanti en réalisant une recherche exhaustive sur une série de déplacements discrets, dans un rayon de déplacement maximum. Ce type de technique est appelée « recherche complète par correspondance de blocs » (*full-search block matching*).

Afin de réduire la charge calculatoire de l'algorithme de recherche complète, des techniques de recherches rapides ont été proposées [Jain 81, Srinivasan 85, Liu 93]. Ces méthodes réduisent les temps de calcul de manière significative en comparaison à la méthode de référence. Cependant, la convergence vers le minimum global est garantie uniquement lorsque le critère de correspondance est une fonction monotone de  $\mathbf{s}$ .

Les algorithmes par correspondance de blocs ont été initialement conçus pour estimer des déplacements avec une précision d'un pixel. Toutefois, une précision sous-pixellique peut être obtenue. Pour cela, l'intensité de l'image doit être interpolée dans des régions fractionnelles de pixels.

En pratique, cette étape est implémentée dans un traitement postérieur où les vecteurs de déplacement de précision un pixel sont raffinés à une précision fractionnelle de pixel. Ce post-traitement augmente significativement la complexité calculatoire.

Les méthodes décrites jusqu'ici dans cette partie traitent des images à une résolution donnée, observées à la même échelle. Pour réduire la complexité calculatoire, et pour prendre en considération le caractère multi-échelle du mouvement dans une scène, des algorithmes hiérarchiques [Anandan 87, Bierling 88] et multi-grilles [Dufaux 94] ont été proposés.

Dans la technique standard par correspondance de blocs, le mouvement est limité à la translation. Un modèle affine pour la correspondance d'image est proposé dans [Fuh 91]. Avec ce modèle, chaque bloc suit une transformation affine au lieu d'une translation dans la méthode de base.

Un algorithme généralisé par correspondance de blocs a été proposé dans [Seferidis 93]. Ce modèle inclut des mouvements complexes tels que la rotation ou des déformations non-linéaires.

Bien entendu, un modèle de mouvement plus complexe entraîne une estimation de mouvement plus précise. Néanmoins cela a pour inconvénient d'augmenter considérablement la complexité calculatoire, et une estimation précise des paramètres du mouvement devient difficile.

D'autre part, les résultats obtenus avec ces techniques d'estimation ne sont pas toujours très satisfaisants en comparaison au mouvement réel dans la scène.

### 3.2.4 Méthodes du domaine fréquentiel

Les techniques du domaine fréquentiel se basent sur un modèle de mouvement, et elles utilisent les relations entre les coefficients des images dans le domaine « transformé » (de Gabor ou de Fourier).

Fleet et Jepson se sont intéressés à l'utilisation de la transformée de Gabor dans la résolution de problème de recalage dans [Fleet 90, Fleet 91].

La méthode de recalage par corrélation de phase (CP) est intéressante pour sa bonne précision sous-pixellique dans le cas d'une translation, sa vitesse d'exécution sur un ordinateur grâce à la transformée de Fourier rapide (TFR), et la facilité de sa mise en pratique. Elle illustre bien ce type de méthode.

Cette méthode a été choisie pour réaliser le recalage dans le processus de SR décrit dans le **Chapitre 5**. Elle est décrite en détails, ainsi qu'une extension intéressante utilisant le gradient des images au lieu des images elles-mêmes, dans la section **3.3** de ce chapitre, p. 38.

D'autres extensions, permettant d'étendre cette famille de méthodes à des mouvements plus complexes de rotation et d'homothétie, sont décrites dans la partie **3.5**, p. 57.

Contrairement aux méthodes précédemment présentées dans cet état de l'art, la méthode par CP ne permet pas d'obtenir un champ de mouvement dense. Toutefois, elle peut être utilisée sur des sections d'images de manière à estimer des paramètres de mouvement locaux.

### 3.2.5 Techniques à base d'ensemble de niveaux

Les méthodes basées sur les ensembles de niveaux sont présentées pour compléter cet état de l'art. Un ensemble de niveau  $\mathcal{L}_n$  représente un groupe de pixel ayant la même intensité  $n$  dans une image  $I$  :

$$\mathcal{L}_n(I) = \{\mathbf{r} \in \mathcal{R}; I(\mathbf{r}) = n\}.$$

Le but de ces méthodes est de faire correspondre des sections de deux images en utilisant des mesures de surface et de moments d'inertie. Monasse présente un travail de référence utilisant cette approche [Monasse 00b].

La méthode proposée dans cet article peut être qualifiée de semi-locale dans la mesure où elle ne permet pas d'obtenir un champ de mouvement dense.

L'idée principale est de faire correspondre les intensités des image en se basant sur le voisinage, c'est-à-dire en se basant sur des régions fermées, dont le niveau de gris se situe entre deux valeurs. Ces régions sont appelées **sections**.

Si le contraste d'une image est modifié, alors l'ensemble de toutes les sections est globalement invariant : on parle d'**invariance par contraste**.

Un changement de contraste est modélisé comme une fonction réelle continue  $c$  strictement croissante. Si  $I_1$  et  $I_2$  sont deux images montrant une même scène avec des contrastes différents, alors une écriture de l'altération de  $I_1$  en  $I_2$  est  $I_2 = c(I_1)$ .

Les ensembles de niveaux supérieurs et inférieurs de l'image  $I$  sont définis formellement comme :

$$\chi^\mu(I) = \{\mathbf{r}; I(\mathbf{r}) \leq \mu\} \quad \text{et} \quad \chi_\lambda(I) = \{\mathbf{r}; I(\mathbf{r}) \geq \lambda\},$$

et les **sections** par  $\chi_\lambda^\mu = \chi^\mu \cap \chi_\lambda$ .

On peut voir que :

$$\chi^{c(\mu)}(c(I)) = \chi^\mu(I) \quad \text{et} \quad \chi_{c(\lambda)}(c(I)) = \chi_\lambda(I),$$

ce qui montre que ces ensembles de caractéristiques sont invariants par contraste de manière globale. Les composantes fermées, ou ensembles de niveaux supérieurs et inférieurs, sont appelées « **formes** » (*shapes*).

Afin de réaliser une correspondance rapide, des caractéristiques globales sont calculées pour chaque forme : leur surface ( $A_S$ ), leur barycentre (moments d'ordre 1), et leurs moments d'ordre 2. Ces moments peuvent être combinés pour obtenir des caractéristiques invariante par translation et rotation.

On considère la matrice d'inertie  $I_S$  propre à la forme  $S$  :

$$I_S = \begin{pmatrix} \mu_S^{2,0} & \mu_S^{1,1} \\ \mu_S^{1,1} & \mu_S^{0,2} \end{pmatrix} \quad \text{avec} \quad \mu_S^{i,j} = \int_S (x - x_S)^i (y - y_S)^j dx dy,$$

et où  $(x_S, y_S)$  est le barycentre de  $S$ .

Soit  $R_\theta S + \mathbf{s}$  la région  $S$  translatée de  $\mathbf{s}$  et ayant subi une rotation d'un angle  $\theta$ .

Alors  $I_{R_\theta S + \mathbf{s}} = R_\theta I_S R_{-\theta}$  et ainsi,

- $|I_{R_\theta S + \mathbf{s}}| = |I_S|$  où  $|\cdot|$  représente le déterminant,
- et  $\text{tr}(I_{R_\theta S + \mathbf{s}}) = \text{tr}(I_S)$  où  $\text{tr}(\cdot)$  représente la trace.

Deux formes  $S_1$  et  $S_2$  extraites respectivement de  $I_1$  et  $I_2$  sont considérées comme **candidats à une correspondance** si :

$$S_1 \leftrightarrow S_2 \quad \Leftrightarrow \quad \begin{cases} A_{S_1} & \cong & A_{S_2} \\ |I_{S_1}| & \cong & |I_{S_2}| \\ \text{tr}(I_{S_1}) & \cong & \text{tr}(I_{S_2}) \end{cases}$$

Une stricte égalité n'est pas imposée afin de permettre de petites déformations des formes, et pour éviter les effets de la quantification. Une différence typique entre ces caractéristiques est comprise entre 10 et

20 %.

Enfin une procédure de vote permet d'éliminer les « mauvaises » correspondances dues au déplacement de petits objets. Il permet également de sélectionner les paramètres de rotation et translation qui ont reçu le plus de votes.

Une précision sous-pixellique peut être obtenue avec cette méthode.

Une extension de la méthode présentée a été proposée par Cao et Collin dans [Cao 01]. Le principe de leur méthode consiste à représenter une image par sa carte topographique, qui peut être représentée par un arbre [Monasse 00a]. Chaque nœud de l'arbre représente une structure dans l'image. La comparaison de deux images se fait par celle de leur carte topographique [Guichard 97], par le biais de leur arbre associé.

Le calcul de cet arbre est réalisé à partir de la transformée rapide en ensemble de niveaux (*fast level set transform*) [Monasse 00a, Monasse 00b].

Cette extension permet d'évaluer une translation et une rotation globale, ainsi qu'une homothétie.

### 3.3 Recalage d'images par corrélation de phase

L'état de l'art présenté a montré différentes méthodes permettant d'évaluer les transformations inter-images. La plupart d'entre elles permettent d'obtenir un champ de mouvement dense, et elles peuvent nécessiter de longs calculs.

Le contexte de ce travail permet de se limiter à une translation globale de la scène. C'est pourquoi la méthode de corrélation de phase (CP) s'est avérée être un choix intéressant (précise, rapide en temps de calcul et robuste au bruit).

#### 3.3.1 Principe de fonctionnement

Une manière élégante d'aligner deux images, dont l'une est la translatée de l'autre, est d'utiliser la CP [Kuglin 75]. Cette méthode s'appuie sur la propriété de translation de la TF, parfois qualifiée de **théorème de déplacement** (*Shift Theorem*). Une translation est représentée par une exponentielle complexe dans le domaine de Fourier.

Soit  $I_1$  et  $I_2$  deux images qui diffèrent uniquement l'une de l'autre par un déplacement  $(d_x, d_y)$ , c'est-à-dire que l'on a en chaque pixel :

$$I_2(x, y) = I_1(x - d_x, y - d_y).$$

Soit  $\widehat{I}_1$  et  $\widehat{I}_2$  leur TF. Le théorème de déplacement donne la relation :

$$\widehat{I}_2(u, v) = \widehat{I}_1(u, v) e^{-j(u*d_x + v*d_y)}.$$

En d'autres termes, ces deux images ont la même amplitude dans le domaine de Fourier, mais surtout une différence de phase directement liée à leur déplacement. Cette différence de phase est équivalente

à la phase du spectre de puissance croisée (SPC, *cross-power spectrum*). Ainsi, en multipliant  $\widehat{I}_2$  par le conjugué de  $\widehat{I}_1$  et en divisant le tout par le module du produit de ces deux images on obtient :

$$\frac{\widehat{I}_2(u, v) \widehat{I}_1^*(u, v)}{|\widehat{I}_2(u, v) \widehat{I}_1^*(u, v)|} = e^{-j(u*d_x + v*d_y)}$$

En appliquant la TF inverse (TFI) à cette exponentielle complexe on trouve  $\delta(x - d_x, y - d_y)$ . Il s'agit d'une fonction de Dirac centrée sur la valeur du déplacement  $(d_x, d_y)$  : on l'appelle le point de recalage. Ces coordonnées se retrouvent aisément par recherche du maximum. En pratique la TF continue est remplacée par la TF discrète (TFD), et la fonction de Dirac devient la fonction de Kronecker (définie par  $\delta(0) = 1$ , et  $\delta(t) = 0$  pour  $t \neq 0$ ).

Puisque la différence de phase contribue de manière égale pour chaque fréquence, cette technique est particulièrement bien adaptée aux images qui ont du bruit dans une bande-passante étroite (on peut alors appliquer le principe de CP dans la portion basse fréquence du spectre). Il s'agit donc aussi d'une technique efficace pour des images obtenues sous différentes conditions d'éclairage, étant donné qu'habituellement les fonctions d'éclairage varient lentement, et sont donc concentrées aux fréquences spatiales basses. Cette propriété d'utiliser uniquement l'information de phase pour la corrélation est parfois qualifiée de blanchiment (*whitening*) de chaque image. Entre autres choses, le blanchiment est invariant aux changements linéaires dans l'intensité lumineuse et rend la mesure de corrélation relativement indépendante de la scène. La propriété la plus remarquable de la CP, en comparaison à la méthode classique de la corrélation croisée (*cross-correlation*), est la précision avec laquelle le pic de la fonction de corrélation peut être détecté. La méthode de CP donne un pic singulier au point de recalage, tandis que la corrélation croisée standard donne plusieurs pics, dont un pic principal qui n'est pas toujours centré exactement au bon endroit.

### 3.3.2 Extension au recalage sous-pixellique

La méthode précédente telle qu'elle est décrite dans [Kuglin 75], ou rappelée dans [Brown 92], suppose un déplacement pixellique entre les deux images. Dans le cas d'un déplacement sous-pixellique, cette méthode « brute » donnera la valeur entière la plus proche du déplacement réel. Par exemple pour un déplacement  $(1,65, 2,30)$ , on trouvera  $(2,0, 2,0)$ . Ce résultat limite considérablement les performances des traitements à suivre.

Foroosh, Zerubia et Berthod proposent une extension à cette méthode dans [Foroosh 02]. Leur modèle est basé sur l'**hypothèse que les images qui ont subi un déplacement sous-pixellique ont été, à l'origine, déplacées d'une valeur entière. Puis, elles ont été ramenées à une valeur sous-pixellique par le biais d'un sous-échantillonnage.** On s'intéresse donc au SPC normalisé de deux images sous-échantillonnées.

On pose que  $I_1(x, y) = I(x, y)$  et  $I_2(x, y) = I(x - x_0, y - y_0)$  avec  $(\mathbf{x}_0, \mathbf{y}_0)$  un vecteur de 2 valeurs entières.

On pose également que  $\widehat{I}_1(u, v) = \widehat{I}(u, v)$  et  $\widehat{I}_2(u, v) = \widehat{I}(u, v) e^{-j(u*x_0 + v*y_0)}$  sont leur TFD respective.



Ainsi, leur TFD correspondant au sous-échantillonnage par des **facteurs entiers M et N** respectivement le long des axes  $x$  et  $y$  sera donné par :

$$\widehat{I}_{s1}(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \widehat{I}\left(\frac{u+2\pi m}{M}, \frac{v+2\pi n}{N}\right)$$

et

$$\widehat{I}_{s2}(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \widehat{I}\left(\frac{u+2\pi m}{M}, \frac{v+2\pi n}{N}\right) \exp\left(-j\left(\frac{u+2\pi m}{M}x_0, \frac{v+2\pi n}{N}y_0\right)\right).$$

En posant

$$\widehat{h}_{mn}(u, v) = \frac{\widehat{I}\left(\frac{u+2\pi m}{M}, \frac{v+2\pi n}{N}\right)}{\sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \widehat{I}\left(\frac{u+2\pi m'}{M}, \frac{v+2\pi n'}{N}\right)}, \quad (3.7)$$

le SPC des images sous-échantillonnées est donné par :

$$\widehat{C}(u, v) = \frac{\widehat{I}_{s2}}{\widehat{I}_{s1}} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \widehat{h}_{mn}(u, v) \exp\left(-j\left(\frac{u+2\pi m}{M}x_0, \frac{v+2\pi n}{N}y_0\right)\right). \quad (3.8)$$

À partir de ces résultats on peut dire que le SPC normalisé de deux images sous-échantillonnées est, simplement, une décomposition polyphase [Vetterli 95] d'une impulsion unitaire filtrée. Ceci signifie que le SPC consiste en la somme de toutes les composantes polyphases d'une impulsion unitaire filtrée.

Malgré son apparence trompeuse, en absence de recouvrement de spectre (*aliasing*),  $\widehat{h}_{mn}(u, v)$  ne dépend pas de  $\widehat{I}(u, v)$ . Il s'agit en fait d'un filtre passe-bande idéal.

Ce modèle suppose donc désormais qu'il n'y a pas d'*aliasing*, ou qu'il est évité par un pré-filtrage. À partir de cette hypothèse, nous allons maintenant montrer que  $\{\widehat{h}_{mn}(u, v)\}_{m=0, \dots, M-1, n=0, \dots, N-1}$  sont des filtres passes-bandes idéaux quelle que soit l'entrée  $\widehat{I}(u, v)$ . Remarquons d'abord qu'en absence d'*aliasing* les supports spectraux de :

$$\left\{ \widehat{I}\left(\frac{u+2\pi m}{M}, \frac{v+2\pi n}{N}\right) \right\}_{m=0, \dots, M-1, n=0, \dots, N-1} \quad (3.9)$$

ne se superposent pas (le support spectral d'une fonction est le domaine en dehors duquel le spectre de la fonction est identiquement zéro). Ceci se montre par le fait que :

$$\left(\frac{u+2\pi m}{M}, \frac{v+2\pi n}{N}\right) \neq \left(\frac{u+2\pi m'}{M}, \frac{v+2\pi n'}{N}\right) \quad \text{si} \quad (m, n) \neq (m', n').$$

Par exemple, prenons  $u$  et  $v$  dans  $[0, 2\pi[$  et  $M = N = 2$ .

Lorsque  $(m, n) = (0, 0)$  on obtient :

$$\frac{u+2\pi m}{M} \in [0, \pi[ \quad \text{et} \quad \frac{v+2\pi n}{N} \in [0, \pi[$$

pour  $(m, n) = (1, 0)$  on a :

$$\frac{u + 2\pi m}{M} \in [\pi, 2\pi[ \quad \text{et} \quad \frac{v + 2\pi n}{N} \in [0, \pi[$$

pour  $(m, n) = (0, 1)$  on a :

$$\frac{u + 2\pi m}{M} \in [0, \pi[ \quad \text{et} \quad \frac{v + 2\pi n}{N} \in [\pi, 2\pi[$$

pour  $(m, n) = (1, 1)$  on a :

$$\frac{u + 2\pi m}{M} \in [\pi, 2\pi[ \quad \text{et} \quad \frac{v + 2\pi n}{N} \in [\pi, 2\pi[$$

Ceci implique que si après le sous-échantillonnage aucun recouvrement de spectre n'a lieu, alors les supports des fonctions  $\left\{ \hat{I}\left(\frac{u+2\pi m}{M}, \frac{v+2\pi n}{N}\right) \right\}$   $m \in \{0, 1\}$ ,  $n \in \{0, 1\}$  seront disjoints. En fait dans le cas général, les supports de ces fonctions sont obtenus par le morcellement total de la bande passante à 2 dimensions (2-D)  $[0, 2\pi[ \times [0, 2\pi[$  en  $M \times N$  supports rectangulaires égaux le long des deux axes en fréquence.

Maintenant, nous allons esquisser une simple preuve du fait que  $\hat{h}_{mn}(u, v)$  ne dépend pas de  $\hat{I}(u, v)$ . Soit  $u$  et  $v$  à valeurs dans  $[0, 2\pi[$ , et soit le support 2-D que l'on note  $B$ . Soit également  $B_{mn}$  les supports rectangulaires disjoints des fonctions définies par l'équation (3.9) (on a donc  $\hat{I}\left(\frac{u+2\pi m}{M}, \frac{v+2\pi n}{N}\right) = 0$  en dehors de  $B_{mn}$ , et  $\cup B_{mn} = B$ ). Puisque les supports spectraux sont disjoints, à l'intérieur du support  $B_{mn}$ , tous les termes du dénominateur de l'équation (3.7) sont nuls excepté le terme correspondant à  $(m', n') = (m, n)$ . D'autre part, en dehors du support  $B_{mn}$  le numérateur est nul. Ainsi, en repartant de l'équation (3.7) :

$$\hat{h}_{mn}(u, v) = \frac{\hat{I}\left(\frac{u + 2\pi m}{M}, \frac{v + 2\pi n}{N}\right)}{\sum_{m'=0}^{M-1} \sum_{n'=0}^{N-1} \hat{I}\left(\frac{u + 2\pi m'}{M}, \frac{v + 2\pi n'}{N}\right)}$$

on a :

$$\hat{h}_{mn}(u, v) = \begin{cases} \frac{\hat{I}\left(\frac{u + 2\pi m}{M}, \frac{v + 2\pi n}{N}\right)}{\hat{I}\left(\frac{u + 2\pi m}{M}, \frac{v + 2\pi n}{N}\right)} = 1 & \text{dans } B_{mn} \\ 0 & \text{en dehors de } B_{mn} \end{cases} \quad (3.10)$$

À partir de la réponse fréquentielle des filtres décrits dans l'équation (3.10), on peut déduire que le SPC de deux images sous-échantillonnées peut être vu comme la version sous-échantillonnée d'une impulsion unitaire filtrée, où le filtre a une réponse fréquentielle rectangulaire. En d'autres termes, la CP mène à un noyau (*kernel*) de Dirichlet 2-D sous-échantillonné, lequel peut être approximé par une fonction en sinus cardinal (*sinc*) 2-D.

Le principal résultat de ce qui précède peut être résumé ainsi : en général, dans le cas discret, la CP donne une impulsion unitaire. Cependant, dans le cas d'images sous-échantillonnées, le résultat est une version sous-échantillonnée d'une impulsion unitaire filtrée.

Ce résultat implique que la CP donne un noyau de Dirichlet 2-D sous-échantillonné. Ainsi dans le cas d'images sous-échantillonnées, la CP ne contient pas un seul pic cohérent, mais plusieurs pics dont les plus importants sont proches les uns des autres.

Idéalement, l'équation (3.8) fournit le SPC normalisé de deux images sous-échantillonnées. Toutefois, puisque la relation s'obtient à partir de l'hypothèse de périodicité pour des images de taille finie discrète, en pratique du bruit additif sera introduit à cause des régions qui ne sont pas identiques dans les deux images. On suppose que ce bruit additif est distribué normalement autour d'une valeur moyenne de zéro. En d'autres termes, on suppose :

$$C_\eta(x, y) = C(x, y) + \eta(x, y) \quad (3.11)$$

où  $\eta(x, y)$  est un bruit gaussien de moyenne nulle, et  $C_\eta$  est la CP bruitée.

La TFI discrète de l'équation (3.8) donne :

$$C(x, y) = \frac{1}{WH} \frac{\sin\left(\pi(Mx - x_0)\right)}{\sin\left(\pi(Mx - x_0)/W\right)} \frac{\sin\left(\pi(Ny - y_0)\right)}{\sin\left(\pi(Ny - y_0)/H\right)} \quad (3.12)$$

où **W** et **H** sont des valeurs entières représentant la largeur et la hauteur des images avant le sous-échantillonnage. On peut montrer que pour les tailles d'images utilisées en pratique, la fonction de Dirichlet de l'équation (3.12) peut être approximée correctement par une fonction sinc, et cette approximation est sans conséquence. Ainsi nous avons :

$$\tilde{C}(x, y) \approx \frac{\sin\left(\pi(Mx - x_0)\right)}{\pi(Mx - x_0)} \frac{\sin\left(\pi(Ny - y_0)\right)}{\pi(Ny - y_0)} \quad (3.13)$$

À partir des équations (3.13) et (3.11) on obtient :

$$C_\eta(x, y) \approx \frac{\sin\left(\pi(Mx - x_0)\right)}{\pi(Mx - x_0)} \frac{\sin\left(\pi(Ny - y_0)\right)}{\pi(Ny - y_0)} + \eta(x, y) \quad (3.14)$$

En général, pour des déplacements sous-pixelliques, le niveau de bruit est très faible et on peut facilement identifier le pic principal et les pics immédiatement adjacents de la fonction sinc 2-D, par une simple inspection.

Une fois que  $C(x, y)$  a été calculé,  $(x_0, y_0)$  peut être obtenu par une application directe de l'équation (3.13) à un ensemble de points. Notons que les résultats sont d'autant plus fiables que cette équation est appliquée à des points où la puissance du signal est concentrée. En fait, pour deux images dont l'une est l'autre translatée d'une valeur sous-pixellique, la puissance du signal est largement concentrée dans le voisinage du pic principal de la CP. Afin d'être plus précis, pour des déplacements sous-pixelliques, la puissance du signal de la CP est habituellement concentrée dans un pic principal de coordonnées  $(x_p, y_p)$ , et deux pics à côté de coordonnées  $(x_s, y_p)$  et  $(x_p, y_s)$

où  $x_s = x_p \pm 1$  (modulo  $X$ ) et  $y_s = y_p \pm 1$  (modulo  $Y$ ).

**X et Y sont des valeurs entières qui représentent les dimensions des images après sous-échantillonnage.** En utilisant ces trois points on peut alors trouver les valeurs du déplacement sous-pixellique.

On peut appliquer cette méthode dans le cas général. On cherche la valeur du déplacement selon l'axe des  $x$ . Pour ce faire, on part de l'équation (3.13) appliquée aux points  $(x_p, y_p)$  et  $(x_s, y_p)$ . On obtient ainsi :

$$\frac{\sin\left(\pi(Mx_s - x_0)\right)}{\pi(Mx_s - x_0)} = \frac{C(x_s, y_p)}{C(x_p, y_p)} \frac{\sin\left(\pi(Mx_p - x_0)\right)}{\pi(Mx_p - x_0)}.$$

En remarquant que  $\frac{\sin\left(\pi(Mx_s - x_0)\right)}{\sin\left(\pi(Mx_p - x_0)\right)} = \pm 1$ , on a :

$$\pm C(x_p, y_p) (Mx_p - x_0) = C(x_s, y_p) (Mx_s - x_0),$$

$$M \left( C(x_s, y_p) x_s \pm C(x_p, y_p) x_p \right) = x_0 \left( C(x_s, y_p) \pm C(x_p, y_p) \right),$$

ce qui donne enfin :

$$\Delta x = \frac{x_0}{M} = \frac{C(x_s, y_p) x_s \pm C(x_p, y_p) x_p}{C(x_s, y_p) \pm C(x_p, y_p)}, \quad (3.15)$$

les deux solutions possibles étant :

$$\frac{C(x_s, y_p) x_s + C(x_p, y_p) x_p}{C(x_s, y_p) + C(x_p, y_p)} \quad \text{et} \quad \frac{C(x_s, y_p) x_s - C(x_p, y_p) x_p}{C(x_s, y_p) - C(x_p, y_p)}. \quad (3.16)$$

$\Delta x = \frac{x_0}{M}$  est le déplacement sous-pixellique selon l'axe des  $x$ . On peut remarquer qu'aucune connaissance sur le taux de sous-échantillonnage  $M$  n'est nécessaire. On constate également qu'il y a deux solutions. Cette ambiguïté est due à la forme isotrope de  $C(x, y)$ . Cependant, la solution correcte est facile à identifier puisqu'elle est dans l'intervalle  $[x_p, x_s]$  dans le cas  $x_s = x_p + 1$ , et dans l'intervalle  $[x_s, x_p]$  dans le cas  $x_s = x_p - 1$ . De la même manière on peut trouver  $\Delta y = \frac{y_0}{N}$ . Les deux solutions sont alors :

$$\frac{C(x_p, y_s) y_s + C(x_p, y_p) y_p}{C(x_p, y_s) + C(x_p, y_p)} \quad \text{et} \quad \frac{C(x_p, y_s) y_s - C(x_p, y_p) y_p}{C(x_p, y_s) - C(x_p, y_p)}. \quad (3.17)$$

### 3.3.3 Extension avec utilisation du gradient

En s'appuyant sur la méthode développée dans [Feroosh 02] et décrite ci-dessus, Argyriou et Vlachos proposent une amélioration dans [Argyriou 03].

Leur idée consiste à transformer les images  $I_1(x, y)$  et  $I_2(x, y)$  avant leur passage dans le domaine de Fourier. Un des premiers axes de recherche concernant le recalage des images s'orientait vers des

informations obtenues à partir des gradients des images. Le principe d'utiliser les gradients est bien justifié par le fait que cette opération isole précisément certaines caractéristiques importantes des images, telles que des transitions dominantes. Celles-ci fournissent des points de référence utiles dans le cadre de l'estimation du mouvement.

Pour chaque pixel d'une trame donnée  $I_t(x, y)$  (excepté pour les pixels de bord), les approximations discrètes des gradients horizontal et vertical sont estimées en utilisant la différence centrale, c'est-à-dire :

$$K_t^h(x, y) = I_t(x + 1, y) - I_t(x - 1, y) \quad \text{et} \quad K_t^v(x, y) = I_t(x, y + 1) - I_t(x, y - 1).$$

Ces deux termes sont alors combinés en une seule image complexe de la forme :

$$K_t(x, y) = K_t^h(x, y) + j K_t^v(x, y).$$

Cette représentation contient une information de phase et d'amplitude en chaque pixel. Ainsi, soient deux images  $I_t$  et  $I_{t+1}$  de gradient discret  $K_t$  et  $K_{t+1}$ . On utilise alors la méthode décrite dans la partie précédente en appliquant la TF à  $K_t$  et  $K_{t+1}$ . Les auteurs de cet article pensent obtenir une meilleure précision dans l'évaluation du décalage sous-pixellique que dans le cas précédent.

### 3.3.4 Équation de recalage considérant les bords

Dans le but de pouvoir reconstruire une image translatée, la question se pose dans le sens inverse, à savoir : soit une image initiale  $I_0$  à déplacer selon le vecteur  $(d_x, d_y)$  afin d'obtenir l'image  $I_1$ , **quelle est l'équation de recalage dans le cas d'une approximation bilinéaire ?**

On peut commencer par traiter le cas d'un déplacement d'une valeur entière de pixel. Dans ce cas particulier, la nouvelle image  $I_1$  s'obtient par une expression simple, fonction de  $I_0$  :

$$I_1(x, y) = I_0(x - d_x, y - d_y) \tag{3.18}$$

Le cas des bords sera traité ensuite.

Afin d'obtenir cette équation de recalage pour une translation sous-pixellique (c'est-à-dire  $(d_x, d_y) \in ]-1; 1[$ ), partons de l'observation d'une image représentant un tel décalage (cf. FIG. 3.1).

L'image de droite représente l'image de gauche translatée de  $d_x = \frac{3}{4}$  pixel et  $d_y = \frac{1}{4}$  pixel. On s'aperçoit que pour chaque nouvelle valeur de pixel  $I_1(x, y)$ , quatre pixels de l'image initiale  $I_0$  interviennent avec une importance proportionnelle à la surface qu'ils occupent dans la nouvelle image. Ces surfaces peuvent être facilement exprimées en fonction de  $d_x$  et  $d_y$ . Dans le cas d'un déplacement positif en  $x$  et en  $y$ , soit  $d_x > 0$  et  $d_y > 0$ , on a :

$$I_1(x, y) = \begin{cases} (1 - d_x) & (1 - d_y) & I_0(x, y) & + & d_x & (1 - d_y) & I_0(x - 1, y) & + \\ (1 - d_x) & d_y & I_0(x, y - 1) & + & d_x & d_y & I_0(x - 1, y - 1) & \end{cases}$$

Dans le cas d'un déplacement négatif, soit  $d_x < 0$  et  $d_y < 0$ , on a :

$$I_1(x, y) = \begin{cases} (1 - |d_x|) & (1 - |d_y|) & I_0(x, y) & + & |d_x| & (1 - |d_y|) & I_0(x + 1, y) & + \\ (1 - |d_x|) & |d_y| & I_0(x, y + 1) & + & |d_x| & |d_y| & I_0(x + 1, y + 1) & \end{cases}$$

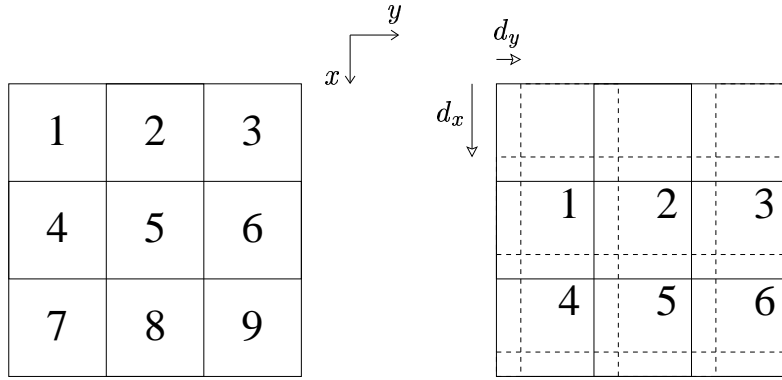


FIG. 3.1: Exemple d'un déplacement sous-pixelique sur une image de 9 pixels.

L'observation de ces deux cas permet de généraliser cette équation. Soit un déplacement quelconque défini par :

$$\begin{cases} x_0 = X + d_x & (x, y) \in \mathbb{Z}^2 \\ y_0 = Y + d_y & (d_x, d_y) \in ]-1; 1[ \end{cases}^2$$

Posons  $sig_{x_0}$  la valeur du signe de  $x_0$ , et  $sig_{y_0}$  la valeur du signe de  $y_0$ . Les valeurs possibles de  $sig_{x_0}$  et  $sig_{y_0}$  sont donc  $-1$  et  $1$ . Dans un premier temps il suffit d'appliquer l'équation (3.18) au déplacement entier  $(x, y)$ , soit :

$$I_1(x, y) = I_0(x - X, y - Y).$$

Ensuite on pose :

$$\begin{cases} d'_x = |x_0| - X \\ d'_y = |y_0| - Y \end{cases}$$

On remarque que  $d'_x$  et  $d'_y$  sont toujours positifs. On peut alors appliquer l'équation de recalage généralisée suivante :

$$I_2(x, y) = \begin{cases} (1 - d'_x) & (1 - d'_y) & I_1(x, y) & + & d'_x & (1 - d'_y) & I_1(x - sig_{x_0}, y) & + \\ (1 - d'_x) & d'_y & I_1(x, y - sig_{y_0}) & + & d'_x & d'_y & I_1(x - sig_{x_0}, y - sig_{y_0}) \end{cases} \quad (3.19)$$

Pour le cas des bords, deux solutions sont possibles. La première consiste à dire que **l'image est périodique**. Ainsi, dans le cas d'un déplacement positif en  $x$  et en  $y$ , ce qui apparaît sur le haut et la gauche de l'image peut être construit à partir de ce qui a disparu en bas et à droite de l'image. Cette solution n'est pas très satisfaisante. En effet, on peut imaginer l'image d'un désert, avec le sable jaune en bas et le ciel bleu en haut. L'utilisation de cette hypothèse consisterait à dire qu'après le déplacement, le ciel inconnu en haut peut être construit à partir du sable jaune qui a disparu en bas. On constate très bien ici la limite de cette méthode.

La seconde solution consiste à utiliser la condition de limite de Neumann (*Neumann Boundary Condition*). Celle-ci consiste à dire que **la scène qui se trouve juste à l'extérieur de l'image est la réflexion de la trame originale à la frontière**. On sent bien que cette hypothèse nous donnera de

meilleurs résultats que l'hypothèse précédente. Dans notre exemple, le ciel inconnu en haut sera construit à partir du symétrique du ciel par rapport à l'axe représenté par le bord de l'image.

Pour les simulations de la présente étude, la condition de limite de Neumann, qui semblait plus pertinente, a été choisie.

La représentation mathématique de ces 2 hypothèses est établie pour un vecteur de taille  $H$ ,  $x = \{0, 1, 2, \dots, H - 1\}$ .

Dans l'hypothèse d'un vecteur périodique ou « circulant », les valeurs limites peuvent être obtenues ainsi :

$$\begin{cases} x = x + H & \text{si } x < 0 \\ x = x - H & \text{si } x > H - 1 \end{cases}$$

Dans le cas de l'hypothèse de condition de limite de Neumann, on a :

$$\begin{cases} x = -x - 1 & \text{si } x < 0 \\ x = -x + 2H - 1 & \text{si } x > H - 1 \end{cases}$$

## 3.4 Exemple d'applications, résultats

### 3.4.1 Cas d'un déplacement précis réalisé à partir d'une fonction numérique

Ces deux méthodes de recalage par CP (cf. **3.3.2** et **3.3.3**) ont été testées sur deux images créées à partir d'une fonction numérique afin d'être évaluées. Ainsi, la valeur du vecteur déplacement entre ces deux images peut être fixée précisément. D'autre part, la taille des images est  $256 \times 256$  pixels.

Le choix de la fonction numérique s'est porté initialement sur la somme de deux fonctions gaussiennes de variance différente, centrée en  $(73, 88)$  pour celle ayant un écart-type  $\sigma_1 = 30$ , et centrée en  $(173, 183)$  pour celle dont l'écart-type  $\sigma_2 = 23$ . L'équation de cette fonction à deux variables est donc, pour l'image de référence :

$$I(x, y) = \frac{1}{2\pi\sigma_1^2} e^{-\left(\frac{(x-73)^2 + (y-88)^2}{2\sigma_1^2}\right)} + \frac{1}{2\pi\sigma_2^2} e^{-\left(\frac{(x-173)^2 + (y-183)^2}{2\sigma_2^2}\right)} \quad (3.20)$$

Pour obtenir la seconde image, celle de référence translatée de  $(d_x, d_y)$ , la fonction (3.20) a été utilisée en calculant  $I(x - d_x, y - d_y)$ . On peut voir l'image de référence sur FIG. 3.2 (a). Les images (c) et (d) montrent respectivement ses gradients horizontal  $K^h$  et vertical  $K^v$ . La norme du gradient, calculée par  $\sqrt{(K^h)^2 + (K^v)^2}$ , est montrée sur (b).

Pour cette évaluation, tous les déplacements sous-pixelliques possibles entre 1 et 2 ont été réalisés en utilisant un pas de 0,05. Cette simulation représente donc 400 cas définis par  $(d_x, d_y) \in [1, 2]^2$ . À chaque fois, la valeur du déplacement a été calculée à partir des deux méthodes.

Lors de la construction des images numériques, un problème pratique est apparu. Ce problème est décrit ici à des fins pédagogiques.

Chaque capteur<sup>1</sup> de la matrice CCD de la caméra a un rôle intégrateur. Le modèle d'intégration des capteurs est [Samson 04] :

$$\mathbf{s}[i, j] = \int_{i-0,5}^{i+0,5} \int_{j-0,5}^{j+0,5} h_0(u, v) du dv \quad (3.21)$$

où  $\mathbf{s}[i, j]$  représente le pourcentage d'intensité lumineuse au pixel  $(i, j)$ , et  $h_0$  est la fonction de transfert optique (FTO, *optical point-spread function*) de chaque pixel.

Lors de la création synthétique des images, cet aspect intégrateur a d'abord été négligé. Ainsi, au lieu de prendre la valeur de l'intégrale de la fonction sur la surface du capteur (3.21), la valeur de la fonction au centre de chaque pixel était utilisée comme valeur du pixel. Cette mauvaise construction entraînait des erreurs importantes d'estimation comme cela va être montré.

Afin de considérer cet aspect d'intégration, la longueur d'un côté de pixel a été introduite. En subdivisant ce côté en un nombre impair  $q$  de valeurs (de 1 à 11 dans les simulations), on obtient un quadrillage régulier de  $q^2$  cases recouvrant la surface de chaque pixel. Il suffit alors de calculer la moyenne des valeurs de la fonction au centre de ces  $q^2$  cases pour en obtenir une estimation plus précise au pixel  $(i, j)$ .

La FIG. 3.3 montre la valeur absolue de la différence entre cette même image obtenue d'une part en prenant pour chaque pixel la valeur de la fonction en son centre ( $q = 1$ ), et d'autre part en intégrant la fonction sur la surface de chaque pixel (avec  $q = 11$ ). La couleur noire représente une différence nulle ou très faible entre les images, et plus le blanc de l'image est intense, plus la différence entre les images est importante. Une légère variation de 2,65 % des valeurs entre ces deux images est observée. La variation maximum des valeurs de la différence entre  $q = 1$  et  $q = \{3, 5, 9, 11\}$  est de 2,84 %.

La variation des valeurs est calculée ainsi :

$$\text{variation}(I_1, I_2) = \frac{\max_{\mathbf{r} \in \mathcal{R}} (I_1(\mathbf{r}) - I_2(\mathbf{r})) - \min_{\mathbf{r} \in \mathcal{R}} (I_1(\mathbf{r}) - I_2(\mathbf{r}))}{\max_{\mathbf{r} \in \mathcal{R}} (I_1(\mathbf{r})) - \min_{\mathbf{r} \in \mathcal{R}} (I_1(\mathbf{r}))}.$$

Le tableau récapitulatif TAB. 3.1 résume les résultats d'estimation des déplacements évalués à partir des deux méthodes.  $\bar{E}_{d_x}$  représente l'erreur moyenne commise lors de l'estimation de  $d_x$ , et  $\max d_x$  représente l'erreur maximum réalisée lors de toutes les estimations de  $d_x$ .

Dans la création des images numériques, la prise en compte de l'aspect intégrateur des pixels donne de meilleurs résultats que lorsqu'on prend la valeur de la fonction au centre du pixel. On observe jusqu'à un rapport 6 entre les performances obtenues à partir de ces deux variantes de construction.

On remarque également que les résultats s'améliorent sensiblement lorsqu'on augmente le nombre  $q^2$  des valeurs utilisées pour calculer la valeur de chaque pixel.

Enfin, on note que les résultats obtenus à partir de **l'utilisation des gradients permettent d'avoir une meilleure précision dans l'estimation du mouvement**. En effet, dans le meilleur des cas, l'erreur maximum commise par la première méthode est de 2,4 % tandis qu'elle est de 1,3 % pour la

---

<sup>1</sup>Par la suite, on parlera indifféremment d'un capteur de la matrice d'acquisition et de son pixel analogue dans une image ainsi acquise.



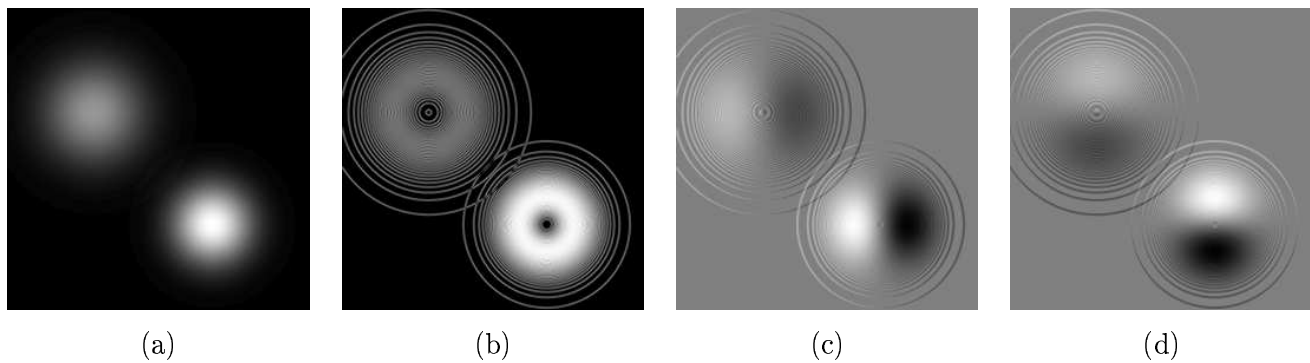


FIG. 3.2: Image numérique utilisée pour l'évaluation des méthodes de recalage.  
 (a) image correspondant à (3.20), (c) et (d) respectivement ses gradients horizontal et vertical,  
 (b) la norme du gradient.



FIG. 3.3: Différence entre les images obtenues en prenant  $q = 1$  et  $q = 11$ .

	Méthode 1				Méthode 2 (gradient)			
	$\bar{E}_{d_x}$	$\bar{E}_{d_y}$	$\max d_x$	$\max d_y$	$\bar{E}_{d_x}$	$\bar{E}_{d_y}$	$\max d_x$	$\max d_y$
Valeur au centre du pixel ( $q = 1$ )	3,36	2,93	15,5	12,4	3,17	3,00	12,9	13,2
Intégration ( $q = 3$ )	2,77	1,55	8,8	4,1	2,47	2,22	6,5	5,5
Intégration ( $q = 5$ )	1,05	0,43	5,0	2,6	0,76	0,64	2,7	1,8
Intégration ( $q = 7$ )	0,86	0,78	4,1	<b>2,4</b>	0,57	0,47	1,6	1,4
Intégration ( $q = 9$ )	0,78	0,92	3,7	2,5	0,52	<b>0,45</b>	1,7	<b>1,3</b>
Intégration ( $q = 11$ )	<b>0,73</b>	1,06	3,5	<b>2,4</b>	0,50	0,48	1,8	1,5

TAB. 3.1: Moyenne et maximum des erreurs (réalisés à partir de 400 estimations, exprimés en %).

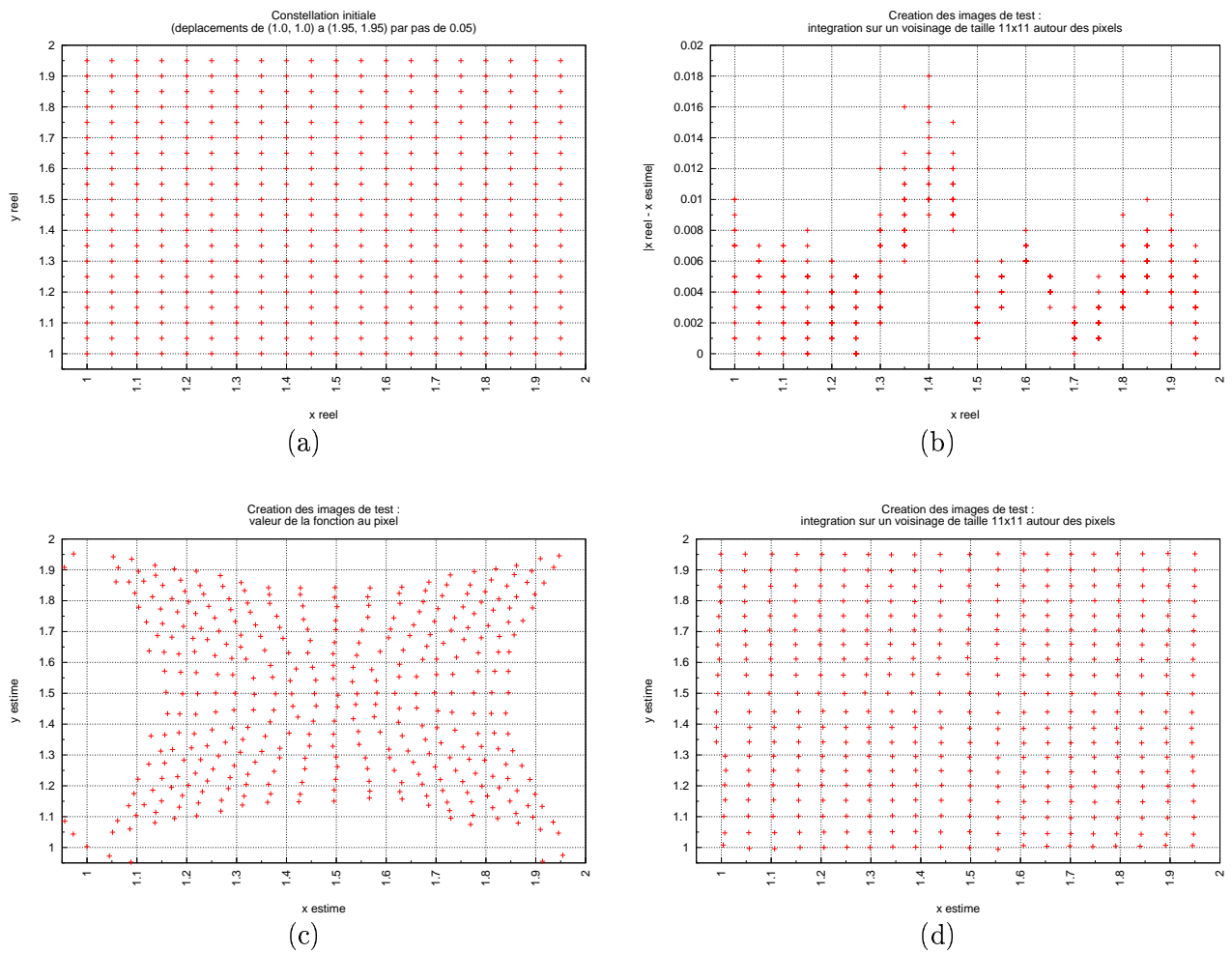


FIG. 3.4: Résultats d'estimation des déplacements sous la forme de constellation.

(a) constellation initiale, (b) erreur sur l'estimation de  $d_x$ ,  
 (c) et (d) constellations obtenues après estimations respectives pour  $q = 1$  et  $q = 11$ .

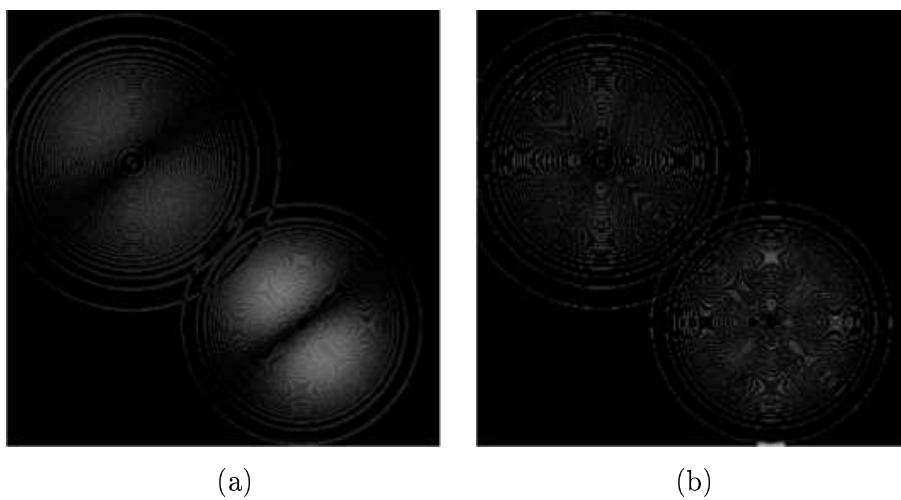


FIG. 3.5: Effet du recalage observé à partir d'images de différence.

(a) différence entre images réelle et translatée, (b) différence entre images réelle et translatée corrigée.

seconde méthode. De même, toujours dans le meilleur des cas, l'erreur moyenne commise sur les 400 estimations a été de 0,73 % avec la première méthode alors qu'elle a été de 0,45 % avec la seconde méthode.

Si on prend comme exemple le vecteur déplacement  $(\mathbf{1,35}, \mathbf{1,85})$ , et pour des images obtenues par intégration avec  $q = 11$ , la première méthode calcule un déplacement de  $(\mathbf{1,342}, \mathbf{1,857})$ , et la seconde de  $(\mathbf{1,340}, \mathbf{1,847})$ . Ceci correspond à des erreurs de 0,8 % et 0,7 % pour la première méthode, et 1,0 % et 0,3 % pour la seconde. Le temps de calcul pour obtenir ces deux résultats est de 16,8 secondes sur une machine LINUX utilisant un processeur Celeron à 600 MHz, et du code écrit en langage C. Ce temps inclut la création des images.

Dans le cas « grossier » où on prend la valeur de la fonction au centre du pixel, le temps de calcul est alors de 1,4 seconde, et on obtient  $(1,321, 1,797)$  (soit 2,9 % et 5,3 % d'erreur) pour la première méthode, et  $(1,322, 1,792)$  (soit 2,8 % et 5,8 % d'erreur) pour l'autre.

La FIG. 3.4 (a) montre la constellation initiale. Chaque point représente une valeur du vecteur déplacement. Il s'agit donc de tous les déplacements tels que  $(d_x, d_y) \in [1, 2]^2$  par pas de 0,05.

La figure (c) montre la constellation obtenue après estimation des déplacements, avec la méthode du gradient, et en construisant les images avec la valeur de la fonction au centre de chaque pixel ( $q = 1$ ).

La figure (d) montre la constellation obtenue à partir de la même méthode, mais en construisant les images en intégrant la surface de chaque pixel sur 121 valeurs ( $q = 11$ ).

Enfin, la figure (b) montre l'erreur commise par la méthode du gradient lors de l'estimation de  $d_x$  pour le cas où on intègre avec  $q = 11$ . Pour ce dernier résultat, à chaque déplacement  $d_x$  correspondent 20 valeurs estimées, et pour chaque estimée on observe la différence entre la valeur réelle et la valeur calculée. Les deux constellations (c) et (d) permettent à nouveau de réaliser le biais introduit lorsqu'on fait l'approximation  $q = 1$ , au lieu d'intégrer la surface du pixel.

Pour finir avec ce premier test, les deux images de FIG. 3.5 (obtenues avec  $q = 11$ ) montrent la qualité de ce recalage. Le déplacement réel entre les deux images est de  $(\mathbf{1,35}, \mathbf{1,85})$ . L'équation de recalage (3.19) dans le cas d'une approximation bilinéaire est ensuite utilisée sur l'image translatée, et pour un déplacement de  $(-\mathbf{1,340}, -\mathbf{1,847})$ .

L'image FIG. 3.5 (a) montre la valeur absolue de la différence entre l'image réelle de référence et sa translatée, (b) la valeur absolue de la différence entre l'image réelle de référence et l'image translatée corrigée.

La différence entre l'image réelle et sa translatée FIG. 3.5 (a) représente une variation de 12,22 % des valeurs, et l'écart-type des valeurs de l'image de différence vaut  $7,5 \times 10^{-1}$ . Ces valeurs peuvent être comparées avec celles obtenues à partir de la différence entre l'image réelle et l'image translatée et corrigée (b). La variation des valeurs est de 2,15 %, et l'écart-type vaut  $5,3 \times 10^{-2}$ .

Pour conclure cette partie, rappelons qu'avec un bon modèle de construction d'images de synthèse, l'erreur sur le résultat est inférieure à 4 % pour la première méthode, et à 2 % pour la seconde.

### 3.4.2 Cas d'un déplacement obtenu par simulation

Dans un deuxième temps, les méthodes de recalage par CP ont été testées à partir d'une image infrarouge (IR) translatée d'une valeur sous-pixellique connue. Pour le faire, l'équation de recalage dans

le cas d'une approximation bilinéaire (cf. **3.3.4**) a été utilisée afin de déplacer l'image réelle de référence de **(10,3, 12,6)**.

Ces deux images sont présentées sur la figure FIG. 3.6. L'image (a) est celle de référence.

L'effet de l'utilisation de la condition de limite de Neumann peut être observée sur l'image FIG. 3.6 (b). Celle-ci se traduit par le fait que la bande supérieure (sur 11 pixels) et la bande de gauche (sur 13 pixels) de l'image translatée ont été obtenues à partir du symétrique par rapport au bord de l'image de référence. On constate également que l'utilisation de l'équation de recalage introduit un léger flou dans l'image translatée.

La méthode de recalage sous-pixelique et la méthode utilisant le gradient de l'image ont été appliquées aux deux images de la FIG. 3.6. Ces deux méthodes donnent des résultats très proches. La première détecte un déplacement de **(10,16, 12,69)**, et la seconde de **(10,17, 12,70)**. Le temps de calcul pour obtenir ces deux résultats est de 1,8 seconde sur une machine LINUX utilisant un processeur Celeron à 600 MHz. Dans ce cas, l'erreur sur le résultat réel est inférieure à 15 %, si on ne considère que la partie sous-pixelique du déplacement.

Ensuite l'équation de recalage est à nouveau utilisée sur l'image translatée, et pour un déplacement de **(-10,17, -12,70)**. L'image résultat obtenue est celle de la FIG. 3.7. Ici aussi on peut remarquer l'effet de la condition de limite de Neumann sur les bandes inférieure et droite de l'image, ainsi qu'un léger flou dans l'image obtenue.

On peut également observer l'information obtenue à partir du passage au gradient de l'image. L'utilisation d'un algorithme rapide de TF impose l'utilisation d'image dont les dimensions soient des puissances de 2. Ainsi nos images initiales, de taille  $220 \times 320$  pixels, sont ramenées à des images de taille  $128 \times 256$  pixels.

La FIG. 3.12 (a) montre un zoom pixelique du pic (point blanc) aux coordonnées **(10, 13)** : il s'agit du point de recalage principal (qui représente une approximation entière du vrai déplacement). La figure (b) présente un tableau récapitulatif des valeurs des niveaux de gris des pixels voisins de ce pic. Les 3 pixels utilisés pour calculer la valeur sous-pixelique du déplacement avec (3.16) et (3.17) sont entourés d'un trait rouge épais.

Les valeurs indiquées sur FIG. 3.12 (b) sont celles obtenues en pratique (comprises entre -33,0 et 113,2) ré-échelonnées sur une échelle entière comprise entre 0 et 255.

Enfin, la FIG. 3.8 présente deux images illustrant la qualité de ce recalage. Il s'agit pour l'image (a) de la valeur absolue de la différence entre l'image réelle de référence et sa translatée (celles de la FIG. 3.6). Et l'image (b) montre la valeur absolue de la différence entre l'image réelle de référence et l'image translatée corrigée (FIG. 3.7).

L'écart-type des valeurs des pixels de la différence entre l'image réelle et sa translatée (a) est 32,8. Celui de la différence entre l'image réelle et l'image translatée et corrigée (b) est 10,6.

### 3.4.3 Cas d'un déplacement réel sur images de nuages

Pour ce troisième test, deux images IR réelles d'une séquence de 350 images ont été utilisées. La première image est en fait le coin haut droit de l'image de référence FIG. 3.6 (a).



FIG. 3.6: Exemple de déplacement sous-pixellique connu.

(a) image IR réelle de référence ( $220 \times 320$  pixels), (b) même image tradatée de  $(10,3, 12,6)$ .



FIG. 3.7: Image tradatée corrigée par une translation de  $(-10,17, -12,70)$ .

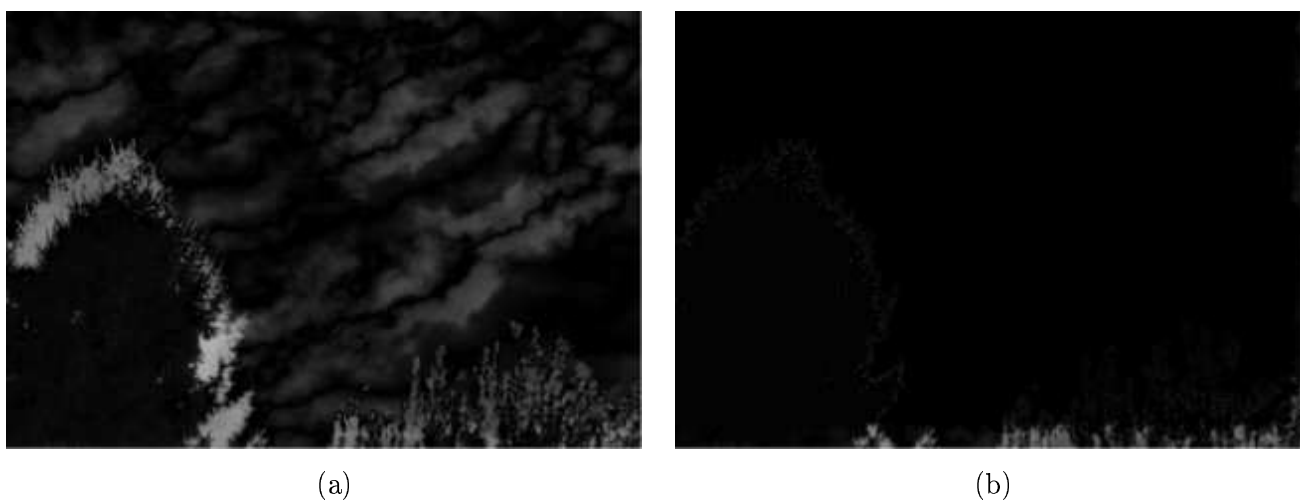


FIG. 3.8: (a) Différences entre images réelle et tradatée, (b) et entre images réelle et recalée.

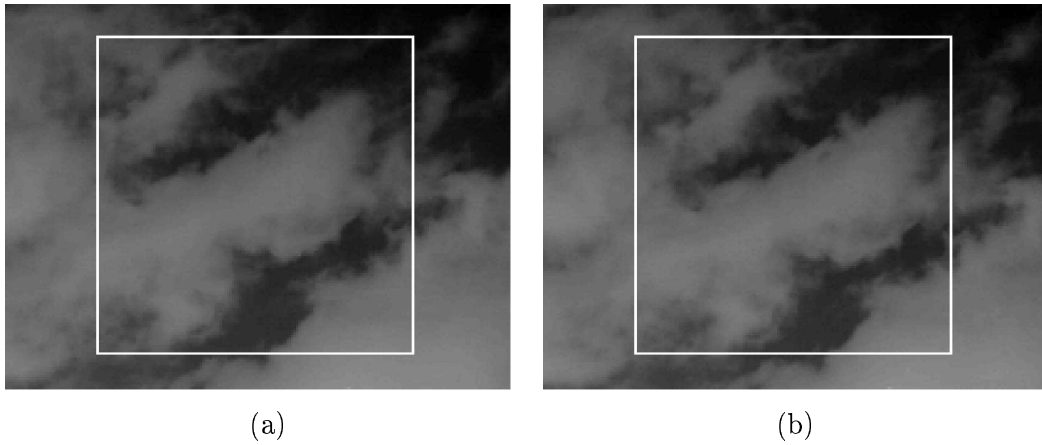


FIG. 3.9: Deux images IR réelles montrant un déplacement de nuages.  
 (a) 1<sup>ère</sup> (b) et 350<sup>e</sup> image d'une séquence vidéo montrant le ciel ( $156 \times 205$  pixels).

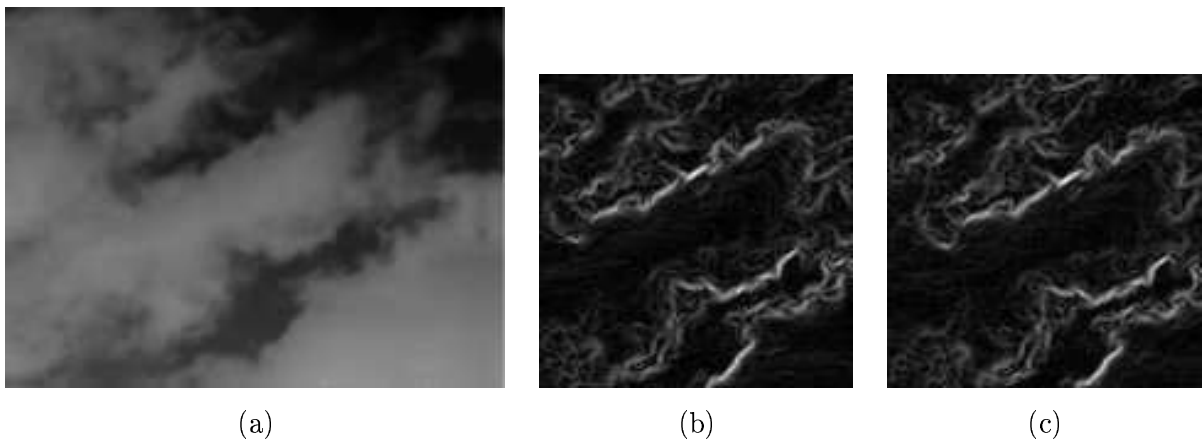


FIG. 3.10: Image recalée et normes du gradient des deux images de la FIG. 3.9.  
 (a) 350<sup>e</sup> image corrigée par une translation de  $(-9, 51, -2, 48)$ ,  
 (b) et (c) norme du gradient des images (a) et (b) de FIG. 3.9.

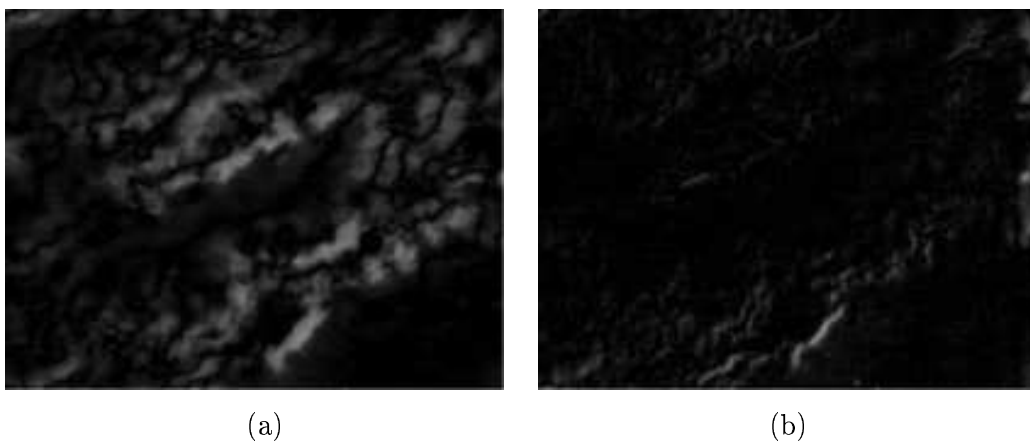


FIG. 3.11: (a) Différences entre 1<sup>ère</sup> et 350<sup>e</sup> image, (b) et entre 1<sup>ère</sup> et 350<sup>e</sup> image recalée.

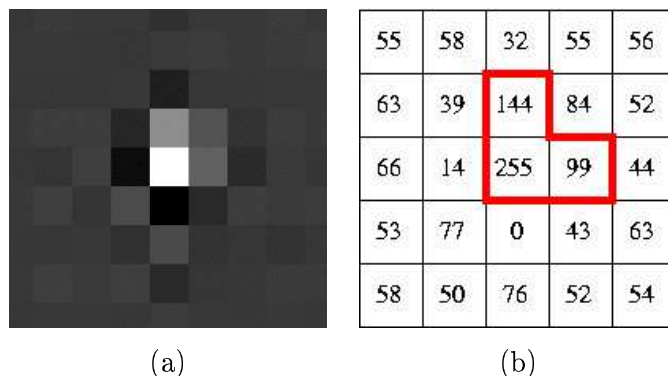


FIG. 3.12: Zoom sur le point de recalage et valeur des niveaux de gris des pixels voisins.

La seconde image est le même coin, mais pris sur la 350<sup>e</sup> image.

La caméra était fixe pendant la prise de cette séquence, mais les nuages se déplaçaient.

Ces images ont une taille réelle de  $156 \times 205$  pixels. L'utilisation de la TF rapide implique leur réduction à une taille carrée de  $128 \times 128$  pixels pour les traitements relatifs au domaine de Fourier. La valeur du déplacement inter-images est inconnu, mais il est visible que les nuages se sont déplacés de plusieurs pixels (cf. FIG. 3.9 (a) et (b), et FIG. 3.10 (b) et (c)).

La première méthode de recalage sous-pixelique détecte un déplacement de  $(-0,07, -0,12)$ , tandis que la méthode utilisant le gradient de l'image de  $(9,51, 2,48)$ . À partir des images de différence de la FIG. 3.11, on peut conclure que le déplacement réel entre ces deux images est très proche de  $(9,51, 2,48)$ . **On constate également la limite de la première méthode, et l'intérêt apporté par l'utilisation des gradients de l'image.** Le temps de calcul pour obtenir ces deux résultats est de 0,8 seconde avec un processeur Celeron à 600 MHz.

La figure FIG. 3.10 permet d'observer l'information obtenue à partir du passage au gradient de l'image. L'image (b) montre la norme du gradient de la 1<sup>ère</sup> image, et (c) celle de la 350<sup>e</sup> image. L'image (a) correspond à la 350<sup>e</sup> image recalée à partir du déplacement évalué par la méthode du gradient.

Enfin, la figure FIG. 3.11 montre deux images permettant d'évaluer la qualité de ce recalage. Il s'agit pour l'image (a) de la différence entre les deux images de la FIG. 3.9. L'image (b) montre la différence entre la 1<sup>ère</sup> image de la séquence et la 350<sup>e</sup> image recalée (FIG. 3.10 (a)).

L'écart-type des valeurs des pixels de la différence entre les deux images initiales FIG. 3.9 (a) est 15,5. Celui de la différence entre la 1<sup>ère</sup> et la 350<sup>e</sup> image recalée (b) est 4,1.

### 3.4.4 Cas d'un grand déplacement réel vertical

Pour ce quatrième et dernier test, deux images réelles extraites d'une séquence d'images IR ont été utilisées. Les images (a) et (b) de FIG. 3.13 montrent les 1<sup>ère</sup> et 21<sup>e</sup> image de cette séquence représentant un important mouvement vertical.

Il s'agit d'images de taille  $220 \times 320$  pixels.

L'image FIG. 3.14 (b) montre la norme du gradient de la 1<sup>ère</sup> image, et (c) celle de la 21<sup>e</sup>. La taille de ces deux images est  $128 \times 256$  pixels.

La première méthode de recalage détecte un déplacement de **(0,19, 48,64)**, et la méthode qui utilise le gradient de l'image de **(0,22, 48,67)**. L'image FIG. 3.14 (a) montre la 1<sup>ère</sup> image recalée à partir des informations obtenues à partir de la méthode utilisant le gradient.

Enfin, la figure FIG. 3.15 présente deux images permettant d'évaluer la qualité de ce recalage. Il s'agit pour (a) de la valeur absolue de la différence entre les deux images de la FIG. 3.13. Et pour l'image (b), de la valeur absolue de la différence entre la 1<sup>ère</sup> image de la séquence et la 21<sup>e</sup> image recalée (FIG. 3.14 (a)).

L'écart-type de la différence FIG. 3.15 (a) est 30,6. Celui de la différence FIG. 3.15 (b) est 14,6.

Ce dernier test permet de constater la limite induite par la condition de limite de Neumann. Celle-ci peut être observée en comparant l'image de référence FIG. 3.13 (a) et l'image recalée FIG. 3.14 (a). Une conclusion peut être tirée de cette observation. Étant donné que notre tâche est de recalibrer une image translatée par rapport à une image de référence, la meilleure idée pour reconstruire une partie manquante (c'est-à-dire, si on se rapporte au dernier exemple, pour reconstruire correctement la bande des 49 pixels du bas de l'image FIG. 3.14 (a)) est de : **reprendre à l'identique cette zone dans l'image de référence.**





FIG. 3.13: 1<sup>ère</sup> et 21<sup>e</sup> image d'une séquence montrant un mouvement vertical. (220 × 320 pixels).

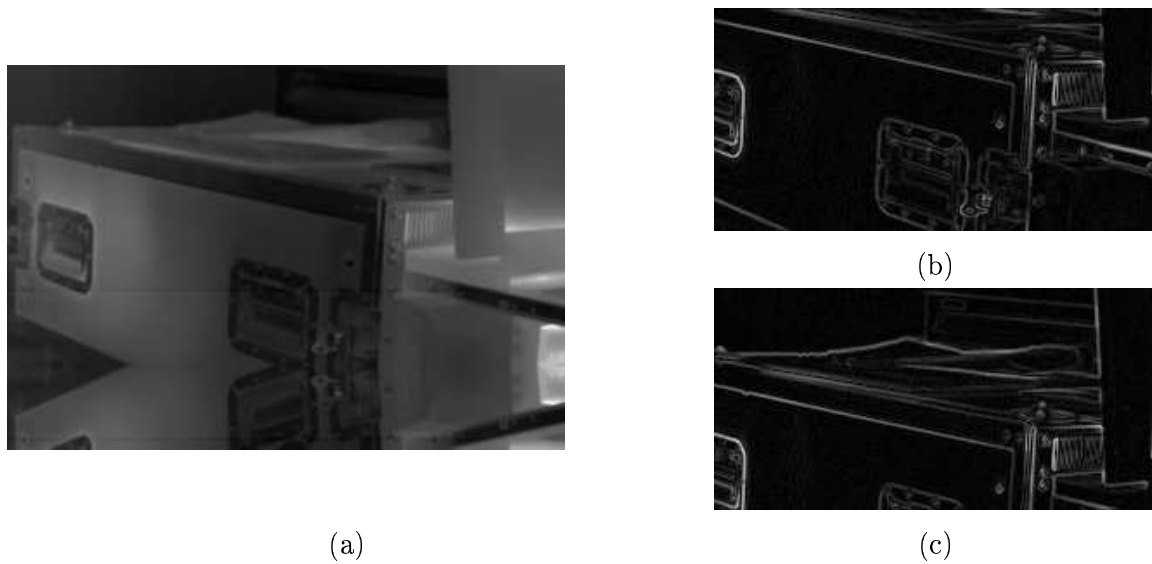


FIG. 3.14: Image recalée et normes du gradient des deux images de la FIG. 3.13.

(a) 21<sup>e</sup> image corrigée par une translation de  $(-0,22, -48,67)$ ,  
 (b) et (c) norme du gradient des images (a) et (b) de FIG. 3.13.



FIG. 3.15: (a) Différences entre 1<sup>ère</sup> et 21<sup>e</sup> image, (b) et entre 1<sup>ère</sup> et 21<sup>e</sup> image recalée.

### 3.5 Extension de ce modèle à la rotation et à l'homothétie

Des extensions à la méthode de recalage par CP ont été proposées afin de détecter une rotation selon l'axe de la caméra, et une homothétie entre deux images. Cette extension a été abordée dans [Chen 94], et elle a été détaillée dans [Reddy 96]. Je vais en donner une présentation rapide.

Si  $I_2$  est une réplique de  $I_1$  ayant subi une rotation d'angle  $\theta_0$ , alors on peut écrire :

$$I_2(x, y) = I_1(x \cos \theta_0 + y \sin \theta_0, y \cos \theta_0 - x \sin \theta_0).$$

La propriété de rotation de Fourier montre que leurs transformées sont reliées par une relation équivalente dans les domaines spatial et fréquentiel :

$$\widehat{I}_2(u, v) = \widehat{I}_1(u \cos \theta_0 + v \sin \theta_0, v \cos \theta_0 - u \sin \theta_0).$$

Les modules de  $\widehat{I}_1$  et  $\widehat{I}_2$  sont identiques à une rotation près. En utilisant le système de coordonnées polaires, on retrouve le cas de la translation déjà traité en détails au cours de ce chapitre :

$$\widehat{I}_2(\rho, \theta) = \widehat{I}_1(\rho, \theta - \theta_0).$$

Dans le cas d'une homothétie, si  $I_2$  est une réplique de  $I_1$  avec des facteurs d'échelle  $(a, b)$ , alors les deux images sont reliées par :

$$I_2(x, y) = I_1(ax, by).$$

Il existe une relation entre leur TF respective :

$$\widehat{I}_2(u, v) = \frac{1}{|ab|} \widehat{I}_1\left(\frac{u}{a}, \frac{v}{b}\right).$$

Si on ignore les facteurs multiplicatifs et que l'on utilise la fonction  $\log$  :

$$\widehat{I}_2(\log u, \log v) = \widehat{I}_1(\log u - \log a, \log v - \log b).$$

Ainsi, en utilisant des axes logarithmiques, on peut évaluer des changements d'échelles entre deux images en trouvant une translation dans le domaine fréquentiel.

Des détails sur la mise en pratique de cette extension peuvent être trouvés dans [Wolberg 00].

### 3.6 Conclusions

Récemment, les méthodes d'estimation du mouvement opérant dans le domaine fréquentiel ont connu un regain d'intérêt. Elles offrent un avantage en terme d'efficacité de calcul lié à l'utilisation d'algorithmes rapides, lesquels sont bien documentés. La technique la plus connue dans cette catégorie est certainement celle de la corrélation de phase (CP). Elle est devenue la méthode d'estimation du mouvement de choix pour un large ensemble de studios professionnels, et pour des applications liées à la télévision. En plus de son efficacité de calcul, la CP offre des avantages clés en ce qui concerne sa réponse robuste aux bords et à certaines caractéristiques importantes des images, son immunité aux changements d'illumination et aux ombres en mouvement, et sa capacité à mesurer de grands déplacements sans sacrifier la précision sous-pixellique.

Cette étude a permis de comparer deux extensions sous-pixelliques de la méthode de recalage d'images par CP. Ces deux méthodes donnent des résultats très proches. Cependant, **celle qui utilise le gradient des images s'avère être beaucoup plus efficace lorsqu'il s'agit d'image dont les zones ne sont pas très bien dessinées** comme dans l'exemple de la partie 3.4.3, sur des images de nuages. En effet dans ce cas, la première méthode s'est avérée complètement inefficace. On constate que le fait d'utiliser les gradients permet de mieux délimiter les différentes zones de l'image (mise en évidence de contours). Voir les images (b) et (c) de FIG. 3.10 p. 53 pour avoir une illustration de cette affirmation. En fait, cette deuxième méthode prend totalement en considération l'amplitude et la phase des gradients de l'image, assurant ainsi la sélection de caractéristiques utiles et fiables de l'image. Celles-ci sont extraites en vertu de l'effet de filtrage passe-bande de l'opérateur **gradient** lorsqu'il est appliqué à une image. De plus, cette méthode réalise la CP en considérant une certaine immunité contre le bruit gaussien additif.

Une série de tests effectués à partir d'une fonction numérique a permis d'estimer la précision avec laquelle ces méthodes retrouvaient la valeur du déplacement. Ils ont montré qu'**avec un bon modèle numérique de construction d'images de synthèse, l'erreur commise sur l'estimation est inférieure à 4 % de la valeur réelle du déplacement pour la première méthode, et à 2 % pour la seconde. En moyenne, l'erreur commise est inférieure à 3,5 %.**

D'autre part, la condition de limite de Neumann aux bords, intéressante dans un premier temps, a montré ses limites dans l'exemple de la partie 3.4.4. La solution la plus judicieuse, dans un cas de recalage, pour reconstruire une partie manquante, est de :

**reconstruire à l'identique cette zone manquante à partir de l'image de référence.**

Enfin, il est intéressant de remarquer que l'utilisation d'un algorithme rapide pour réaliser la transformée de Fourier permet de rendre ces méthodes utilisables pour un traitement en temps réel.

# Chapitre 4

## Super-résolution (SR)

### Sommaire

---

<b>4.1</b>	<b>Définition</b>	<b>61</b>
<b>4.2</b>	<b>Modélisation du lien entre les images BR et l'image HR</b>	<b>63</b>
<b>4.3</b>	<b>Problème inverse mal posé</b>	<b>65</b>
<b>4.4</b>	<b>État de l'art</b>	<b>67</b>
4.4.1	Les méthodes du domaine fréquentiel	68
4.4.1.1	Reconstruction par suppression du recouvrement de spectre	69
4.4.1.2	Méthodes de type moindres carrés (MC) et MC régularisés	71
4.4.1.3	Méthodes utilisant les critères robustes (TLS)	73
4.4.1.4	Méthodes basées sur le théorème d'échantillonnage multicanal	73
4.4.1.5	Résumé sur les méthodes de SR dans le domaine fréquentiel	74
4.4.2	Les méthodes du domaine spatial	76
4.4.2.1	Interpolation à partir d'une grille d'échantillonnage non-uniforme	76
4.4.2.2	Méthode algébrique de rétro-projection filtrée	79
4.4.2.3	Méthodes itératives de rétro-projection	81
4.4.2.4	Méthodes stochastiques	86
4.4.2.5	Méthodes théoriques basées sur des ensembles	93
4.4.2.6	Méthodes hybrides	100
4.4.2.7	Méthodes de filtrage adaptatif et optimal	101
4.4.2.8	Méthodes régularisées de Tikhonov	102
4.4.2.9	Résumé sur les méthodes de SR dans le domaine spatial	102
4.4.3	Comparaisons et conclusions	103
4.4.3.1	Comparaisons des techniques de restauration par SR	103
4.4.3.2	Estimation du mouvement et restauration par SR	104
<b>4.5</b>	<b>Les limites</b>	<b>105</b>

---

**Table des figures**


---

4.1	Génération d'une séquence HR à partir d'une séquence BR. . . . .	62
4.2	Illustration au niveau pixelique des opérateurs $D$ , $M$ , $D^\top$ et $M^\top$ dans le processus de SR. . . . .	64
4.3	Illustration du problème direct en images. . . . .	65

---

**Liste des tableaux**


---

4.1	Résumé et comparaisons des méthodes de SR dans les domaines fréquentiel et spatial. .	104
4.2	Comparaisons qualitatives de deux méthodes de SR : MAP et POCS. . . . .	105

---

Le plus souvent, les algorithmes de super-résolution (SR) sont conçus et utilisés dans un objectif d'amélioration de la qualité visuelle d'une image, du point de vue d'un observateur humain, au sens de la précision du détail. C'est par exemple le cas pour une application d'identification des caractères de plaques d'immatriculation de véhicules éloignés. C'est également le cas lorsque l'on se place dans le cadre de l'imagerie médicale, ou celui des applications militaires telles que la détection, la reconnaissance et l'identification (DRI) de cibles. Dans le domaine de la défense, la problématique de l'amélioration de la qualité de vidéos obtenues à partir de caméras embarquées sur des drones, et dont la résolution est limitée par les capteurs d'acquisition et les capacités de transmission, est un enjeu très important. Il en est de même pour l'amélioration de la résolution des vidéos infrarouges, limitée par le coût et l'encombrement des capteurs de haute définition. Les techniques de SR peuvent aussi servir à l'amélioration d'images satellitaires, ou à la conversion de format vidéo de définition standard (SD,  $576 \times 720$  pixels pour PAL et SECAM,  $480 \times 720$  pixels pour NTSC) au format de haute définition (HD,  $720 \times 1280$  pixels pour le format 720 lignes, et  $1080 \times 1920$  pixels pour le format CIF, *Common Image Format*). La télévision HD existe déjà aux États-Unis depuis 1998, en Corée du Sud et en Australie depuis 2001, au Japon depuis 2003, et elle est en cours de lancement en Europe depuis 2004. En France, elle est connue sous l'appellation de Télévision Numérique Terrestre (TNT). Aujourd'hui, très peu d'émissions proposées par les chaînes avec un label HD sont vraiment obtenues par des caméras de haute résolution, le reste étant acquis en SD. Il est donc nécessaire pour les diffuseurs d'améliorer la qualité de ces vidéos afin de fournir une forte valeur ajoutée à leurs canaux HD. Dans les différents contextes évoqués, il s'agit d'obtenir des images de meilleure qualité, et pour lesquelles le niveau de détail est souvent primordial.

La SR est définie dans la première partie de ce chapitre ; elle constitue un « **problème inverse mal posé** ».

La modélisation de son « **problème direct** » réciproque est présentée dans la section 4.2, et des éléments sur l'aspect « **mal posé** » du problème sont donnés dans la partie 4.3.

Un **état de l'art** détaillé consacré aux méthodes de SR est présenté dans la partie 4.4. Ces techniques sont classées en deux familles : celles du domaine **fréquentiel** détaillées dans 4.4.1, et celles du domaine **spatial** présentées dans 4.4.2. Des éléments de **comparaisons** et de **conclusions** sur ces différentes méthodes sont donnés dans 4.4.3.

Enfin, ce chapitre termine en donnant des éléments sur les **limites** de la SR dans 4.5.

## 4.1 Définition

La reconstruction d'une image par une méthode de SR consiste à produire une image de haute-résolution (HR, *high-resolution* en anglais) à partir de plusieurs images de résolution plus faible. On parle alors d'images de basse résolution (BR, ou LR en anglais pour *low-resolution*) [Baker 02]. Il s'agit donc d'augmenter la résolution spatiale des images, c'est-à-dire le nombre de pixels utilisés pour leur représentation. Ainsi, réaliser la reconstruction par SR d'un facteur  $k_0$  d'une image de taille  $H \times L$  pixels consiste à obtenir une image de taille  $k_0 H \times k_0 L$  pixels.

Ces images BR peuvent être extraites d'une séquence vidéo obtenue soit à partir d'une caméra fixe montrant une scène en mouvement, soit en bougeant une caméra qui filmerait une scène fixe, soit encore à partir de plusieurs caméras fixes ou appareils photographiques observant une même scène. **Le**

**mouvement et l'information non redondante sont ce qui rend le processus de SR possible.** Chaque image BR doit fournir une vue différente de la même scène. Théoriquement, les informations non redondantes au sujet de la scène peuvent être obtenues à partir de conditions d'éclairage différentes ou à partir de différents capteurs. Il s'agit alors du problème de SR par fusion de données multicanaux, non traité dans la présente étude. Cependant, à partir d'un seul appareil d'imagerie et sous les mêmes conditions d'éclairage, il est nécessaire qu'il y ait des mouvements relatifs entre les images. Ainsi, l'image HR obtenue contient plus d'information utile qu'une seule image BR prise dans la séquence vidéo initiale.

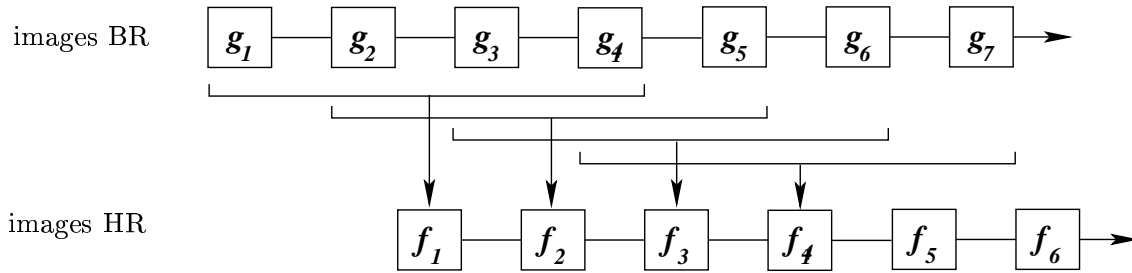


FIG. 4.1: Génération d'une séquence HR à partir d'une séquence BR.

Ce type de procédé peut également être appliqué de manière récursive dans le but de réaliser une vidéo HR à partir d'une vidéo BR, et pourquoi pas, appliqué de nouveau à la séquence HR. En effet, la grande majorité des algorithmes de SR utilise une courte séquence d'images BR en entrée afin de produire une seule image HR super-résolue en sortie (on parle d'approche MISO pour *Multi-Input Single-Output*). Cependant, toutes ces techniques peuvent être appliquées à la restauration vidéo en utilisant une « fenêtre glissante » sur les trames BR à traiter, comme illustré par la FIG. 4.1. Pour une trame HR donnée, une fenêtre mouvante détermine le sous-ensemble de trames BR correspondant et devant être traité pour produire une trame de sortie HR. La fenêtre avance avec le temps et produit successivement les trames HR de la séquence de sortie (approche MIMO pour *Multi-Input Multi-Output*).

Une autre manière de définir la SR consiste à remarquer que la plupart des images contiennent des contours et que ceux-ci sont représentés par de hautes fréquences spatiales. Ces images ne sont donc pas strictement limitées en bande passante, et leur échantillonnage ne respecte alors pas le critère de Nyquist. Les images numériques sont ainsi dégradées par des phénomènes de recouvrement de spectre (*aliasing*), en plus de la perte de hautes fréquences due au flou. À cause de ces effets d'*aliasing*, et d'après le théorème d'échantillonnage de Shannon [Shannon 49], la reconstruction de l'image originale ne peut pas être assurée par interpolation. Un procédé de SR consiste donc à retrouver les hautes fréquences perdues d'une image. Pour cela on exploite les nouvelles données apportées par les autres images de la séquence. La connaissance du mouvement sous-pixellique (c'est à dire de l'ordre d'une fraction de pixel) nous permet de sur-échantillonner l'image tout en réduisant les erreurs liées au recouvrement de spectre. Plus généralement, la SR fait référence à la suppression du flou causé par le système d'acquisition des images (c'est le problème classique de restauration), et à la récupération d'information sur les fréquences spatiales se trouvant au-delà de la limite de diffraction du système optique. La SR consiste donc à retrouver de l'information perdue [Borman 04].

La conséquence de l'utilisation du mouvement dans la reconstruction par SR est que l'une des étapes fondamentales de ce processus est celle du recalage des différentes images BR entre elles, dans un repère commun. C'est pourquoi un état de l'art sur les méthodes de recalage a été présenté dans la partie 3.2, et plus précisément, la méthode utilisée et son évaluation ont été décrites dans le **Chapitre 3**.

## 4.2 Modélisation du lien entre les images BR et l'image HR

Pour pouvoir construire une image HR à partir d'un ensemble d'images BR, il faut d'abord modéliser leurs liens. Cette modélisation décrit comment on obtient des images BR à partir d'une image HR. On parle alors de la modélisation du « **problème direct** ». L'opération consistant à estimer  $\mathbf{f}$  à partir des  $\mathbf{g}_i$  est appelée « **problème inverse** ».

Le choix de modélisation du problème direct fait dans ce travail est décrit par l'équation (4.1). Il s'agit donc de modéliser le lien qui unit les images BR à une image HR en sachant qu'une image BR est une version filtrée, bougée et échantillonnée d'une image HR :

$$\mathbf{g}_i = \mathbf{D} \mathbf{M}_i \mathbf{B} \mathbf{f} + \boldsymbol{\epsilon}_i = \mathbf{H}_i \mathbf{f} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M \quad (4.1)$$

avec

- $\mathbf{f}$  l'image HR,
- $\{\mathbf{g}_i, i = 1, \dots, M\}$  les  $M$  images BR,
- $\mathbf{B}$  un opérateur de filtrage passe-bas nécessaire avant l'échantillonnage de l'image HR ; il symbolise l'**optique des capteurs BR** de l'appareil d'acquisition et l'**effet d'intégration** de la lumière sur ces capteurs,
- $\mathbf{M}_i$  des opérateurs représentant les **déplacements du capteur** pendant l'acquisition des données,
- $\mathbf{D}$  un opérateur d'**échantillonnage** de paramètre  $d$  représentant le facteur de décimation,
- $\boldsymbol{\epsilon}_i$  des bruits additifs représentant toutes les **différentes erreurs** liées à cette modélisation,
- et enfin,  $\mathbf{H}_i = \mathbf{D} \mathbf{M}_i \mathbf{B}$  un opérateur composite reliant l'image HR  $\mathbf{f}$  aux images BR  $\mathbf{g}_i$ .

Les images  $\mathbf{g}_i$ ,  $\boldsymbol{\epsilon}_i$  et  $\mathbf{f}$  peuvent également être écrites à l'aide de la notation  $g_i(\mathbf{r})$ ,  $\epsilon_i(\mathbf{r})$  et  $f(\mathbf{r})$  où  $\mathbf{r} = (x, y) \in \mathbb{N}^2$  représente la position d'un pixel, et  $g_i(\mathbf{r}) = [\mathbf{D} \mathbf{M}_i \mathbf{B} \mathbf{f}](\mathbf{r}) + \epsilon_i(\mathbf{r}) = [\mathcal{H}_i \mathbf{f}](\mathbf{r}) + \epsilon_i(\mathbf{r})$  où  $\mathcal{B}$ ,  $\mathcal{M}_i$ ,  $\mathcal{D}$  et  $\mathcal{H}_i$  sont respectivement les opérateurs continus équivalents de  $\mathbf{B}$ ,  $\mathbf{M}_i$ ,  $\mathbf{D}$  et  $\mathbf{H}_i$ .

On peut aussi réécrire l'équation (4.1) comme ceci :

$$\mathbf{g} = \mathbf{H} \mathbf{f} + \boldsymbol{\epsilon}$$

où  $\mathbf{g}$ ,  $\mathbf{H}$  et  $\boldsymbol{\epsilon}$  sont les combinaisons suivantes :

$$\mathbf{g} = \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_M \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_M \end{bmatrix} \quad \text{et} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_M \end{bmatrix}.$$

Calculer  $\mathbf{g}$  en connaissant  $\mathbf{f}$  et l'opérateur  $\mathbf{H}$  consiste à utiliser le **modèle direct** ; dans la SR, il s'agit d'obtenir les images BR à partir de l'image HR. Estimer l'image HR  $\mathbf{f}$  en utilisant la connaissance des observations  $\mathbf{g}$  et l'opérateur  $\mathbf{H}$  consiste à résoudre le **problème inverse** réciproque, ou à résoudre le problème de SR.



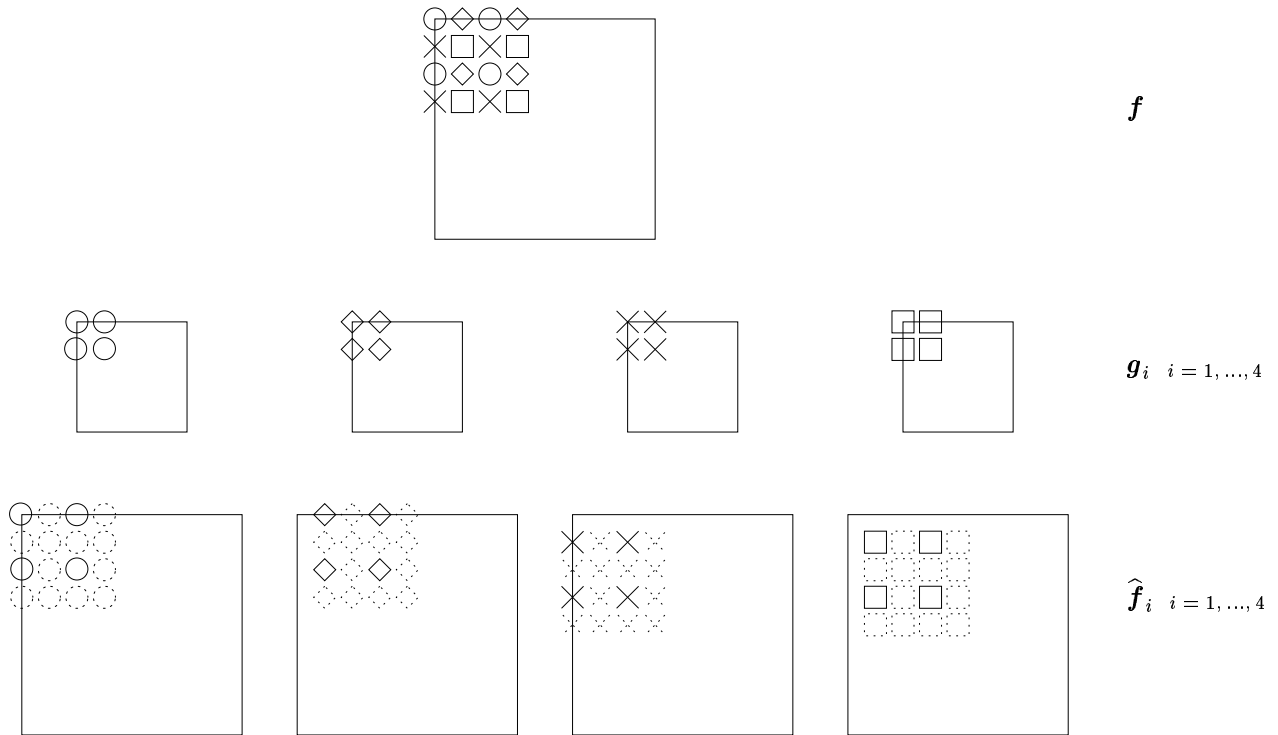


FIG. 4.2: Illustration au niveau pixelique des opérateurs  $\mathbf{D}$ ,  $\mathbf{M}$ ,  $\mathbf{D}^\top$  et  $\mathbf{M}^\top$  dans le processus de SR. Image HR  $\mathbf{f}$  (haut), images BR  $\mathbf{g}_i = \mathbf{D}\mathbf{M}_i\mathbf{f}$ ,  $i=1, \dots, 4$  (milieu), et images recalées et interpolées  $\hat{\mathbf{f}}_i = \mathbf{M}_i^\top \mathbf{D}^\top \mathbf{g}_i$  (bas). Une méthode classique d'obtention d'une image HR consiste à calculer  $\hat{\mathbf{f}} = \frac{1}{M} \sum_{i=1}^M \hat{\mathbf{f}}_i$

La FIG. 4.2 permet d'illustrer le rôle des opérateurs  $\mathbf{D}$  et  $\mathbf{M}_i$  dans le cas d'un mouvement global de translation. Cette figure est constituée de 3 parties, chacune d'entre elles pouvant être vue comme une ligne. La première représente une image HR, chaque symbole représentant un pixel différent de l'image. La deuxième ligne représente 4 images BR différentes obtenues à partir de l'image HR précédente et un facteur de décimation  $d = 2$ . Enfin, les images BR recalées et interpolées sur la grille HR peuvent être vues sur la troisième et dernière ligne. Les symboles dessinés en traits pleins sont les valeurs des pixels originaux obtenus à partir des images BR, et ceux en pointillés sont des valeurs interpolées de pixel.

Prendre un pixel sur deux dans l'image HR originale afin de créer les images BR montre l'effet de l'opérateur  $\mathbf{D}$  pour un facteur de décimation  $d = 2$ . L'effet que deux pixels voisins de l'image HR, par exemple le cercle et le losange en haut à gauche, soient situés au même endroit dans les images BR correspondantes (le premier pixel en haut à gauche) montre l'effet des opérateurs  $\mathbf{M}_i$ . Ainsi, la deuxième ligne de la FIG. 4.2 montre le passage de  $\mathbf{f}$  à  $\mathbf{g}_i$ ,  $i=1, \dots, 4$  en omettant l'opérateur  $\mathbf{B}$  et le bruit  $\epsilon_i$ .

La présence de valeurs interpolées de pixel sur la troisième ligne (en pointillés) illustre l'effet de l'opérateur adjoint de  $\mathbf{D}$ , que l'on note  $\mathbf{D}^\top$ . Et le fait que les cercle et losange supérieurs gauches aient retrouvés leur position initiale (en comparaison à leur position dans l'image HR de la première ligne) montre l'effet de l'opérateur adjoint de  $\mathbf{M}_i$ , que l'on note  $\mathbf{M}_i^\top$ .

La FIG. 4.3 montre un exemple de l'utilisation du modèle direct (4.1) appliqué à l'image initiale (a) de taille  $250 \times 250$  pixels. Il s'agit d'une photo satellitaire montrant une zone rurale. On peut y distinguer des routes, des parcelles de terre homogènes et même une zone faiblement urbanisée en bas à droite de l'image. L'optique des capteurs utilisés par l'appareil d'acquisition, et l'intégration de la lumière sur

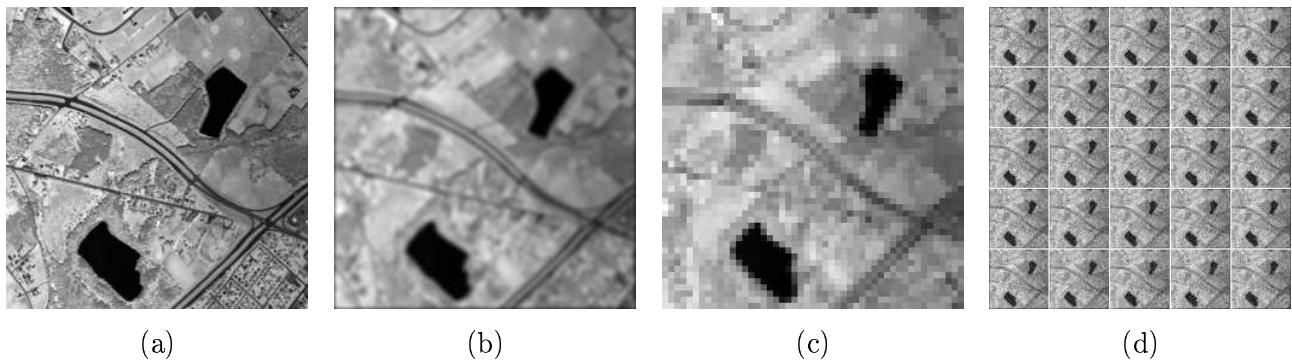


FIG. 4.3: Illustration du problème direct en images.

(a) image HR  $\mathbf{f}$  ( $250 \times 250$  pixels), (b)  $\mathbf{B}\mathbf{f}$  avec  $\mathbf{B}$  un noyau gaussien passe-bas de taille  $7 \times 7$ , (c) image BR non bruitée ( $50 \times 50$  pixels,  $d = 5$ )  $\mathbf{g}_1 = \mathbf{H}_1 \mathbf{f}$ , (d)  $M = 25$  images BR bruitées  $\mathbf{g}_i = \mathbf{H}_i \mathbf{f} + \epsilon_i$ .

leur surface est considérée par le biais de l'opérateur  $\mathbf{B}$ . En pratique, cette opération est réalisée par une convolution de l'image HR par un noyau gaussien représentant un filtrage passe-bas. L'image (b) représente  $\mathbf{B}\mathbf{f}$  avec  $\mathbf{B}$  un noyau gaussien de taille  $7 \times 7$ . L'image (c) est l'une des petites images BR de taille  $50 \times 50$  pixels obtenues avec un opérateur de décimation de paramètre  $d = 5$ , avant ajout du bruit. Elle est affichée à la même résolution que (a) et (b) dans le but de comparer leur résolution spatiale réelle. Enfin l'image (d) montre l'ensemble des  $M = d^2$  images BR ainsi obtenus, après ajout de bruits blancs gaussiens centrés indépendants, ayant tous un rapport signal à bruit (RSB) de 10 dB. Les opérateurs de mouvements  $\mathbf{M}_i$  utilisés sont des transformations de translation globale. Soit  $\epsilon_n$  un bruit et  $\mathbf{f}$  une image non bruitée, la définition utilisée par la suite pour le RSB en dB est :

$$\text{RSB} = 10 \times \log_{10} \left( \frac{\sum_{\mathbf{r} \in \mathcal{R}} (f(\mathbf{r}) + \epsilon_n(\mathbf{r}))^2}{\sum_{\mathbf{r} \in \mathcal{R}} \epsilon_n(\mathbf{r})^2} \right) \quad (4.2)$$

où  $\mathcal{R}$  représente l'ensemble des positions des pixels d'une image. D'un point de vue « image », le problème de SR consiste à utiliser les 25 images que l'on voit sur (d) pour reconstruire (a).

### 4.3 Problème inverse mal posé

Dans son travail sur les équations différentielles [Hadamard 23], Hadamard définit un problème comme étant « **bien posé** » si la solution à ce problème :

1. existe pour toutes données,
2. est unique et,
3. dépend continûment des données.

Cette troisième condition est, par nature, une condition de stabilité de la solution. Par opposition, un problème qui ne satisfait pas une de ces trois conditions est qualifié de « **mal posé** ».

En pratique, les solutions des problèmes inverses sont déterminées en utilisant des méthodes de calculs informatiques. Par conséquent les données observées, les solutions et les modèles utilisés doivent

être numérisés, et donc, prendre des valeurs discrètes. Dans ce cas discret, qui peut être exprimé par une écriture matricielle, on s'intéresse alors à la « **singularité** » et au « **conditionnement** » du problème :

- un problème est dit **singulier** s'il n'y a pas existence et unicité de la solution,
- un problème est dit **mal conditionné** si sa solution exacte est inacceptable d'un point de vue pratique, c'est-à-dire qu'un faible bruit sur les données est propagé et amplifié sur la solution. Un problème peut d'ailleurs être régulier tout en étant mal conditionné.

Il a été montré que la discrétisation de problèmes mal posés conduit à des problèmes numériques mal conditionnés [Bertero 98]. En particulier plus la discrétisation est fine, plus le nombre de conditions du système matriciel correspondant est important, ce qui conduit en fin de compte à une instabilité numérique de la solution calculée. Dans le cas des problèmes mal conditionnés, un certain nombre de méthodes ont été développées afin de déterminer des solutions approchées et stables. Ces méthodes ont été unifiées par la théorie de la régularisation des problèmes mal posés. Dans la pratique, quasiment tous les problèmes inverses discrets sont dans le meilleur des cas, mal conditionnés, et dans le pire des cas, singuliers.

Ainsi, le terme de « **régularisation** » fait référence aux méthodes qui utilisent de l'information additionnelle pour compenser la perte d'information qui caractérise les problèmes mal posés. Cette information supplémentaire est typiquement qualifiée d'information *a priori* puisqu'elle ne peut pas être déduite des observations ou du processus d'observation, et qu'elle doit être connue « avant le fait ». Typiquement, cette information *a priori* est choisie pour représenter des caractéristiques désirées de la solution comme par exemple l'énergie totale, la douceur des dégradés, la positivité, etc... Le rôle de cette information est de contraindre ou de réduire l'espace des solutions qui sont compatibles avec les données observées. De cette manière, l'ajout d'information *a priori* permet d'obtenir un nouveau problème bien posé qui est très lié au problème mal posé initial. Il est très important de s'assurer que les contraintes *a priori* utilisées reflètent précisément les caractéristiques requises sur la solution. Par exemple, en imagerie, les intensités lumineuses sont toujours positives. Une théorie déterministe sur les solutions régularisées des problèmes mal posés a été proposé pour la première fois par Tikhonov et Arsenin dans [Tikhonov 77].

La restauration par SR est en général un problème inverse mal posé [Borman 04]. Comme cela a été vu dans la partie 4.2, une séquence d'images BR est obtenue à partir d'un système d'acquisition filmant une scène. Le processus d'imagerie qui fournit cette séquence représente le **problème direct**. Le **problème inverse** correspondant est celui de déterminer l'estimée de la scène à partir des images observées et de la caractérisation du processus d'acquisition. Le problème direct est donc celui de la simulation, tandis que le problème inverse est celui de la restauration. D'autre part, le problème de reconstruction par SR peut échouer à satisfaire une ou plusieurs des conditions d'Hadamard. Cet échec peut provenir des caractéristiques du système d'acquisition ou des données observées.

- La présence de bruit dans le processus d'observation peut fournir une séquence d'images observées qui, à partir des caractérisations données du système d'acquisition, n'a pas de contenu utile quelle que soit la scène. Dans ce cas, le système est non-inversible et la scène ne peut pas être estimée à partir des observations : il y a alors **inexistence de la solution**.
- Dans le cas où l'opérateur d'imagerie est régulier, un simple manque de données – lesquelles représentent les contraintes sur l'espace solution – est suffisant pour aboutir à la **non-unicité de la solution**. Par exemple dans un cas discret constitué de  $M$  images BR, chacune d'elles comportant  $H_1$  pixels, il y a un maximum de  $MH_1$  contraintes indépendantes. On suppose qu'une seule image

HR est constituée de  $H_2$  pixels devant être estimés à partir des données, avec  $H_2 > MH_1$ . Puisque le nombre d'inconnues est supérieur au nombre de contraintes, il est clair qu'il n'y a pas une seule et unique solution au problème inverse. D'autre part, puisque par définition la SR consiste à retrouver de l'information perdue dans le processus d'acquisition, on peut s'attendre à ce que la solution au problème de restauration ne soit pas unique.

- En fonction des caractéristiques du système d'imagerie, le problème inverse peut être très sensible aux perturbations dans les données, générant **une dépendance discontinue de la solution par rapport aux données**. Considérons par exemple un système d'acquisition avec une réponse spectrale qui décroît de manière asymptotique vers zéro lorsque les fréquences augmentent. En théorie un tel système est inversible, mais en pratique, son inverse est instable. En effet, une composante du bruit arbitrairement petite présente à une fréquence suffisamment élevée peut provoquer un large signal artificiel arbitraire dans la solution calculée. En pratique, de telles restaurations sont typiquement écrasées par l'amplification du bruit.

Même s'il arrive que dans de rares circonstances les conditions d'Hadamard soient satisfaites, en pratique les problèmes de reconstruction par SR sont mal posés. Malgré les difficultés liées à ce constat, les méthodes régularisées permettent d'obtenir des images super-résolues de grande qualité, comme nous le verrons dans le chapitre suivant. L'ajout d'information *a priori* est primordial à cette réalisation.

## 4.4 État de l'art

Les techniques de SR peuvent être divisées en deux classes principales en fonction du domaine dans lequel la méthode de restauration est réalisée : celles du domaine **fréquentiel** et celles du domaine **spatial**.

Bien que les approches dans le domaine **fréquentiel** soient intuitivement simples et attractives par leur aspect implémentation sur un ordinateur, elles sont peu adaptées aux modèles généraux de mouvement et d'observation de scène. D'autre part, elles ne peuvent pas utiliser facilement de contraintes générales *a priori* ou de modèles de bruit.

Les formulations dans le domaine **spatial** fournissent une flexibilité considérable en ce qui concerne les modèles de dégradation et d'observation qui peuvent être représentés. De plus, elles sont bien adaptées à l'inclusion de connaissances *a priori*. Ces avantages apparaissent au détriment de l'augmentation de la complexité calculatoire des algorithmes et donc, de leur temps d'exécution.

Comme nous l'avons vu, la reconstruction par SR utilise simultanément l'information provenant de plusieurs images différentes obtenues à partir d'une séquence d'images observées. Cette disponibilité de plusieurs images donne la possibilité de tirer avantage du mouvement relatif existant entre elles, pour les recalculer dans un repère commun et ainsi fournir des contraintes ou de l'information complémentaires, qui aideront la reconstruction des images super-résolues. Dans de nombreuses méthodes de SR, l'utilisation de contraintes résultantes du recalage des images ou de la compensation du mouvement comme contraintes *a priori* est évidente. D'autre part, l'utilisation de contraintes *a priori* non-linéaires permettent une extrapolation de la largeur de la bande fréquentielle au-delà de la limite de la bande du système d'acquisition.

En ce qui concerne les méthodes de restauration du domaine **spatial**, deux techniques permettant

de calculer des solutions régularisées au problème inverse mal posé de SR ont gagné en importance : les méthodes basées sur l'estimateur du maximum *a posteriori* (MAP) de probabilité et qui sont présentées dans la partie 4.4.2.4, et les méthodes de projection sur des ensembles convexes (POCS, *projection onto convex sets*) qui sont discutées dans la partie 4.4.2.5.

Dans les approches utilisant un estimateur MAP, le problème inverse de SR est vu comme un problème d'inférence statistique, et une utilisation explicite d'information sous la forme de densité de probabilité *a priori* sur la solution est faite. Cette approche statistique fournit un cadre de travail théorique rigoureux. Les approches de type POCS définissent l'espace des solutions possibles comme des régions d'intersection d'ensembles de contraintes convexes. Une conséquence de cette restriction est l'existence d'une procédure itérative qui détermine les solutions satisfaisant cette intersection de contraintes. Une caractéristique attractive de l'approche POCS est qu'elle propose un cadre de travail pratique pour l'ajout d'information *a priori*.

D'une certaine manière, l'estimateur MAP et les approches POCS sont complémentaires. Les méthodes hybrides tentent de combiner les caractéristiques les plus intéressantes de chaque technique ; elles sont abordées dans la partie 4.4.2.6.

Enfin, dans la partie 4.4.3, les avantages et inconvénients des méthodes du domaine fréquentiel et spatial sont résumés. Et une comparaison des méthodes de type MAP et POCS est également proposée pour la catégorie des méthodes de restauration du domaine spatial.

Cette étude est inspirée du travail réalisé par Borman dans [Borman 04], et elle a été enrichie de quelques nouvelles références à des travaux récents.

#### 4.4.1 Les méthodes du domaine fréquentiel

Cette partie présente plus particulièrement des techniques utilisant une formulation fréquentielle du problème de reconstruction par SR. Pour réaliser cette restauration, ces méthodes de SR s'appuient sur la propriété de translation de la transformée de Fourier (TF) afin de modéliser le mouvement global de translation présent dans la scène (propriété expliquée précisément dans le **Chapitre 3**), ainsi que sur la théorie d'échantillonnage. La plupart de ces méthodes sont basées sur des hypothèses permettant l'utilisation de procédures efficaces pour le calcul des images à reconstruire, la plus importante étant la transformée de Fourier rapide, ce qui rend ce type d'approche attractif. Toutefois, les propriétés de la TF sont également **responsables des limites** liées à l'utilisation de modèles de mouvement et d'observation plus généraux.

Le premier travail de SR est basé sur ce type de formulation. Il s'agit du papier fondateur de Tsai et Huang [Tsai 84] par lequel cette étude commence. Bien que ce travail précurseur soit limité par la restriction des modèles de mouvement possibles envisagés par la méthode proposée, celui-ci a servi de catalyseur au développement de formulations plus sophistiquées à la fois dans le domaine fréquentiel, et également dans le domaine spatial.

Les algorithmes du domaine fréquentiel présentés ici sont qualifiés de méthodes de SR car ils permettent de retrouver de l'information fréquentielle au-delà de la limite de Nyquist. Ceci est réalisé en restaurant les termes d'*aliasing* liés au sous-échantillonnage des images observées.

#### 4.4.1.1 Reconstruction par suppression du recouvrement de spectre

La première formulation du problème de restauration d'image à partir de plusieurs observations présentée dans [Tsai 84], fut motivée par le besoin d'améliorer la résolution des images des satellites *Landsat*. Ce type de satellite prend des images de la même région terrestre au cours de ses révolutions successives autour de la Terre. Bien entendu, en général, ces images ne coïncident pas parfaitement. Le problème est alors formulé comme une étape de recalage des images, suivie d'une étape de restauration. Cette dernière étape est réalisée sans aucune considération pour le bruit et les dégradations, tel le flou lié au mouvement du satellite, qui sont introduits au cours du processus d'acquisition. Ainsi les images observées sont modélisées comme des images sous-échantillonnées d'une scène inconnue statique, avec des variables spatiales continues, et en ne considérant que les mouvements de translations globaux. Un échantillonnage idéal impulsif est supposé, mais avec une fréquence d'échantillonnage inférieur à la fréquence de Nyquist. Les propriétés de translation et de recouvrement de spectre [Roberts 87] sont utilisées afin d'écrire un système d'équations reliant les coefficients de la transformée de Fourier discrète (TFD) des images observées, aux échantillons de la transformée de Fourier continue (TFC) de la scène inconnue que les observations représentent. Le système d'équations est résolu dans le domaine fréquentiel et l'image est retrouvée grâce à la TFD inverse. L'écriture du système d'équations nécessite la connaissance préalable des mouvements de translation inter-images avec une précision sous-pixellique.

On note une image continue  $f(x, y)$  et  $f_i(x, y)$ ,  $i = 1, 2, \dots, M$ , un ensemble de  $M$  versions de  $f(x, y)$  translatées spatialement. À  $i$  fixé, on a :

$$f_i(x, y) = f(x + d_x, y + d_y),$$

où  $d_x$  et  $d_y$  sont les composantes en  $x$  et  $y$  de la translation globale entre les images  $f(x, y)$  et  $f_i(x, y)$ . On note  $\hat{f}(u, v)$  et  $\hat{f}_i(u, v)$  les TFC de  $f(x, y)$  et  $f_i(x, y)$ . D'après la propriété de translation de la TF,

$$\hat{f}_i(u, v) = e^{-j(u*d_x + v*d_y)} \hat{f}(u, v) \quad (4.3)$$

décrit la conséquence dans le domaine fréquentiel d'une translation  $(d_x, d_y)$  réalisée dans le domaine spatial. Les images translatées sont alors échantillonnées de manière à produire  $M$  images observées  $g_i$  :

$$g_i[h, l] = f(hT_x + d_x, lT_y + d_y), \quad \text{avec} \quad h = 0, 1, \dots, H-1 \quad \text{et} \quad l = 0, 1, \dots, L-1,$$

et où  $H$  et  $L$  sont les dimensions des images échantillonnées.  $T_x$  et  $T_y$  sont les périodes d'échantillonnage selon les axes  $x$  et  $y$ . Les TFD des  $M$  images observées sont notées  $\hat{g}_i[u, v]$ . La TFD de la  $i^e$  image observée à la paire de fréquence  $(u, v)$  est liée à la TFC de la  $i^e$  image translatée continue par la relation de recouvrement de spectre suivante :

$$\hat{g}_i[u, v] = \frac{1}{T_x T_y} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \hat{f}_i\left(\frac{u}{H T_x} + m w_x, \frac{v}{L T_y} + n w_y\right), \quad (4.4)$$

où  $w_x = 1/T_x$  et  $w_y = 1/T_y$  sont les taux d'échantillonnage selon les axes  $x$  et  $y$ . Si  $f(x, y)$  est à bande limitée, il existe des entiers  $L_x$  et  $L_y$  tels que  $|\hat{f}(u, v)| = 0$  pour  $|u| \geq L_x w_x$  et  $|v| \geq L_y w_y$ . Dans ce cas, la somme infinie de l'équation (4.4) devient finie. Et en combinant ce résultat avec (4.3), on peut réécrire la relation de recouvrement de spectre pour chaque paire de fréquence  $(u, v)$  :

$$\hat{g}_i[u, v] = \frac{1}{T_x T_y} \sum_{m=-L_x - \frac{u}{H}}^{L_x - \frac{u}{H}} \sum_{n=-L_y - \frac{v}{L}}^{L_y - \frac{v}{L}} \exp\left(-j(u'(m) d_x + v'(n) d_y)\right) \hat{f}(u'(m), v'(n)), \quad (4.5)$$

avec  $u'(m) = \frac{u}{HT_x} + mw_x$  et  $v'(n) = \frac{v}{LT_y} + nw_y$ . La relation (4.5) peut également s'écrire sous la forme matricielle suivante :

$$\mathbf{G} = \mathbf{\Phi} \mathbf{F}. \quad (4.6)$$

Dans cette expression  $\mathbf{G}$  est un vecteur colonne de taille  $M \times 1$ , et dont le  $i^e$  élément est le coefficient de la TFD  $\hat{g}_i[u, v]$ .  $\mathbf{\Phi}$  est une matrice de taille  $M \times 4L_xL_y$  qui lie les coefficients de la TFD des images observées aux coefficients inconnus à estimer de la TFC de  $f(x, y)$ , et qui sont contenus dans le vecteur colonne  $\mathbf{F}$  de taille  $4L_xL_y \times 1$ . Dans ce cas, la méthode de SR revient à calculer les TFD des  $M$  images observées, déterminer la matrice  $\mathbf{\Phi}$ , résoudre le système d'équation (4.6) afin de trouver les coefficients de la TFC dans  $\mathbf{F}$ , et enfin utiliser la TFD inverse pour obtenir l'image restaurée. Tsai et Huang montre également qu'il est possible de factoriser  $\mathbf{\Phi}$  de manière à réduire le coût calculatoire global. D'autre part, la construction de la matrice  $\mathbf{\Phi}$  nécessite la connaissance des paramètres de translation  $(d_x, d_y)$  pour chaque image. Il est donc primordial de les estimer avant d'envisager l'étape de restauration.

Pour le recalage des images, Tsai et Huang utilisent une nouvelle approche qui n'a peut être pas bénéficié de suffisamment d'attention. Les auteurs proposent un algorithme de recalage multi-images simultanée, lequel délivre des paramètres de recalage fiables, même dans des conditions sévères de sous-échantillonnage, du moment qu'une quantité assez grande d'images est disponible. Cette approche ne semble pas avoir été suffisamment explorée dans la communauté de SR, et elle pourrait proposer un moyen prometteur d'amélioration des performances de ces méthodes. Toutefois, cette méthode ne s'intéresse qu'au cas limité de l'estimation du mouvement global de translation.

Bien que cette approche dans le domaine fréquentiel soit intéressante du point de vue du temps de calcul, elle présente plusieurs inconvénients. Tout d'abord, l'hypothèse d'échantillonnage idéal n'est pas réaliste. En effet, les systèmes d'acquisition sont caractérisés par des capteurs d'échantillonnage qui réalisent une intégration à la fois spatiale et temporelle (pendant le temps d'ouverture). De plus, aucune considération n'a été faite sur le système optique qui a permis d'enregistrer les images. Il en est de même en ce qui concerne les dégradations liées au bruit. C'est un défaut important car la présence de bruit a un effet de détérioration important sur la résolution du système d'équation (4.6). Enfin, l'hypothèse de mouvement global de translation considéré par les auteurs limite le champ du domaine applicatif. Ce modèle est approprié à une approche dans le domaine **fréquentiel** grâce à la dualité entre sa formulation dans le domaine spatial et dans le domaine de Fourier. Cependant, des modèles de mouvement plus généraux n'ont pas de formulation analogue simple dans le domaine de Fourier. Ils ne sont donc pas aussi pratiques pour l'utilisation des approches fréquentielles. Cette limitation intrinsèque à ce type de méthodes représente certainement leur plus sérieux inconvénient. Ceci est malencontreux puisque de nombreuses applications de la SR nécessitent des modèles de mouvement plus généraux.

Deux des limitations de la méthode de Tsai et Huang sont abordées par Tekalp, Ozkan et Sezan dans [Tekalp 92]. Une extension de la précédente méthode est proposée en considérant un bruit d'observation, ainsi que les effets d'une fonction de transfert optique (FTO. On parle de *point spread function* ou PSF en anglais)  $h(x, y)$  qu'ils supposent être linéaire et invariante par translation. Un échantillonnage périodique est encore supposé, mais il n'est plus impulsif de par l'utilisation d'une FTO dans le modèle. L'hypothèse d'un mouvement global de translation de la scène est conservée. Les auteurs disposent de  $M = k_0^2$  images observées, chacune d'entre elles étant sous-échantillonnée d'un facteur  $k_0$  selon les axes  $x$  et  $y$ . Pour chaque paire de fréquence échantillonnée  $(u, v)$  un système de  $k_0^2$  équations à  $k_0^2$  inconnues est formulé. Ces équations sont résolues au sens des moindres carrés du fait de la présence

d'un bruit d'observation. Le modèle d'observation qu'ils utilisent dans le domaine spatial est :

$$g_i[h, l] = \sum_{m=0}^{H-1} \sum_{n=0}^{L-1} \left[ f(x - d_x, y - d_y) * h(x, y) \right] \delta(x - mT_x, y - nT_y) + \epsilon_i(mT_x, nT_y)$$

où  $*$  représente la convolution 2-D,  $\delta(\cdot, \cdot)$  est la fonction de Kronecker, et  $\epsilon_i$  le bruit d'observation de la  $i^e$  image.

Comme pour [Tsai 84], le système d'équation dans le domaine fréquentiel reliant les coefficients de la TFD des images observées sous-échantillonnées de la scène originale translatée et floue, aux échantillons de la TFC de l'image originale inconnue, est établi. Après le calcul de la solution au sens des moindres carrés obtenue à partir des équations dans le domaine fréquentiel, une estimation de la scène originale est obtenue en appliquant la TFD inverse aux coefficients échantillonnés de la TFC de l'image originale. En plus de cette méthode, les auteurs proposent dans ce même papier une méthode de SR dans le domaine spatial de type POCS étudiée plus bas.

Dans le livre [Tekalp 95], Tekalp consacre un chapitre à la présentation de techniques de restauration multi-images par SR. L'auteur propose une introduction au problème de SR et il expose plusieurs modèles d'observation capables d'introduire des effets de mouvements de la scène, ainsi que des FTO de capteurs et de systèmes optiques. Les méthodes [Tsai 84, Tekalp 92] y sont présentées de manière détaillée dans un *tutorial*.

Dans [Kaltenbacher 96], les auteurs Kaltenbacher et Hardie utilisent la même approche que [Tsai 84], mais ils proposent une méthode alternative d'estimation des paramètres du mouvement global de translation inter-images. Leur méthode d'estimation sous-pixelique est basée sur un développement en série de Taylor appliqué à une seule image. Cette méthode est plus rapide que celle proposée dans [Tsai 84]. Toutefois les bases théoriques de cette méthode fondée sur une information de gradient mono-image restent une question ouverte.

#### 4.4.1.2 Méthodes de type moindres carrés (MC) et MC régularisés

Cette partie aborde les solutions de type moindres carrés récursifs (RLS, *recursive least squares*) à l'équation (4.6). Ces méthodes sont efficaces en termes de temps de calcul.

Kim, Bose et Valenzuela [Kim 90b] utilisent le même cadre fréquentiel de restauration et le même modèle de mouvement global de translation que [Tsai 84], mais en étendant cette formulation à la considération du bruit d'observation et aux effets liés au flou spatial. Les auteurs proposent une solution de type RLS et RLS pondérés afin de résoudre le système d'équations (4.6) en présence de bruits d'observation. Des conditions suffisantes pour assurer que la matrice  $\Phi$  est régulière sont obtenues par la modélisation des liens existants entre les paramètres du mouvement et les images sous-échantillonnées de la scène. L'approche récursive proposée est efficace d'un point de vue calculatoire, et la formulation des moindres carrés a l'avantage d'améliorer les solutions en présence de bruit, et lorsque le système (4.6) est sous ou sur-déterminé.

Les auteurs abordent également des aspects qui n'avaient pas été traités dans [Tsai 84], à savoir en considérant un flou linéaire et invariant par translation qui supprime la restriction à un échantillonnage impulsif, un flou linéaire et variable par translation, et un bruit d'observation. Cependant l'estimation des paramètres du mouvement global de translation n'est pas traité, mais ceux-ci sont supposés connus *a priori*. Les auteurs remarquent que la présence de zéros dans la représentation spectrale d'une



FTO linéaire et invariante par translation rend le problème mal posé. L'aspect mal posé est dû à la non-unicité de la solution, conséquence de la perte d'information liée à la réponse des zéros.

Kim et Su abordent l'aspect mal posé du problème dans [Kim 91, Kim 93a]. Ils introduisent une FTO linéaire et invariante par translation dans le modèle d'observation du domaine fréquentiel. Puisque des zéros dans le spectre de la FTO sont responsables de l'aspect mal posé du problème inverse, les auteurs remplacent la précédente solution de type RLS par une méthode basée sur la régularisation de Tikhonov [Tikhonov 77].

En présence d'un bruit d'observation, le système d'équations  $\mathbf{G} = \Phi \mathbf{F}$  devient inconsistant puisque  $\mathbf{F}$  se retrouve à l'extérieur du domaine spatial de  $\Phi$ . En outre,  $\Phi$  est habituellement mal conditionné ou singulier à cause des zéros dans la représentation spectrale de la FTO. Une solution régularisée peut être obtenue en trouvant la solution  $\mathbf{F}$  qui minimise l'expression :

$$\|\Phi \mathbf{F} - \mathbf{G}\|^2 + \Gamma(\mathbf{F}), \quad (4.7)$$

où  $\Gamma(\cdot)$  est une fonction « régularisante » ou « stabilisante ». Dans [Kim 91, Kim 93a],

$$\Gamma(\mathbf{F}) = \lambda \|\mathbf{F} - \mathbf{c}\|^2,$$

où  $\mathbf{c}$  est une approximation de la solution. Dans ce cas, le minimum de l'équation (4.7) s'écrit alors :

$$\hat{\mathbf{F}} = (\Phi^t \Phi + \lambda \mathbf{I})^{-1} (\Phi^t \mathbf{G} + \lambda \mathbf{c}). \quad (4.8)$$

L'équation (4.8) peut être résolue par des algorithmes itératifs, en évitant l'étape délicate d'inversion matricielle. Comme la solution est inconnue,  $\mathbf{c}$  est initialement fixé à zéro, puis prend ensuite la valeur de  $\hat{\mathbf{F}}$  à l'itération précédente.

Ces papiers ont l'intérêt d'aborder l'aspect mal posé du problème inverse de SR. Toutefois la fonction stabilisante, basée sur l'erreur quadratique, n'est pas très réaliste pour les images puisqu'elle conduit à obtenir des solutions trop lissées. D'autre part, l'utilisation d'une estimée de la solution inconnue  $\mathbf{c}$  pose le problème de la stabilité et de la convergence des solutions obtenues, et ces questions ne sont pas traitées par ces articles.

Su et Kim proposent dans [Su 94] d'introduire un modèle de mouvement non global à la méthode de référence [Tsai 84]. Le cadre fréquentiel du problème reste le même, mais la nouvelle méthode qu'ils proposent permet l'utilisation de mouvements de translations non globaux. Les images observées sont décomposées en blocs qui se superposent, et c'est à partir de ceux-ci que le mouvement de translation inter-images est estimé en utilisant la procédure décrite dans [Kim 93b]. La méthode de SR de Tsai et Huang est alors appliquée à chaque bloc de manière indépendante, et une interpolation est utilisée pour compléter les parties qui se recoupent ou ne se recoupent pas dans les images observées. Les blocs restaurés sont assemblés afin de produire l'image finale.

Cette extension de [Tsai 84] qui permet d'introduire des mouvements de translations non globaux est utile au cadre de travail de SR dans le domaine fréquentiel, mais la limitation au mouvement de translation demeure. De plus, le problème non trivial du choix de la meilleure décomposition en bloc n'est pas traité, tout comme les cas de mouvements rapides ou indépendants et multiples dans les blocs.

#### 4.4.1.3 Méthodes utilisant les critères robustes (TLS)

Une extension de l'approche RLS est celle des méthodes totales des moindres carrés récursifs (RTLS, *recursive total least squares*). Ces méthodes fournissent une robustesse aux erreurs dans la matrice  $\Phi$ , ainsi qu'aux erreurs dans le vecteur d'observation  $\mathbf{G}$  dans le calcul des solutions des systèmes linéaires de la forme  $\mathbf{G} = \Phi \mathbf{F}$ . Dans le cas de la restauration par SR, les erreurs dans  $\Phi$  sont dues aux erreurs d'estimation du mouvement inter-images, et aux incertitudes liées au modèle d'observation. Les méthodes RTLS utilisent une implémentation efficace d'un point de vue calculatoire de la théorie des moindres carrés globaux (TLS, *total least squares*) qui est présentée dans [Golub 80], et dans la partie 12.3 de [Golub 96].

Bose, Kim et Valenzuela dans [Bose 93a, Bose 93b] étendent les idées développées dans [Kim 90b, Kim 93a] afin d'inclure un degré de robustesse aux erreurs à la fois dans la matrice  $\Phi$  et dans le vecteur d'observation  $\mathbf{G}$  de l'équation (4.6) en utilisant l'approche TLS. Les erreurs d'estimation du mouvement qui se manifestent par des erreurs dans le système matriciel  $\Phi$  tendent à avoir des effets sérieux sur la qualité de la restauration. C'est pourquoi l'utilisation des méthodes TLS qui fournissent de la robustesse à de telles erreurs est envisagée. Le modèle d'observation utilisé est de la forme :

$$\mathbf{G} = [\Phi + \mathbf{E}] \mathbf{F} + \epsilon,$$

où  $\mathbf{E}$  est l'erreur sur  $\Phi$  qui résulte des erreurs dans l'estimation du mouvement. La formulation TLS est alors considérée comme un problème de minimisation sous contrainte :

$$\begin{aligned} &\text{minimiser} && \|\epsilon\| && \text{et} && \|\mathbf{E}\| \\ &\text{avec la condition} && \mathbf{G} - \epsilon = && [\Phi + \mathbf{E}] \mathbf{F}. \end{aligned} \quad (4.9)$$

La norme matricielle de Frobenius (définie comme la racine carrée de la somme du module au carré de tous les éléments de la matrice) est utilisée. Un algorithme RTLS basé sur [Davila 91, Davila 94] est utilisé pour résoudre l'équation (4.9) de manière efficace. Comme la formulation de cette équation dans le domaine fréquentiel utilise des coefficients de TFD complexes, la minimisation (4.9) est ramenée à une expression équivalente à valeurs réelles. Cette approche peut également s'appliquer lorsqu'on introduit une FTO linéaire et invariante par translation dans le modèle d'observation, comme c'est le cas dans [Kim 90b]. Bose, Kim et Zhou proposent une analyse théorique des performances de leur algorithme de restauration RTLS basée sur les travaux existants consacrés à la théorie TLS dans [Bose 94].

#### 4.4.1.4 Méthodes basées sur le théorème d'échantillonnage multicanal

Ur et Gross proposent une méthode de restauration par SR [Ur 92] basée sur le théorème d'échantillonnage généralisé de Papoulis [Papoulis 77], et sur une de ses variantes, le théorème d'échantillonnage multicanal de Brown [Brown 81]. Bien que l'implémentation de cette méthode soit réalisée dans le domaine spatial, la technique utilisée est fondamentalement du domaine fréquentiel puisqu'elle est basée sur la propriété de translation de la transformée de Fourier pour modéliser la translation de l'image originale.

Les auteurs considèrent une fonction  $f(x)$  à bande limitée sur  $]-\sigma, \sigma[$ , que l'on fait passer à travers  $M$  canaux linéaires « mutuellement indépendants ». Les signaux en sortie de ces canaux sont échantillonnés

à la fréquence  $\frac{2\sigma}{M}$  (et ils sont donc sous-échantillonnés d'un facteur  $\frac{1}{M}$  par rapport à la fréquence de Nyquist) afin de produire  $M$  signaux discrets  $g_i[nT]$  avec  $T = \frac{M}{2\sigma}$ ,  $n \in \mathbb{Z}$  et  $i = 1, 2, \dots, M$ . D'après le théorème d'échantillonnage multicanal [Brown 81],  $f(x)$  peut être parfaitement reconstruite à partir des signaux  $g_i[nT]$  que l'on passe dans  $M$  filtres linéaires, en additionnant les sorties, puis en effectuant une interpolation à bande limitée. En particulier :

$$\hat{f}[x] = \sum_{i=1}^M \hat{f}_i[x] = \sum_{i=1}^M \sum_{n=-\infty}^{\infty} g_i[nT] h_i(x - nT),$$

où  $\hat{f}[x]$  est une version échantillonnée de  $f(x)$  qui satisfait le critère de Nyquist et qui peut être interpolée de manière à retrouver exactement  $f(x)$ . Les  $M$  filtres  $g_i[n]$  à réponse impulsionnelle peuvent être définis à partir des équations données dans [Papoulis 77] et [Brown 81]. En particulier, dans la formulation de Brown, les matrices qui contiennent des échantillons fréquentiels des  $M$  signaux échantillonnés sont utilisées afin de déterminer l'ensemble des filtres  $h_i(\cdot)$  qui doivent être appliqués dans le domaine temporel aux  $M$  signaux dégradés et sous-échantillonnés afin d'obtenir une version échantillonnée de  $f(x)$ . Cette approche est semblable à celle de [Tsai 84] dans laquelle les systèmes d'équations dans le domaine fréquentiel sont résolus pour trouver une représentation dans le domaine fréquentiel de  $f(x)$ . Le théorème d'échantillonnage multicanal fournit une méthode dans le domaine temporel qui calcule une restauration équivalente à celle de la méthode de Tsai et Huang, laquelle consiste à résoudre un système linéaire d'équations dans le domaine fréquentiel.

Ur et Gross considèrent que les canaux de dégradations linéaires correspondent aux effets de flous liés à la FTO, et au mouvement global de translation modélisé comme un retard. En observant que les opérations de flou et de translation commutent, et en supposant une FTO commune à tous les canaux, ils montrent que le problème de SR peut être séparé en deux processus distincts : la fusion des signaux sous-échantillonnés en une seule fonction à bande limitée sur  $]-\sigma, \sigma[$ , suivie par la déconvolution de ce signal fusionné. N'importe quel algorithme standard de restauration d'images peut être appliqué au signal fusionné pour corriger les dégradations liées au flou optique.

Cette méthode nécessite de connaître les paramètres de mouvements globaux de translation inter-images de façon *a priori*. Leur estimation n'est pas abordée dans cet article.

L'approche de restauration proposée dans [Ur 92] est une formulation dans le domaine spatial qui est analogue à celle de [Tsai 84] dans le domaine fréquentiel. La seule différence significative se trouve dans l'ajout d'une seule FTO commune à toutes les observations. Le bruit d'observation et le flou de mouvement ne sont pas considérés. Étant donnée la similitude des deux approches, elles souffrent des mêmes limitations concernant les modèles de mouvement. Tout comme les formulations fréquentielles, ce travail est limité au modèle de mouvement global de translation.

#### 4.4.1.5 Résumé sur les méthodes de SR dans le domaine fréquentiel

- Les avantages des méthodes de restauration par SR dans le domaine fréquentiel sont les suivants :
- **simplicité.** Les principes théoriques sous-jacents à ces approches peuvent facilement être appréhendés en terme de théorie de Fourier. Même si des complexités sont présentes au niveau de leur implémentation, les idées fondamentales sont claires.
  - **complexité calculatoire.** L'utilisation de mécanisme rapide de calcul tels que la TF rapide rend ces méthodes rapides en temps de calcul. D'autre part avec ce type de formulation, le système

d'équations obtenu est découplé, et un seul échantillon fréquentiel de l'image super-résolue peut être calculé indépendamment des autres. Ceci permet l'utilisation d'une implémentation parallèle.

- **mécanisme de SR intuitif.** Les techniques abordées sont toutes basées sur le principe de suppression du recouvrement de spectre de [Tsai 84]. Cette formulation explicite le mécanisme sous-jacent à l'amélioration de la résolution : la restauration de composantes fréquentielles au-delà de la limite de Nyquist d'une image observée individuellement. Dans de nombreuses méthodes analysées par la suite, le mécanisme de SR est moins transparent.

Ces méthodes de SR ont également des inconvénients significatifs qui doivent être traités si elles doivent être appliquées à des problèmes plus généraux de restauration vidéo :

- **limitation au modèle de translation.** Toutes les méthodes abordées dans cette partie utilisent un modèle de mouvement de translation global. Même [Su 94], qui propose une décomposition des images BR observées par blocs avec une estimation de mouvements locaux, impose des mouvements globaux de translation aux blocs. Cette limitation est dû à la formulation de ce type de transformation dans le domaine de Fourier, et constitue un problème assez fondamental des approches fréquentielles.
- **inflexibilité des modèles de mouvement.** Les méthodes du domaine fréquentiel nécessitent une transformation duale dans le domaine de Fourier, du modèle de mouvement décrit dans le domaine spatial. Cette exigence impose des limitations sévères dans le domaine des modèles possibles de mouvement. Pourtant, ces modèles sont nécessaires à des applications sophistiquées de restauration multi-images.
- **modèles de dégradation limités.** Pour des raisons similaires à celles évoquées dans le point précédent, il est aussi difficile d'introduire des modèles de dégradation variant spatialement dans la formulation du domaine fréquentiel.
- **ajout de connaissance *a priori* du domaine spatial limité.** Puisque la SR est un problème inverse typiquement mal posé, les méthodes régularisées sont souvent requises à l'obtention de solutions acceptables. Même si des méthodes décrites utilisent des fonctions régularisantes, celles-ci doivent être spatialement invariantes en raison de l'utilisation de la TF. Souvent, les contraintes *a priori* les plus utiles sont exprimées dans le domaine spatial et peuvent être variables d'un point de vue spatial. Or, les méthodes fréquentielles sont faiblement adaptées à l'ajout de telles contraintes.

Lorsque le modèle de mouvement de translation global est approprié, il y a un intérêt non négligeable à utiliser des approches du domaine fréquentiel pour réaliser une restauration par SR. Cependant, de nombreuses applications émergentes sont caractérisées par leur besoin d'utiliser des modèles de mouvement et de dégradation plus généraux. Dans ce cas, les méthodes fréquentielles ne sont pas adaptées.

Dans la partie suivante, les méthodes de SR dans le domaine spatial sont abordées. Ces techniques considèrent tous les inconvénients propres aux méthodes fréquentielles, mais souvent, au prix d'une augmentation significative du temps de calcul des algorithmes.

#### 4.4.2 Les méthodes du domaine spatial

La seconde grande famille de méthodes de restauration par SR est basée sur une formulation et une résolution dans le domaine **spatial**. Les deux principaux avantages de ces méthodes sont qu'ils fournissent :

- **un choix important de modèles d'observation.** Formuler le modèle d'observation directement dans le domaine spatial permet une énorme flexibilité dans la modélisation des déplacements inter-images et des dégradations. En particulier, il est possible de considérer :
  - des modèles arbitraires de mouvements globaux et locaux,
  - le flou lié au mouvement,
  - les dégradations du système optique invariant ou variant spatialement,
  - un échantillonnage non-uniforme.
- **une prédisposition à l'ajout de contraintes *a priori*.** Afin de traiter l'aspect mal posé du problème de SR, il est recommandé d'utiliser des méthodes qui facilitent l'ajout de contraintes *a priori*. La formulation et la résolution du problème de restauration dans le domaine spatial permet l'utilisation de modèles d'images puissants tels que :
  - des modèles à base de champs de Markov aléatoires (MRF, *Markov Random Field*),
  - des modèles de contraintes basés sur une formulation ensembliste (POCS),
  - des modèles non-linéaires capables d'extrapoler la bande passante,
  - des modèles non-stationnaires.

Cet état de l'art sur les méthodes de SR du domaine spatial s'étend sur un vaste domaine et commence avec les méthodes simples et intuitives basées sur l'interpolation. Suivra la présentation d'une méthode en provenance de la communauté de l'imagerie tomographique, et appelée méthode algébrique de rétro-projection filtrée. Cette approche a été étendue avec les méthodes itératives de rétro-projection. Une revue détaillée des deux approches les plus importantes sera présentée : les méthodes stochastiques, puis les méthodes théoriques basées sur les ensembles.

Une partie sera ensuite consacrée aux méthodes hybrides qui combinent les éléments des approches stochastiques et ensemblistes pour la restauration.

Enfin les méthodes dérivées de la théorie du filtrage adaptatif et optimal seront abordées, ainsi que les méthodes basées sur la régularisation de Tikhonov. En conclusion, des éléments de synthèse seront présentés.

##### 4.4.2.1 Interpolation à partir d'une grille d'échantillonnage non-uniforme

Cette partie est consacrée à une approche simple de construction d'images super-résolues, basée sur une interpolation dans le domaine **spatial**. Toutes les images observées BR sont recalées géométriquement par rapport à une image fixe de référence. Celles-ci forment une image composite d'échantillons non-uniformément espacés. Ces échantillons sont alors interpolés puis ré-échantillonnés sur un quadrillage d'échantillonnage HR régulièrement espacé. Cette approche très intuitive ne prend cependant pas en considération le fait que les échantillons dans les images BR sont non-idéaux. En fait, ces échantillons sont une fonction des caractéristiques du système d'acquisition optique, et ils peuvent comporter des effets de flou temporel ou spatial liés à l'objectif et aux capteurs. En conséquence, l'image reconstruite ne contient pas tout le champ du contenu fréquentiel qui pourrait être restauré à partir d'un modèle

d'imagerie plus réaliste.

Keren, Peleg et Brada décrivent une approche du domaine **spatial** dans [Keren 88] qui comporte un recalage d'images utilisant un modèle de rotation et translation globales, suivies d'une restauration en deux étapes. La procédure de recalage suppose  $M$  images observées à partir desquelles une technique basée sur les gradients estime les paramètres de translation et de rotation pour chacune des  $M - 1$  images, par rapport à une image choisie comme référence. La première étape de restauration consiste alors à recalculer les images BR dont on a augmenté la surface des pixels, de manière à correspondre à la grille HR sur laquelle on « empile » ces images. Chaque pixel de l'image HR résultante est une moyenne de valeurs choisies parmi les valeurs de pixels correspondants des images recalées, en ne considérant pas les valeurs extrêmes (ce qui supprime le bruit non-linéaire, tandis que la moyenne élimine le bruit linéaire). L'image résultante composée de  $M$  images BR peut être sujette à d'autres traitements tels que de suppression du flou, ou servir d'estimée initiale dans un algorithme itératif de restauration comme c'est le cas dans la seconde phase de la méthode proposée. Cette procédure itérative est décrite dans la partie **4.4.2.3**.

Bien que le travail [Keren 88] contribue aux premiers efforts concernant le **recalage** des images, la méthode par interpolation proposée est simple dans son approche du problème de SR. Rien n'est fait pour supprimer les effets de recouvrement de spectre dans les images observées, comme dans la plupart des autres méthodes de SR. En effet, seule une fraction de l'information disponible est utilisée dans la restauration. Cette méthode ne propose pas un cadre de travail systématique pour réaliser la SR. Elle comporte peu d'indications sur la nature de l'amélioration apportée à l'image, et aucune analyse de la méthode n'est présentée.

Une autre approche basée sur l'**interpolation** est proposée par Aizawa, Komatsu et Saito dans [Aizawa 91]. Elle intègre le traitement de l'acquisition d'images HR à partir de caméras stéréo. Bien que ce traitement soit différent de celui de SR à partir d'une séquence d'images, on peut trouver une analogie dans l'utilisation d'une paire d'images acquise à partir d'une seule caméra avec un mouvement relatif de la scène, comme c'est le cas pour une paire d'images stéréo, avec des vues similaires mais non-identiques de la scène.

En considérant la réponse fréquentielle spatiale d'une matrice à deux dimensions régulièrement espacée à ouverture échantillonnée, les auteurs montrent qu'il y a une limite, contrôlée par le théorème de Nyquist, à la mesure des composantes fréquentielles reconstruites à partir de l'image échantillonnée. Si on considère la possibilité d'échantillonner à des positions spatiales « entre » les pixels de la matrice, il est démontré que la réponse fréquentielle efficace du système combinée de deux images est augmentée.

En pratique cette méthode est réalisée en utilisant deux caméras (et davantage dans leurs articles postérieurs) positionnées de manière à fournir des images translattées d'une distance sous-pixelique par rapport à l'image de référence. L'image de la seconde caméra est recalée par rapport à la matrice d'échantillonnage de l'image de référence avec une précision sous-pixelique, en utilisant des techniques basées sur le gradient ou de correspondance de bloc (*block-matching*). Des effets liés à la projection perspective ou à l'alignement différent des caméras imposent en général aux échantillons de la seconde caméra de ne pas tomber de manière régulière sur la grille d'échantillonnage de référence, et sont donc espacés non-uniformément. Pour reconstruire une image HR, les auteurs ont dû faire face au problème

d'interpolation et de ré-échantillonnage de manière à obtenir l'image désirée. Ceci est réalisé à partir d'une méthode de restauration appliquée à des échantillons espacés non-uniformément [Clark 85]. Le principe est démontré à partir du cas unidimensionnel. On suppose que la fonction  $f(x)$  est échantillonnée de manière non-uniforme en des points  $x_n$ ,  $n \in \mathbb{Z}$ . Si une fonction de correspondance point-à-point  $\gamma(\cdot)$  et une constante positive  $T$  existent tels que  $nT = \gamma(x_n)$ , et si  $f(\gamma^{-1}(x))$  est à bande limitée à  $w_0 = \frac{\pi}{T}$  (c'est-à-dire sur  $]-\frac{\pi}{T}; \frac{\pi}{T}[$ ), alors on peut montrer que  $f(x)$  peut être reconstruit ainsi :

$$f(x) = \sum_{n=-\infty}^{\infty} f(x_n) \frac{\sin[w_0(\gamma(x) - nT)]}{w_0(\gamma(x) - nT)}.$$

On reconnaît la fonction d'interpolation à bande limitée *sinc* avec une fonction de correspondance non-linéaire utilisée afin de corriger l'espacement non-uniforme des échantillons. La généralisation à deux dimensions occasionne quelques complications additionnelles mineures en ce qui concerne le choix de  $\gamma(\cdot, \cdot)$ , mais le principe reste le même.

Dans des publications postérieures, Aizawa *et al.* étendent ces méthodes au cas de plusieurs caméras, et une approche itérative de type simulation-et-correction basée sur la méthode de Landweber [Landweber 51] est utilisée pour estimer les échantillons espacés uniformément de l'image super-résolue. Ces articles seront étudiés plus bas.

Tekalp, Ozkan et Sezan proposent dans [Tekalp 92], parmi d'autres méthodes, une procédure en deux étapes où un sur-échantillonnage des images BR est suivi par une restauration afin de corriger l'effet de la FTO du capteur. Les images BR sont recalées puis combinées pour créer une image HR intermédiaire constituée d'échantillons espacés non-uniformément. Cette image est alors interpolée et ré-échantillonnée sur une grille uniforme pour produire une image HR avec des échantillons disposés uniformément. Pour ce faire, les auteurs suggèrent la méthode à base de « *thin-plate spline* » de Franke [Franke 82b], la méthode bayésienne itérative de Sauer et Allebach [Sauer 87], ou la méthode basée sur une formulation POCS de Yeh et Stark [Yeh 90]. Comme l'image interpolée est encore dégradée par le flou optique et les erreurs résultant du processus d'interpolation lui-même, une dernière étape de restauration est réalisée. Les auteurs suggèrent l'utilisation de n'importe quelle méthode de déconvolution usuelle qui prend en considération la présence du bruit.

Aizawa, Komatsu et Saito étendent la méthode présentée dans [Aizawa 91] en introduisant une procédure itérative de restauration par SR. Ils considèrent le problème de l'acquisition d'images HR à partir de plusieurs caméras dans [Aizawa 92, Komatsu 93a], et ils étendent leurs méthodes dans [Komatsu 93b, Saito 94] afin de considérer des caméras dont la taille des pixels n'est pas homogène. L'objectif est de construire des images à une résolution qui dépasse les limites de la résolution physique des matrices individuelles de capteurs. Comme nous l'avons déjà vu, cette problématique est étroitement liée à celle de la SR à partir d'une séquence d'images obtenue à partir d'une seule caméra.

Comme dans [Aizawa 91], les images observées BR sont fusionnées en utilisant la compensation de mouvement pour former une image HR  $\tilde{\mathbf{f}}$  échantillonnée non-uniformément. Néanmoins, on cherche à obtenir une image  $\mathbf{f}$  échantillonnée uniformément. Plusieurs techniques basées sur l'interpolation sont considérées [Franke 82a, Carlotto 82, Clark 85, Kim 90a], cependant une approche basée sur la méthode de Landweber [Landweber 51] est utilisée pour estimer itérativement les échantillons uniformes

de l'image HR à partir des échantillons non-uniformes produits au cours du processus de recalage. L'image uniformément échantillonnée  $\mathbf{f}$  est reliée à l'image fusionnée  $\tilde{\mathbf{f}}$  par la relation :

$$\tilde{\mathbf{f}} = \mathbf{A} \mathbf{f},$$

où  $\mathbf{A}$  représente le processus d'échantillonnage non-uniforme. L'inversion de cette équation n'est pas possible à cause de sa dimension et de la singularité probable de  $\mathbf{A}$ . À la place, une procédure itérative qui corrige le processus d'échantillonnage non-uniforme est utilisée. L'image HR est initialisée à une valeur  $\mathbf{f}^{(0)}$ , puis elle est mise à jour de manière itérative en utilisant la formule de récurrence suivante :

$$\mathbf{f}^{(n+1)} = \mathbf{f}^{(n)} + \alpha \mathbf{A}^* (\tilde{\mathbf{f}} - \mathbf{A} \mathbf{f}^{(n)}), \quad (4.10)$$

où  $\mathbf{A}$  est une approximation linéaire du processus d'échantillonnage non-uniforme,  $\mathbf{A}^*$  est l'opérateur adjoint de  $\mathbf{A}$ ,  $\alpha$  est un paramètre de contrôle et  $\tilde{\mathbf{f}}$  est l'image fusionnée constituée d'échantillons espacés non-uniformément. Le paramètre  $\alpha$  est choisi de manière à assurer la convergence des itérations décrites par l'équation (4.10).

Dans la partie 4.4.2.3, une méthode itérative identique de type Landweber est utilisée. Cependant dans ce cas, l'opérateur  $\mathbf{A}$  modélise le processus d'imagerie tandis qu'ici, il représente la correspondance entre une paire d'images super-résolues.

Les auteurs proposent également une méthode d'acquisition par intégration temporelle à partir d'une seule caméra qui utilise une séquence d'images pour calculer une seule image HR [Nakazawa 95]. À la différence de leurs précédents travaux, cet article aborde le problème de la restauration par SR multi-images. La méthode proposée utilise l'estimation du mouvement et le recalage des images en utilisant des blocs quadrilatéraux déformables, suivis d'une phase de restauration qui réalise une interpolation à partir d'échantillons espacés de manière non-uniforme. Une de leurs innovations est que la méthode de restauration inclut des contraintes de lissage spatial et par conséquent, peut être considérée comme une méthode de la classe des solutions régularisées. Une procédure itérative est utilisée pour résoudre le problème variationnel résultant. Toutefois, cette méthode ne prend pas en compte les effets du bruit ou le flou de l'image.

Les techniques décrites ci-dessus présentent l'avantage de la simplicité, mais elles ne traitent pas de nombreux problèmes importants. Par exemple, les modèles d'observation utilisés sont souvent peu réalistes car ils ne considèrent pas correctement les effets liés au flou optique, au flou de mouvement et au bruit. De plus, peu de ces méthodes incluent de l'information *a priori* pour essayer de traiter l'aspect mal posé du problème de restauration par SR multi-images.

#### 4.4.2.2 Méthode algébrique de rétro-projection filtrée

Les **méthodes algébriques de rétro-projection filtrée** (FBP, *filtered back-projection*), généralement utilisées en imagerie tomographique, ont été appliquées au problème de SR. La restauration (on parle plus souvent de reconstruction dans la littérature tomographique) est considérée comme un problème inverse linéaire discret, résolu en une seule étape. Un opérateur de « rétro-projection », souvent l'opérateur adjoint de l'opérateur linéaire chargé de l'approximation du modèle d'observation, est utilisé pour obtenir la solution. Il est aussi habituel d'inclure une étape de filtrage inverse afin d'améliorer la



qualité de l'image restaurée.

Frieden et Aumann dans [Frieden 87] sont les premiers à avoir adapté ce type de méthode à la restauration par SR. Ils ne considèrent par la SR à partir d'une séquence d'images, mais à partir de plusieurs mesures 1-D d'une scène stationnaire. En remarquant que la FTO d'un système d'acquisition 1-D représente une intégrale linéaire, tandis que dans le cas de SR multi-images, il s'agit d'une intégrale spatiale, la formulation du problème ne diffère que dans la forme de la FTO du système d'acquisition. Le point important est que des techniques de restauration semblables sont applicables dans les deux cas. Dans [Frieden 87], les dimensions des éléments du détecteur dans la matrice de capteurs linéaires sont supposées être plus grandes que la limite de résolution du système optique. D'autre part, la géométrie du système d'acquisition fournit des mesures qui se superposent sur une région de la scène donnée, permettant une restauration à une résolution plus élevée que la limite de la fréquence d'échantillonnage spatiale de la matrice de capteurs. Le processus d'imagerie est décrit par le modèle linéaire discret suivant :

$$\mathbf{g} = \mathbf{H} \mathbf{f}, \quad (4.11)$$

où  $\mathbf{g}$  est un vecteur de données observées obtenu à partir des acquisitions d'images successives,  $\mathbf{H}$  est le système matriciel modélisant le processus optique et  $\mathbf{f}$  est un vecteur contenant les images inconnues à estimer. Un bruit d'observation et/ou un manque de données, ainsi que la taille considérable de la matrice  $\mathbf{H}$  excluent l'utilisation d'une inversion directe pour estimer  $\mathbf{f}$ . À la place, les auteurs proposent la création d'une estimée à partir de la rétro-projection des données observées :

$$\hat{\mathbf{f}} = \mathbf{H}^\top \mathbf{g}$$

où  $\mathbf{H}^\top$  est l'opérateur adjoint de  $\mathbf{H}$ . On parle de matrice de rétro-projection dans la littérature tomographique.

En général, cette matrice de rétro-projection  $\mathbf{H}^\top$  n'est pas équivalente à la matrice inverse  $\mathbf{H}^{-1}$ . L'équation (4.11) permet d'écrire :

$$\hat{\mathbf{f}} = \mathbf{H}^\top \mathbf{H} \mathbf{f}.$$

Ainsi, la matrice de rétro-projection sera équivalente à la matrice inverse si et seulement si  $\mathbf{H}^\top \mathbf{H} = \mathbf{I}$ . Dans l'application donnée de SR, la matrice  $\mathbf{H}^\top \mathbf{H}$  est Toeplitz, et elle est associée à une FTO bien définie, qui permet un filtrage inverse dans le domaine de Fourier afin de restaurer les données observées rétro-projetées. Le fait d'appliquer un filtre inverse pour  $\mathbf{H}^\top \mathbf{H}$  donne l'estimée :

$$\hat{\mathbf{f}} = \left( \mathbf{H}^\top \mathbf{H} \right)^{-1} \mathbf{H}^\top \mathbf{g},$$

estimée au sens de l'erreur quadratique moyenne du minimum linéaire (LMMSE, *linear minimum mean square error*) de  $\mathbf{f}$ .

Frieden et Aumann ne considèrent pas le bruit d'observation. Cet aspect a des conséquences sérieuses puisque le filtrage inverse est très sensible au bruit à cause de la réponse du filtre inverse qui, typiquement, augmente les amplitudes lorsque les fréquences augmentent. La méthode proposée ne s'inscrit pas dans la catégorie des méthodes régularisées. On ne peut donc pas en attendre des performances acceptables en présence de bruit ou de données manquantes. En dépit de ces inconvénients, cet article offre

l'intérêt de présenter une première tentative de l'application des techniques de reconstruction d'images tomographiques au problème de restauration par SR. D'autres publications ont étendu les principes de bases qui ont été présentés ci-dessus, de manière à inclure des solutions régularisées et itératives à la formulation linéaire de ce problème inverse de restauration par SR. Ces méthodes sont présentées dans la partie suivante.

#### 4.4.2.3 Méthodes itératives de rétro-projection

Cette section présente une vaste classe de méthodes de SR multi-images, qui ont en commun une approche itérative simple et puissante de type simulation-et-correction pour la restauration. Comme pour le cas précédent, cette classe tient son nom d'une famille de méthodes très utilisées en imagerie tomographique et que l'on qualifie de **méthodes itératives de rétro-projection** (IBP, *iterative back-projection*). Le cadre général de ces méthodes est d'abord introduit de façon informelle, puis des approches plus spécifiques seront abordées.

À partir de la définition d'un processus d'acquisition reliant une scène HR à une séquence d'images observées, il est possible de simuler la sortie du système optique en utilisant l'estimée de la scène originale. Les images simulées peuvent alors être comparées aux images observées, et une erreur résiduelle peut être formulée. Une faible erreur résiduelle suggère que l'estimée de la scène est précise, tandis qu'une erreur résiduelle élevée indique que l'estimée est pauvre. Cependant, l'information sur l'erreur peut servir à améliorer l'estimée de la scène. Une technique qualifiée de « rétro-projection » permet d'utiliser l'erreur résiduelle afin de mettre à jour l'estimée de la scène HR qui deviendra ainsi une meilleure approximation de la scène originale. Cette approximation est meilleure dans la mesure où le résultat du processus d'imagerie simulé qui utilise l'estimée mise à jour a réduit l'erreur résiduelle, en comparaison à l'estimée précédente. Ce mécanisme peut être réitéré dans l'espoir de réduire l'erreur résiduelle de simulation jusqu'à un minimum. De cette manière, le processus itératif comporte deux étapes : la simulation des images observées, et la rétro-projection de l'erreur afin de corriger l'estimée de la scène originelle.

Plusieurs méthodes de cette catégorie vont être présentées dans cette partie, et des méthodes plus sophistiquées qui étendent les idées de base présentées ici seront développées dans des parties suivantes. On verra qu'il est possible de formaliser ces méthodes de restauration dans un cadre de travail théorique commun, avec des différences manifestes dans les approches en ce qui concerne le choix des opérateurs de projection et de rétro-projection, de même que dans la fonction d'erreur à minimiser.

On considère le modèle d'observation linéaire reliant les images BR observées (décrit par le vecteur  $\mathbf{g}$  constitué des pixels de chaque image ordonnés de façon lexicographique) à la scène inconnue constituée de valeurs discrètes  $\mathbf{f}$ , par le biais de l'opérateur  $\mathbf{H}$  et qui, typiquement, modélise la compensation du mouvement et le sous-échantillonnage. De manière algébrique, on a :

$$\mathbf{g} = \mathbf{H} \mathbf{f}. \quad (4.12)$$

À partir d'une estimée  $\hat{\mathbf{f}}^{(n)}$  de la restauration super-résolue de la scène  $\mathbf{f}$ , il est possible de simuler les images observées BR en appliquant l'équation (4.12) à cette estimée :

$$\hat{\mathbf{g}}^{(n)} = \mathbf{H} \hat{\mathbf{f}}^{(n)}.$$

Les procédures IBP mettent à jour l'estimée de la restauration super-résolue en « rétro-projetant » l'erreur entre les images BR simulées  $\hat{\mathbf{g}}^{(n)}$  et les images BR observées  $\mathbf{g}$  par le biais de l'opérateur de rétro-projection  $\mathbf{H}^{\text{RP}}$ . En général,  $\mathbf{H}^{\text{RP}}$  est conçu comme une approximation de l'inverse de l'opérateur  $\mathbf{H}$ . Cette procédure itérative est décrite formellement par :

$$\begin{aligned}\hat{\mathbf{f}}^{(n+1)} &= \hat{\mathbf{f}}^{(n)} + \mathbf{H}^{\text{RP}}(\mathbf{g} - \hat{\mathbf{g}}^{(n)}), \\ \hat{\mathbf{f}}^{(n+1)} &= \hat{\mathbf{f}}^{(n)} + \mathbf{H}^{\text{RP}}(\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}^{(n)}).\end{aligned}\tag{4.13}$$

Il est intéressant de remarquer le parallèle entre cette approche de type simulation-et-correction, et de nombreuses autres méthodes bien connues de résolution de problèmes en apparence sans rapport avec celui étudié. Par exemple en reconstruction tomographique, l'équation (4.13) est connue sous le terme de rétro-projection itérative (IBP). L'opérateur  $\mathbf{H}$  est une matrice de projection décrivant l'interaction du faisceau de radiation pénétrant avec l'objet observé, et  $\mathbf{H}^{\text{RP}}$  est un opérateur de rétro-projection. Cependant, dans la solution des équations intégrales, l'équation (4.13) est de la forme de l'itération de Landweber [Landweber 51]. Le choix des opérateurs de projection et de rétro-projection de l'équation (4.13) varient très largement.

Peleg, Keren *et al.* introduisent initialement ces idées dans [Peleg 87, Keren 88], puis ils les développeront dans leurs travaux suivants sous le nom d'algorithmes IBP pour la restauration par SR. Dans [Peleg 87], les auteurs considèrent la SR à partir d'un ensemble d'images translattées globalement représentant une scène 2-D immuable; tandis que dans [Keren 88], un modèle plus général de translation et rotation globales est utilisé. Pour ces deux travaux, les effets d'une FTO invariante spatialement sont considérés.

Les auteurs proposent une méthode itérative pour estimer la scène inconnue. À partir d'une estimée de l'image super-résolue, un processus d'imagerie simulé produit un ensemble d'images BR comparées avec les images BR observées. L'estimée super-résolue est alors modifiée de manière à réduire l'erreur entre les images observées et simulées. L'opération s'arrête lorsque l'erreur a atteint un niveau prédéterminé, ou après un nombre donné d'itérations.

Les images observées sont notées  $g_i[h, l]$  avec  $i = 1, \dots, M$ . La  $n^e$  estimée de l'image super-résolue est notée  $\hat{f}^{(n)}[x, y]$ . En supposant que le processus d'imagerie qui résulte des images observées est connu, il est possible de simuler le processus d'acquisition en utilisant l'estimée  $\hat{f}^{(n)}[x, y]$  comme une approximation de la scène originale, produisant  $M$  images simulées BR  $\hat{g}_i^{(n)}[h, l]$ . L'erreur entre les images simulées à la  $n^e$  itération et les images originales observées est définie comme :

$$E^{(n)} = \sum_{i=1}^M \sum_{h,l} |\hat{g}_i^{(n)}[h, l] - g_i[h, l]|.\tag{4.14}$$

L'objectif étant que les images simulées soient identiques aux images observées, l'erreur  $E^{(n)}$  doit être minimisée. Peleg *et al.* proposent les modifications suivantes de l'estimée courante  $\hat{f}^{(n)}[x, y]$ . Soit un pixel  $[x, y]$ , si  $\hat{f}^{(n)}[x, y] = m$ , on considère les modifications  $\hat{f}^{(n)}[x, y] = m - 1$  et  $\hat{f}^{(n)}[x, y] = m + 1$ . Pour les trois valeurs de pixels  $\{m - 1, m, m + 1\}$ , on calcule les images BR simulées correspondantes et l'erreur donnée par l'équation (4.14) associée. Finalement, on choisit la valeur de pixel pour laquelle l'erreur est minimum. On répète ce processus pour chaque pixel de l'image HR après plusieurs itérations.

Cette simple optimisation tente de réduire l'erreur de façon monotone sur une base de pixel par pixel. À partir de leurs résultats expérimentaux, les auteurs observent que leur système itératif tend à se finir dans des minima locaux. Ce problème est abordé en utilisant une approche de recuit simulé (*simulated annealing*) [Kirkpatrick 83] pour l'optimisation. Bien qu'il ait été montré dans [Geman 84] que le recuit simulé converge probablement vers le minimum global si on dispose d'un « programme de cuisson » (*annealing schedule*) adapté, le programme requis est trop lent pour être d'une utilité pratique. Choisir un programme de cuisson approprié pour fournir des performances utiles dans un temps acceptable est souvent difficile, et puisque le processus fonctionne pendant un temps fini, il conserve une probabilité finie d'atteindre un minimum local, même avec cet algorithme. La convergence de l'optimisation proposée est donc lente.

Le problème de l'unicité de la solution n'est pas traité. Puisque les dimensions de la scène inconnue sont plus grandes que celles des images observées, il est vraisemblable qu'il existe plusieurs images qui, étant donné le processus d'imagerie, produisent des ensembles d'images observées identiques. Le cadre de travail proposé ne fournit pas de moyens de favoriser certaines solutions par rapport à d'autres. En outre, la solution particulière atteinte dans l'espace des solutions possibles sera dépendant de l'estimée initiale de l'image super-résolue, de même que l'ordre dans lequel les valeurs des pixels sont mises à jours. Les effets du bruit sur la stabilité et la qualité de la solution sont insuffisamment traités. La méthode proposée pour la rétro-projection de l'erreur est très simpliste en comparaison des méthodes de reconstructions tomographiques classiques, par exemple. Le mécanisme de correction proposé utilise de manière très limitée la connaissance disponible concernant les caractéristiques du système d'acquisition.

Irani et Peleg dans [Irani 90, Irani 91] étendent les travaux précédents [Peleg 87, Keren 88] en améliorant le méthode de rétro-projection de l'erreur entre les images BR simulées et les images observées. Ils modifient la fonction d'erreur de l'équation (4.14) en :

$$E^{(n)} = \sqrt{\sum_{i=1}^M \sum_{h,l} \left( \hat{g}_i^{(n)}[h, l] - g_i[h, l] \right)^2}. \quad (4.15)$$

Ils considèrent également un processus d'imagerie simulé qui inclut une FTO modélisant les dégradations du système optique. Toutefois, la plus importante modification par rapport à [Peleg 87, Keren 88] est l'utilisation d'un noyau de rétro-projection dans la mise à jour de l'estimée  $\hat{f}^{(n)}[x, y]$ . Dans la procédure de mise à jour proposée, la valeur de chaque pixel HR  $\hat{f}^{(n)}[x, y]$  est mise à jour d'après tous les pixels BR dépendants de sa valeur (via le processus d'imagerie simulé). La contribution d'un pixel BR  $\hat{g}_i^{(n)}[h, l]$  est l'erreur  $\hat{g}_i^{(n)}[h, l] - g_i[h, l]$  multiplié par un facteur qui mesure la contribution relative du pixel HR  $\hat{f}^{(n)}[x, y]$  au pixel BR  $g_i[h, l]$ . Il en résulte que les erreurs dans les pixels BR fortement influencés par un pixel HR particulier, contribuent plus fortement au terme correctif appliqué à l'estimée  $\hat{f}^{(n)}[x, y]$ . La correction peut être écrite sous la forme :

$$\hat{f}^{(n+1)}[x, y] = \hat{f}^{(n)}[x, y] + \sum_{i=1}^M \sum_{\xi} \left( \hat{g}_i^{(n)}[h, l] - g_i[h, l] \right) h_0^{\text{RP}}[x - h, y - l],$$

où  $\xi$  est l'ensemble de pixels BR  $[h, l]$  dans les  $n^e$  images BR dépendant des pixels HR  $\hat{f}^{(n)}[x, y]$ .  $h_0^{\text{RP}}[\cdot, \cdot]$  est un noyau de rétro-projection qui ajuste la correction de l'estimée HR selon l'influence d'un pixel HR donné  $[x, y]$  sur le pixel BR  $[h, l]$  pour laquelle l'erreur est calculée. Les auteurs observent que  $h_0^{\text{RP}}$  peut

être choisi de manière à influencer la solution vers laquelle le processus itératif converge. De plus, le noyau de rétro-projection peut être utilisé de manière à incorporer des contraintes supplémentaires sur la solution telles que la douceur. Cependant, il est important de noter que ces contraintes sont linéaires par rapport aux données. Des contraintes non-linéaires seront rencontrées plus tard. Un théorème fournissant une condition suffisante sur  $h_0^{\text{RP}}$ , pour assurer la convergence de la solution dans le cas spécial de la suppression du flou, est démontré.

Irani et Peleg continuent de développer leurs idées présentées dans [Irani 90, Irani 91] en proposant une procédure très générale de restauration par SR de scènes contenant des mouvements indépendants arbitraires dans [Irani 93]. Leur principale contribution ne concerne pas l'algorithme de SR, inchangé par rapport à [Irani 90, Irani 91], mais l'incorporation d'un algorithme chargé de suivre plusieurs mouvements indépendants. Ce développement permet la restauration par SR pour des objets partiellement cachés, des objets transparents, ou un objet d'intérêt donné.

La détection de mouvement, et des algorithmes d'estimation, sont analysés en détail dans [Irani 94], et les principes sont résumés ici. L'approche proposée utilise une procédure itérative pour identifier et suivre plusieurs mouvements indépendants qui surviennent dans la scène. Un seul mouvement est d'abord estimé, et un objet correspondant à ce mouvement est identifié et suivi. On parle de mouvement « dominant » et d'objet. Lorsque l'objet dominant est identifié et suivi, il est exclu de la région d'analyse, et le processus est répété afin d'identifier d'autres objets ainsi que leur mouvement. La procédure s'interrompt lorsqu'aucun objet supplémentaire n'est identifié. Le mouvement des objets peut être modélisé en utilisant n'importe quel modèle de mouvement paramétrique 2-D. Cette méthode est également étendue afin d'inclure une approche de type « intégration temporelle » qui construit une représentation interne pour chaque objet en mouvement dans plusieurs trames.

Shah et Zakhor utilisent une méthode de restauration de Landweber [Landweber 51] semblable à celle d'Irani et Peleg dans [Shah 96, Shah 99]. Néanmoins, leur principale contribution ne concerne pas les techniques de restauration, mais une nouvelle approche pour résoudre les problèmes causés par des estimées de mouvements incertains dans la restauration par SR. Une méthode d'estimation du mouvement est suggérée, dans laquelle un ensemble de vecteurs de mouvements possibles est stocké pour chaque pixel, dans chaque image BR.

Les auteurs montrent à quel point il est courant que les techniques par correspondance de blocs (*block-matching*) puissent sélectionner des estimés incorrects de mouvement, même si elles maximisent des mesures de corrélation. Ce fait est bien admis : l'estimation du mouvement, tout comme la restauration par SR, est un problème mal posé. Souvent, plusieurs candidats de mouvements sont « optimaux » de manière équivalente, en respectant la mesure de corrélation de bloc. Puisque l'efficacité d'un algorithme de SR dépend de manière critique de la précision des vecteurs de mouvement estimés [Schultz 96], des estimations erronées de mouvement sont indésirables puisqu'elles se manifestent sous la forme d'artefacts désagréables dans l'image restaurée. Étant donné la vraisemblance des erreurs dans l'estimation du mouvement, la situation où le vrai mouvement est rejeté en faveur d'une estimée incorrecte qui a une mesure de corrélation de bloc plus élevée se produit souvent. Cependant en choisissant de maintenir un ensemble de résultats « proches de l'optimal », un recours à d'autres estimées de mouvements possibles est envisageable, si l'estimée du mouvement initial pour un pixel particulier est considéré comme inconsistent par rapport au mouvement des pixels qui l'entourent.

Plus récemment Dekeyser, Bouthemy et Pérez ont proposé une autre extension de [Irani 93] dans [Dekeyser 00]. Les auteurs considèrent un modèle bruité par un bruit blanc additif gaussien, avec une FTO linéaire et invariante par translation. Le modèle de l'équation (4.12) devient :

$$\mathbf{g} = \mathbf{H} \mathbf{f} + \boldsymbol{\epsilon}.$$

Le débruitage est effectué lors d'un pré-traitement utilisant un filtrage spatio-temporel de Wiener. En plus d'être rapide en temps de calcul, ce filtre préserve les détails fins de l'image et n'ajoute pas d'effet de flou dans l'image. Le mouvement inter-images est considéré comme constitué d'un mouvement dominant global qui suit un modèle affine à 6 paramètres. Ce type de modèle de mouvement paramétrique, dont les paramètres peuvent être estimés rapidement, donne de bons résultats. Pour l'initialisation de l'image HR, les auteurs utilisent la première image de la séquence qu'ils interpolent spatialement de façon cubique [Keys 81].

Les méthodes IBP fournissent un cadre de travail utile pour résoudre le problème de SR en proposant un mécanisme pour contraindre la restauration par SR, en se conformant aux images observées. Bien que cette approche soit intuitivement séduisante, elle présente plusieurs inconvénients.

- **Précision du modèle d'imagerie.** La procédure de simulation-et-correction est bonne si le modèle d'imagerie est bon. Si le modèle d'acquisition est imprécis ou inexact, il est alors impossible de retrouver la scène inconnue à partir des images observées. Au mieux, il sera seulement possible de trouver une approximation de la scène originelle avec des erreurs conséquentes de la précision du modèle d'observation. Ceci reste vrai quelle que soit la méthode de restauration avec un modèle du système d'acquisition.
- **Critère de détermination de la métrique de l'erreur.** Il existe plusieurs choix pour considérer la manière dont l'erreur entre les images simulées et observées est calculée. Rien ne permet de dire lequel est optimal.
- **Méthode de correction de l'estimée de la scène HR à partir de l'erreur de simulation.** Même lorsqu'une métrique d'erreur est déterminée, la question de comment corriger l'estimée de la scène inconnue étant donné l'erreur entre les images observées et simulées demeure. La rétro-projection de l'erreur pour corriger l'estimée nécessite une approximation de l'inverse du modèle du système d'acquisition, qui peut être difficile à obtenir, ou présenter des instabilités.
- **Existence de la solution.** Peu d'articles présentés dans cette partie abordent la question de l'existence de la solution. Plusieurs scénarios peuvent conduire à la non-existence de solution. Par exemple, le bruit d'observation mène à des contraintes contradictoires et à un système inconsistant d'équations d'observation. Dans ce cas, il est impossible de trouver une solution au problème inverse de restauration.
- **Unicité de la solution.** Il est souvent supposé, à tort, que minimiser l'erreur de simulation implique une solution raisonnable (si toutefois il en existe une). En fait, ceci n'est pas systématique. Par exemple en restauration d'image par SR, le nombre des inconnues à estimer est souvent bien plus grand que le nombre de données mesurées. Par conséquent, il existe plusieurs images inconnues cohérentes avec les données observées. Ainsi, minimiser (ou même annuler) simplement l'erreur entre les images observées et simulées est insuffisant pour différencier les nombreuses images solutions super-résolues qui satisfont les contraintes sur les données observées.

Corollairement, une méthode itérative convergente ne conduit pas forcément à une solution unique. Des différences dans l'initialisation, ou même dans l'ordre de mise à jour des pixels, peut amener la méthode à converger vers des solutions très différentes, toutes satisfaisant pourtant les contraintes sur les données observées. Il est donc nécessaire de déterminer les conditions sous lesquelles le système itératif converge, et si la solution obtenue est unique ou non.

- **Stabilité de la solution par rapport aux données.** Dans la plupart des méthodes qui viennent d'être présentées, il était supposé qu'il n'y avait pas de bruit d'observation, et que l'ajout de petites quantités de bruit n'affecterait pas grandement la solution. Cependant, ces hypothèses sont souvent fausses. Même des systèmes physiques réalisables sont capables de présenter des variations illimitées dans la solution pour des variations limitées dans les données observées.

Le lecteur aura reconnu les conditions d'Hadamard pour un problème bien posé [Hadamard 23] dans les trois derniers points évoqués. Une méthode puissante pour aborder les problèmes de non-existence, non-unicité et de continuité des problèmes mal-posés passe par l'intermédiaire de l'utilisation de connaissances *a priori* qui contraignent l'espace solution. Ceci n'est pas facilement réalisable à partir du cadre de travail de type simulation-et-correction qui vient d'être traité. Même si Irani et Peleg dans [Irani 90, Irani 91] reconnaissent que le choix du noyau de rétro-projection peut imposer une contrainte de douceur sur l'image restaurée, cette approche n'est ni flexible, ni générale.

Toutefois, les outils pour traiter les problèmes mal posés de manière élégante et efficace existent. Dans les parties suivantes, on considère les méthodes utilisant des contraintes *a priori* pour obtenir des solutions régularisées aux problèmes inverses mal posés. Les méthodes statistiques sont discutées dans la partie suivante, et la formulation POCS basée sur la théorie des ensembles dans la partie 4.4.2.5. Les approches qui utilisent une régularisation de type Tikhonov sont également rapidement mentionnées dans la partie 4.4.2.8.

On peut noter que les méthodes stochastiques et POCS ne sont pas dérivées des méthodes itératives de Landweber ou de type simulation-et-correction dont on vient de parler. Cependant, ces méthodes incluent la méthode itérative de Landweber comme un cas spécifique puisqu'elle permet d'incorporer des contraintes garantissant que la reconstruction par SR est cohérente avec les données observées. Néanmoins, ces méthodes diffèrent sur le fait que la restauration a besoin également de vérifier des contraintes *a priori*. L'ajout de contraintes *a priori* permet l'obtention de solutions de grande qualité, même pour les problèmes inverses mal posés.

#### 4.4.2.4 Méthodes stochastiques

Nous avons vu dans la partie 4.3 que la restauration par SR est typiquement un problème inverse mal posé. Nous avons également noté que l'inclusion d'information *a priori* dans le processus de résolution fournit un mécanisme qui permet d'obtenir des solutions ayant un sens physique pour cette catégorie de problèmes. Récemment, les chercheurs ont reconnu l'utilité des méthodes stochastiques pour la résolution des problèmes mal posés. Contrairement aux précédentes déterministes, ces techniques utilisent un modèle statistique pour définir le modèle d'observation, et elles appliquent des résultats bien connus de la théorie de l'estimation statistique afin d'obtenir des solutions régularisées au problème inverse. Ces méthodes sont qualifiées de « bayésiennes ». Elles sont particulièrement bien adaptées à la résolution de problèmes inverses mal posés, car elles fournissent un cadre de travail naturel à l'ajout de contraintes *a priori* sous la forme de fonction de densité *a priori* sur les inconnues à estimer. Ces techniques bayé-

siennes ont rencontré un très grand succès, et elles sont maintenant centrales pour la résolution de problèmes inverses mal posés dans une grande variété d'applications. L'approche bayésienne est presque synonyme d'estimation MAP, laquelle constitue le point de départ de cette discussion consacrée aux méthodes stochastiques. D'autres méthodes seront également rapidement abordées.

Un modèle linéaire d'observation dans le domaine spatial a déjà été introduit auparavant afin de décrire le processus d'acquisition des images observées à partir d'une scène inconnue. Ainsi, l'estimation de la scène inconnue est devenue un problème de résolution d'un système linéaire inverse. La différence essentielle entre les travaux présentés jusqu'à maintenant et la formulation stochastique dont il est question ici, réside dans la nature des observations et des inconnues. Elles étaient **déterministes** jusqu'alors, et elles sont considérées maintenant comme **statistiques**. Cela suppose que les observations, les inconnues et le bruit sont des réalisations de processus stochastiques, associés à des distributions de probabilités, et reliés entre eux par des équations également stochastiques. Cette présentation commence avec un résumé de la formulation bayésienne MAP appliquée aux problèmes inverses linéaires.

### Méthodes basées sur l'estimateur du Maximum *A Posteriori* (MAP)

On suppose que le modèle d'observation stochastique linéaire qui relie une inconnue  $\mathbf{f}$  aux observations  $\mathbf{g}$ , et auquel on ajoute un bruit additif  $\boldsymbol{\epsilon}$ , est décrit par :

$$\mathbf{g} = \mathbf{H} \mathbf{f} + \boldsymbol{\epsilon}. \quad (4.16)$$

L'approche MAP pour estimer le vecteur inconnu  $\mathbf{f}$ , a pour but de trouver l'estimée  $\hat{\mathbf{f}}_{\text{MAP}}$  pour laquelle la probabilité *a posteriori*  $p(\mathbf{f} | \mathbf{g})$  est maximum, soit :

$$\hat{\mathbf{f}}_{\text{MAP}} = \arg \max_{\mathbf{f}} [p(\mathbf{f} | \mathbf{g})]. \quad (4.17)$$

En appliquant la règle de Bayes à (4.17) on obtient :

$$\hat{\mathbf{f}}_{\text{MAP}} = \arg \max_{\mathbf{f}} \left[ \frac{p(\mathbf{g} | \mathbf{f}) p(\mathbf{f})}{p(\mathbf{g})} \right],$$

où  $p(\mathbf{g} | \mathbf{f})$  est appelée la fonction de vraisemblance,  $p(\mathbf{f})$  est la distribution *a priori* sur la solution et  $p(\mathbf{g})$  est appelée l'évidence. En remarquant que  $\mathbf{g}$  est connu, cette maximisation peut également s'écrire :

$$\hat{\mathbf{f}}_{\text{MAP}} = \arg \max_{\mathbf{f}} [p(\mathbf{g} | \mathbf{f}) p(\mathbf{f})].$$

Et comme le logarithme est une fonction monotone croissante, cette expression est équivalente à la maximisation suivante sous une contrainte de positivité :

$$\hat{\mathbf{f}}_{\text{MAP}} = \arg \max_{\mathbf{f}} [\log(p(\mathbf{g} | \mathbf{f})) + \log(p(\mathbf{f}))],$$

où  $\log(p(\mathbf{g} | \mathbf{f}))$  est appelé la fonction de log-vraisemblance, et  $\log(p(\mathbf{f}))$  est le logarithme de la densité *a priori* de  $\mathbf{f}$ . Ainsi, maximiser la probabilité *a posteriori* nécessite de maximiser une somme de deux termes. Le premier terme,  $\log(p(\mathbf{g} | \mathbf{f}))$ , décrit de manière statistique les contraintes qui relient les



observations à l'inconnue que l'on cherche à estimer. Le second terme,  $\log(p(\mathbf{f}))$ , décrit également de manière stochastique la connaissance sur la forme de l'inconnue. C'est ce deuxième terme qui donne sa puissance à la formulation bayésienne, parce qu'il permet l'ajout de contraintes qui favorisent des solutions ayant un sens physique, dans l'espace de toutes les solutions possibles.

D'autre part, l'équation (4.16) montre que la fonction de vraisemblance est déterminée par la densité de probabilité du bruit  $f_{\epsilon}(\cdot)$ , tel que :

$$p(\mathbf{g} | \mathbf{f}) = f_{\epsilon}(\mathbf{g} - \mathbf{H} \mathbf{f}).$$

Le bruit est souvent supposé être gaussien, et dans ce cas, l'utilisation du logarithme naturel dans l'expression précédente supprime l'exponentielle dans la densité  $f_{\epsilon}(\cdot)$ . Il y a également beaucoup de flexibilité en fonction du choix de la densité *a priori* sur  $\mathbf{f}$ . Ces dernières années les modèles à base de champs de Markov aléatoires (MRF, *Markov Random Field*), bien adaptés à la modélisation des images (voir la partie 5.3.1), ont gagné en importance. Les MRF possèdent une densité de probabilité de Gibbs de la forme :

$$p(\mathbf{f}) = \frac{1}{Z} \exp\left(-\frac{1}{\beta} U(\mathbf{f})\right),$$

où  $Z$  est une constante de normalisation,  $\beta$  est appelé le paramètre de « température », et  $U(\mathbf{f})$  est « l'énergie » de la configuration  $\mathbf{f}$ . D'un point de vue manipulation algébrique et calculatoire, la forme exponentielle de la densité de Gibbs est grandement simplifiée par l'utilisation du logarithme naturel dans la formulation de la solution MAP.

Si le bruit  $\epsilon$  est supposé gaussien, et si l'*a priori* sur la solution  $\mathbf{f}$  est choisi convexe, alors il peut être montré [Schultz 95b] que l'équation d'optimisation (4.17) est convexe, et que la solution  $\hat{\mathbf{f}}_{\text{MAP}}$  existe et est unique. C'est un résultat très important puisqu'il implique que même un problème mal posé peut, avec suffisamment de contraintes sur l'inconnue à estimer, avoir une solution et une seule.

L'étude sur les méthodes MAP de restauration par SR commence avec les travaux de Schultz et Stevenson. Le développement détaillé fournira un cadre de travail théorique clair, représentatif de nombreuses méthodes existantes de restauration MAP. Ceci permettra ensuite de mentionner d'autres articles issus de ces travaux fondateurs.

Schultz et Stevenson étendent leurs précédentes publications [Schultz 92, Schultz 94] consacrées à l'interpolation bayésienne des images en utilisant un modèle *a priori* à base de MRF de Huber (HMRF) au problème de restauration d'image par SR dans [Schultz 95b, Schultz 95c, Schultz 96]. Ils proposent un modèle d'observation linéaire caractérisé par une matrice de sous-échantillonnage et de compensation de mouvement. Celle-ci considère à la fois le mouvement de la scène et de la caméra. Le flou dû au mouvement pendant le temps d'ouverture non nul du capteur n'est pas examiné dans leur travail initial. Le modèle d'observation qu'ils utilisent, et inchangé dans leurs travaux suivants, est maintenant décrit.

On suppose l'observation d'un nombre impair  $M$  d'images BR  $\mathbf{g}[h, l, i]$  avec  $i = c - \frac{M-1}{2}, \dots, c, \dots, c + \frac{M-1}{2}$ ,  $h = 1, \dots, H$  et  $l = 1, \dots, L$ . L'objectif est de restaurer une image super-résolue  $\mathbf{f}[x, y, c]$  avec  $x = 1, \dots, k_0 H$ ,  $y = 1, \dots, k_0 L$  et  $k_0 \in \mathbb{N}$  coïncidant avec l'image centrale  $\mathbf{g}[h, l, c]$  de la séquence des images observées. Un modèle de sous-échantillonnage décrivant l'intégration spatiale des capteurs de la matrice CCD d'acquisition est proposé pour la  $c^e$  image observée :

$$\mathbf{g}[h, l, c] = \frac{1}{k_0^2} \left( \sum_{x=k_0 h - k_0 + 1}^{k_0 h} \sum_{y=k_0 l - k_0 + 1}^{k_0 l} \mathbf{f}[x, y, c] \right).$$

Ainsi, le sous-échantillonnage de l'image centrale est réalisé en « moyennant » un bloc carré de pixels HR. Cette expression modélise l'intégration spatiale de l'intensité lumineuse sur une surface carrée représentant un capteur de la matrice CCD. Cette relation peut être réécrite plus succinctement en ordonnant les pixels des images BR et HR de manière lexicographique :

$$\mathbf{g}_c = \mathbf{H}_c \mathbf{f}_c,$$

avec  $\mathbf{H}_c$  la matrice de sous-échantillonnage de taille  $HL \times k_0^2 HL$ . Chaque ligne de  $\mathbf{H}_c$  fait correspondre un bloc carré d'échantillons HR de taille  $k_0 \times k_0$  à un seul pixel BR. Les autres images observées  $\mathbf{g}_i$  sont reliées à  $\mathbf{f}_c$  par des matrices de sous-échantillonnage et de compensation de mouvement  $\mathbf{H}_i$  dont le rôle est, comme leur nom l'indique, de compenser le mouvement entre les images :

$$\mathbf{g}_i = \mathbf{H}_i \mathbf{f}_c + \mathbf{u}_i, \quad \forall i \neq c.$$

$\mathbf{H}_i$  est également de taille  $HL \times k_0^2 HL$ , et elle relie l'image HR  $\mathbf{f}_c$  à la  $i^e$  image BR observée qui coïncide temporairement avec l'image centrale de la séquence BR. Le vecteur  $\mathbf{u}_i$  contient les pixels non observables à partir de  $\mathbf{f}_c$ , mais présents dans  $\mathbf{f}_i$ . Les éléments de  $\mathbf{u}_i$  ne sont pas connus puisque  $\mathbf{f}_c$  est inconnu. Les lignes de  $\mathbf{H}_i$  qui contiennent de l'information utile sont celles pour lesquelles les éléments de  $\mathbf{g}_i$  sont entièrement observés à partir des éléments compensés de  $\mathbf{f}_c$ . Pour améliorer la robustesse, on supprime les lignes pour lesquelles la précédente affirmation n'est pas vraie, ce qui produit un ensemble réduit d'équations noté :

$$\mathbf{g}'_i = \mathbf{H}'_i \mathbf{f}_c.$$

En pratique,  $\mathbf{H}'_i$  est inconnu et doit être estimé à partir des images observées BR  $\mathbf{g}_i$  et  $\mathbf{g}_c$ . L'estimation de  $\mathbf{H}'_i$  nécessitant l'estimation du mouvement à partir de la séquence BR, des erreurs seront probablement introduites dans le modèle. Pour les prendre en considération, un terme de bruit additif est introduit et une matrice d'observation estimée  $\widehat{\mathbf{H}}'_i$  permet d'étendre le modèle d'observation à :

$$\mathbf{g}'_i = \widehat{\mathbf{H}}'_i \mathbf{f}_c + \boldsymbol{\epsilon}_i,$$

où  $\boldsymbol{\epsilon}_i$  contient les erreurs résultantes de l'utilisation de l'estimée  $\widehat{\mathbf{H}}'_i$ . Schultz et Stevenson supposent que les éléments de  $\boldsymbol{\epsilon}_i$  sont des variables aléatoires gaussiennes indépendantes et identiquement distribuées (i.i.d.). Ainsi, avec  $M$  images observées, le modèle décrit est équivalent au système d'équations suivant :

$$\left\{ \begin{array}{l} \mathbf{g}'_{c-\frac{M-1}{2}} = \widehat{\mathbf{H}}'_{c-\frac{M-1}{2}} \mathbf{f}_c + \boldsymbol{\epsilon}_{c-\frac{M-1}{2}} \\ \vdots \\ \mathbf{g}'_{c-1} = \widehat{\mathbf{H}}'_{c-1} \mathbf{f}_c + \boldsymbol{\epsilon}_{c-1} \\ \mathbf{g}_c = \mathbf{H}_c \mathbf{f}_c \\ \mathbf{g}'_{c+1} = \widehat{\mathbf{H}}'_{c+1} \mathbf{f}_c + \boldsymbol{\epsilon}_{c+1} \\ \vdots \\ \mathbf{g}'_{c+\frac{M-1}{2}} = \widehat{\mathbf{H}}'_{c+\frac{M-1}{2}} \mathbf{f}_c + \boldsymbol{\epsilon}_{c+\frac{M-1}{2}}. \end{array} \right.$$

Celui-ci peut être réécrit en utilisant des vecteurs empilés et des matrices partitionnées sous la forme simplifiée suivante :

$$\mathbf{g} = \mathbf{H} \mathbf{f}_c + \boldsymbol{\epsilon}.$$

L'image inconnue  $\mathbf{f}_c$  est alors estimée en utilisant le critère MAP, à partir de l'ensemble des images observées  $\{\mathbf{g}_i\}$ , soit :

$$\hat{\mathbf{f}}_c = \arg \max_{\mathbf{f}_c} \left[ \log \left( p(\mathbf{f}_c | \{\mathbf{g}_i\}) \right) \right].$$

Cette expression peut être réécrite en appliquant la règle de Bayes :

$$\hat{\mathbf{f}}_c = \arg \max_{\mathbf{f}_c} \left[ \log \left( p(\{\mathbf{g}_i\} | \mathbf{f}_c) \right) + \log \left( p(\mathbf{f}_c) \right) \right].$$

Pour le terme *a priori* sur la solution  $\log(p(\mathbf{f}_c))$ , Schultz et Stevenson utilisent le modèle d'image HMRF préservant les discontinuités. Ce modèle impose des contraintes de douceur sur la restauration, tout en préservant des caractéristiques de contours visuellement importantes (qu'on appelle aussi les discontinuités).

Les erreurs d'estimation sur le mouvement sont supposées indépendantes, ce qui permet d'écrire le terme de vraisemblance sous la forme  $p(\{\mathbf{g}_i\} | \mathbf{f}_c) = \prod_{i=1}^M p(\mathbf{g}_i | \mathbf{f}_c)$ . Si  $\mathbf{H}_c$  est exactement connu, trouver  $\hat{\mathbf{f}}_c$  nécessite de résoudre l'optimisation contrainte suivante :

$$\hat{\mathbf{f}}_c = \arg \max_{\mathbf{f}_c \in \mathcal{F}} \left[ \log \left( p(\{\mathbf{g}_i\}_{i \neq c} | \mathbf{f}_c) \right) + \log \left( p(\mathbf{f}_c) \right) \right]$$

où l'ensemble de contraintes  $\mathcal{F}$  est défini par :

$$\mathcal{F} = \{ \mathbf{f} : \mathbf{H}_c \mathbf{f} = \mathbf{g}_c \}.$$

Cette optimisation contrainte est résolue par une technique de projection par gradient [Bazaraa 93]. Les détails de l'implémentation de cette méthode peuvent être trouvés dans [Schultz 96].

Le modèle utilisé par Schultz et Stevenson suppose un mouvement local de translation. Le mouvement de la scène est évalué en utilisant un estimateur hiérarchique de résolution sous-pixellique par correspondance de blocs, à la fois dans un but d'efficacité calculatoire, et afin d'assurer une cohérence dans les estimées de mouvement.

Convaincus de la nécessité d'une information de mouvement précise pour la restauration par SR, Schultz, Meng et Stevenson ont étudié des techniques sous-pixelliques d'estimation de mouvement. Dans [Schultz 97a, Schultz 98] la procédure MAP de restauration par SR de [Schultz 96] est revue, et plusieurs méthodes d'estimation du mouvement sont étudiées. En particulier, une nouvelle procédure d'estimation du mouvement par bloc qui utilise un modèle projectif à 8 paramètres (basé sur [Mann 94]), et la méthode par flux optique de Horn et Schunck [Horn 81], sont discutées et utilisées dans plusieurs expériences de restauration par SR. Des méthodes permettant de détecter et d'éliminer des estimées incorrectes de vecteur mouvement sont également présentées. Les résultats de SR sont améliorés par la minimisation des artéfacts désagréables résultants de mauvais vecteurs mouvement.

Ayant constaté que l'estimation du mouvement est également un problème mal posé, Schultz et Stevenson proposent d'utiliser une méthode régularisée d'estimation du mouvement dans [Schultz 97b], laquelle favorise la douceur des champs de mouvement tout en autorisant les discontinuités. Ceci est réalisé en utilisant des techniques bayésiennes d'estimation du mouvement, une contrainte *a priori* de douceur de type HMRF, et la méthode d'estimation du flux optique de Horn et Schunck. Les restaurations d'image par SR à partir d'une séquence d'image BR, avec le mouvement évalué à partir de l'estimateur de mouvement bayésien, sont comparées avec celles utilisant les techniques d'estimation de mouvement par bloc de [Schultz 96]. Les résultats dans le premier cas sont trouvés visuellement supérieurs par les auteurs.

Il est intéressant de noter que dans [Schultz 97a, Schultz 97b, Schultz 98] les estimées de mouvement sont calculées seulement à partir de paires d'images. Il est raisonnable d'anticiper une meilleure fiabilité dans les estimées de mouvement calculées en utilisant d'avantage d'images. On peut noter également que les techniques d'estimation de mouvement abordées ne sont pas basées sur des objets comme c'était le cas dans [Irani 93]. Rien n'est tenté pour construire des représentations internes pour plusieurs mouvements indépendants dans la scène. Ces aspects apparaissent comme des problèmes de recherche ouverts.

Hardie, Barnard et Armstrong présentent une méthode MAP de SR dans [Hardie 97b] presque identique à celle présentée dans [Schultz 96]. Ils considèrent les cas d'estimation de mouvements globaux et locaux. Pour le cas global, ils utilisent le modèle de rotation et de translation de [Irani 91], et pour le cas local, ils se basent sur les techniques de flux optique décrites dans [Tekalp 95] avec segmentation du premier plan et du fond. La formulation du problème de SR est équivalente à celle de Schultz et Stevenson sous l'hypothèse de bruit gaussien, et en utilisant un modèle *a priori* MRF gaussien. Ce travail est étendu dans [Hardie 97a] en considérant une estimation simultanée de l'image super-résolue  $\hat{\mathbf{f}}$  et des paramètres de mouvement  $\hat{\mathbf{s}}$ . Une formulation MAP similaire à celle de [Hardie 97b] est utilisée, mais en incluant les paramètres inconnus de mouvement dans l'optimisation :

$$(\hat{\mathbf{f}}, \hat{\mathbf{s}}) = \arg \max_{\mathbf{f}, \mathbf{s}} [p(\mathbf{f}, \mathbf{s} | \mathbf{g})]. \quad (4.18)$$

Puisque l'estimation de  $\hat{\mathbf{f}}$  nécessite les estimées de mouvement  $\hat{\mathbf{s}}$ , une procédure est suggérée dans laquelle  $\hat{\mathbf{f}}$  et  $\hat{\mathbf{s}}$  sont estimés alternativement. Cela signifie que  $\hat{\mathbf{s}}$  est fixé et  $\hat{\mathbf{f}}$  est estimé, puis que  $\hat{\mathbf{f}}$  est fixé et  $\hat{\mathbf{s}}$  est estimé. Bien que cette procédure puisse être longue à converger, elle possède des avantages. Le mouvement n'est plus estimé directement à partir des images BR comme dans la plupart des algorithmes de SR. À la place, les paramètres de mouvement sont choisis pour minimiser l'erreur entre les images BR observées et les images BR simulées. Ces dernières sont calculées à partir du modèle d'observation et de l'estimée HR courante. Des *a priori* gaussiens sont utilisés à la fois pour l'image HR et le mouvement. L'optimisation de l'équation (4.18) est réalisée en utilisant une méthode itérative de descente de coordonnées. Les auteurs observent qu'il n'est pas possible d'utiliser la méthode du gradient conjugué, plus rapide, puisque le coût fonctionnel n'est pas constant d'une itération à l'autre, à cause de l'estimation jointe des paramètres de mouvement et de l'image super-résolue. L'estimation alternée des paramètres de mouvement  $\hat{\mathbf{s}}$  et de l'image restaurée  $\hat{\mathbf{f}}$  évoque l'algorithme EM (*Expectation Maximization*) [Dempster 77] qui fournit un cadre de travail général pour l'estimation statistique sous les conditions de données incomplètes.

La présentation plus détaillée d'un modèle d'observation qui inclut une translation et une rotation globale, une FTO du système optique et du capteur peut être trouvée dans [Hardie 98]. Le recalage des

images y est réalisé en utilisant la méthode d'Irani et Peleg [Irani 91], et de Lucas et Kanade [Lucas 85]. Dans ce papier, le mouvement est estimé séparément de la phase de restauration par SR. Les méthodes de descente de gradient et de descente de gradient conjugué sont utilisées pour maximiser la probabilité *a posteriori*.

Lorette, Shekarforoush et Zerubia présentent également une formulation MAP du problème de SR dans [Lorette 97]. Dans un autre travail [Shekarforoush 95], ils considèrent le problème de restauration 3-D par SR. Dans [Lorette 97], les auteurs étudient trois fonctions de pénalités non-linéaires pour leur modèle *a priori* MRF qui permet la restauration de contours, tout en assurant la douceur. Cette approche est semblable au modèle d'image HMRF préservant les contours utilisé par Schultz et Stevenson.

Dans [Cheeseman 94], Cheeseman, Kanefsky, Kraft et Stutz ont appliqué les techniques MAP de restauration par SR aux images reçues des véhicules distants explorant le système solaire, en particulier en provenance de *Viking*, *Voyager* et plus récemment des missions de *Mars Pathfinder*. Leur formulation suppose un bruit gaussien et utilise des termes *a priori* qui mène à un système linéaire d'équations. Ces équations sont résolues en utilisant les méthodes de Gauss-Jacobi.

Hunt, Sementilli, Nadar et Yuen ont réalisé des travaux sur une formulation Poisson MAP de la SR, et ils ont fourni une analyse théorique des performances de cette technique de SR dans [Hunt 91, Hunt 92, Sementilli 93, Sementilli 94, Yuen 94, Hunt 95]. Ces travaux se concentrent sur la restauration d'images astronomiques, et ils font des hypothèses concernant le support borné des objets reconstruits non valides dans le problème plus général de la restauration par SR de vidéo. L'hypothèse de statistiques de type Poisson pour les photons dans le modèle d'observation, bien qu'essentiel pour une restauration efficace d'images astronomiques, est moins approprié pour les applications de restauration vidéo où de faible quantité de photons ne sont généralement pas un problème. Une grande partie de leurs travaux se concentre sur des techniques de restauration par SR bien connues dans la communauté astronomique, mais sont d'une application limitée pour la restauration générale de séquences vidéos. Ces travaux complètent l'étude est sont une source potentielle d'idées théoriques.

### Méthodes basées sur l'estimateur du maximum de vraisemblance

Tom et Katsaggelos s'intéressent au problème de restauration par SR dans [Tom 94b, Tom 94c]. Ils considèrent que sa résolution passe par trois étapes : le recalage des images BR, leur restauration, suivie par une interpolation qui permet d'obtenir l'image super-résolue. Afin d'améliorer la restauration, une approche simultanée de recalage et de restauration est proposée dans [Tom 94c]. Ce travail s'appuie sur les idées d'identification et de restauration d'image multicanal développées dans [Tom 94a]. Dans [Tom 95], ces idées sont étendues pour réaliser un recalage, une restauration et une interpolation combinés.

Ces papiers sont essentiellement basés sur la formulation de maximum de vraisemblance (MV, ou ML en anglais pour *Maximum Likelihood*) pour le traitement des étapes de recalage, de restauration et d'interpolation. Un problème fondamental de cette approche est que l'estimation MV est faiblement adaptée à la résolution de problèmes inverses mal posés, puisqu'elle ne permet pas d'inclure d'information *a priori*. En fait, l'estimation MV est un cas particulier de l'estimation MAP, sous l'hypothèse d'*a priori* uniforme. Or tout au long de cette partie, nous avons vu que la régularisation en utilisant des contraintes *a priori* constitue une condition nécessaire à l'obtention de résultats de restauration de haute qualité. Ces travaux n'abordent pas suffisamment ce sujet, en dépit de leur contribution à la problématique de restauration et de recalage simultanés.

Elad et Feuer dans [Elad 96, Elad 97] ont également considéré (parmi d'autres techniques qui seront

traitées plus tard) une formulation basée sur l'estimation MV, mais les mêmes critiques peuvent être appliquées à ces travaux. Plus précisément dans [Elad 96], ils modifient l'estimé MV pour inclure une fonction de régularisation qui, dans ses effets, est équivalent à une formulation MAP en utilisant un *a priori* gaussien ou à une régularisation de type Tikhonov [Tikhonov 77]. Ils suggèrent aussi que les méthodes de type simulation-et-correction produisent des estimées MV de la SR.

### Résumé sur les méthodes de restauration stochastiques

Les techniques de restauration statistique, en particulier la formulation MAP qui utilise des modèles *a priori* de MRF, sont devenues importantes dans la littérature de la restauration par SR, pour plusieurs raisons :

- Le cadre d'estimation MAP permet **l'incorporation directe de contraintes *a priori* sur la solution**. Ceci s'avère essentiel à l'obtention de solutions de haute qualité au problème inverse mal posé de restauration par SR.
- Des *a priori* MRF sont utilisés d'une façon **considérable** car ils fournissent **une méthode intuitive, pratique et réaliste à la modélisation de caractéristiques spécifiques des images, en utilisant seulement un modèle d'interaction de voisinage local**. Avec un choix judicieux de la fonction de pénalité et de la mesure d'activité spatiale dans le modèle MRF, des *a priori* peuvent être définis qui fournissent les contraintes de douceur nécessaires à assurer des restaurations acceptables sous des conditions de données limitées ou pauvres, mais permet également la restauration efficace de discontinuités (contours).
- **La restauration et l'estimation simultanées du mouvement sont possibles** dans le cadre de l'estimation MAP.

La formulation de l'estimation MAP représente l'une des approches des plus prometteuses, et des plus flexibles, à la restauration d'image par SR. Ses seuls concurrents en terme d'efficacité et d'adaptabilité sont les méthodes de restauration théoriques basées sur des ensembles.

#### 4.4.2.5 Méthodes théoriques basées sur des ensembles

Cette partie traite des méthodes de résolution du problème de restauration par définition des ensembles de contraintes à satisfaire par les solutions potentielles. Ceci inclut au minimum le respect de contraintes assurant que ces solutions sont logiques avec les données observées. Ainsi il est commun (surtout pour les problèmes mal posés) d'inclure des contraintes représentant une connaissance *a priori* concernant l'espace des solutions réalistes. Dans le but de trouver des solutions qui satisfont simultanément la totalité des ensembles de contraintes, des procédures itératives sont souvent utilisées. L'utilisation d'ensemble de contraintes convexes mène aux méthodes de projection sur des ensembles convexes (POCS).

#### Méthodes de projection sur des ensembles convexes (POCS)

Nous avons remarqué précédemment l'importance des techniques d'estimation bayésiennes pour la restauration d'images par SR. Seules les méthodes POCS offrent des résultats comparables dans la littérature sur les méthodes de SR. Dans la formulation POCS, des ensembles de contraintes sont utilisés pour définir l'espace solution réaliste pour les images super-résolues à estimer. Les contraintes sont

définies comme des ensembles convexes dans l'espace vectoriel  $\mathbb{R}^{kH \times kL}$  qui représente l'espace contenant toutes les images HR. Les ensembles offrant des caractéristiques souhaitées de la solution sont définis sur cet espace. Ils comprennent des contraintes telles que la fidélité aux données observées, la positivité, l'énergie limitée, la douceur, etc. L'espace solution du problème de SR est, par construction, l'intersection des ensembles de contraintes convexes. POCS est une méthode itérative qui, à partir de n'importe quel point dans l'espace vectoriel, localise un point qui satisfait simultanément toutes les contraintes. Soient  $q$  ensembles de contraintes convexes dans  $\mathbb{R}^{kH \times kL}$  tels que l'intersection des ensembles est non vide. POCS projette un point dans l'espace vectoriel sur chaque ensemble de contraintes à la suite, en répétant le processus jusqu'à ce qu'un point se trouvant dans l'intersection des  $q$  ensembles soit atteint. Ce mécanisme itératif converge à condition que les ensembles de contraintes soient convexes. Une étude théorique détaillée des méthodes POCS peut être trouvée dans [Youla 78, Youla 82, Combettes 93]. Des analyses sur l'utilisation de techniques POCS en restauration d'images sont présentées dans [Sezan 84, Oskoui-Fard 88].

Les méthodes POCS ont suscité beaucoup d'intérêt, dans une multitude d'applications de reconstruction et de restauration d'images, pour trois principales raisons.

- La formulation POCS est très intuitive, et généralement **simple à implémenter**. La seule source potentielle de difficulté réside dans la détermination des opérateurs de projection.
- La méthode POCS étant formulée dans le domaine spatial, **des modèles très généraux de mouvement et d'observation peuvent être utilisés**. La complexité de ces modèles a peu de conséquences sur les procédures de résolution POCS.
- La **facilité avec laquelle de l'information a priori peut être incluse** est certainement l'aspect le plus utile de la formulation POCS. En général, il est simple de définir des ensembles de contraintes convexes incorporant les caractéristiques souhaitées de la solution. Ces ensembles peuvent imposer des restrictions telles que la positivité ou l'énergie limitée, difficiles à représenter en termes de fonctions de coût.

L'étude consacrée aux méthodes POCS commence par l'analyse détaillée d'un des premiers papiers consacrés à une technique POCS de restauration par SR. Celle-ci démontre de nombreuses idées fondamentales. Cette présentation permettra une explication plus rapide des travaux qui ont suivi.

Stark et Oskoui proposent une des premières approches POCS pour la résolution du problème de SR dans [Stark 89]. En particulier, leur papier s'intéresse au problème d'une matrice linéaire d'acquisition initialement abordé par Frieden et Aumann (voir la partie 4.4.2.2), et au problème de restauration d'une image super-résolue à partir de plusieurs images obtenues par des matrices planes. Cette présentation commence par un rapide aperçu du modèle d'imagerie proposé. Le cadre théorique est développé pour le cas d'une image invariante dans le temps. L'extension de ce modèle au cas de l'imagerie variant dans le temps avec plusieurs acquisitions d'images peut être déduit du cas précédent.

Le système optique projette une image continue  $f(h, l)$  sur la matrice de capteurs d'acquisitions, qui comporte une matrice régulière constituée de  $H \times L$  détecteurs. La sortie  $g_m$  du  $m^e$  capteur de la matrice d'acquisition, dont la réponse spatiale caractéristique est  $\sigma_m(h, l)$ , est donnée par :

$$g_m = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(h, l) \sigma_m(h, l) dh dl, \quad 1 \leq m \leq HL.$$

Cette intégration sur les variables spatiales continues  $(h, l)$  peut être discrétisée sur la grille de SR  $[x, y]$

produisant :

$$g_m = \sum_{x=1}^X \sum_{y=1}^Y f[x, y] \sigma_m[x, y], \quad (4.19)$$

où  $X > H$  et  $Y > L$  sont les dimensions de la matrice de SR. La réponse spatiale caractéristique du  $m^e$  détecteur  $\sigma_m(h, l)$  prend la forme discrète  $\sigma_m[x, y]$ , et elle constitue la région fractionnelle du pixel super-résolu  $[x, y]$  contenu dans la région de réponse du  $m^e$  détecteur BR. Ceci suppose que la réponse du détecteur est uniforme dans sa région de perception. Ainsi  $\sigma_m[x, y]$  prend les valeurs 0, 1 et  $r_m$  avec  $r_m \in ]0, 1[$ , si le pixel super-résolu  $[x, y]$  se trouve respectivement entièrement à l'extérieur, entièrement à l'intérieur, et partiellement dans la région de réponse du  $m^e$  détecteur.

Si on ordonne de façon lexicographique  $f[x, y]$  et  $\sigma_m[x, y]$  de manière à obtenir les vecteurs colonnes  $\mathbf{f}$  et  $\boldsymbol{\sigma}_m^\top$ , l'équation (4.19) peut être réécrite sous la forme compacte suivante :

$$g_m = \boldsymbol{\sigma}_m^\top \mathbf{f}.$$

On considère ensuite les ensembles  $C_m$  définis par :

$$C_m = \{\mathbf{f}; \boldsymbol{\sigma}_m^\top \mathbf{f} = g_m\}, \quad 1 \leq m \leq HL. \quad (4.20)$$

À  $m$  fixé,  $C_m$  représente l'ensemble de toutes les images SR discrétisées pour lesquelles la réponse du  $m^e$  capteur de la matrice de détecteurs est  $g_m$ , la valeur observée. Puisque cette matrice est constituée de  $H \times L$  détecteurs, à chaque image observée correspond un ensemble  $\{C_m\}$  avec  $m \in [1; HL] \subset \mathbb{N}$ . Ces ensembles contraignent les valeurs possibles que peuvent prendre une estimée de la solution  $\mathbf{f}$ . En particulier, les contraintes décrites par l'équation (4.20) assurent que l'image super-résolue  $\mathbf{f}$  est cohérente avec les données observées  $g_m$ . En observant que les contraintes sur  $\mathbf{f}$  sont linéaires, il est possible en principe d'obtenir un nombre suffisant d'équations, de manière à ce qu'une solution pour  $\mathbf{f}$  puisse être trouvée par inversion matricielle.

Pour restaurer une image HR de  $X \times Y$  pixels par inversion matricielle,  $XY$  équations d'observations indépendantes sont requises. Et puisque  $X > H$  et  $Y > L$ , on remarque immédiatement que plus d'une image observée est nécessaire afin d'obtenir un nombre suffisant d'équations permettant de trouver une solution par inversion matricielle. Pour obtenir  $XY > HL$  équations indépendantes, de l'information nouvelle doit être obtenue, à partir d'images observées supplémentaires. Cette nouvelle information est dérivée des différences entre les images observées, et elle est souvent liée au mouvement de la scène ou de la caméra. Dans les environnements d'acquisition, où les mouvements de la scène ou de la caméra sont incontrôlés, obtenir exactement le bon mouvement qui assurera une matrice d'observation de rang plein peut être impossible, ou peut nécessiter un grand nombre d'images observés. Même lorsque cette matrice d'observation est de rang plein, la présence de bruit dans les observations peut encore empêcher l'obtention d'une solution par inversion matricielle, puisque les données observées peuvent impliquer un système linéaire illogique et inconsistant.

En réalité, les images observées supplémentaires contribuent à des ensembles supplémentaires de la forme de ceux définis par l'équation (4.20). Ces contraintes additionnelles augmentent le système d'équations sous-déterminé, dérivé pour une seule image, de manière à ce que le système approche le cas du rang plein.



La collection d'ensembles  $\{C_m\}$  est souvent considérée comme des contraintes de cohérence sur les données. Ceci permet la définition d'un opérateur de projection  $\mathcal{P}_m$  :

$$\mathcal{P}_m \mathbf{f} = \begin{cases} \mathbf{f} & \text{si } \boldsymbol{\sigma}_m^\top \mathbf{f} = g_m \\ \mathbf{f} + \frac{g_m - \boldsymbol{\sigma}_m^\top \mathbf{f}}{\boldsymbol{\sigma}_m^\top \boldsymbol{\sigma}_m} \boldsymbol{\sigma}_m & \text{sinon.} \end{cases}$$

$\mathcal{P}_m \mathbf{f}$  est la projection du point  $\mathbf{f}$  sur l'ensemble  $C_m$ . Dans la formulation POCS, une estimation initiale  $\mathbf{f}^{(0)}$  de l'image super-résolue est projetée sur chacun des ensembles de contraintes  $C_m$  pour assurer la cohérence avec les données mesurées  $g_m$ . Soit  $K$  le nombre total des ensembles de contraintes sur les pixels observés, la  $n^e$  estimée  $\mathbf{f}^{(n)}$  est mise à jour en utilisant la séquence de projection suivante :

$$\mathbf{f}^{(n+1)} = \mathcal{P}_1 \mathcal{P}_2 \dots \mathcal{P}_K \mathbf{f}^{(n)}. \quad (4.21)$$

Cette séquence de projection est appliquée de manière récurrente pour produire une estimée mise à jour de l'image super-résolue. L'aspect fermé et convexe des ensembles de contraintes assure la convergence de la procédure itérative à un point satisfaisant toutes les contraintes [Youla 82]. Il est important de remarquer que ce point est, en général, non-unique. Il se situe sur la surface convexe définie par l'intersection des ensembles de contraintes. Généralement, la solution est également dépendante de l'estimation initiale.

Il est intéressant de remarquer que l'application de la méthode POCS en utilisant uniquement la contrainte de cohérence des données de l'équation (4.20) est identique à la technique de reconstruction algébrique dans la littérature de l'imagerie tomographique [Herman 80].

En plus des contraintes de cohérence sur les données, des contraintes supplémentaires représentant la connaissance *a priori* sur la forme de la solution peuvent être ajoutées. Par exemple, le domaine des valeurs dans l'image solution peut être contraint à l'ensemble  $C_A$  :

$$C_A = \{\mathbf{f}; \alpha \leq f_m \leq \beta, \alpha < \beta\}.$$

Une autre possibilité est d'exiger que la solution soit à énergie finie en utilisant l'ensemble de contraintes :

$$C_E = \{\mathbf{f}; \|\mathbf{f}\|_2^2 \leq E\}.$$

Si on sait que l'image est semblable à une certaine image de référence  $\mathbf{f}_R$ , alors la solution peut être contrainte de manière à se situer à une distance  $\epsilon_R$  de  $\mathbf{f}_R$  en utilisant les ensembles,

$$C_R = \{\mathbf{f}; \|\mathbf{f} - \mathbf{f}_R\| \leq \epsilon_R\}.$$

Un support limité peut également être imposé en définissant un ensemble de points  $A$  pour lesquels on exige que la solution soit zéro, soit :

$$C_S = \{\mathbf{f}; f_m = 0, \forall m \in A\}.$$

Des contraintes supplémentaires sur la solution peuvent être définies d'une manière similaire. L'itération définie par l'équation (4.21) est alors augmentée pour inclure ces ensembles de contraintes supplémentaires. L'ajout de connaissance *a priori* contraint l'espace solution, permettant ainsi d'obtenir des performances robustes en présence de bruit, de données contradictoires ou manquantes.

Le travail de Stark et Oskoui [Stark 89] constitue la première application de la méthode POCS à la restauration d'images par SR. L'approche présentée dans leur article contient plusieurs inconvénients, et certains d'entre eux sont abordés dans d'autres travaux. Pour l'instant le modèle d'observation proposé n'inclut pas de bruit, et pour la  $i^e$  image observée il est de la forme :

$$\mathbf{g}_i = \mathbf{H}_i \mathbf{f},$$

où  $\mathbf{H}_i = [\boldsymbol{\sigma}_1^{i\top} \ \boldsymbol{\sigma}_2^{i\top} \ \dots \ \boldsymbol{\sigma}_{HL}^{i\top}]^t$ . Dans le cas où l'on a  $M$  images observées, le modèle devient :

$$\mathbf{g} = \mathbf{H} \mathbf{f},$$

avec  $\mathbf{g} = [\mathbf{g}_1^t \ \mathbf{g}_2^t \ \dots \ \mathbf{g}_M^t]^t$  et  $\mathbf{H} = [\mathbf{H}_1^t \ \mathbf{H}_2^t \ \dots \ \mathbf{H}_M^t]^t$ .

Tekalp, Ozkan et Sezan proposent trois approches différentes à la restauration par SR dans [Tekalp 92]. Seule l'approche basée sur la méthode POCS et qui étend le travail [Stark 89] est abordée ici. Les deux autres méthodes qui sont une formulation dans le domaine fréquentiel et une méthode d'interpolation/restauration ont déjà été présentées respectivement dans les parties 4.4.1.1 et 4.4.2.1. La première contribution de ce papier réside dans l'utilisation d'un modèle d'observation incluant le bruit. Les auteurs proposent un modèle d'observation mono-image de la forme :

$$\mathbf{g} = \mathbf{H} \mathbf{f} + \boldsymbol{\epsilon},$$

où le système matriciel  $\mathbf{H}$  inclut les effets de la FTO d'un capteur, et un mouvement global de translation. Pour rendre compte du bruit d'observation, la contrainte de cohérence des données pour le  $m^e$  pixel d'observation qui était une contrainte d'égalité dans l'équation (4.20) du modèle de Stark et Oskoui, est modifiée pour permettre une erreur  $\delta_0$  telle que :

$$C_m = \{\mathbf{f}; |r_m| < \delta_0\}, \quad (4.22)$$

où  $r_m$  est le  $m^e$  élément du résidu  $\mathbf{r} = \mathbf{g} - \mathbf{H} \mathbf{f}$ . Dans la mesure où  $\mathbf{g} - \mathbf{H} \mathbf{f} = \boldsymbol{\epsilon}$ , le résidu  $\mathbf{r}$  n'est que le bruit d'observation  $\boldsymbol{\epsilon}$ , et  $\delta_0$  représente la confiance dans l'observation, et peut être fixé d'après les statistiques du bruit. Les auteurs définissent l'opérateur de projection nécessaire pour l'ensemble défini par l'équation (4.22), et montrent des résultats de simulation dans lesquels une image super-résolue est restaurée à partir de plusieurs images sous-échantillonnées et globalement translatées. Bien que ce travail aborde le problème du bruit d'observation, le modèle de mouvement utilisé suppose seulement un mouvement global de translation entre les images, et ne considère pas les effets de flou liés au mouvement. Toutefois il ne s'agit pas d'un constat lié à une limitation propre de la méthode POCS, et cet aspect est traité dans le travail [Patti 94] présenté ci-dessous.

Patti, Sezan et Tekalp abordent dans [Patti 94] les manques de [Tekalp 92]. Ils utilisent le même modèle d'observation général  $\mathbf{g} = \mathbf{H} \mathbf{f} + \boldsymbol{\epsilon}$ , ainsi qu'une formulation POCS pour résoudre ces équations. Leur contribution réside dans le modèle de mouvement compris dans le système matriciel  $\mathbf{H}$ , et qui diffère des travaux précédents. Plus précisément, la matrice  $\mathbf{H}$  incorpore une FTO variant spatialement

qui modélise les dégradations causées par un échantillonnage avec un capteur BR, ainsi que les effets de mouvement relatif entre le capteur et la scène (ce qui inclut le flou lié à une ouverture non nulle en temps). La FTO variant spatialement résulte de trois étapes de modélisation : d'abord l'ajout du flou causé par le mouvement relatif entre la scène et le capteur pendant le temps d'ouverture ; suivi par les effets des dimensions physiques et des caractéristiques de la réponse des capteurs BR (on parle d'effets de flou lié au capteur) ; et enfin, les effets d'échantillonnage d'une image BR sur une grille HR, en rapport avec la matrice de récepteurs. Ce modèle combiné définit le système matriciel  $\mathbf{H}$ .

Le modèle de mouvement proposé, inspiré de celui de Sawchuk [Sawchuk 72], permet l'inclusion d'une classe très générale de mouvements de la scène. En particulier, les auteurs montrent comment il est possible en théorie d'inclure un mouvement de scène général, tant qu'aucune occlusion de l'objectif n'est présente. Une extension théorique illustre comment la FTO variant spatialement peut être trouvée pour ce cas général. Ce modèle est illustré à partir de l'exemple de mouvements à vitesse temporairement constante par morceaux, avec le cas du mouvement global de translation complètement développé, puis utilisé dans les simulations. Un chemin à vitesse constante par morceau est défini comme ayant des vitesses de mouvement qui peuvent changer d'une image à une autre, mais restent constantes pendant le temps d'ouverture. La restauration de l'image HR est réalisée à partir du cadre de travail POCS, en utilisant ce qui a été développé dans [Tekalp 92].

Pour résumer, [Patti 94] fournit un cadre de travail pour la restauration par SR basé sur POCS qui considère le recouvrement de spectre, le flou de capteur, le flou de mouvement et le bruit additif. À noter que le flou spatial dû au mouvement pendant le temps d'ouverture n'avait pas été traité auparavant. Ce travail aborde également les restrictions liées à l'utilisation de flous linéaires et invariants par translation, et de bruits invariants spatialement.

En plus de leur travail sur le problème de restauration par SR, Patti, Sezan et Tekalp ont également publié dans le domaine de la conversion de standards vidéos [Patti 95]. L'objectif est la conversion de format vidéos entre différentes matrices d'échantillonnages spatio-temporelles définies par des standards. Leur travail étend [Tekalp 92, Patti 94] de manière à inclure une matrice d'échantillonnage spatio-temporelle arbitraire en entrée. La forme du modèle d'observation discret reste inchangée par rapport au modèle proposé qui utilise une FTO variant spatialement dans [Patti 94]. Une formulation POCS identique à celle de [Patti 94] est utilisée pour la restauration. Dans un travail plus récent [Patti 97a], la restauration d'images super-résolues à partir de vidéo entrelacée est traitée. Ce papier contient plusieurs extensions en terme de détection et d'analyse de mouvement, en utilisant à la fois un modèle de mouvement dominant global et affine estimés à partir de techniques basées sur le gradient, et des régions segmentées de mouvement locaux.

Les détails complets du travail de Patti, Sezan et Tekalp peuvent être trouvés dans [Patti 97b]. Ce papier consolide et perfectionne les précédents travaux [Tekalp 92, Patti 94, Patti 95]. Le modèle d'observation incorpore une FTO variant spatialement qui prend en considération les mouvements de scène et de caméra, le flou de mouvement dû à un temps d'ouverture non nul, les dimensions physiques des éléments du capteur BR et le flou causé par le système optique d'acquisition. Le bruit de capteur est également considéré. Une innovation de [Patti 97b] réside dans l'ajout d'un échantillonnage de la scène continue sur une matrice arbitraire d'espace-temps. Une présentation détaillée du modèle de mouvement est fournie, ainsi que les techniques utilisées pour l'estimation du mouvement. La restauration se fait toujours par la méthode POCS présentée. Les résultats de plusieurs simulations sont exposés pour montrer les performances de la méthode proposée.

Il est intéressant de noter que les méthodes IBP présentées dans la partie 4.4.2.3 constituent un cas particulier de la procédure POCS où la contrainte de cohérence des données de l'équation (4.20) nécessite l'égalité sans autoriser une tolérance d'erreur.

Les extensions des modèles d'observation pour inclure des mouvements locaux arbitraires rencontrent le problème de l'apparition d'artéfacts provoquée par des estimées de mouvement incertaines. Estimer un mouvement local est généralement moins fiable que d'estimer un mouvement global. Ceci est dû à un manque de données. Un mouvement global peut souvent être estimé de manière fiable car un grand nombre de pixels est utilisé dans le calcul. Cependant pour un mouvement local, une estimée doit être calculée en utilisant seulement un petit voisinage local de pixels, et par conséquent, la fiabilité de l'estimation de mouvement est réduite. Si l'on souhaite réaliser une restauration par SR de haute qualité, il est nécessaire d'obtenir des estimées de mouvement justes et précises, avec une résolution sous-pixelle. Il est impératif que les estimées de mouvement soient fiables, pour éviter que des artéfacts visuels ne dégradent la restauration. Dans [Eren 97], Eren, Sezan et Tekalp étendent [Patti 97b] afin de considérer les moyens d'améliorer la robustesse des méthodes de SR aux erreurs dans les estimations de mouvement. Pour ce faire, deux nouveaux concepts sont introduits : une « carte de validité » et une « carte de segmentation ». La carte de validité empêche les projections basées sur les observations pour lesquelles les estimations de mouvement sont imprécises, tandis que la carte de segmentation permet une poursuite basée sur les objets et un traitement où des modèles de mouvements plus précis basés sur les objets peuvent être utilisés pour améliorer la qualité de la restauration. La carte de segmentation permet également la restauration d'objets spécifiques d'une manière similaire à [Irani 93].

### Méthode d'ellipsoïdes englobants

Une variante de la formulation POCS qui utilise un ellipsoïde pour délimiter les ensembles de contraintes a été examinée par Tom et Katsaggelos dans [Tom 96a, Tom 96b], et a été brièvement mentionnée par Elad et Feuer dans [Elad 96, Elad 97]. Un ellipsoïde englobant est calculé à partir d'ensembles de contraintes ellipsoïdales [Katsaggelos 90]. Le centre de gravité de cet ellipsoïde est choisi comme estimée de l'image HR super-résolue. Comme un calcul direct de ce point est habituellement impossible, une méthode itérative est utilisée. On peut remarquer que cette approche prend une forme proche de celle des méthodes régularisées. Le modèle d'observation utilisé dans ce travail est semblable à celui proposé par Schultz et Stevenson dans [Schultz 96].

### Résumé sur les méthodes de restauration théoriques sur des ensembles

Cette partie a abordée essentiellement les premières méthodes basées sur la formulation POCS, et a donné un rapide aperçu de la méthode d'ellipsoïdes englobants. Les avantages de l'approche de restauration POCS inclut :

- la simplicité,
- un modèle d'observation dans le domaine spatial flexible,
- la possibilité d'inclure de l'information *a priori* facilement.

L'étude présentée a clairement montré ces aspects. Toutefois, la formulation POCS présente également des inconvénients tels que :

- la non unicité de la solution. L'espace des solutions possibles pour les méthodes POCS est défini comme l'intersection des ensembles de contraintes convexes. Par conséquent, à moins que cette

intersection ne soit un point, la solution n'est pas unique. Cet aspect constitue un inconvénient sérieux. Dans les formulations MAP considérées, une optimisation convexe était favorisée afin d'assurer l'existence et l'unicité de la solution. Pour POCS, ce n'est en général pas possible.

- **la dépendance de la solution avec la valeur d'initialisation** constitue un autre inconvénient majeur. La solution obtenue à partir d'une méthode itérative est dépendante de l'initialisation arbitraire, ainsi que de l'ordre d'application des opérateurs de projection. Les méthodes POCS projettent de manière itérative l'estimée courante de l'image HR sur les ensembles de contraintes. La solution est le premier point de l'espace vectoriel des solutions possibles qui satisfont tous les ensembles de contraintes. Lorsqu'on utilise des projecteurs simples, la solution se trouve sur la surface du volume d'intersection des ensembles de contraintes.
- **le coût calculatoire.** Les méthodes POCS nécessitent des calculs considérables et un grand nombre d'itérations pour atteindre la convergence. Une analyse de ces méthodes montre que sans l'utilisation de projections « relaxées » ou sans modifications semblables, POCS peut être très lent à converger au fur et à mesure que la solution est approchée. C'est pourquoi en général, une méthode POCS ne fonctionne pas jusqu'à convergence, mais jusqu'à un point où l'estimée est acceptable visuellement.
- **les difficultés d'implémentation des opérateurs nécessaires de projection.** Une hypothèse fondamentale de la formulation POCS est l'existence d'opérateurs de projection pour chaque ensemble de contraintes. Ces opérateurs devraient projeter un point sur le point le plus proche de l'ensemble de contraintes, et devraient être idempotents ( $\mathcal{P} \circ \mathcal{P} = \mathcal{P}$ ). Cependant, ce n'est pas toujours possible dans les situations pratiques, comme par exemple si on considère la contrainte convexe qu'un signal soit à bande limitée. Le cadre POCS nécessite un opérateur qui projette le signal sur l'ensemble convexe représentant les signaux à bande limitée. Un tel opérateur de projection est un filtre passe-bas idéal. Étant donné qu'un tel filtre n'est pas réalisable, l'opérateur de projection ne peut pas être idempotent. Sous ces conditions, la théorie POCS n'est pas applicable. Une analyse plus détaillée de cet aspect se trouve dans [Reeves 93].

On peut noter que la méthode d'ellipsoïdes englobants n'a pas de solution unique. Le problème est que la solution choisie, le centre de gravité de cet ellipsoïde, est très arbitraire. Il est improbable que le centre de gravité doive être favorisé parmi toutes les solutions possibles dans l'ellipsoïde englobant. Par contraste avec la formulation MAP, sous des conditions appropriées facilement satisfaites en pratique, la solution est unique et maximise la probabilité *a posteriori*.

#### 4.4.2.6 Méthodes hybrides

Des travaux combinant les approches de restauration par SR de type MV, MAP et POCS ont été réalisés. En particulier, les caractéristiques souhaitées de la formulation MAP et celles de la formulation POCS très flexible sont associées, créant une optimisation hybride.

Elad et Feuer dans [Elad 96, Elad 97], après avoir passé en revue plusieurs techniques existantes de SR, proposent une méthode hybride de type MV/POCS. Celle-ci utilise la formulation MV pour aborder la SR comme un problème d'inférence statistique, tout en utilisant des contraintes basées sur la projection pour réaliser la régularisation. Les auteurs revendiquent que leur algorithme de restauration hybride est une généralisation des méthodes existantes. Toutefois, on peut noter qu'une approche d'estimation statistique combinée avec des contraintes projetées a été présentée auparavant par Schultz et Steven-

son dans [Schultz 95a, Schultz 95b, Schultz 95c, Schultz 96]. Dans ces travaux les auteurs utilisent un algorithme de SR contraint de type MAP, dans lequel ils appliquent de façon alternative des contraintes basées sur les projections, et une méthode itérative de type MAP.

L'idée générale de ces méthodes hybrides est de maximiser la probabilité *a posteriori* ou la fonction de vraisemblance (respectivement dans les contextes MAP et MV), tout en s'assurant que la solution demeure dans un espace contraint formulé dans le cadre de la théorie des ensembles. Dans [Schultz 95a, Schultz 95b, Schultz 95c, Schultz 96], l'ensemble de contraintes assure que le sous-échantillonnage de l'image super-résolue correspond exactement à la trame de référence de la séquence d'images BR.

Les méthodes hybrides qui combinent l'inférence statistique et les contraintes projetées sont prometteuses car elles associent les caractéristiques les plus favorables des méthodes stochastiques (solution par estimation optimale, rigueur mathématique, ajout direct d'information *a priori*) et des approches basées sur la formulation POCS (mécanisme puissant pour l'inclusion de contraintes linéaires et non-linéaires *a priori* dans des ensembles théoriques).

#### 4.4.2.7 Méthodes de filtrage adaptatif et optimal

Plusieurs chercheurs ont proposé des approches basées sur le filtrage inverse pour résoudre le problème de SR. Les efforts dans ce domaine sont rapidement présentés. Ces techniques sont abordées afin de compléter cette étude, mais elles sont souvent sous-optimales en terme d'ajout de contraintes *a priori*.

Dans [Jacquemod 92], Jacquemod, Odet et Goutte proposent une approche simple de restauration par déconvolution, qui fait l'hypothèse d'un mouvement de translation sous-pixellique. Ainsi, un filtre de déconvolution approprié à la restauration d'images observées fusionnées est déterminé. Cette approche est peu adaptée à l'incorporation de modèles d'observation plus généraux, et elle est limitée en terme d'ajout de contraintes *a priori*.

Erdem, Sezan et Ozkan ont proposé une approche par filtrage LMMSE dans [Erdem 92]. Les auteurs utilisent un filtrage de Wiener multi-images, avec mouvement compensé, pour la restauration de séquences d'images dégradées par un flou spatial linéaire et invariant par translation, et un bruit additif. Ils supposent un modèle de mouvement de translation global, mais le flou lié au mouvement n'est pas incorporé. Bien que les modèles de mouvement et de dégradation sont limités, et que des contraintes *a priori* non-linéaires sont difficiles à incorporer dans leur formulation, l'intérêt de cette étude résulte de la présentation d'une restauration multi-images simultanée. Une formulation semblable est présentée par Srinivas et Srinath dans [Srinivas 90].

Les techniques basées sur le filtrage adaptatif, en particulier le filtrage de Kalman, ont également été appliquées en restauration par SR. Patti et Sezan dans [Patti 98], et Elad et Feuer dans [Elad 95, Elad 99] en sont des exemples. Dans [Patti 98], les auteurs proposent un filtrage par modèle de Kalman avec mouvement compensé, capable de restaurer par SR, même en présence de flous variant spatialement. Bien que la formulation du filtrage de Kalman proposée soit efficace d'un point de vue calculatoire, elle reste un estimateur LMMSE. Des contraintes de modèles non-linéaires d'images, permettant d'extrapoler la bande passante, ne peuvent pas facilement être incorporées dans ce cadre de travail. Les mêmes remarques s'appliquent au travail d'Elad et Feuer dans [Elad 95, Elad 99], réduit à une estimée LMMSE de l'image HR. Kim consacre un chapitre de son manuscrit de thèse [Kim 94] à une approche similaire par filtrage de Kalman, appliqué à la restauration par SR.

#### 4.4.2.8 Méthodes régularisées de Tikhonov

En raison de l'aspect mal posé du problème de restauration par SR, Hong, Kang et Katsaggelos ont proposé dans [Hong 97a, Hong 97b] d'utiliser la régularisation de Tikhonov [Tikhonov 77] pour la restauration par SR. Il s'agit d'une approche déterministe de régularisation, qui utilise des fonctions de régularisation pour imposer des contraintes de douceur sur l'espace des solutions possibles. La fonction de régularisation utilisée dans ces papiers est équivalente à un modèle *a priori* MRF gaussien dans la formulation MAP. Les *a priori* MRF gaussiens sont bien connus pour produire des restaurations trop lissées [Bouman 93]. En effet, les faibles performances du modèle *a priori* MRF gaussien sont la raison pour laquelle des *a priori* tels que le HMRF et le MRF généralisé gaussien [Bouman 93], qui imposent la douceur tout en permettant la restauration des contours, furent introduits. L'approche régularisée de Tikhonov est un cas particulier de la formulation plus générale bayésienne sous les hypothèses de bruit et *a priori* gaussiens. Le modèle d'observation utilisé dans [Hong 97a, Hong 97b] est semblable à la matrice de sous-échantillonnage avec mouvement compensé de [Schultz 96].

#### 4.4.2.9 Résumé sur les méthodes de SR dans le domaine spatial

La restauration par SR en utilisant un formalisme dans le domaine spatial comble de nombreux déficits des approches du domaine fréquentiel.

- Les méthodes du domaine spatial, en utilisant le modèle d'observation linéaire de la forme  $\mathbf{g} = \mathbf{H} \mathbf{f}$ , sont capables d'inclure un **champ quasi-illimité de modèles de mouvements**. Puisque la matrice  $\mathbf{H}$  modélise les pixels des images observées comme une combinaison linéaire de n'importe quelle configuration des pixels de l'image HR super-résolue  $\mathbf{f}$ , il y a une énorme flexibilité dans la formulation du modèle de mouvement. Il n'y a pas de limitation à des modèles globaux comme pour les approches du domaine fréquentiel. Il est aussi simple de considérer un modèle de mouvement local qu'un modèle global lorsqu'on utilise le formalisme du domaine spatial.
- Le système matriciel  $\mathbf{H}$  permet également **l'ajout de dégradations linéaires** telles que le flou de mouvement résultant d'un temps d'ouverture non nul (modélisé comme une intégration spatiale sur la trajectoire du mouvement), le flou variant ou invariant spatialement, des pixels manquants, etc. Il est très difficile, sinon impossible, d'inclure de telles dégradations dans le formalisme fréquentiel.
- Comme cela a déjà été dit à plusieurs reprises, l'ajout d'information *a priori* est nécessaire à la résolution de problèmes inverses mal posés tels que la restauration par SR. Des modèles *a priori* stochastiques, tout comme des ensembles de contraintes convexes permettent à des méthodes simples, mais également très puissantes, d'**incorporer des connaissances a priori dans le processus de restauration**. Introduire des contraintes dans le domaine spatial est intuitif et direct. En utilisant des modèles MRF définis en utilisant seulement des interactions spatiales locales, il est possible de fournir des contraintes de douceur, tout en préservant la possibilité de reconstruire des caractéristiques de contours francs (ce qui est impossible dans le domaine fréquentiel).
- La combinaison de données à partir de plusieurs images, de même que l'utilisation de contraintes *a priori* réalistes et souvent non-linéaires sur l'image à reconstruire, donnent aux méthodes du domaine spatial un **mécanisme puissant pour l'extrapolation de la bande passante**. Les modèles MRF qui utilisent des fonctions de pénalités non-linéaires sont particulièrement utiles car elles sont capables d'introduire de l'information fréquentielle nouvelle.

- Les méthodes stochastiques, en particulier les méthodes d'estimation MAP, sont dérivées d'un **formalisme mathématique très solide** dans lequel des développements théoriques supplémentaires peuvent être réalisés. La théorie qui s'applique au caractère optimal des estimateurs est directement applicable, de même que pour les limites sur les erreurs, etc. Il s'agit d'un cadre de travail abondant, ce qui n'est pas le cas des approches fréquentielles.

Toutefois, les méthodes de SR dans le domaine spatial ont un certain coût :

- Contrairement aux approches du domaine fréquentiel pour lesquelles la restauration était un processus relativement simple, les **optimisations intrinsèques** dans la plupart des méthodes du domaine spatial sont **plus complexes** que pour leurs homologues du domaine fréquentiel.
- La flexibilité augmentée des méthodes du domaine spatial tend à se faire au prix d'une **augmentation calculatoire importante**. Ceci est particulièrement vrai pour les méthodes qui utilisent des *a priori* non convexes, pour lesquels il est nécessaire de recourir à la méthode lente du recuit simulé, ou à d'autres approches permettant l'optimisation en présence d'extréma locaux.
- **Mécanismes de SR moins intuitifs**. Dans le formalisme fréquentiel, le mécanisme de restauration par SR était clair : suppression du recouvrement de spectre dans des images sous-échantillonnées et translattées. Pour les techniques du domaine spatial, les mécanismes sous-jacents d'amélioration de la résolution ne sont pas aussi transparents, en dépit du fait de la suppression du recouvrement de spectre et de l'extrapolation de la bande passante.

L'étude présentée montre que pour réaliser la restauration par SR de séquences d'images contenant un mouvement plus compliqué qu'un mouvement de translation global, les techniques du domaine spatial sont à préférer, et constituent souvent la seule approche viable.

Les avantages et inconvénients des techniques de restauration du domaine spatial ont été présentés en termes généraux. La partie suivante propose l'orientation la plus prometteuse.

### 4.4.3 Comparaisons et conclusions

Cette partie présente un résumé comparatif des principales familles de méthodes de restauration multi-images par SR. Un premier tableau récapitulatif des différences entre les méthodes du domaine fréquentiel et du domaine spatial est présenté, ainsi qu'un second tableau comparant les deux approches de restauration les plus prometteuses MAP et POCS. Des points importants concernant le lien entre l'estimation du mouvement et la restauration par SR sont également présentés, et des conclusions sont proposées.

#### 4.4.3.1 Comparaisons des techniques de restauration par SR

Un tableau comparatif des deux grandes familles de méthodes de SR, celles du domaine spatial et celles du domaine fréquentiel, est proposé dans TAB. 4.1. Il s'agit de généralisations, et des exceptions sont observables dans la vaste étendue des techniques existantes. Ce tableau comparatif met en valeur le fait que les méthodes de restauration du domaine spatial sont, en général, supérieures aux méthodes du domaine fréquentiel. Toutefois, cette supériorité apparaît au prix d'une augmentation de la complexité calculatoire et théorique.

Dans la classe des méthodes de SR du domaine spatial, deux techniques majeures s'avèrent être les plus prometteuses : l'approche bayésienne MAP, et la formulation POCS relative à la théorie des



	Formulation fréquentielle	Formulation spatiale
Modèle d'observation	dans le domaine fréquentiel	dans le domaine spatial
Modèle de mouvement	translation globale	presque illimité
Modèle de bruit	limité	linéaire, variant et invariant par translation
Mécanisme de SR	suppression du recouvrement de spectre	suppression du recouvrement de spectre et extrapolation de la bande passante avec des contraintes <i>a priori</i>
Simplicité	très simple	généralement complexe
Coût calculatoire	faible	élevé
Contraintes <i>a priori</i>	limitées	presque illimitées
Régularisation	limitée	excellente
Possibilités d'extension	faibles	riches
Performances	bonnes pour des applications spécifiques	bonnes

TAB. 4.1: Résumé et comparaisons des méthodes de SR dans les domaines fréquentiel et spatial.

ensembles. Ces deux approches sont comparées dans TAB. 4.2. Les avantages de ces deux formulations ont déjà été détaillés auparavant dans les parties 4.4.2.4 et 4.4.2.5. D'une certaine manière, ces deux approches sont complémentaires, et nous avons vu dans la partie 4.4.2.6 que les meilleures caractéristiques de chacune d'elles peuvent être combinées.

#### 4.4.3.2 Estimation du mouvement et restauration par SR

L'état de l'art présenté a montré que l'estimation du mouvement est une composante critique de toutes les méthodes de restauration multi-images par SR. L'information de mouvement, et l'estimation du mouvement dans le contexte de la SR, appellent plusieurs observations importantes :

- **Le mouvement relatif de résolution sous-pixellique entre les images observées** contribue à l'apport de nouvelles informations. Il peut également fournir des contraintes pour la résolution du problème inverse de restauration.
- **Les estimées de mouvement doivent être exactes et précises** à une résolution sous-pixellique. Ceci semble évident à la lecture du précédent point.
- **Les estimées de mouvement ne sont pas nécessairement denses** (c'est-à-dire définies pour tous les emplacements spatiaux) pour réaliser la restauration par SR. Il est plus important que ces estimées de mouvement soient précises plutôt que denses. De manière intuitive, les estimées de mouvement calculées dans des régions de valeurs presque constantes ont peu de chance d'être précises, et sont d'une utilité limitée pour les algorithmes de SR puisqu'une faible augmentation

	<b>Approche bayésienne MAP</b>	<b>Approche POCS</b>
Contexte théorique	riche	limité
Contraintes <i>a priori</i>	densité de probabilité <i>a priori</i> , contraintes projetées possibles	ensembles convexes, projeté sur des contraintes
Solution super-résolue	unique, estimée MAP	non-unique, volume d'intersection d'ensembles
Optimisation	itérative, méthodes de descente variées, bonne convergence	itérative, méthodes de projection, souvent lent à converger
Coût calculatoire	élevé	élevé
Complications	optimisation difficile pour des <i>a priori</i> non-convexes	opérateurs de projection potentiellement problématiques

TAB. 4.2: Comparaisons qualitatives de deux méthodes de SR : MAP et POCS.

de résolution est possible dans ces régions. Les estimées de mouvement sont les plus utiles dans les régions où elles peuvent être estimées de la manière la plus fiable, à savoir, dans les régions de grande variance spatiale.

- L'estimation du mouvement est aussi un problème mal posé. C'est pourquoi **les méthodes de régularisation peuvent être appliquées avec succès au problème d'estimation de mouvement**.
- **Utilisation de modèles de mouvement généraux.** Plusieurs modèles de mouvement globaux qui incluent la translation, la rotation, les modèles affines et projectifs, de même que différents modèles locaux ont été abordés. Les futurs travaux de SR pourront utiliser des estimations de mouvement basées sur des modèles ou des objets, en permettant la restauration d'objets partiellement cachés, transparents, ou en mouvement.

## 4.5 Les limites

Des travaux complémentaires ont été consacrés à l'étude des limites des méthodes de reconstruction par SR. Pour l'essentiel, il s'agit des travaux de Baker et Kanade dans [Baker 02], et ceux de Lin et Shum [Lin 01, Lin 04]. Les problèmes de modélisation du problème de SR sont également abordés dans [Wang 04].

Dans [Baker 02], les auteurs montrent que la reconstruction par SR devient fondamentalement plus difficile quand le facteur de SR  $k_0$  augmente. Ils commencent par modéliser l'expression qui relie les images BR à l'image HR. Ils font également les hypothèses suivantes pour parvenir à des conclusions :

la même FTO est utilisée pour l'acquisition de toutes les images BR, les capteurs photosensibles de la matrice CCD d'acquisition sont carrés et de taille identique, et la seule transformation inter-images considérée est la translation globale. Bien entendu, la qualité de l'image reconstruite est fonction de la précision avec laquelle on connaît le déplacement entre les images BR, et du nombre d'images utilisées pour cette reconstruction, mais les auteurs n'abordent pas ces aspects.

Finalement en se basant sur leur modèle, Baker et Kanade montrent que pour des valeurs élevées de  $k_0$ , il existe un volume considérable de solutions vérifiant les contraintes discrétisées de reconstruction par SR. Ainsi, les contraintes de reconstruction fournissent de moins en moins d'information utile lorsque  $k_0$  augmente. Ceci est essentiellement lié à la taille des capteurs, et à l'effet d'intégration de la lumière sur la surface photosensible non-nulle. Mais c'est cette physique des capteurs, en provoquant un effet de moyenne de la lumière reçue à sa surface, qui implique l'existence de limites dans la reconstruction par SR.

On peut signaler qu'un *a priori* de douceur, utilisé pour traiter l'ambiguïté du nombre de solutions vérifiant les contraintes de reconstruction, assure seulement que des solutions lissées soient choisies. Ce résultat est souvent préférable au fait que sans *a priori*, n'importe quelle solution peut être choisie, ce qui peut donner des résultats moins satisfaisants.

Le travail [Lin 04] de Lin et Shum s'inscrit dans la continuité de [Baker 02], et il se base sur le même modèle. Les auteurs reformulent la question de l'existence de limites fondamentales à la SR de la façon suivante : trouver la distribution des facteurs de SR (*magnification factor*)  $\mathcal{K}_0$  au delà de laquelle les images HR reconstruites devient significativement de la vérité terrain.  $\mathcal{K}_0$  est envisagé dans un cas réel, tandis qu'à ma connaissance, toutes les études précédentes ne s'intéressaient qu'au cas entier  $k_0$ . L'enjeu soulevé est celui du gain de temps dans les expérimentations.

Les auteurs donnent des limites à  $\mathcal{K}_0$ , pour des algorithmes classiques de SR tels que ceux basés sur un estimateur MAP, et avec une hypothèse de mouvement local de translation entre les images BR. Si le recalage et la suppression du bruit ne sont pas très bons, la limite pratique du facteur de SR est de  $\mathcal{K}_0 = 1,6$ . La valeur  $\mathcal{K}_0 = 2,5$  est vivement conseillée, et la valeur de la limite théorique est estimée à  $\mathcal{K}_0 = 5,7$  sous des conditions de simulation. D'autre part, des éléments concernant le nombre d'images suffisant à une bonne résolution sont donnés. Si la partie fractionnelle de  $\mathcal{K}_0$  est 0,5 alors le nombre suffisant d'image BR est  $M = 4\mathcal{K}^2$ . Et pour un facteur de SR entier,  $M = k_0^2$  est suffisant. L'obtention de ces résultats est expliquée dans [Lin 04], et des simulations en vérifient la validité.

Wang and Qi dans [Wang 04] abordent le problème de l'utilisation d'un bon modèle de SR. Ils comparent deux modèles basés sur (4.1), en se posant la question de l'ordre à privilégier dans le modèle direct : 1,  $\mathbf{M}_i\mathbf{B}$  ou 2,  $\mathbf{B}\mathbf{M}_i$ , avec  $\mathbf{M}_i$  les opérateurs de mouvement inter-images et  $\mathbf{B}$  l'opérateur représentant l'optique du capteur. Ils basent leur étude sur des opérateurs arbitraires de mouvement, et des systèmes optiques variant par translation. Ainsi les auteurs remarquent que lorsque le mouvement n'est pas connu de façon *a priori*, et qu'il doit donc être estimé à partir des images BR, utiliser le modèle 1 peut conduire à des erreurs systématiques, tandis que le modèle 2 est plus approprié, et qu'il conduit à de meilleures performances. Toutefois ils signalent que si le flou optique est invariant par translation spatio-temporelle, et que le mouvement dans l'image est de type translation pure, le choix de l'un des deux modèles par rapport à l'autre ne fait pas de différence.

# Chapitre 5

## Approche bayésienne en SR

### Sommaire

---

<b>5.1</b>	<b>Cadre général</b> . . . . .	<b>109</b>
<b>5.2</b>	<b>Modèle gaussien et lien avec la régularisation</b> . . . . .	<b>110</b>
<b>5.3</b>	<b>Modélisation markovienne</b> . . . . .	<b>111</b>
5.3.1	Modélisation d'une image par champ de Markov . . . . .	111
5.3.2	Modèle de Gauss–Markov . . . . .	115
5.3.3	Modèle markovien avec variable de contours . . . . .	116
5.3.4	Modèle markovien avec variable de régions . . . . .	117
<b>5.4</b>	<b>Mise en œuvre du calcul de la solution</b> . . . . .	<b>121</b>
5.4.1	L'échantillonneur de Gibbs . . . . .	121
5.4.2	Algorithmes proposés . . . . .	122
<b>5.5</b>	<b>Liens avec les méthodes existantes</b> . . . . .	<b>126</b>
<b>5.6</b>	<b>Méthodes d'évaluation</b> . . . . .	<b>127</b>
<b>5.7</b>	<b>Évaluation des performances</b> . . . . .	<b>128</b>
5.7.1	Effets du bruit . . . . .	129
5.7.2	Effets des erreurs dans les estimées du mouvement . . . . .	134
5.7.3	Importance du nombre d'images BR utilisées vs. répartition déplacements . . . . .	136
5.7.4	Résultats obtenus à partir de données réelles . . . . .	138
5.7.5	Bilan de l'évaluation et interface Matlab de simulation . . . . .	139

---

**Table des figures**


---

5.1	Les cliques associées à deux systèmes de voisinage en dimension 2. . . . .	112
5.2	Comparaison des deux modèles d'homogénéité de régions : constant et bilinéaire. . . . .	121
5.3	Comparaison d'images HR et BR dans un exemple synthétique. . . . .	129
5.4	Résultats de SR obtenus dans un cas faiblement bruité, RSB = 30 dB. . . . .	132
5.5	Résultats de SR obtenus dans un cas bruité, RSB = 15 dB. . . . .	132
5.6	Résultats de segmentation pour $K$ allant de 2 à 6 avec modèle constant, RSB = 30 dB. . . . .	133
5.7	Résultats de segmentation pour $K$ allant de 4 à 6 avec modèle constant, RSB = 15 dB. . . . .	133
5.8	Résultats de segmentation pour $K$ allant de 4 à 6 avec modèle bilinéaire, RSB = 30 dB. . . . .	134
5.9	Résultats de SR avec mauvaises données de recalage $\hat{\mathbf{s}}_0$ , RSB = 30 dB. . . . .	135
5.10	Résultats de SR en fonction du nombre des images utilisées, RSB = 30 dB. . . . .	137
5.11	Résultats de SR sur la séquence réelle « réveil » et agrandissement d'une région. . . . .	138
5.12	Interface Matlab de l'outil de simulation de SR. . . . .	140

---

**Liste des tableaux**


---

5.1	Évaluation des performances en milieu bruité. Modèle constant d'homogénéité. . . . .	130
5.2	Évaluation des performances en milieu bruité. Modèle bilinéaire d'homogénéité. . . . .	130
5.3	Évaluation des performances en fonction des données de recalage. . . . .	135
5.4	Évaluation des performances en fonction du nombre et de la disposition des images. . . . .	137

---

Ce chapitre présente une nouvelle approche de restauration d'image par SR à partir de plusieurs images. Cette méthode s'inscrit dans le contexte de l'estimation bayésienne, et dans la famille des techniques stochastiques (cf. 4.4.2.4). L'approche proposée considère que l'image HR à reconstruire est constituée de régions homogènes. La méthode permet d'obtenir à la fois une image HR représentant la scène observée, mais également un résultat de segmentation en correspondance avec l'image restaurée. Ce résultat de segmentation peut s'avérer très utile à un traitement postérieur d'extraction de caractéristiques géométriques dans l'image, ou de détection.

Le cadre de cette méthode a été présenté avec de nombreux détails dans [Humblot 05c]. La définition de l'homogénéité utilisée dans cette première publication suit un modèle constant. Une extension de cette définition selon un modèle bilinéaire a ensuite été proposée dans [Humblot 05b]. Ce chapitre reprend les principes généraux déjà présentés dans ces articles, en précisant certains aspects. La partie 5.7 propose également des éléments d'évaluation de ces méthodes liés à la qualité de l'information de recalage, à l'importance du bruit dans la séquence BR initiale, et à l'importance du nombre d'images utilisées pour la reconstruction.

Après un rappel du **cadre général** de SR dans la première partie de ce chapitre, le **modèle gaussien** et son lien avec la **régularisation** sont présentés dans la section 5.2.

Des éléments sur la **modélisation markovienne** d'une image, et sur la **description du modèle** de SR proposé, sont donnés dans la partie 5.3.

Le **calcul de la solution** recherchée est décrit dans la section 5.4.

Les **liens** entre la **méthode** de SR proposée et les **méthodes existantes** sont abordés dans 5.5.

Enfin, les **critères d'évaluations** sont décrits dans la partie 5.6, et une **évaluation des performances** de la méthode proposée ainsi qu'un **bilan** de cette évaluation sont présentés dans 5.7.

## 5.1 Cadre général

La formulation du problème direct, réciproque à celui de SR et utilisée pour aborder ce problème, a été détaillée dans la partie 4.2. Le contexte applicatif (cf. 1.1) a permis de limiter le modèle de mouvement entre les images (lié aux opérateurs  $\mathbf{M}_i$ ) à celui de translation globale. La méthode utilisée pour le recalage est l'extension à base de gradient, de la méthode fréquentielle sous-pixellique de corrélation de phase, présentée en détails et évaluée dans le chapitre 3.

En ce qui concerne l'opérateur de filtrage  $\mathbf{B}$  qui représente l'optique des capteurs de l'appareil d'acquisition, une FTO gaussienne invariante par translation a été choisie. La taille du masque utilisé est fonction du facteur  $k_0$  de SR, et de l'appareil employé pour l'acquisition de la séquence vidéo. Le bruit additif  $\epsilon_i$  est supposé être blanc, gaussien, centré, indépendant et identiquement distribué pour chaque image observée  $\mathbf{g}_i$ .

## 5.2 Modèle gaussien et lien avec la régularisation

Le contexte bayésien de l'approche proposée utilise donc un formalisme probabiliste. Il consiste à formuler des dépendances entre les données observées, les données inconnues à estimer, et les paramètres ou hyper-paramètres du problème, sous la forme de probabilités.

Ainsi, les images observées  $\mathbf{g}_i$  sont modélisées par la fonction de vraisemblance  $p(\mathbf{g}_i | \mathbf{f}, \boldsymbol{\theta}_\epsilon)$ , c'est-à-dire la probabilité des images observées connaissant l'image HR à estimer  $\mathbf{f}$ , et le modèle d'observation (4.1). L'hypothèse d'indépendance des différents bruits additifs  $\boldsymbol{\epsilon}_i$  permet d'écrire :

$$p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta}_\epsilon) = \prod_{i=1}^M p(\mathbf{g}_i | \mathbf{f}, \boldsymbol{\Sigma}_{\epsilon_i}) = \prod_{i=1}^M p_{\boldsymbol{\epsilon}_i}(\mathbf{g}_i - \mathbf{H}_i \mathbf{f}).$$

Les hypothèses qualifiant le bruit de blanc, gaussien et centré consistent à définir :

$$p_{\boldsymbol{\epsilon}_i}(\boldsymbol{\epsilon}_i) = \mathcal{N}(\boldsymbol{\epsilon}_i | \mathbf{0}, \boldsymbol{\Sigma}_{\epsilon_i}).$$

On note  $E[.]$  l'espérance, ou la moyenne, d'une variable aléatoire.

$E[\boldsymbol{\epsilon}_i] = \mathbf{0}$  représente l'absence d'erreur systématique (bruit centré),

et  $E[\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^t] = \boldsymbol{\Sigma}_{\epsilon_i} = \sigma_{\epsilon_i}^2 \mathbf{I}$  avec  $\mathbf{I}$  représentant la matrice identité, signifie que le bruit est non corrélé (ou blanc).  $\boldsymbol{\Sigma}_{\epsilon_i}$  est donc une matrice diagonale [Mohammad-Djafari 01].

Ceci permet de définir le vecteur de paramètres contenant la variance du bruit de chaque image BR :  $\boldsymbol{\theta}_\epsilon = \{\sigma_{\epsilon_i}^2, i = 1, \dots, M\}$ , et ainsi :

$$p(\mathbf{g}_i | \mathbf{f}, \boldsymbol{\Sigma}_{\epsilon_i}) = \mathcal{N}(\mathbf{g}_i | \mathbf{H}_i \mathbf{f}, \sigma_{\epsilon_i}^2 \mathbf{I}) = \left( \frac{1}{2\pi \sigma_{\epsilon_i}^2} \right)^{\frac{R}{2}} \exp \left( -\frac{1}{2\sigma_{\epsilon_i}^2} \|\mathbf{g}_i - \mathbf{H}_i \mathbf{f}\|^2 \right) \quad (5.1)$$

où  $R$  représente le nombre de pixels dans une image BR  $\mathbf{g}_i$  de taille  $H \times L$ , et où  $\mathcal{N}(\mathbf{g}_i | \mathbf{H}_i \mathbf{f}, \sigma_{\epsilon_i}^2 \mathbf{I})$  signifie que la variable  $\mathbf{g}_i$  suit une loi gaussienne de moyenne  $\mathbf{H}_i \mathbf{f}$  et de variance  $\sigma_{\epsilon_i}^2 \mathbf{I}$  dont l'expression exacte est donnée par l'équation (5.1).

La règle de Bayes permet d'écrire :

$$p(\mathbf{f} | \mathbf{g}_i, \boldsymbol{\Sigma}_{\epsilon_i}) = \frac{p(\mathbf{g}_i | \mathbf{f}, \boldsymbol{\Sigma}_{\epsilon_i}) p(\mathbf{f} | \boldsymbol{\Sigma}_{\epsilon_i})}{p(\mathbf{g}_i | \boldsymbol{\Sigma}_{\epsilon_i})}.$$

Si aucune hypothèse particulière n'est faite sur la solution recherchée  $\mathbf{f}$ , cela revient à utiliser un *a priori* uniforme, soit  $p(\mathbf{f} | \boldsymbol{\Sigma}_{\epsilon_i}) = p(\mathbf{f}) = \text{constante}$ . Comme cela a été vu dans l'état de l'art (cf. 4.4), un estimateur couramment utilisé dans les méthodes stochastiques est celui du MAP. Il consiste à évaluer la valeur  $\hat{\mathbf{f}}_{\text{MAP}}$  qui maximise la probabilité *a posteriori*  $p(\mathbf{f} | \mathbf{g}_i, \boldsymbol{\Sigma}_{\epsilon_i})$ . Du point de vue de la maximisation de cette probabilité selon la variable  $\mathbf{f}$ , le terme d'évidence  $p(\mathbf{g}_i | \boldsymbol{\Sigma}_{\epsilon_i})$  est une constante dont on peut s'affranchir. D'autre part, dans le cas de lois gaussiennes, on préfère parfois utiliser le logarithme népérien, fonction monotone croissante.

On passe alors de la maximisation d'une probabilité, à la minimisation d'une fonction :

$$\hat{\mathbf{f}}_{\text{MAP}} = \arg \max_{\mathbf{f}} [p(\mathbf{f} | \mathbf{g}_i, \boldsymbol{\Sigma}_{\epsilon_i})] = \arg \min_{\mathbf{f}} [-\ln(p(\mathbf{f} | \mathbf{g}_i, \boldsymbol{\Sigma}_{\epsilon_i}))] = \arg \min_{\mathbf{f}} [\mathbf{J}(\mathbf{f})]. \quad (5.2)$$

Ainsi, dans le cas particulier exposé, calculer  $\hat{\mathbf{f}}_{\text{MAP}}$  revient à trouver l'image  $\mathbf{f}$  qui minimise la fonction :

$$\mathbf{J}(\mathbf{f} | \mathbf{g}, \boldsymbol{\theta}_\epsilon) = \sum_{i=1}^M \frac{1}{\sigma_{\epsilon_i}^2} \|\mathbf{g}_i - \mathbf{H}_i \mathbf{f}\|^2 + R \ln(\sigma_{\epsilon_i}).$$

On considère maintenant le cas où l'on connaît une information *a priori* sur la solution  $\mathbf{f}$  : sa moyenne  $E[\mathbf{f}] = \mathbf{f}_0$ , et sa matrice de covariance  $E[(\mathbf{f} - \mathbf{f}_0)(\mathbf{f} - \mathbf{f}_0)^t] = \boldsymbol{\Sigma}_f = \sigma_f^2 \mathbf{P}_0$ .

Dans ce cas il est possible de faire l'hypothèse que  $p(\mathbf{f}) = \mathcal{N}(\mathbf{f}_0, \boldsymbol{\Sigma}_f)$  :

$$p(\mathbf{f}) = \left( \frac{1}{2\pi \sigma_f^2} \right)^{\frac{R'}{2}} |\mathbf{P}_0|^{-\frac{1}{2}} \exp \left( -\frac{1}{2\sigma_f^2} (\mathbf{f} - \mathbf{f}_0)^t \mathbf{P}_0^{-1} (\mathbf{f} - \mathbf{f}_0) \right)$$

où  $R'$  représente le nombre de pixels dans l'image  $\mathbf{f}$ . Si on suppose que l'inverse de la matrice de covariance  $\mathbf{P}_0^{-1}$  existe et peut s'écrire  $\mathbf{P}_0^{-1} = \mathbf{D}^t \mathbf{D}$ , alors avec cet *a priori* sur la solution, calculer  $\hat{\mathbf{f}}_{\text{MAP}}$  revient à trouver l'image  $\mathbf{f}$  qui minimise la fonction :

$$\mathbf{J}(\mathbf{f} | \mathbf{g}, \boldsymbol{\theta}_\epsilon) = \sum_{i=1}^M \frac{1}{\sigma_{\epsilon_i}^2} \|\mathbf{g}_i - \mathbf{H}_i \mathbf{f}\|^2 + \frac{1}{\sigma_f^2} \|\mathbf{D}(\mathbf{f} - \mathbf{f}_0)\|^2. \quad (5.3)$$

On retrouve les notions de solution régularisée et de régularisation quadratique.

## 5.3 Modélisation markovienne

### 5.3.1 Modélisation d'une image par champ de Markov

Les notions exposées dans cette partie ont été inspirées de [Tupin 02].

L'information contenue dans une image est plus importante que la simple donnée des niveaux de gris en chaque pixel (ou site), et la description se fait en termes de régions, contours, structures définis par les contrastes, texture, etc. qui peuvent être présents dans l'image. Le niveau de gris en un site n'est donc souvent pas significatif en lui-même, mais dans ses relations et interactions avec les pixels voisins.

Cette propriété des images que constituent les interactions locales entre niveaux de gris voisins pour définir les différentes régions de l'image, permet d'utiliser un formalisme markovien dans de nombreux traitements tels que la **restauration** ou la **segmentation**. Le principe consiste à définir des énergies locales entre groupes de sites reflétant les interactions entre niveaux de gris. L'énergie globale est alors reliée à la probabilité d'apparition de l'image dans le cadre des champs de Gibbs.

Ainsi, une autre manière de modéliser une image consiste à dire qu'elle est formée d'un ensemble fini  $\mathcal{R}$  de sites  $\mathbf{r}$  correspondant aux pixels :  $\mathcal{R} = \{\mathbf{r} = (x, y); 0 \leq x \leq H - 1, 0 \leq y \leq L - 1\}$ .  $\mathcal{R}$  est donc un réseau discret fini, partie de  $\mathbb{N}^2$ .

À chaque site est associé un descripteur, représentant l'état du site et qui peut être son niveau de gris, une étiquette ou une information plus complexe, et prenant ses valeurs dans un ensemble  $E$ .



Par exemple  $E = \{0, \dots, N - 1\}$  représente les valeurs de niveaux de gris d'une image monochrome, et  $E = \{0, \dots, N - 1\}^q$  représente les valeurs susceptibles d'être associées à un pixel d'une image multi-spectrale à  $q$  bandes.

La notion d'interactions locales nécessite de structurer les relations spatiales entre les différents sites du réseau. Pour ce faire, on munit  $\mathcal{R}$  d'un système de voisinage  $\mathcal{V}$  vérifiant les propriétés suivantes :

$$\mathcal{V}(\mathbf{r}) = \{\mathbf{r}'\} \quad \text{tels que} \quad \begin{cases} \mathbf{r} \notin \mathcal{V}(\mathbf{r}) \\ \mathbf{r}' \in \mathcal{V}(\mathbf{r}) \Leftrightarrow \mathbf{r} \in \mathcal{V}(\mathbf{r}') \end{cases}$$

L'ensemble  $\mathcal{V}(\mathbf{r})$  est appelé voisinage de  $\mathbf{r}$ , et  $\mathbf{r}'$  est dit voisin de  $\mathbf{r}$  si  $\mathbf{r}' \in \mathcal{V}(\mathbf{r})$ . La relation de voisinage est symétrique.

Un sous-ensemble  $\mathbf{c}$  de  $\mathcal{R}$  est appelé « clique relative au système de voisinage  $\mathcal{V}$  » si  $\mathbf{c}$  est un singleton de  $\mathcal{R}$ , ou un ensemble de sites tous voisins les uns des autres (au sens de  $\mathcal{V}$ ).

En fonction du système de voisinage utilisé, le système de cliques sera différent et fera intervenir un nombre plus ou moins important de sites, comme illustré sur FIG. 5.1.

On note  $\mathcal{C}$  l'ensemble des cliques de  $\mathcal{R}$  relatif à  $\mathcal{V}$ . Quand cela est nécessaire, on précise l'ordre en notant  $\mathbf{c}_k$  une clique d'ordre  $k$ .

$\mathcal{C}_k$  est alors l'ensemble des cliques d'ordre  $k$  sur  $\mathcal{R}$ .

Une clique d'ordre 2 peut également se noter :  $\mathbf{c}_2 = (\mathbf{r}, \mathbf{r}')$ .

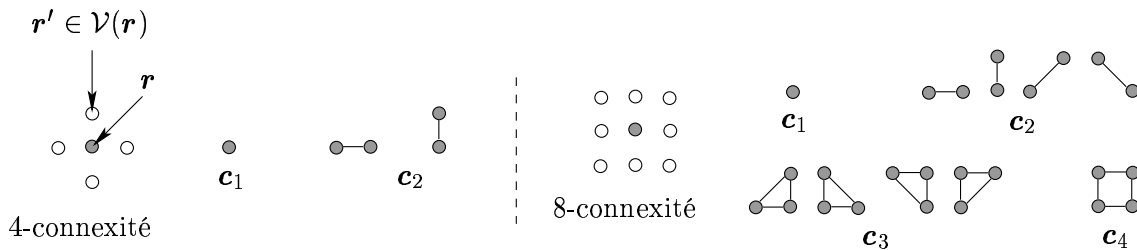


FIG. 5.1: Les cliques associées à deux systèmes de voisinage en dimension 2.

Deux systèmes de voisinage et leurs cliques associées sont représentés sur la FIG. 5.1.

Le voisinage d'ordre 1 d'un pixel contient ses voisins en « 4-connextité », et celui d'ordre 2 contient ses voisins en « 8-connextité ».

Les interactions locales entre niveaux de gris (ou descripteurs) de sites voisins peuvent alors s'exprimer comme un potentiel de cliques.

Soit  $\mathbf{c}$  une clique, on lui associe le potentiel  $U_{\mathbf{c}} = \phi(\mathbf{c})$  avec  $\phi(\cdot)$  une fonction de clique positive (car  $U$  est une énergie qui doit être positive ou nulle), dont la valeur dépend des niveaux de gris des pixels constituant la clique. De la même manière, on peut définir l'énergie globale de l'image comme la somme des potentiels de toutes les cliques :

$$U = \sum_{\mathbf{c} \in \mathcal{C}} U_{\mathbf{c}}$$

et l'énergie locale en un site comme la valeur des potentiels de toutes les cliques auxquelles il appartient :

$$U_{\mathbf{r}} = \sum_{\mathbf{c} \in \mathcal{C} / \mathbf{r} \in \mathbf{c}} U_{\mathbf{c}}$$

On utilise une modélisation probabiliste de l'image pour définir les champs de Markov. Ainsi, on considère maintenant l'image comme une réalisation d'un champ aléatoire.

Soit  $\mathbf{r}$  un site de l'image, on peut lui associer une variable aléatoire (v.a.)  $F_{\mathbf{r}}$  prenant ses valeurs dans  $E$ . Le niveau de gris  $f_{\mathbf{r}}$  en  $\mathbf{r}$  est alors une réalisation de la v.a.  $F_{\mathbf{r}}$ .

On définit ensuite  $\mathbf{F} = (F_{\mathbf{r}}, F_{\mathbf{r}'}, \dots)$  prenant ses valeurs dans  $\Omega = E^{|\mathcal{R}|}$ , avec  $|\mathcal{R}|$  le cardinal de  $\mathcal{R}$ . Généralement, on note les v.a. en lettres majuscules, et leurs réalisations en minuscules.

Dans ce cadre probabiliste, l'image considérée est simplement une réalisation  $\mathbf{f}$  du champ. D'une certaine manière la loi de probabilité jointe de  $\mathbf{f}$  notée  $p(\mathbf{F} = \mathbf{f})$ , ou  $p(\mathbf{f})$  par abus de langage, et les probabilités conditionnelles locales  $p(f_{\mathbf{r}} | f_{\mathbf{r}'}, \mathbf{r}' \in \mathcal{V}(\mathbf{r}))$  d'une valeur en un site conditionnellement aux valeurs de ses voisins, permettent de mesurer le lien statistique existant entre un niveau de gris et le reste de l'image. L'hypothèse markovienne permet d'évaluer ces quantités. On se place dans le cas où  $E$ , l'espace des valeurs des descripteurs, est quantifié, ce qui permet de manipuler des probabilités.

Soit  $f_{\mathbf{r}}$  la valeur prise au site  $\mathbf{r}$  du descripteur, et  $\mathbf{f}^{\mathbf{r}} = (f_{\mathbf{r}'})_{\mathbf{r}' \neq \mathbf{r}}$  la configuration de l'image, excepté le site  $\mathbf{r}$ . On dit que  $\mathbf{F}$  est un champ de Markov si et seulement si la probabilité conditionnelle locale en un site n'est fonction que de la configuration du voisinage du site considéré.

D'une manière mathématique, ceci s'écrit :

$$p(F_{\mathbf{r}} = f_{\mathbf{r}} | \mathbf{f}^{\mathbf{r}}) = p(F_{\mathbf{r}} = f_{\mathbf{r}} | f_{\mathbf{r}'}, \mathbf{r}' \in \mathcal{V}(\mathbf{r}))$$

Ainsi, le niveau de gris en un site ne dépend que des niveaux de gris des pixels voisins de ce site. Cette hypothèse markovienne se justifie bien dans le cas de la plupart des images naturelles constituées de zones homogènes ou texturées, et pour un grand nombre d'images de synthèse.

On définit la mesure de Gibbs de fonction d'énergie  $U : \Omega \rightarrow \mathbb{R}$  comme la probabilité  $p$  définie sur  $\Omega$  par :

$$p(\mathbf{F} = \mathbf{f}) = \frac{1}{Z} \exp(-U(\mathbf{f})) \quad \text{avec} \quad U(\mathbf{f}) = \sum_{\mathbf{c} \in \mathcal{C}} U_{\mathbf{c}}(\mathbf{f})$$

où  $\mathcal{C}$  est le système de cliques associé au système de voisinage  $\mathcal{V}$  de  $U$ .

$Z = \sum_{\mathbf{f} \in \Omega} \exp(-U(\mathbf{f}))$  est une constante de normalisation appelée fonction de partition de Gibbs.

En pratique, il est presque impossible de calculer cette constante à cause du nombre considérable de configurations possibles.

Par exemple, dans le cas d'une image binaire ( $\text{Card}(E) = 2$ ) de taille  $256 \times 256$  pixels, on a  $\text{Card}(\Omega) = 2^{65\,536} \approx 2 \times 10^{19\,728}$  configurations possibles !

La notation couramment utilisée pour  $U_{\mathbf{c}}(\mathbf{f})$  est abusive car elle ne dépend pas de l'ensemble de la configuration  $\mathbf{f}$ , mais seulement de  $\mathbf{f}$  restreinte à la clique  $\mathbf{c}$  :  $U_{\mathbf{c}}(\mathbf{f}) = U_{\mathbf{c}}(f_{\mathbf{r}}, \mathbf{r} \in \mathbf{c})$ .

Le champ de Gibbs de potentiel associé au système de voisinage  $\mathcal{V}$  se définit comme le champ aléatoire  $\mathbf{F}$  dont la probabilité est une mesure de Gibbs associée au système de voisinage  $\mathcal{V}$ , ce qui implique donc :

$$p(\mathbf{F} = \mathbf{f}) = \frac{1}{Z} \exp(-U(\mathbf{f})) = \frac{1}{Z} \exp\left(-\sum_{\mathbf{c} \in \mathcal{C}} U_{\mathbf{c}}(\mathbf{f})\right) \quad (5.4)$$

Ainsi, l'énergie globale d'un champ de Gibbs possède la propriété de se décomposer sous forme d'une somme d'énergies locales. On observe que plus une configuration d'un champ de Gibbs a une énergie faible, plus elle est probable.

Le théorème de Hammersley-Clifford [Besag 74] établit alors le résultat fondamental suivant sous les hypothèses :

- $\mathcal{R}$  est fini ou dénombrable,
- le système de voisinage  $\mathcal{V}$  est borné,
- l'espace des états  $E$  est discret,

$\mathbf{F}$  est un champ de Markov relativement à  $\mathcal{V}$ , et  $\forall \mathbf{f} \in \Omega$ ,  $p(\mathbf{F} = \mathbf{f}) > 0$  si et seulement si  $\mathbf{F}$  est un champ de Gibbs de potentiel associé à  $\mathcal{V}$ .

Par exemple, si on considère un champ de Markov de voisinage 4-connexe, l'énergie de la configuration  $\mathbf{f}$  peut s'écrire sous la forme :

$$U(\mathbf{f}) = \sum_{\mathbf{c} = (\mathbf{r}) \in \mathcal{C}_1} U_{\mathbf{c}}(f_{\mathbf{r}}) + \sum_{\mathbf{c} = (\mathbf{r}, \mathbf{r}') \in \mathcal{C}_2} U_{\mathbf{c}}(f_{\mathbf{r}}, f_{\mathbf{r}'})$$

On remarque que ce théorème n'est valable que lorsqu'aucune configuration n'est interdite (condition  $\forall \mathbf{f}$ ,  $p(\mathbf{F} = \mathbf{f}) > 0$ ). Toutefois, des solutions ont été proposées lorsqu'il est souhaitable de supprimer certaines configurations irréalistes.

On peut noter que rien n'interdit la non-stationnarité du champ, c'est-à-dire la variation des potentiels  $U_{\mathbf{c}}$  en fonction de la localisation de la clique  $\mathbf{c}$  dans l'image. Rien n'impose non plus la symétrie des potentiels, et on peut donc avoir  $U_{\mathbf{c} = (\mathbf{r}, \mathbf{r}')}(\mathbf{0}, \mathbf{1}) \neq U_{\mathbf{c}}(\mathbf{1}, \mathbf{0})$ .

Le théorème de Hammersley-Clifford, et la forme bien spécifique de probabilité de  $\mathbf{F}$  qui en résulte, permet de lier les probabilités globales et locales. Ainsi, la probabilité conditionnelle locale  $p(f_{\mathbf{r}} | \mathbf{F}^{\mathbf{r}} = \mathbf{f}^{\mathbf{r}})$  peut s'écrire :

$$p(F_{\mathbf{r}} = f_{\mathbf{r}} | \mathbf{F}^{\mathbf{r}} = \mathbf{f}^{\mathbf{r}}) = \frac{p(\mathbf{F} = \mathbf{f})}{p(\mathbf{F}^{\mathbf{r}} = \mathbf{f}^{\mathbf{r}})} = \frac{\exp(-U(f_{\mathbf{r}}, \mathbf{f}^{\mathbf{r}}))}{\sum_{n \in E} \exp(-U(n, \mathbf{f}^{\mathbf{r}}))}$$

On définit l'énergie locale  $U_{\mathbf{r}}$  au site  $\mathbf{r}$  par :

$$U_{\mathbf{r}}(f_{\mathbf{r}} | f_{\mathbf{r}'}, \mathbf{r}' \in \mathcal{V}(\mathbf{r})) = \sum_{\mathbf{c} \in \mathcal{C} | \mathbf{r} \in \mathbf{c}} U_{\mathbf{c}}(f_{\mathbf{r}}, f_{\mathbf{r}'}, \mathbf{r}' \in \mathcal{V}(\mathbf{r})) = \sum_{\mathbf{c} \in \mathcal{C} | \mathbf{r} \in \mathbf{c}} U_{\mathbf{c}}(f_{\mathbf{r}}, V_{\mathbf{r}})$$

en posant  $V_{\mathbf{r}} = (f_{\mathbf{r}'}, \mathbf{r}' \in \mathcal{V}(\mathbf{r}))$ . Cette énergie locale ne fait donc intervenir que les voisins de  $\mathbf{r}$ . On peut alors écrire l'énergie globale  $U(\mathbf{f})$  sous la forme :

$$U(\mathbf{f}) = \sum_{\mathbf{c} \in \mathcal{C} | \mathbf{r} \notin \mathbf{c}} U_{\mathbf{c}}(\mathbf{f}) + \sum_{\mathbf{c} \in \mathcal{C} | \mathbf{r} \in \mathbf{c}} U_{\mathbf{c}}(\mathbf{f}) = \sum_{\mathbf{c} \in \mathcal{C} | \mathbf{r} \notin \mathbf{c}} U_{\mathbf{c}}(\mathbf{f}) + U_{\mathbf{r}}(f_{\mathbf{r}} | V_{\mathbf{r}})$$

En simplifiant l'expression de la probabilité conditionnelle locale, en supprimant les termes communs qui font intervenir les cliques ne contenant pas le site  $\mathbf{r}$  au numérateur et au dénominateur, on obtient :

$$\begin{aligned} p(F_{\mathbf{r}} = f_{\mathbf{r}} | \mathbf{F}^{\mathbf{r}} = \mathbf{f}^{\mathbf{r}}) &= \frac{\exp\left(-\sum_{\mathbf{c} \in \mathcal{C} | \mathbf{r} \notin \mathbf{c}} U_{\mathbf{c}}(\mathbf{f}) - U_{\mathbf{r}}(f_{\mathbf{r}} | V_{\mathbf{r}})\right)}{\sum_{n \in E} \exp\left(-\sum_{\mathbf{c} \in \mathcal{C} | \mathbf{r} \notin \mathbf{c}} U_{\mathbf{c}}(\mathbf{f}) - U_{\mathbf{r}}(n | V_{\mathbf{r}})\right)} \\ &= \frac{\exp(-U_{\mathbf{r}}(f_{\mathbf{r}} | V_{\mathbf{r}}))}{\sum_{n \in E} \exp(-U_{\mathbf{r}}(n | V_{\mathbf{r}}))} \end{aligned}$$

Cette expression ne fait intervenir que les potentiels des cliques contenant le site  $\mathbf{r}$  (ce qui, au passage, permet de retrouver l'hypothèse markovienne). Elle est très importante; en effet, s'il est impossible d'accéder à la probabilité d'une configuration  $\mathbf{f}$  à cause de la constante de normalisation, il est également possible de calculer en chaque site sa probabilité conditionnelle locale. Cette expression est à la base de tous les algorithmes de simulation de champs markoviens.

Pour résumer ce qui vient d'être décrit, la définition d'un champs de Markov passe par la définition de sa **fonction d'énergie**  $U$ . Celle-ci nécessite la définition d'un **système de voisinage**  $\mathcal{V}$ , qui définit alors le **système de cliques**, et de **fonctions de potentiel** associées aux cliques. Ces fonctions de potentiel permettent d'accéder à la **probabilité globale d'une configuration**, et aux **probabilités conditionnelles locales**.

### 5.3.2 Modèle de Gauss–Markov

Lorsqu'on modélise une image à l'aide d'un champ de Markov, en considérant par exemple un modèle du premier ordre limité aux quatre voisins les plus proches, utiliser une fonction d'énergie  $U(\mathbf{f})$  quadratique et homogène dans l'équation (5.4) permet d'obtenir un modèle de type Gauss–Markov. Comme cela a été vu en détail dans la partie introductive précédente, ceci revient à définir la vraisemblance de l'image  $\mathbf{f}$  sous la forme :

$$p(\mathbf{f}) \propto \exp\left(\sum_{\mathbf{r} \in \mathcal{R}} \left(f(\mathbf{r}) - \frac{1}{4} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} f(\mathbf{r}')\right)^2\right).$$

Le nom de Gauss vient de l'expression obtenue qui rappelle la forme d'une fonction gaussienne. En effet, cette formulation exprime que chaque valeur de pixel  $\mathbf{r}$  de l'image  $\mathbf{f}$  suit une loi gaussienne de

moyenne égale à la moyenne des valeurs des pixels voisins, et de variance égale à 1 :

$$\mathcal{N}\left(f(\mathbf{r}) \mid \frac{1}{4} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} f(\mathbf{r}'), 1\right).$$

Utiliser cette seule hypothèse *a priori* et choisir un estimateur MAP pour calculer  $\hat{\mathbf{f}}_{MAP}$  comme présenté dans la partie 5.2, revient à chercher l'image  $\mathbf{f}$  qui minimise la fonction :

$$\mathbf{J}(\mathbf{f}) = \sum_{\mathbf{r} \in \mathcal{R}'} \left( f(\mathbf{r}) - \frac{1}{4} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} f(\mathbf{r}') \right)^2.$$

Dans le cadre du problème de SR, avec l'hypothèse d'un bruit additif blanc, gaussien et centré sur les observations,  $\hat{\mathbf{f}}_{MAP}$  est obtenue en minimisant la fonction :

$$\mathbf{J}(\mathbf{f} \mid \mathbf{g}, \boldsymbol{\theta}_\epsilon) = \sum_{i=1}^M \sum_{\mathbf{r} \in \mathcal{R}} \frac{1}{\sigma_{\epsilon_i}^2} \left( g_i(\mathbf{r}) - [\mathcal{H}_i \mathbf{f}](\mathbf{r}) \right)^2 + \sum_{\mathbf{r} \in \mathcal{R}'} \left( f(\mathbf{r}) - \frac{1}{4} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} f(\mathbf{r}') \right)^2.$$

Cette expression peut être réécrite de manière plus compacte :

$$\mathbf{J}(\mathbf{f} \mid \mathbf{g}, \boldsymbol{\theta}_\epsilon) = \sum_{i=1}^M \frac{1}{\sigma_{\epsilon_i}^2} \|\mathbf{g}_i - \mathbf{H}_i \mathbf{f}\|^2 + \|\mathbf{D} \mathbf{f}\|^2. \quad (5.5)$$

Le choix d'un *a priori* de Gauss–Markov est souvent qualifié d'*a priori de douceur*. Il a pour effet d'**adoucir** les contours des objets dans l'image reconstruite en affaiblissant leur intensité. Ceci est dû à l'aspect « moyenne des valeurs des pixels voisins » de ce modèle.

$\mathbf{D}$  est une matrice dont l'action sur  $\mathbf{f}$  est équivalente à une opération de convolution par un noyau de la forme :

$$\begin{pmatrix} 0 & -\frac{1}{4} & 0 \\ -\frac{1}{4} & 1 & -\frac{1}{4} \\ 0 & -\frac{1}{4} & 0 \end{pmatrix}.$$

### 5.3.3 Modèle markovien avec variable de contours

Une autre modélisation markovienne consiste à utiliser une image binaire de contours  $\mathbf{q}$  qualifiée de « variable cachée ». Le terme de variable cachée fait référence au fait qu'elle n'apparaît pas dans l'équation du modèle d'observation (4.1). D'un point de vue pratique et conceptuel, la manipulation de variables cachées fournit une grande variété d'outils pour prendre en considération des connaissances *a priori*. L'image de contours peut être connue par avance, ou estimée de manière jointe à celle de l'image que l'on cherche à reconstruire.

Elle est définie par  $q(\mathbf{r}) = 1$  si le pixel  $\mathbf{r}$  est évalué comme un contour, et  $q(\mathbf{r}) = 0$  dans le cas contraire.

L'estimation de l'image  $\mathbf{f}$  se fait ici encore par le biais d'un modèle de dépendance inter-pixels par champ de Markov du premier ordre. Ce modèle, accompagné d'un choix de fonction d'énergie quadratique, revient à écrire la probabilité conditionnelle de  $\mathbf{f}$  connaissant  $\mathbf{q}$  de la manière suivante :

$$p(\mathbf{f} | \mathbf{q}) \propto \exp \left( \sum_{\mathbf{r} \in \mathcal{R}} \left( f(\mathbf{r}) - \frac{1}{4} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} (1 - q(\mathbf{r})) f(\mathbf{r}') \right)^2 \right).$$

De même que dans le modèle de Gauss–Markov, cette formulation exprime que chaque pixel  $f(\mathbf{r})$  suit une loi gaussienne de la forme :

$$\mathcal{N} \left( f(\mathbf{r}) \mid \frac{1}{4} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} (1 - q(\mathbf{r})) f(\mathbf{r}'), 1 \right).$$

Cette modélisation *a priori* et le choix de l'estimateur MAP reviennent à calculer la solution  $\hat{\mathbf{f}}_{MAP}$  qui minimise la fonction :

$$\mathcal{J}(\mathbf{f} | \mathbf{q}) = \sum_{\mathbf{r} \in \mathcal{R}} \left( f(\mathbf{r}) - \frac{1}{4} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} (1 - q(\mathbf{r})) f(\mathbf{r}') \right)^2.$$

Parfois, une manière légèrement différente d'écrire cette fonction à minimiser est préférée :

$$\mathcal{J}(\mathbf{f} | \mathbf{q}) = \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} (1 - q(\mathbf{r})) \left( f(\mathbf{r}) - \frac{1}{4} f(\mathbf{r}') \right)^2.$$

Ce cas n'a pas fait l'objet d'une étude plus approfondie.

### 5.3.4 Modèle markovien avec variable de régions

La nouvelle méthode de SR présentée a été inspirée et adaptée d'un processus de fusion avec segmentation jointe, objet de l'article [Féron 05]. Le cadre bayésien d'estimation utilisé a été exposé en détail dans [Snoussi 04].

Cette nouvelle approche de restauration d'image par SR est basée sur l'idée que les images BR ont en commun des régions homogènes. Ainsi, on cherche à estimer conjointement une image HR  $\mathbf{f}$  à valeur réelle, et une image de segmentation à valeurs discrètes représentée par la variable cachée  $\mathbf{z} = \{z(\mathbf{r}), \mathbf{r} \in \mathcal{R}\} \in \{1, \dots, K\}^{k_0 \times R}$  où  $K$  est le nombre de régions homogènes différentes. Pour rappel,  $R$  est le nombre de pixel dans une image BR, et  $k_0$  le facteur de SR. Par le choix de cette modélisation, le problème devient l'estimation de l'ensemble de variables  $(\mathbf{f}, \mathbf{z})$ .

Un modèle de Markov caché (HMM, *Hidden Markov Model*) est utilisé, et la poursuite de la présentation nécessite d'introduire les notations suivantes :

$$\mathcal{R}_k = \{\mathbf{r}; z(\mathbf{r}) = k\}, \quad \text{Card}(\mathcal{R}_k) = n_k, \quad \mathcal{R} = \bigcup_{k=1}^K \mathcal{R}_k, \quad \forall i \neq j \quad \mathcal{R}_i \cap \mathcal{R}_j = \emptyset,$$

$$\mathbf{f}_k = \{f(\mathbf{r}); z(\mathbf{r}) = k\}, \quad \mathbf{f} = \bigcup_{k=1}^K \mathbf{f}_k.$$

Dans l'hypothèse où tous les pixels  $\mathbf{f}_k$  d'une région homogène  $k$  sont caractérisés par une moyenne  $m_k$ , et une variance  $\sigma_k^2$ , on suppose que les valeurs du vecteur  $\mathbf{f}_k$  suivent une loi gaussienne de la forme :

$$p(\mathbf{f}_k) = p(f(\mathbf{r}), \mathbf{r} \in \mathcal{R}_k) = \mathcal{N}(\mathbf{f}_k | m_k \mathbf{1}, \mathbf{\Sigma}_k), \quad (5.6)$$

où  $\mathbf{1}$  est un vecteur unitaire de taille  $n_k$ .

Supposer que les pixels dans une région homogène sont indépendants conduirait à pouvoir écrire que  $\mathbf{\Sigma}_k = \sigma_k^2 \mathbf{I}$  avec  $\mathbf{I}$  la matrice identité de taille  $n_k$ . Toutefois, il est plus pertinent de faire l'hypothèse que les pixels appartenant à une région homogène sont dépendants les uns des autres.  $\mathbf{\Sigma}_k$  est donc la matrice de covariance d'un champs de Markov aléatoire (MRF) paramétré par  $\{\sigma_k, \rho_k\}$ , avec une dépendance représentée à l'aide d'un modèle de Markov du premier ordre (cf. FIG. 5.1).

Plus particulièrement, on peut écrire :

$$p(\mathbf{f}_k) \propto \exp \left( \frac{1}{2\pi \sigma_k^2} \sum_{\mathbf{r} \in \mathcal{R}_k} \left( f(\mathbf{r}) - m_k(\mathbf{r}) - \rho_k \beta_{\mathbf{r}} \sum_{\mathbf{r}' \in (\mathcal{V}(\mathbf{r}) \cap \mathcal{R}_k)} (f(\mathbf{r}') - m_k(\mathbf{r}')) \right)^2 \right)$$

et

$$p(f_k(\mathbf{r}) | f(\mathbf{r}'), \mathbf{r}' \in \mathcal{R}) \propto \exp \left( \frac{1}{2\pi \sigma_k^2} \left( f(\mathbf{r}) - m_k(\mathbf{r}) - \rho_k \beta_{\mathbf{r}} \sum_{\mathbf{r}' \in (\mathcal{V}(\mathbf{r}) \cap \mathcal{R}_k)} (f(\mathbf{r}') - m_k(\mathbf{r}')) \right)^2 \right)$$

avec  $\beta_{\mathbf{r}}$  un coefficient qui dépend du pixel  $\mathbf{r}$  et de son voisinage. Soit  $n_{\mathbf{r}} = \text{Card}(\mathcal{V}(\mathbf{r}) \cap \mathcal{R}_k)$ , c'est-à-dire le nombre de pixels voisins de  $\mathbf{r}$  appartenant également à la région  $\mathcal{R}_k$ , alors :

$$\beta_{\mathbf{r}} = \begin{cases} \frac{1}{n_{\mathbf{r}}} & \text{si } n_{\mathbf{r}} \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

De plus, le facteur  $\rho_k = 0$  dans l'hypothèse d'indépendance, et  $\rho_k = 1$  dans celle de dépendance. L'expression de  $m_k(\cdot)$  sera précisée plus loin dans cette partie.

Deux définitions de l'homogénéité présentées lors de ces travaux seront détaillées à la fin de cette partie.

D'autre part, les pixels appartenant à des régions différentes sont supposés indépendants. Cela permet d'écrire la probabilité conditionnelle de connaître l'image HR connaissant son résultat de segmentation joint sous la forme :

$$p(\mathbf{f} | \mathbf{z}) = \prod_{k=1}^K p(\mathbf{f}_k) = \prod_{k=1}^K \mathcal{N}(\mathbf{f}_k | m_k \mathbf{1}, \mathbf{\Sigma}_k).$$

L'équation (5.6) est conditionnée par la valeur  $z(\mathbf{r}) = k$ , et produit un modèle de mélange fini (FMM, *Finite Mixture Model*).

Ceci signifie également qu'un nombre fini de classes est utilisé pour la classification jointe, et on obtient :

$$p(f(\mathbf{r})) = \sum_{k=1}^K p(z(\mathbf{r}) = k) \mathcal{N}(f_k(\mathbf{r}) | m_k, \sigma_k^2).$$

Le cadre bayésien à base de FMM est un outil commun en classification, mais en général, la variable discrète associée à chaque pixel est supposée être i.i.d. [Cheeseman 94].

Ici, un modèle de Potts Markov (PMM, *Potts Markov Model*) a été privilégié pour ces variables car il permet de considérer la corrélation spatiale des pixels tout en étant simple à mettre en œuvre.

Ainsi, la distribution de probabilité de l'image des labels s'écrit :

$$p(\mathbf{z}) = \frac{1}{T(\alpha)} \exp \left( \alpha \sum_{\mathbf{c}=(\mathbf{r}, \mathbf{r}') \in \mathcal{C}_2} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right). \quad (5.7)$$

Ici aussi, un modèle de Markov du premier ordre est utilisé. Ceci suppose donc que chaque label  $z(\mathbf{r})$  est fonction de la valeur du label de ses quatre plus proches voisins  $z(\mathbf{r}')$ ,  $\mathbf{r}' \in \mathcal{V}(\mathbf{r})$ .

$T(\alpha)$  est une constante de normalisation, et le paramètre  $\alpha$  représente un degré de dépendance spatial. Ce dernier contrôle la taille moyenne des classes : prendre  $\alpha$  trop faible conduit à obtenir des images « aléatoires » (équivalentes à des images de bruit), augmenter  $\alpha$  rend les différentes classes plus homogènes, choisir  $\alpha$  trop grand conduit à l'obtention d'images quasiment uniformes [Morris 96]. Il existe une valeur critique théorique sur  $\alpha$ , fonction du nombre de labels  $K$ , en dessous de laquelle les résultats obtenus sont aléatoires. Ainsi, il existe une plage de bon fonctionnement pour ce paramètre. En pratique, on utilise  $\alpha = 2$ .

L'équivalence d'Hammersley-Clifford entre les champs aléatoires de Markov et Gibbs permet d'écrire :

$$p(z(\mathbf{r}) | z(\mathbf{r}'), \mathbf{r}' \in \mathcal{R}) \propto \exp \left( \alpha \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right).$$

Ceci montre que la probabilité d'obtenir un label  $z(\mathbf{r})$  pour un pixel est liée au nombre de pixels voisins ayant le même label.

La fonction  $\delta(\cdot)$  est la fonction de Kronecker définie par  $\delta(0) = 1$ , et  $\delta(t) = 0$  pour  $t \neq 0$ .

Bien que le modèle de Potts ne soit pas bien adapté à des traitements de segmentation pur [Morris 96], il s'avère efficace lorsqu'il est utilisé avec un traitement joint de SR.

Des alternatives à ce choix de modèle sont la variante hétérogène du modèle de Potts ou le modèle Chien proposés dans [Descombes 99].

Pour terminer cette présentation du modèle *a priori*, la dernière phase consiste à assigner des lois de probabilités aux différents hyper-paramètres  $\theta$  du problème. Pour simplifier, des *a priori* conjugués sont utilisés.

Ce choix a deux principaux avantages :

- les *a priori* conjugués peuvent être considérés comme une famille particulière d'*a priori* basée sur la géométrie différentielle [Snoussi 03, Snoussi 04],
- ils présentent l'intérêt pratique que les lois de probabilités *a priori* et *a posteriori* restent dans la même famille.

De manière classique, les lois de probabilités conjuguées pour les moyennes  $m_k$  sont les gaussiennes  $\mathcal{N}(m_k | m_{k0}, \sigma_{k0}^2)$ , celles pour les matrices de covariances  $\Sigma_k$  des lois inverses de Wishart  $\mathcal{IW}(\Sigma_k | \alpha_{k0}, \mathbf{A}_{k0})$ , ce qui revient à utiliser des lois inverses gamma  $\mathcal{IG}(\sigma_k^2 | \alpha_{k0}, \beta_{k0})$  pour les variances  $\sigma_k^2$ .



Enfin, les lois *a priori* conjuguées pour la variance du bruit  $\sigma_{\epsilon_i}^2$  dans les images observées sont également choisies comme des lois inverses gamma  $\mathcal{IG}(\sigma_{\epsilon_i}^2 | \alpha_0^{\epsilon_i}, \beta_0^{\epsilon_i})$ .

Pour rappel, la forme de la loi inverse gamma est :

$$\mathcal{IG}(\sigma_k^2 | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \frac{1}{\sigma_k^2} \right)^{\alpha-1} \exp \left( -\beta \frac{1}{\sigma_k^2} \right),$$

avec  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ , pour  $\alpha$  réel strictement positif,

ce qui permet aussi d'écrire la relation  $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ .

Dans le cas entier, la relation  $\Gamma(n + 1) = n!$  peut également être utilisée, avec  $n!$  représentant la fonction factorielle.

La forme de la loi inverse Wishart est :

$$\mathcal{IW}(\mathbf{\Sigma}_k | \alpha, \mathbf{\Lambda}) = \frac{|\mathbf{\Lambda}|^\alpha}{\Gamma(\alpha)} (\mathbf{\Sigma}_k)^{1-\alpha} \exp(-\mathbf{\Lambda} \mathbf{\Sigma}_k^{-1}).$$

Deux définitions de l'homogénéité des régions ont été présentées au cours du développement de ce modèle. Elles se basent sur l'équation (5.6) qui formule l'hypothèse que tous les pixels de l'image  $\mathbf{f}$  ayant le label  $k$  suivent une loi gaussienne de moyenne  $m_k$  et de variance  $\sigma_k^2$ .

Dans la méthode présentée dans [Humblot 05c], nous avons fait l'hypothèse que les régions homogènes étaient constantes par morceaux. Ceci peut s'écrire  $m_k(\mathbf{r}) = m_k, \forall \mathbf{r} \in \mathcal{R}_k$ .

Ainsi, dans ce cas, chaque région  $k$  est définie par l'ensemble de paramètres  $(m_k, \sigma_k^2, \rho_k)$  selon les lois *a priori* conjuguées détaillées ci-dessus.

Ce modèle a ensuite été étendu dans [Humblot 05b] de manière à considérer des régions homogènes suivant un modèle bilinéaire par morceaux.

Ceci s'écrit  $m_k(\mathbf{r}) = m_k(x, y) = a_k \times x(\mathbf{r}) + b_k \times y(\mathbf{r}) + m_k, \forall \mathbf{r} \in \mathcal{R}_k$ .

L'utilisation d'un tel modèle conduit à caractériser chaque région  $k$  par les paramètres  $(a_k, b_k, m_k, \sigma_k^2, \rho_k)$ . Les lois *a priori* sur  $(a_k, b_k)$  sont choisies comme des lois gaussiennes de moyenne nulle, et de variance élevée (ce qui est donc presque équivalent à un *a priori* uniforme). D'un point de vue *a posteriori*, cela revient à utiliser des lois gaussiennes dont les valeurs moyennes sont estimées à partir des données, et dont les variances sont faibles. La méthode utilisée pour estimer ces valeurs moyennes à partir des données est celle de la régression linéaire, avec un estimateur au sens du maximum de vraisemblance. Comme précédemment, des lois *a priori* conjuguées sont utilisées pour la fixation des paramètres du modèle.

La figure FIG. 5.2 illustre la différence entre le modèle constant (a) et le modèle bilinéaire (c) pour deux images de tailles  $200 \times 200$  pixels.

(b) et (d) montrent les intensités des pixels sur la 170<sup>e</sup> ligne des deux images correspondantes.

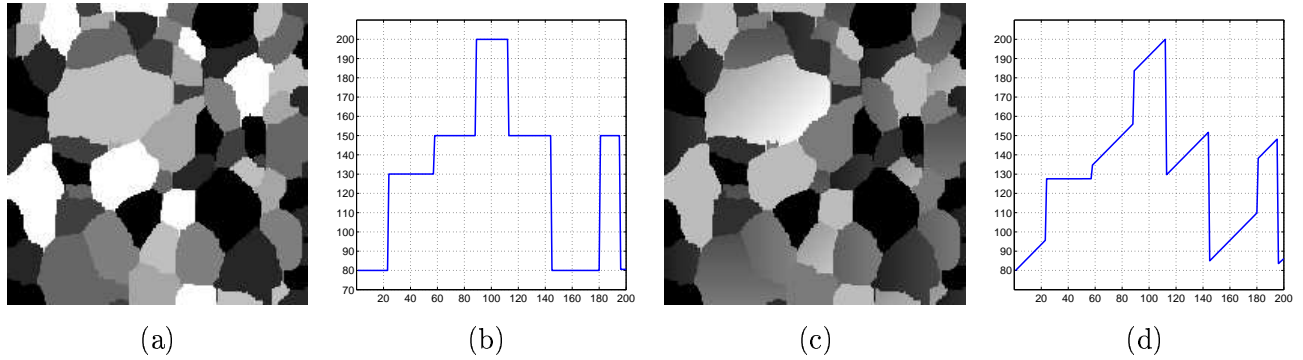


FIG. 5.2: Comparaison des deux modèles d'homogénéité de régions : constant et bilinéaire.

(a) image constituée de régions homogènes constantes, (b) 170<sup>e</sup> ligne de (a),  
(c) image constituée de régions homogènes bilinéaires, (d) 170<sup>e</sup> ligne de (c).

## 5.4 Mise en œuvre du calcul de la solution

### 5.4.1 L'échantillonneur de Gibbs

Un champ de Markov étant défini (cf. 5.3.1), la question se pose de réaliser un tirage d'une configuration (ici, une image) en suivant la loi de probabilité de Gibbs caractéristique de ce champ. Deux algorithmes ont été proposés pour synthétiser des réalisations d'un champ de Markov : l'échantillonneur de Gibbs et l'algorithme de Metropolis. Dans l'approche de ce problème, une méthode hybride utilisant pour partie l'échantillonneur de Gibbs a été choisie.

La méthode de Gibbs, présentée dans [Geman 84], repose sur la construction itérative d'une suite d'images. À la convergence, c'est-à-dire après un nombre d'itérations suffisant, les images construites sont des réalisations tirées selon la loi de Gibbs globale.

La méthode de construction de l'image à l'itération  $k$ , partant de l'image à l'itération  $k - 1$ , se fait par mises à jour successives des sites de l'image. Ainsi, à l'étape  $k$  :

- on choisit un site  $\mathbf{s}$ ,
- au site  $\mathbf{s}$ , selon la configuration des voisins  $V_{\mathbf{s}} = (f_{\mathbf{t}}, \mathbf{t} \in \mathcal{V}(\mathbf{s}))$  pour l'image  $\mathbf{f}^{(k-1)}$ , on calcule la probabilité locale :

$$p(F_{\mathbf{s}} = f_{\mathbf{s}} | V_{\mathbf{s}}) = \frac{\exp(-U_{\mathbf{s}}(f_{\mathbf{s}} | V_{\mathbf{s}}))}{\sum_{n \in E} \exp(-U_{\mathbf{s}}(n | V_{\mathbf{s}}))}$$

- on met à jour le site  $\mathbf{s}$  par tirage aléatoire selon la loi  $p(F_{\mathbf{s}} = f_{\mathbf{s}} | V_{\mathbf{s}})$ .

On considère que l'algorithme a convergé après un grand nombre d'itérations ou lorsque le nombre de changements est faible (uniquement pour de faibles températures). En pratique, on utilise ce deuxième critère. Le choix du site  $\mathbf{s}$  considéré à l'étape  $k$  peut se faire de n'importe quelle façon à condition de balayer tous les sites un très grand nombre de fois (théoriquement un nombre infini de fois). Les méthodes usuelles consistent à tirer un site selon une loi uniforme, ou effectuer un balayage classique, ligne par ligne de l'image.

Une autre manière de procéder, utilisée pour l'implémentation de l'algorithme, consiste à considérer les sites de l'image comme un échiquier, avec des cases noires et blanches alternativement.

Le système de voisinage en « 4-connexité » permet de paralléliser les calculs, de traiter une fois toutes les cases blanches, puis la fois suivante les cases noires.

Cet algorithme construit en réalité une suite d'images  $\mathbf{f}^{(k)}$ , observation d'une suite  $\mathbf{F}^{(k)}$  de champs aléatoires constituant une chaîne de Markov pour un certain noyau de transition.

On peut démontrer le théorème suivant, lorsque la séquence balaye chaque site une infinité de fois :

$$\forall \mathbf{f}^{(0)}, \forall \mathbf{f} \in \Omega, \lim_{k \rightarrow \infty} p(\mathbf{F}^{(k)} = \mathbf{f} \mid \mathbf{F}^{(0)} = \mathbf{f}^{(0)}) = p(\mathbf{f})$$

où  $p$  est la mesure de Gibbs associée au champ de Markov considéré.

Ainsi, après un grand nombre d'itérations, les images  $\mathbf{f}^{(k)}$  générées sont des réalisations de la loi globale  $p(\mathbf{f})$ , indépendamment de la configuration initiale  $\mathbf{f}^{(0)}$ .

La preuve de cet algorithme, et donc de sa convergence, est donnée dans [Geman 84, Winkler 95].

D'autre part, on parle de l'échantillonneur de Gibbs comme d'un algorithme de **relaxation**, car il procède par mises à jour successives des sites, et **probabiliste** car cette mise à jour est fondée sur un **tirage aléatoire**.

### 5.4.2 Algorithmes proposés

Le cadre bayésien d'estimation pour la SR peut être résumé comme suit :

- utiliser le modèle direct d'observation (4.1) et des hypothèses sur le bruit afin d'obtenir le terme de vraisemblance  $p(\mathbf{g} \mid \mathbf{f}, \boldsymbol{\theta}_\epsilon)$  où  $\boldsymbol{\theta}_\epsilon$  représente les paramètres de la distribution de probabilités du bruit.

Ceci a été fait dans la partie 5.2, et on a noté  $\boldsymbol{\theta}_\epsilon = \{\sigma_{\epsilon_i}^2, i = 1, \dots, M\}$ .

- Utiliser toutes les informations *a priori* ou toutes les propriétés désirées sur la solution pour déterminer une distribution de probabilité *a priori*  $p(\mathbf{f} \mid \mathbf{z}, \boldsymbol{\theta}_f)$  où  $\boldsymbol{\theta}_f$  représente ses paramètres. Cette tâche a été accomplie dans la partie 5.3.4. Et selon le cas constant ou bilinéaire, on a respectivement  $\boldsymbol{\theta}_f = \{(m_k, \sigma_k^2, \rho_k), k = 1, \dots, K\}$  ou  $\boldsymbol{\theta}_f = \{(a_k, b_k, m_k, \sigma_k^2, \rho_k), k = 1, \dots, K\}$ .

- Utiliser la règle de Bayes pour obtenir la distribution de probabilités *a posteriori* jointe :

$$p(\mathbf{f}, \mathbf{z} \mid \mathbf{g}, \boldsymbol{\theta}) \propto p(\mathbf{g} \mid \mathbf{f}, \boldsymbol{\theta}_\epsilon) p(\mathbf{f} \mid \mathbf{z}, \boldsymbol{\theta}_f) p(\mathbf{z}) \quad (5.8)$$

dans le cas supervisé où les paramètres  $\boldsymbol{\theta} = (\boldsymbol{\theta}_\epsilon, \boldsymbol{\theta}_f)$  sont connus ; ou pour obtenir la distribution de probabilité *a posteriori* jointe :

$$p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} \mid \mathbf{g}) \propto p(\mathbf{g} \mid \mathbf{f}, \boldsymbol{\theta}_\epsilon) p(\mathbf{f} \mid \mathbf{z}, \boldsymbol{\theta}_f) p(\mathbf{z}) p(\boldsymbol{\theta}_f) p(\boldsymbol{\theta}_\epsilon)$$

dans le cas non-supervisé où les paramètres  $\boldsymbol{\theta}$  sont inconnus.

Les hyper-paramètres de ce problème étant inconnus au début de la résolution, on se place par la suite dans le cas non-supervisé. La définition de la loi *a posteriori* sera traitée dans cette partie.

- Enfin, définir des estimateurs  $\hat{\mathbf{f}}$  pour  $\mathbf{f}$  et  $\hat{\boldsymbol{\theta}}$  pour  $\boldsymbol{\theta}$  basés sur ces lois de probabilités *a posteriori*. Cet aspect sera également abordé dans cette partie.

Les résultats obtenus auparavant, qui permettent de définir la probabilité *a posteriori* jointe recherchée, sont rappelés ci-après.

L'expression complète de la vraisemblance, abordée dans la partie 5.2, est la suivante :

$$p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta}_\epsilon) = \prod_{i=1}^M \left( \frac{1}{2\pi\sigma_{\epsilon_i}^2} \right)^{\frac{R}{2}} \exp \left( -\frac{1}{2\sigma_{\epsilon_i}^2} \sum_{\mathbf{r} \in \mathcal{R}} \left( g_i(\mathbf{r}) - [\mathcal{H}_i \mathbf{f}](\mathbf{r}) \right)^2 \right).$$

L'*a priori* défini dans la partie 5.3.4 permet d'écrire la probabilité conditionnelle relative à la solution recherchée  $\mathbf{f}$  suivante :

$$p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f) = \prod_{k=1}^K \left( \frac{1}{2\pi\sigma_k^2} \right)^{\frac{n_k}{2}} \exp \left( \frac{1}{2\pi\sigma_k^2} \sum_{\mathbf{r} \in \mathcal{R}_k} \left( f(\mathbf{r}) - m_k(\mathbf{r}) - \rho_k \beta \mathbf{r} \sum_{\mathbf{r}' \in (\mathcal{V}(\mathbf{r}) \cap \mathcal{R}_k)} \left( f(\mathbf{r}') - m_k(\mathbf{r}') \right) \right)^2 \right).$$

Ensuite, le choix d'un modèle de Potts Markov pour l'image des labels a permis d'écrire l'expression de  $p(\mathbf{z})$  selon l'équation (5.7).

Enfin, le choix des *a priori* conjugués a également été détaillé à la fin de la partie 5.3.4.

L'ensemble de ces choix permet de pouvoir écrire la loi *a posteriori* jointe de toutes les inconnues à estimer  $\mathbf{f}$ ,  $\mathbf{z}$  et  $\boldsymbol{\theta}$  :

$$p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}) \propto p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta}_\epsilon) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f) p(\mathbf{z}) p(\boldsymbol{\theta}). \quad (5.9)$$

Une première approche peut consister à utiliser l'estimateur MAP, déjà présenté précédemment. Pour ce faire, il faut se placer sous l'hypothèse que  $\boldsymbol{\theta}$  et  $\mathbf{z}$  sont connus.

Comme cela a déjà été présenté sous une forme légèrement différente dans l'équation (5.2), on a :

$$\hat{\mathbf{f}}_{MAP} = \arg \max_{\mathbf{f}} [p(\mathbf{f} | \mathbf{g}, \mathbf{z}, \boldsymbol{\theta})] = \arg \min_{\mathbf{f}} \left[ -\ln(p(\mathbf{f} | \mathbf{g}, \mathbf{z}, \boldsymbol{\theta})) \right] = \arg \min_{\mathbf{f}} [J(\mathbf{f} | \mathbf{g}, \mathbf{z}, \boldsymbol{\theta})]$$

La formule de Bayes donne :

$$p(\mathbf{f} | \mathbf{g}, \mathbf{z}, \boldsymbol{\theta}) = \frac{p(\mathbf{g} | \mathbf{z}, \mathbf{f}, \boldsymbol{\theta}) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f)}{p(\mathbf{g} | \mathbf{z}, \boldsymbol{\theta})},$$

expression dans laquelle le terme  $p(\mathbf{g} | \mathbf{z}, \boldsymbol{\theta})$  joue le rôle d'une constante du point de vue d'une maximisation en  $\mathbf{f}$ .

$\mathbf{z}$  n'apportant aucune information sur  $\mathbf{g}$ , on a  $p(\mathbf{g} | \mathbf{z}, \mathbf{f}, \boldsymbol{\theta}) = p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta})$ .

Ainsi, la maximisation de  $p(\mathbf{f} | \mathbf{g}, \mathbf{z}, \boldsymbol{\theta})$  devient la minimisation de la fonction en  $\mathbf{f}$  :

$$\mathbf{J}(\mathbf{f} | \mathbf{g}, \mathbf{z}, \boldsymbol{\theta}) = \sum_{i=1}^M \frac{1}{\sigma_{\epsilon_i}^2} \sum_{\mathbf{r} \in \mathcal{R}} \left( g_i(\mathbf{r}) - [\mathcal{H}_i \mathbf{f}](\mathbf{r}) \right)^2 + \frac{1}{\sigma_k^2} \sum_{\mathbf{r} \in \mathcal{R}_k} \left( \tilde{f}(\mathbf{r}) - \rho_k \beta_{\mathbf{r}} \sum_{\mathbf{r}' \in (\mathcal{V}(\mathbf{r}) \cap \mathcal{R}_k)} \tilde{f}(\mathbf{r}') \right)^2 \quad (5.10)$$

où  $\tilde{f}(\mathbf{r}) = f(\mathbf{r}) - m_k(\mathbf{r})$ .

On peut constater que lorsqu'on suppose l'image composée d'une seule région statistiquement homogène ( $K = 1$ ), le critère de l'équation (5.10) peut s'écrire de manière compacte :

$$\mathbf{J}(\mathbf{f}) = \sum_{i=1}^M \|\mathbf{g}_i - \mathbf{H}_i \mathbf{f}\|^2 + \left( \frac{\sigma_{\epsilon_i}}{\sigma_f} \right)^2 \|\mathbf{D}(\mathbf{f} - m \mathbf{1})\|^2 \quad (5.11)$$

où  $m$  et  $\sigma_f$  sont estimés au cours du processus de maximisation.

L'équation (5.11) obtenue peut être comparée à celles formulées par les équations (5.3), et (5.5) du cas lié à la régularisation de Gauss–Markov (cf. 5.3.2).

Ainsi, comparée à une méthode classique de régularisation utilisée en SR (voir par exemple [Hong 97a]), l'estimateur MAP correspondant permet de détailler davantage la modélisation de l'image HR, avec des régions homogènes par morceaux.

Les trois algorithmes proposés pour la résolution du problème de SR reposent ainsi sur la loi *a posteriori* jointe (5.9).

Il s'agit de méthodes itératives basées sur le schéma d'un échantillonnage de Gibbs par chaîne de Markov de Monte-Carlo (MCMC, *Markov Chain Monte-Carlo*).

Un algorithme basé totalement sur ce schéma aurait consisté à échantillonner chacune des inconnues à estimer selon leur loi marginale *a posteriori*. Théoriquement, ce type d'échantillonnage fait converger chaque inconnue vers la moyenne *a posteriori* de leur loi marginale, après un très grand nombre d'itérations, ou lorsque le nombre de changements est faible.

Ce type de critère pouvant être long à atteindre, des schémas hybrides d'échantillonnage de Gibbs MCMC utilisant un estimateur MAP pour l'estimation de l'image HR ont été proposés :

– Estimateurs MAP joints (algorithme 1) :

$$\begin{cases} \hat{\mathbf{f}}^n = \arg \max_{\mathbf{f}} \left[ p(\mathbf{f} | \hat{\mathbf{z}}^{n-1}, \hat{\boldsymbol{\theta}}^{n-1}, \mathbf{g}) \right] = \arg \min_{\mathbf{f}} [\mathbf{J}(\mathbf{f} | \mathbf{g}, \mathbf{z}, \boldsymbol{\theta})] \\ \hat{\boldsymbol{\theta}}^n = \arg \max_{\boldsymbol{\theta}} \left[ p(\boldsymbol{\theta} | \hat{\mathbf{f}}^n, \hat{\mathbf{z}}^{n-1}, \mathbf{g}) \right] \\ \hat{\mathbf{z}}^n = \arg \max_{\mathbf{z}} \left[ p(\mathbf{z} | \hat{\mathbf{f}}^n, \hat{\boldsymbol{\theta}}^n, \mathbf{g}) \right] \end{cases}$$

– Échantillonnage de Gibbs / Estimateur MAP (algorithme 2) :

$$\begin{cases} \hat{\mathbf{f}}^n = \arg \max_{\mathbf{f}} \left[ p(\mathbf{f} | \hat{\mathbf{z}}^{n-1}, \hat{\boldsymbol{\theta}}^{n-1}, \mathbf{g}) \right] \\ \hat{\boldsymbol{\theta}}^n = \arg \max_{\boldsymbol{\theta}} \left[ p(\boldsymbol{\theta} | \hat{\mathbf{f}}^n, \hat{\mathbf{z}}^{n-1}, \mathbf{g}) \right] \\ \text{échantillonner } \hat{\mathbf{z}}^n \text{ à partir de } p(\mathbf{z} | \hat{\mathbf{f}}^n, \hat{\boldsymbol{\theta}}^n, \mathbf{g}) \text{ ou à partir de } p(\mathbf{z} | \hat{\boldsymbol{\theta}}^n, \mathbf{g}) \end{cases}$$

– Échantillonnage de Gibbs / Estimateur MAP (algorithme 3) :

$$\left\{ \begin{array}{l} \hat{\mathbf{f}}^n = \arg \max_{\mathbf{f}} \left[ p(\mathbf{f} | \hat{\mathbf{z}}^{n-1}, \hat{\boldsymbol{\theta}}^{n-1}, \mathbf{g}) \right] \\ \text{échantillonner } \hat{\boldsymbol{\theta}}^n \text{ à partir de } p(\boldsymbol{\theta} | \hat{\mathbf{f}}^n, \hat{\mathbf{z}}^{n-1}, \mathbf{g}) \\ \text{échantillonner } \hat{\mathbf{z}}^n \text{ à partir de } p(\mathbf{z} | \hat{\mathbf{f}}^n, \hat{\boldsymbol{\theta}}^n, \mathbf{g}) \text{ ou à partir de } p(\mathbf{z} | \hat{\boldsymbol{\theta}}^n, \mathbf{g}) \end{array} \right.$$

Pour l'initialisation de ces algorithmes, on commence le processus en posant  $K = 1$  et ainsi  $\mathbf{z} = \mathbf{1}$ , et  $\boldsymbol{\theta} = [m_1, \sigma_1^2] = [0, 1]$ . On obtient ainsi une solution régularisée  $\mathbf{f}^{(0)}$  en utilisant (5.11), à partir de laquelle on utilise n'importe quelle méthode de segmentation basée sur l'histogramme pour obtenir  $\mathbf{z}^{(0)}$ . Ce résultat de segmentation permet alors d'estimer  $\boldsymbol{\theta}^{(0)}$  à partir d'une méthode classique au sens du maximum de vraisemblance. Par la suite, on peut appliquer l'un des processus itératifs proposés. Ces opérations sont répétées par itération jusqu'à ce que l'un des deux critères d'arrêt suivants soient atteints :

$$\left\{ \begin{array}{l} \sum_{\mathbf{r} \in \mathcal{R}} \frac{|\hat{\mathbf{f}}^n(\mathbf{r}) - \hat{\mathbf{f}}^{n-1}(\mathbf{r})|}{\hat{\mathbf{f}}^n(\mathbf{r})} < \epsilon_{\mathbf{f}} \\ \text{ou} \\ \frac{|J(\hat{\mathbf{f}}^n) - J(\hat{\mathbf{f}}^{n-1})|}{J(\hat{\mathbf{f}}^n)} < \epsilon_{\mathbf{J}} \end{array} \right.$$

avec  $J(\hat{\mathbf{f}})$  calculé à partir de (5.10).

On peut remarquer qu'il n'y a aucune garantie théorique de convergence pour aucun des algorithmes proposés.

En particulier, le point critique des deux premiers algorithmes est le calcul de  $\hat{\boldsymbol{\theta}}$  :

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \left[ p(\boldsymbol{\theta} | \hat{\mathbf{f}}, \hat{\mathbf{z}}, \mathbf{g}) \right].$$

En effet, ce calcul peut donner des valeurs très faibles de  $\sigma_k$  ( $10^{-6}$  ou encore moins) si les pixels d'une région  $\mathcal{R}_k$  ont tous une intensité voisine. Cette difficulté est atténuée dans le dernier algorithme, et c'est pourquoi seul ce schéma hybride a été conservé par la suite.

Des précisions sur l'obtention des lois marginales *a posteriori* sont données dans l'annexe **A**, p. 145.

## 5.5 Liens avec les méthodes existantes

La méthode proposée s'inscrit dans le contexte de l'**estimation bayésienne**, et dans la famille des **techniques stochastiques** présentées de manière détaillée dans la partie 4.4.2.4.

De nombreux travaux reposant sur une estimation MAP dans un cas supervisé ont été proposés récemment. Les lois *a posteriori* obtenues sont de la forme de (5.8); une fonction  $\mathbf{J}(\mathbf{f})$  à minimiser en est dérivée.

Le papier [Hardie 97a], qui propose une méthode de SR basée sur l'estimation jointe de l'image HR  $\mathbf{f}$  et des paramètres inter-images de recalage  $\mathbf{s}$ , a été présenté dans 4.4.2.4. Les auteurs se placent dans le contexte bayésien, et utilisent la loi *a posteriori* jointe  $p(\mathbf{f}, \mathbf{s} | \mathbf{g})$  pour obtenir une fonction  $\mathbf{J}(\mathbf{f}, \mathbf{s})$  à minimiser alternativement selon chacune des deux variables. Le modèle utilisé correspond à celui présenté dans l'algorithme 1.

Shekarforoush *et al.* ont également présenté une approche bayésienne avec deux variables inconnues à estimer dans [Shekarforoush 95, Lorette 97], en s'intéressant au problème de restauration 3-D par SR. La formulation de leur problème consiste à estimer deux images indépendantes, une première d'intensité  $\mathbf{f}$  et une autre d'altitude  $\mathbf{z}$ . Dans leur modélisation, ces deux images sont considérées comme indépendantes, ce qui n'est pas le cas dans notre approche. Ils cherchent à maximiser la loi *a posteriori* jointe  $p(\mathbf{f}, \mathbf{z} | \mathbf{g})$ , et obtiennent une fonction à minimiser de la forme :

$$\mathbf{J}(\mathbf{f}, \mathbf{z}) = \sum_{i=1}^M \|\mathbf{g}_i - \mathbf{H}_i \mathbf{f}\|^2 + \frac{1}{\sigma_f^2} \|\mathbf{D} \mathbf{f}\|^2 + \frac{1}{\sigma_z^2} \|\mathbf{D} \mathbf{z}\|^2$$

où au niveau pixellique,  $\mathbf{D} \mathbf{f}$  représente l'opération :

$$\sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \phi(f(\mathbf{r}) - f(\mathbf{r}')).$$

Plusieurs choix de fonction de pénalisation  $\phi(\cdot)$  ont été étudiés dans ces papiers, à savoir :

$$\begin{aligned} \phi_1(u) &= u^2 \\ \phi_2(u) &= 2 \times \sqrt{1 + u^2} - 2 \\ \phi_3(u) &= \log(1 + u^2) \\ \phi_4(u) &= \frac{u^2}{1 + u^2}. \end{aligned}$$

Dans notre cas, seule la fonction quadratique représentée par  $\phi_1$  a été expérimentée. La méthode d'optimisation choisie s'apparente aussi à celle de notre algorithme 1.

Dans [Cheeseman 94], Cheeseman *et al.* ont utilisé une technique bayésienne de SR, avec estimateur MAP, sur des images reçues de satellites explorant le système solaire. Dans leur modèle, la variable discrète associée à chaque pixel est supposée i.i.d. Le modèle de Potts Markov utilisé pour ces variables permet de considérer la corrélation spatiale des pixels.

Enfin on peut citer les travaux de Molina *et al.*. Dans [Molina 03a], une approche basée sur un modèle de Gauss–Markov avec mélange de gaussiennes et des variables cachées « lignes ». L'objectif des auteurs est d'obtenir une image HR  $\mathbf{f}$ , et une image  $\mathbf{l}$  modélisant des transitions dans l'image HR. Ce second résultat ressemble à une image de contours. Il s'agit d'une méthode de restauration mono-image et non du traitement de SR multi-images, et d'une modélisation selon un traitement « ligne », et non « région ».

Dans [Molina 03b], les auteurs considèrent le problème de SR dans le cadre bayésien, mais en s'intéressant plus particulièrement à l'estimation des paramètres du problème. De plus, *a priori* utilisé pour modéliser les images HR est de type Gauss–Markov, et n'utilise pas un modèle de mélange de gaussiennes avec variable cachée.

## 5.6 Méthodes d'évaluation

Comment évaluer une image reconstruite par une méthode de SR ?

Dans un cas réel, il est nécessaire d'avoir deux appareils d'acquisition : l'un HR et l'autre BR. Ceux-ci doivent observer une scène identique, avec des points de vue proches, de manière à pouvoir comparer une image  $\mathbf{f}_0$  obtenue par l'appareil HR, avec une image HR  $\hat{\mathbf{f}}$  reconstruite en utilisant une technique de SR, et les images  $\mathbf{g}_i$  provenant de l'appareil BR.

En l'absence d'une seconde caméra HR, l'évaluation de la méthode de SR ne peut être que subjective à un observateur humain, ou basée sur un taux de succès à un traitement postérieur à celui de SR.

Ces traitements ultérieurs peuvent être de type détection ou suivi d'objets, et le résultat joint de segmentation  $\hat{\mathbf{z}}$  pourrait aider à améliorer le taux de réussite. En pratique, une seule caméra a pu être utilisée, et un système plus élaboré d'évaluation par traitement postérieur n'a pas pu être mis en œuvre. Toutefois, quelques résultats obtenus à partir de séquences réelles seront présentés dans la partie 5.7.

Dans un cas de simulation, on dispose d'une image HR de référence  $\mathbf{f}_0$ , et d'un résultat de segmentation  $\mathbf{z}_K$  obtenu à l'aide d'une méthode classique basée sur l'histogramme, ou par construction. On dégrade alors  $\mathbf{f}_0$  selon le modèle d'observation défini avec (4.1) pour obtenir  $M = k_0^2$  images  $\mathbf{g}_i$  BR.

On peut alors appliquer les méthodes de SR en utilisant la connaissance exacte du mouvement inter-images  $\mathbf{s}_0$ , ou une estimation  $\hat{\mathbf{s}}_0$  de celui-ci, pour obtenir une image HR  $\hat{\mathbf{f}}$  et sa segmentation jointe  $\hat{\mathbf{z}}_K$ .

Ces différentes notations étant précisées, quelques critères d'évaluation des méthodes de SR peuvent être présentés.

- L'image de différence  $\Delta$  entre l'image HR de référence  $\mathbf{f}_0$  et une image HR  $\hat{\mathbf{f}}$  reconstruite à partir des images bruitées  $\mathbf{g}_i$  :

$$\Delta(\hat{\mathbf{f}}, \mathbf{f}_0) = \hat{\mathbf{f}} - \mathbf{f}_0 = (\hat{f}(\mathbf{r}) - f_0(\mathbf{r})), \forall \mathbf{r} \in \mathcal{R}$$

dont on peut examiner quelques caractéristiques telles que la moyenne (qui représente le biais de l'estimateur HR) et l'écart-type de l'image de différence  $\Delta(\hat{\mathbf{f}}, \mathbf{f}_0)$  :  $\Delta_{\text{moy}}$  et  $\Delta_{\text{ET}}$ . Plus ces deux paramètres sont faibles, plus l'image reconstruite est statistiquement proche de l'image HR d'ori-



gine, et donc plus la méthode donne de bons résultats.

- Une mesure d'erreur relative normalisée  $\Delta_\alpha(\hat{\mathbf{f}}, \mathbf{f}_0)$  pour  $\alpha = \{1, 2\}$  :

$$\Delta_\alpha(\hat{\mathbf{f}}, \mathbf{f}_0) = \frac{\sum_{\mathbf{r} \in \mathcal{R}} |\hat{f}(\mathbf{r}) - f_0(\mathbf{r})|^\alpha}{\sum_{\mathbf{r} \in \mathcal{R}} |f_0(\mathbf{r})|^\alpha},$$

dont on peut tirer une mesure du rapport signal à bruit en dB :

$$\text{RSB}_{\Delta_\alpha} = 10 \times \log_{10} \left( \frac{1}{\Delta_\alpha(\hat{\mathbf{f}}, \mathbf{f}_0)} \right).$$

Pour des images constituées de régions ayant des contours francs, il semble plus adapté de s'intéresser à la mesure  $L_1 : \Delta_1(\hat{\mathbf{f}}, \mathbf{f}_0)$ . Plus l'image reconstruite  $\hat{\mathbf{f}}$  est proche de  $\mathbf{f}_0$ , plus le RSB est élevé.

- Un coefficient de corrélation normalisé 2-D permettant de mesurer la similarité entre deux fonctions discrètes 2-D :

$$\text{CCN}_{2\text{D}}(\hat{\mathbf{f}}, \mathbf{f}_0) = \frac{\sum_{\mathbf{r} \in \mathcal{R}} \left( (\hat{f}(\mathbf{r}) - \hat{f}_{\text{moy}})(f_0(\mathbf{r}) - f_{0 \text{ moy}}) \right)}{\sqrt{\sum_{\mathbf{r} \in \mathcal{R}} (\hat{f}(\mathbf{r}) - \hat{f}_{\text{moy}})^2 \sum_{\mathbf{r} \in \mathcal{R}} (f_0(\mathbf{r}) - f_{0 \text{ moy}})^2}} \quad (5.12)$$

où  $\hat{f}_{\text{moy}}$  et  $f_{0 \text{ moy}}$  sont les valeurs moyennes respectives des images  $\hat{\mathbf{f}}$  et  $\mathbf{f}_0$ . Plus ce coefficient est proche de 1, plus l'image est bien reconstruite (au sens qu'elle est plus proche de l'image initiale).

- Afin d'avoir un critère de qualité sur le résultat de segmentation obtenue, il est possible de comparer le pourcentage d'erreur dans l'estimation des labels,  $P_{\text{Err Z}}(\hat{\mathbf{z}}_K, \mathbf{z}_K)$ .

## 5.7 Évaluation des performances

L'évaluation a été réalisée à partir d'une image HR de  $256 \times 256$  pixels, codée en niveaux de gris, montrant une région rurale prise depuis un avion (voir FIG. 5.3 (a)). On peut distinguer des régions homogènes représentant des champs, des autoroutes qui se croisent, et une région urbanisée en bas à droite. Cette image  $\mathbf{f}_0$  qui servira de référence, a été dégradée à partir du modèle défini par l'équation (4.1), pour un facteur de décimation  $d = 4$ . Cette image HR peut également être segmentée en  $K$  régions à partir d'une méthode simple de seuils basés sur l'histogramme, pour donner un résultat de segmentation  $\mathbf{z}_K$  de référence qui sera utilisé dans le calcul de  $P_{\text{Err Z}}$ .

Le vecteur de mouvement inter-images  $\mathbf{s}_0$  est complètement déterminé.

L'image FIG. 5.3 (b) montre une des images ( $M = 16$ ) obtenues après dégradation avec un bruit défini par un RSB de 30 dB (cf. éq. (4.2)), et (c) est l'une des images obtenues avec un RSB égal à 15 dB. Ces images BR seront utilisées par la suite, afin d'obtenir des images HR  $\hat{\mathbf{f}}$  comparées à  $\mathbf{f}_0$ .

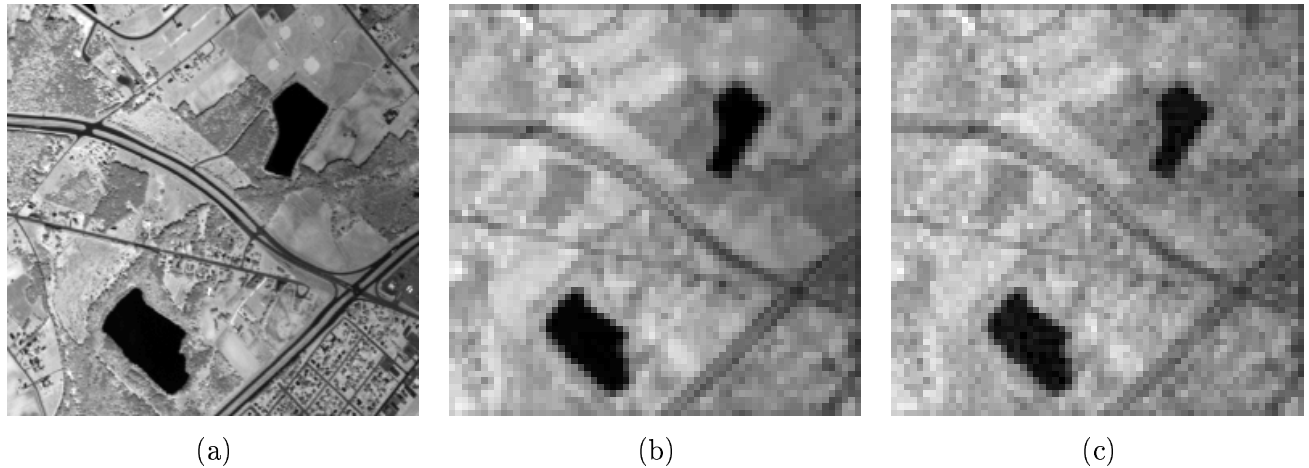


FIG. 5.3: Comparaison d'images HR et BR dans un exemple synthétique.

(a) image HR  $f_0$  de référence,  $256 \times 256$  pixels, (b) une image BR obtenue avec  $d = 4$  ( $64 \times 64$  pixels) et un RSB de 30 dB, (c) image BR obtenue avec  $d = 4$  et un RSB de 15 dB.

### 5.7.1 Effets du bruit

Cette première évaluation consiste à comparer les résultats de reconstruction obtenus, avec des images très peu bruitées (RSB de 30 dB) d'une part, et des images bruitées (RSB de 15 dB) d'autre part.

Cette étude est réalisée en utilisant les données de mouvement inter-images  $s_0$ , et la totalité des 16 images différentes disponibles. Elle a consisté à faire varier le nombre de régions homogènes  $K$  constituant l'image HR, afin d'en déterminer l'importance pour l'image reconstruite. Les résultats obtenus avec la méthode utilisant un modèle constant d'homogénéité sont présentés dans TAB. 5.1 ; ceux utilisant le modèle homogène bilinéaire dans TAB. 5.2.

On peut remarquer que quelle soit la méthode utilisée, quelle soit le nombre  $K$  de régions utilisées pour la segmentation, et à RSB fixé, les ordres de grandeurs de  $\Delta_{\text{moy}}$ ,  $\Delta_{\text{ET}}$ ,  $\Delta_1$  (dont les valeurs ne sont pas précisées ici) et  $\Delta_2$  sont sensiblement les mêmes.

En revanche, le critère  $\text{CCN}_{2\text{D}}$  permet de constater des nuances plus particulières en fonction du nombre  $K$  de régions.

Ainsi, à modèle et RSB fixés, on constate une diminution sensible de ce critère avec l'augmentation de  $K$ . Ceci signifie que l'image reconstruite est plus proche de la vérité « terrain » lorsque le nombre de régions utilisé pour la segmentation jointe est plus faible.

La diminution de qualité est à relativiser car elle est quasi-imperceptible pour l'œil humain. Elle est de 0,25 % entre les choix  $K = 1$  et  $K = 6$  pour un RSB de 30 dB, et de 0,99 % pour un RSB de 15 dB dans le cas du modèle constant. Elle est encore moins perceptible dans le cas du modèle bilinéaire, avec une baisse de 0,16 % pour un RSB de 30 dB et de 0,10 % pour un RSB de 15 dB. Les différences de qualité peuvent être observées à partir des images de la figure FIG. 5.4, (b) et (e) pour le modèle constant (variation de 0,22 %), (c) et (f) pour le modèle bilinéaire (variation de 0,11 %).

D'autre part, à  $K$  et RSB fixés, les critères numériques présentés dans les tableaux TAB. 5.1 et TAB. 5.2 montrent que le modèle bilinéaire donne de meilleurs résultats que le modèle constant. Ceci prouve que ce modèle d'homogénéité, plus complexe, traduit mieux la réalité que le modèle constant en

	RSB = 30 dB					RSB = 15 dB				
	CCN <sub>2D</sub>	$\Delta_{\text{moy}}$	$\Delta_{\text{ET}}$	$\Delta_2$	P <sub>Err Z</sub>	CCN <sub>2D</sub>	$\Delta_{\text{moy}}$	$\Delta_{\text{ET}}$	$\Delta_2$	P <sub>Err Z</sub>
		$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	%		$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	%
$\hat{\mathbf{f}}_{Tikhonov}$	0,8892	3,16	9,19	2,95	×	0,8878	4,46	9,29	3,31	×
$(\hat{\mathbf{f}}_1, \hat{\mathbf{z}}_1)$	0,9139	3,14	8,14	2,38	0	0,9120	4,48	8,34	2,80	0
$(\hat{\mathbf{f}}_2, \hat{\mathbf{z}}_2)$	0,9136	3,14	8,15	2,18	15,0	0,9118	4,48	8,36	2,81	17,6
$(\hat{\mathbf{f}}_3, \hat{\mathbf{z}}_3)$	0,9136	3,15	8,15	2,38	20,8	0,9066	4,38	8,59	2,90	23,1
$(\hat{\mathbf{f}}_4, \hat{\mathbf{z}}_4)$	0,9130	3,14	8,18	2,20	29,8	0,9099	4,47	8,44	2,85	34,0
$(\hat{\mathbf{f}}_5, \hat{\mathbf{z}}_5)$	0,9120	3,15	8,22	2,42	34,1	0,9105	4,48	8,41	2,84	34,3
$(\hat{\mathbf{f}}_6, \hat{\mathbf{z}}_6)$	0,9116	3,14	8,24	2,43	36,7	0,9030	4,31	8,73	2,96	39,8

TAB. 5.1: Évaluation des performances en milieu bruité. Modèle constant d'homogénéité.

	RSB = 30 dB					RSB = 15 dB				
	CCN <sub>2D</sub>	$\Delta_{\text{moy}}$	$\Delta_{\text{ET}}$	$\Delta_2$	P <sub>Err Z</sub>	CCN <sub>2D</sub>	$\Delta_{\text{moy}}$	$\Delta_{\text{ET}}$	$\Delta_2$	P <sub>Err Z</sub>
		$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	%		$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	%
$(\hat{\mathbf{f}}_1, \hat{\mathbf{z}}_1)$	0,9140	3,14	8,13	2,37	0	0,9116	4,47	8,36	2,80	0
$(\hat{\mathbf{f}}_2, \hat{\mathbf{z}}_2)$	0,9135	3,13	8,16	2,38	17,7	0,9116	4,47	8,36	2,80	18,7
$(\hat{\mathbf{f}}_3, \hat{\mathbf{z}}_3)$	0,9135	3,14	8,16	2,38	23,5	0,9117	4,47	8,36	2,80	23,7
$(\hat{\mathbf{f}}_4, \hat{\mathbf{z}}_4)$	0,9132	3,14	8,17	2,39	31,4	0,9113	4,47	8,37	2,81	32,6
$(\hat{\mathbf{f}}_5, \hat{\mathbf{z}}_5)$	0,9127	3,14	8,19	2,40	35,9	0,9111	4,47	8,38	2,28	36,4
$(\hat{\mathbf{f}}_6, \hat{\mathbf{z}}_6)$	0,9125	3,14	8,20	2,41	40,7	0,9107	4,47	8,40	2,83	42,8

TAB. 5.2: Évaluation des performances en milieu bruité. Modèle bilinéaire d'homogénéité.

ce qui concerne la reconstruction de l'image HR.

À modèle fixé, les estimateurs choisis montrent que les performances en milieu bruité (RSB = 15 dB) sont sensiblement inférieures à celles en milieu faiblement bruité. Ce résultat était attendu puisque la présence de bruit vient dégrader la qualité des données observées. Toutefois, il est intéressant de remarquer la robustesse des méthodes proposées en présence de bruit. Ainsi pour les deux cas distincts définis par un RSB de 30 et de 15 dB, le  $CCN_{2D}$  diminue de 0,21 % pour  $K = 1$  et 0,94 % pour  $K = 6$  avec le modèle constant. Et il diminue de 0,26 % pour  $K = 1$  et de 0,20 % pour  $K = 6$  avec le modèle bilinéaire. Ces différences peuvent être observées, avec  $K = 2$ , entre FIG. 5.4 (b) et FIG. 5.5 (b) pour le modèle constant, et entre FIG. 5.4 (c) et FIG. 5.5 (c) pour le modèle bilinéaire.

Les résultats  $\hat{\mathbf{f}}_{\text{Tikhonov}}$  obtenus avec une méthode classique de régularisation à partir d'un estimateur MAP basé sur l'équation (5.5), sont montrés pour comparaison sur FIG. 5.4 (d) et FIG. 5.5 (a). Les critères numériques du tableau TAB. 5.1 montrent que notre approche donne de meilleurs résultats.

Le résultat joint de classification en zones homogènes de la scène HR est également un résultat important. Les images obtenues en utilisant le modèle constant, dans le cas faiblement bruité, sont montrées sur la figure FIG. 5.6 pour  $K$  allant de 2 à 6.

Celles obtenues à partir du modèle constant avec un RSB de 15 dB, sont montrées sur la figure FIG. 5.7 pour  $K$  allant de 4 à 6 ; et celles obtenues en utilisant le modèle bilinéaire avec un RSB de 30 dB, sont montrées sur la figure FIG. 5.8 pour  $K$  allant de 4 à 6.

En comparant les résultats de segmentation pour  $K = \{4, 5, 6\}$ , un RSB de 30 dB, obtenus avec chacun des deux modèles (cf. FIG. 5.6 (d), (e), (f) et FIG. 5.8 (a), (b), (c)), on constate que les régions formées dans le cas bilinéaire sont plus homogènes et plus grandes que dans le cas constant. D'une certaine manière, ceci limite le niveau de détails de l'image segmentée. Par exemple, si on regarde le rendu des autoroutes qui se croisent, il est beaucoup moins remarquable dans le cas bilinéaire. Ce constat est renforcé par le critère  $P_{\text{Err Z}}$  dans les tableaux TAB. 5.1 et TAB. 5.2. Il s'agit de la mesure du pourcentage de labels erronés dans le résultat de segmentation, par rapport à la référence  $\mathbf{z}_K$ . À nombre de labels  $K$  et RSB fixés, dans tous les cas présentés sauf un,  $P_{\text{Err Z}}$  est plus faible pour le modèle constant que pour le modèle bilinéaire.

Lorsqu'on compare les résultats de segmentation dans les cas bruité et peu bruité, pour le modèle constant (cf. FIG. 5.6 (d), (e), (f) et FIG. 5.7 (a), (b), (c)), il est difficile de remarquer des différences particulières. Cependant,  $P_{\text{Err Z}}$  est logiquement supérieur lorsque le bruit est plus important dans les images observées. Cet effet est observé dans tous les cas.

Ainsi, les résultats de segmentation obtenus avec le modèle constant semblent plus riches d'information que ceux obtenus à partir du modèle bilinéaire. De plus, utiliser une valeur  $K$  importante permet d'obtenir plus d'informations sur les régions homogènes de la scène.

D'un point de vue calculatoire, notre méthode est **stochastique**. Ceci implique qu'une autre réalisation de notre algorithme peut donner un autre résultat, proche du précédent, en un nombre d'itérations différent. Avec  $H \times L$  la taille d'une image BR,  $M$  le nombre d'images BR utilisées,  $K$  le nombre de labels utilisés dans la segmentation jointe de  $\mathbf{f}_K$ , et  $k_0$  le facteur de SR : la complexité calculatoire de notre méthode est  $O(H L M k_0^2 K)$ .

Elle est de  $O(H L M k_0^2)$  pour la régularisation permettant d'obtenir  $\hat{\mathbf{f}}_{\text{Tikhonov}}$ .

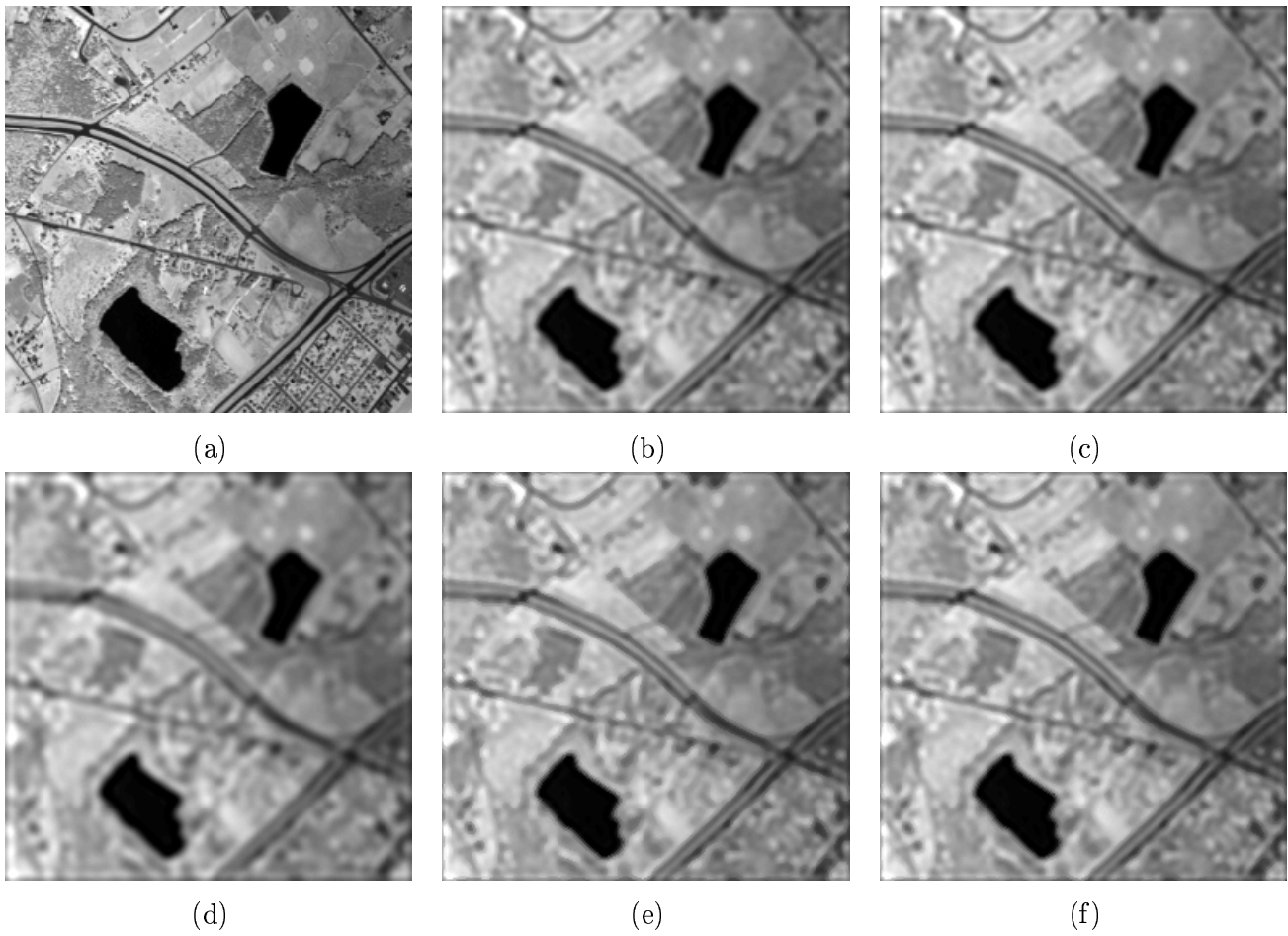


FIG. 5.4: Résultats de SR obtenus dans un cas faiblement bruité, RSB = 30 dB.  
 (a) image  $\mathbf{f}_0$  de référence (b)  $\hat{\mathbf{f}}_2, \text{const.}$  (c)  $\hat{\mathbf{f}}_2, \text{bil.}$  (d)  $\hat{\mathbf{f}}_{\text{Tikhonov}}$  (e)  $\hat{\mathbf{f}}_6, \text{const.}$  (f)  $\hat{\mathbf{f}}_6, \text{bil.}$

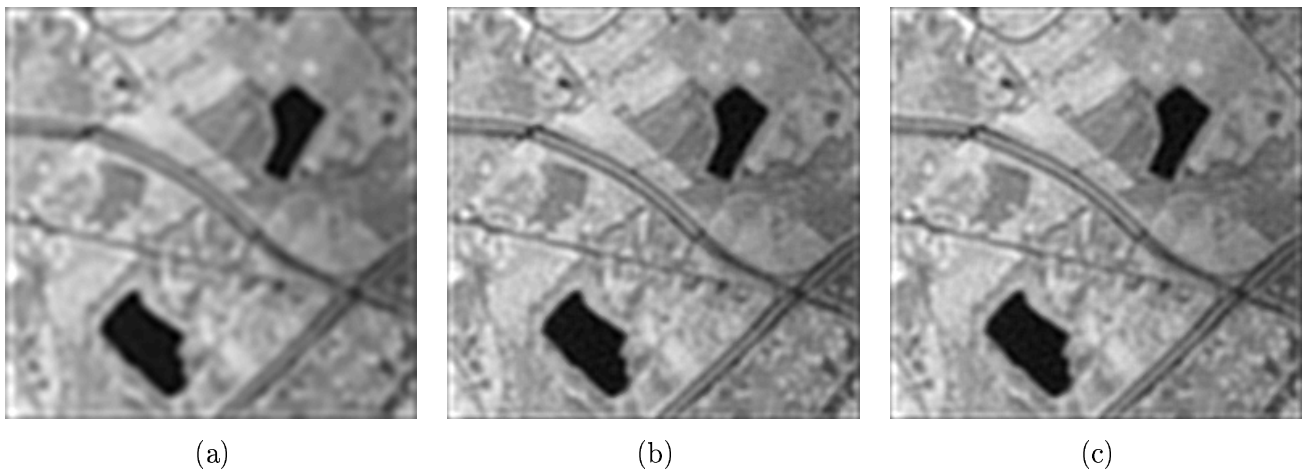


FIG. 5.5: Résultats de SR obtenus dans un cas bruité, RSB = 15 dB.  
 (a)  $\hat{\mathbf{f}}_{\text{Tikhonov}}$  (b)  $\hat{\mathbf{f}}_2, \text{const.}$  (c)  $\hat{\mathbf{f}}_2, \text{bil.}$

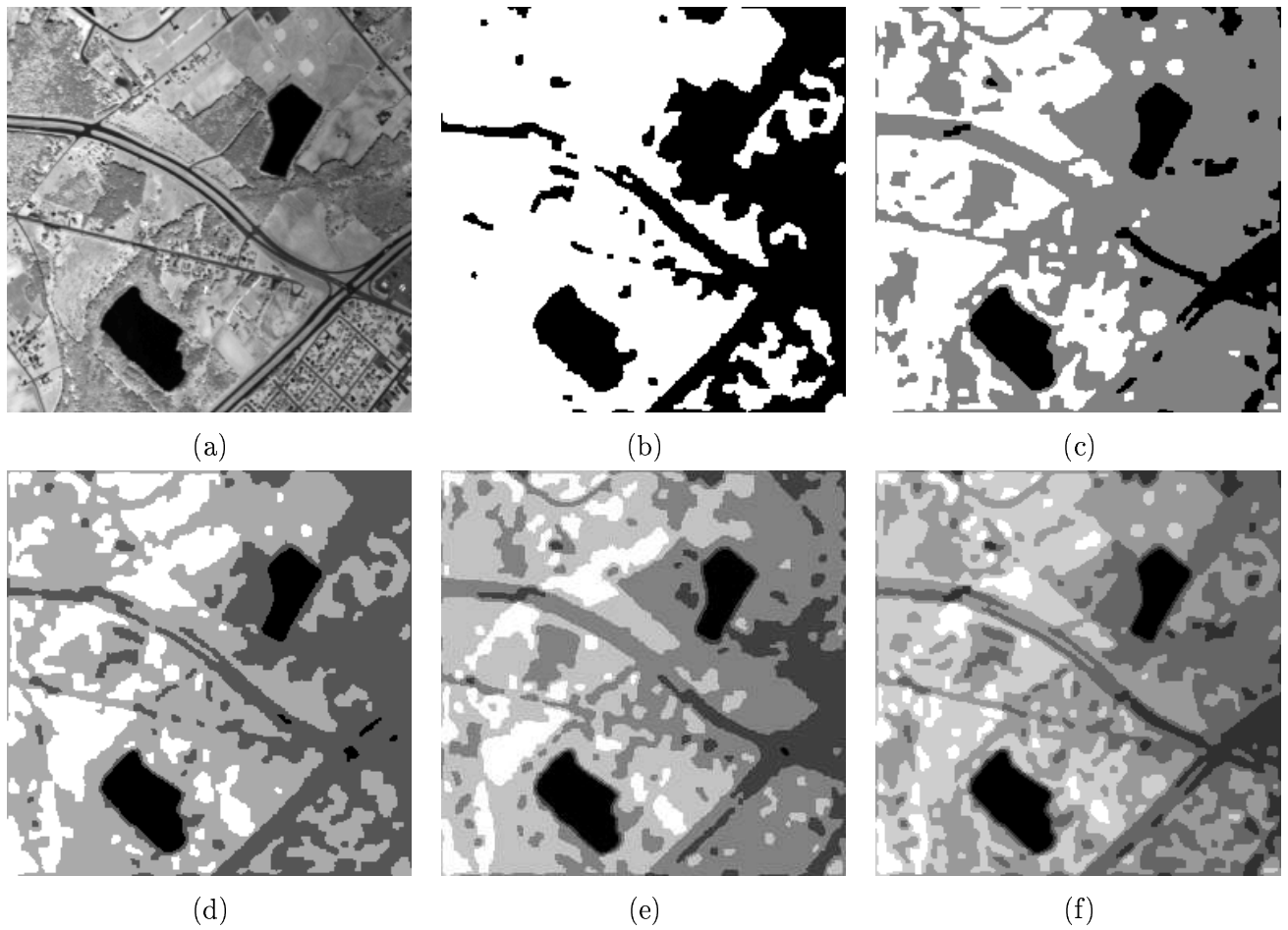


FIG. 5.6: Résultats de segmentation pour  $K$  allant de 2 à 6 avec modèle constant,  $\text{RSB} = 30$  dB.

(a) image  $\mathbf{f}_0$  de référence (b)  $\hat{\mathbf{z}}_{2, \text{const.}}$ . (c)  $\hat{\mathbf{z}}_{3, \text{const.}}$ . (d)  $\hat{\mathbf{z}}_{4, \text{const.}}$ . (e)  $\hat{\mathbf{z}}_{5, \text{const.}}$ . (f)  $\hat{\mathbf{z}}_{6, \text{const.}}$ .

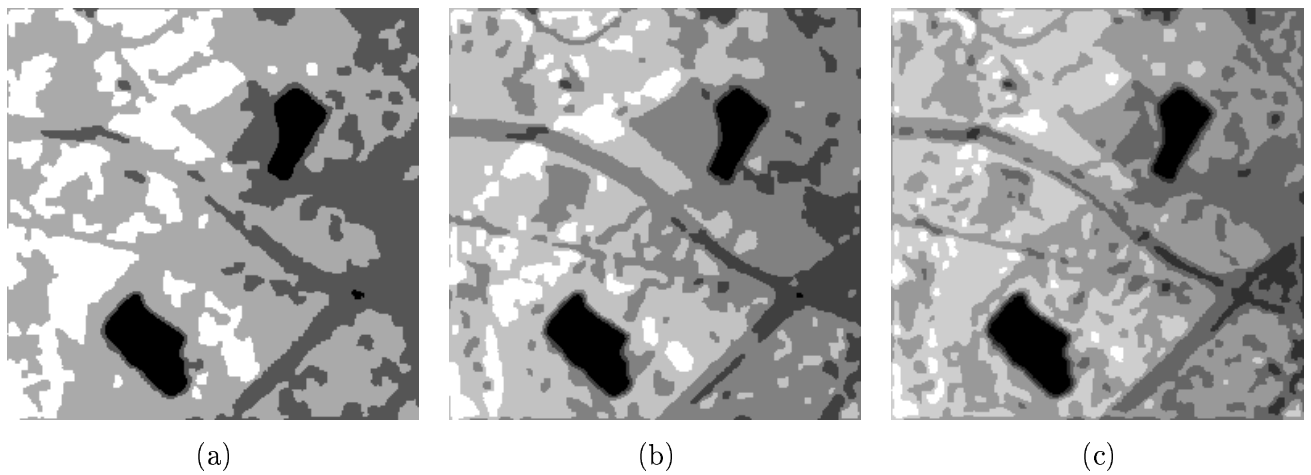


FIG. 5.7: Résultats de segmentation pour  $K$  allant de 4 à 6 avec modèle constant,  $\text{RSB} = 15$  dB.

(a)  $\hat{\mathbf{z}}_{4, \text{const.}}$ . (b)  $\hat{\mathbf{z}}_{5, \text{const.}}$ . (c)  $\hat{\mathbf{z}}_{6, \text{const.}}$ .

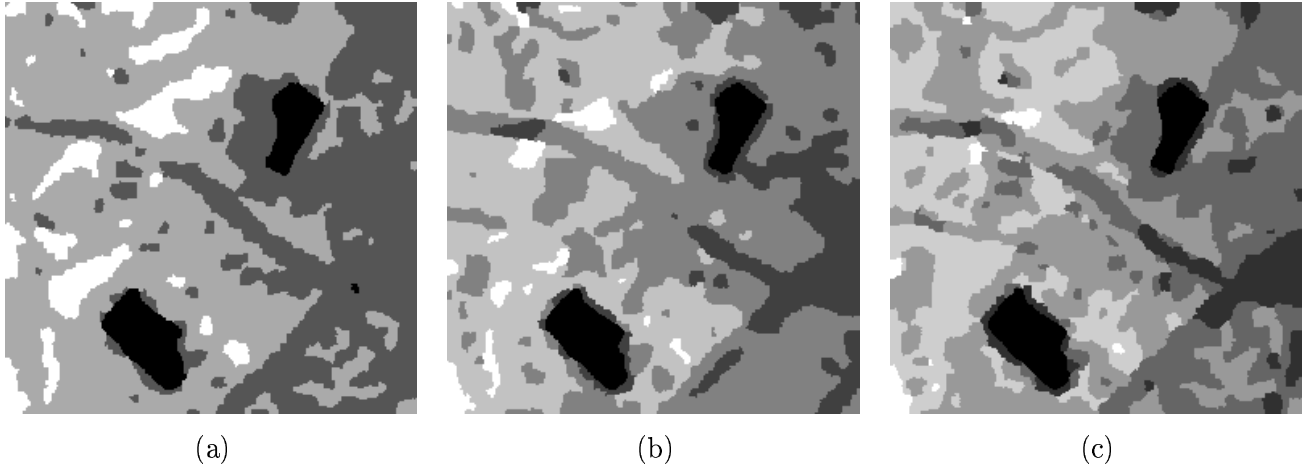


FIG. 5.8: Résultats de segmentation pour  $K$  allant de 4 à 6 avec modèle bilinéaire,  $\text{RSB} = 30$  dB.

(a)  $\hat{z}_4$ , bil. (b)  $\hat{z}_5$ , bil. (c)  $\hat{z}_6$ , bil.

L'aspect itératif de la méthode ne permet pas d'évaluer rigoureusement son coût calculatoire. Quelques éléments à ce sujet sont donnés dans [Humblot 05c]. Pour donner un ordre de grandeur, le cas présenté ici utilise  $(H, L, M, k_0, K) = (64, 64, 16, 4, K = \{1, \dots, 6\})$ .

La plateforme de calcul utilise un processeur Pentium IV à 2,6 GHz, un système d'exploitation Windows XP, et l'application Matlab.

La méthode utilisant le modèle constant a mis 40 et 72 secondes à s'exécuter respectivement pour  $K$  égales 2 et 6.

Celle utilisant le modèle bilinéaire a mis 113 et 769 secondes pour  $K$  égales 2 et 6.

### 5.7.2 Effets des erreurs dans les estimées du mouvement

Cette seconde évaluation a pour but d'observer les effets d'un **mauvais recalage** des images BR sur le traitement de SR. L'évaluation s'est portée sur la séquence faiblement bruitée ( $\text{RSB}$  de 30 dB), et sur 5 cas utilisant de manière différente les données de mouvement inter-images  $\mathbf{s}_0$ . Les 5 cas étudiés sont ceux où le pourcentage d'erreur sur l'estimation sous-pixelle de  $\mathbf{s}_0$  est de 0, 25, 50, 75 et 100 %. On utilise toujours 16 images différentes, avec des vecteurs de déplacements  $\hat{\mathbf{s}}_0$  comprenant un pourcentage donné d'erreur, et on se place dans le cas  $K = 2$ . Le cas  $K = 1$  n'a pas d'intérêt comparatif puisqu'il donne les mêmes résultats quel soit le modèle utilisé.

Le tableau TAB. 5.3 présente les résultats obtenus avec les méthodes utilisant les deux modèles.

À nouveau, on remarque qu'à pourcentage d'erreur identique sur l'estimation du mouvement  $\mathbf{s}_0$ ,  $P_{\text{Err Z}}$  est plus faible lorsque le modèle constant est utilisé. D'autre part, pour un pourcentage d'erreur inférieur ou égal à 50 %, le modèle bilinéaire donne de meilleurs résultats sur l'image HR reconstruite que le modèle constant. Au delà, les performances obtenues à partir des deux modèles sont quasi-identiques. Cette conclusion est illustrée par les images de la FIG. 5.9.

On peut y voir les résultats HR correspondant au modèle bilinéaire du tableau TAB. 5.3 sur (a), (c), (d), (e) et (f). Le  $\text{CCN}_{2\text{D}}$  a diminué de 2,88 % entre (a) et (e), et de 5,15 % entre (a) et (f). À titre de comparaison avec (c) qui utilise le modèle bilinéaire dans les mêmes conditions, l'image (b) montre le

Pourcentage d'erreur dans $\hat{\mathbf{s}}_0$	Modèle constant					Modèle bilinéaire				
	CCN <sub>2D</sub>	$\Delta_{\text{moy}}$	$\Delta_{\text{ET}}$	$\Delta_2$	P <sub>Err Z</sub>	CCN <sub>2D</sub>	$\Delta_{\text{moy}}$	$\Delta_{\text{ET}}$	$\Delta_2$	P <sub>Err Z</sub>
		$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	%		$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	%
$(\hat{\mathbf{f}}_2, \hat{\mathbf{z}}_2), 0$	0,9136	3,14	8,15	2,18	15,0	0,9135	3,13	8,16	2,38	17,7
$(\hat{\mathbf{f}}_2, \hat{\mathbf{z}}_2), 25$	0,8964	2,94	8,87	2,54	15,7	0,9033	3,15	8,59	2,41	19,9
$(\hat{\mathbf{f}}_2, \hat{\mathbf{z}}_2), 50$	0,8950	2,93	8,93	2,57	15,8	0,9016	3,16	8,66	2,45	20,7
$(\hat{\mathbf{f}}_2, \hat{\mathbf{z}}_2), 75$	0,8874	3,18	9,23	2,75	16,0	0,8872	3,17	9,24	2,75	21,0
$(\hat{\mathbf{f}}_2, \hat{\mathbf{z}}_2), 100$	0,8674	3,18	9,97	3,16	16,8	0,8665	3,18	10,0	3,18	21,3

TAB. 5.3: Évaluation des performances en fonction des données de recalage.

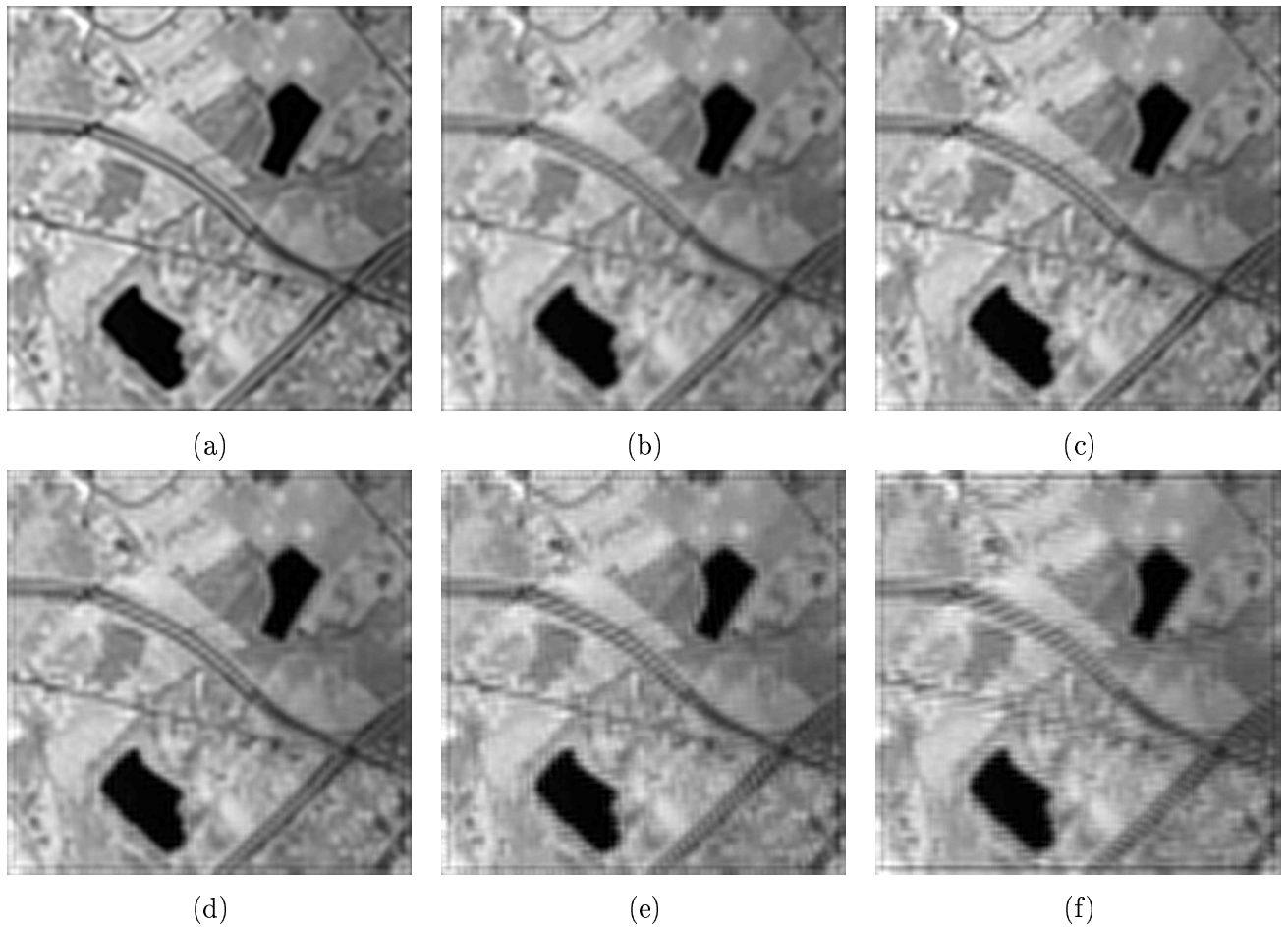


FIG. 5.9: Résultats de SR avec mauvaises données de recalage  $\hat{\mathbf{s}}_0$ , RSB = 30 dB.  
 (a)  $\mathbf{f}_2$ , bil., 0 % d'erreur entre  $\mathbf{s}_0$  et  $\hat{\mathbf{s}}_0$  (b)  $\hat{\mathbf{f}}_2$ , const., 25 % d'erreur (c)  $\hat{\mathbf{f}}_2$ , bil., 25 % d'erreur  
 (d)  $\hat{\mathbf{f}}_2$ , bil., 50 % d'erreur (e)  $\hat{\mathbf{f}}_2$ , bil., 75 % d'erreur (f)  $\hat{\mathbf{f}}_2$ , bil., 100 % d'erreur.



résultat obtenu en utilisant le modèle constant, et 25 % d'erreur dans l'estimation du mouvement.

Lorsque le mouvement entre les images BR n'est pas correctement estimé, des artéfacts peuvent apparaître dans les images reconstruites. En particulier, lorsqu'il y a plus de 50 % d'erreur dans l'estimation de  $\mathbf{s}_0$  (cf. FIG. 5.9 (e) et (f)), l'image reconstruite est considérablement dégradée en comparaison au cas parfait (a).

### 5.7.3 Importance du nombre d'images BR utilisées vs. répartition déplacements

Cette troisième évaluation a pour objectif de quantifier l'**importance du nombre  $M$  d'images BR** utilisées dans la **reconstruction**, ainsi que de leur **position relative** sur la grille HR.

À partir du modèle de mouvement choisi, il faut  $k_0^2$  images différentes (représentant tous les déplacements sous-pixelles possibles) pour reconstruire au mieux une image utilisant un facteur de SR  $k_0$ .

Cette étude comparative est faite pour différentes configurations avec  $M = \{16, 8, 4, 2\}$ . Comme précédemment on utilise  $k_0 = 4$ , la séquence définie par un RSB de 30 dB, et la connaissance de  $\mathbf{s}_0$  pour les déplacements inter-images.

La configuration des  $k_0^2$  déplacements sous-pixelles, nécessaires à une reconstruction idéale, peut être représentée par un quadrillage  $\mathcal{Q}$  de taille  $k_0 \times k_0$  (représentant les déplacements entiers inter-images  $(x, y) \in [0, k_0 - 1]^2$  sur la grille HR, par rapport à une image de référence  $\mathbf{g}_1$ , avec  $(x_1, y_1) = (0, 0)$ ).

Ainsi, dans TAB. 5.4, le cas  $M = 16$  consiste à connaître 16 images, toutes représentatives de déplacements sous-pixelles différents.

La configuration optimale (« opt. ») pour  $M = 8$  consiste à remplir  $\mathcal{Q}$  à la façon d'un échiquier, une case sur deux :  $(0, 0)$ ,  $(0, 2)$ ,  $(1, 1)$ ,  $(1, 3)$ ,  $(2, 0)$ , etc.

Une manière d'utiliser 8 images de façon « déséquilibrée » consiste, par exemple, à les choisir avec des déplacements correspondants au remplissage des deux premières lignes de  $\mathcal{Q}$  (« 2 lig. », défini par  $\{(x, y) : x = \{0, 1\}, \forall y\}$ ), ou les deux premières colonnes (« 2 col. »).

Pour  $M = 4$ , une manière optimale peut consister à remplir la première diagonale de  $\mathcal{Q}$  :  $(0, 0)$ ,  $(1, 1)$ ,  $(2, 2)$  et  $(3, 3)$ .

Une manière moins bien répartie, à remplir la première ligne de  $\mathcal{Q}$  (« 1 lig. », défini par  $\{(x, y) : x = 0, \forall y\}$ ).

Enfin pour  $M = 2$ , le cas « optimal » peut consister à remplir le premier et le troisième élément de la première diagonale :  $(0, 0)$  et  $(2, 2)$ .

Les résultats de SR obtenus avec certains des cas mentionnés dans TAB. 5.4 sont présentés dans FIG. 5.10.

À partir de cet exemple synthétique, on observe que **le nombre d'images utilisées dans le processus de SR est moins important que la position relative des images entre elles**. Une mauvaise répartition, selon les deux axes, des mouvements sous-pixelles entre les images, fait apparaître des artéfacts dans les images reconstruites. Dans FIG. 5.10 (b) avec  $M = 8$ , il s'agit de lignes verticales, et dans (c) avec  $M = 4$ , de lignes horizontales. Le cas de (f) dans lequel transparait un quadrillage est un peu à part puisque c'est un résultat obtenu à partir de 2 images BR seulement sur 16. On remarque néanmoins que ce cas, qui devrait être critique, donne des résultats satisfaisants en terme visuel et de CCN<sub>2D</sub>.

$\mathbf{K} = 2$ $(\hat{\mathbf{f}}_2, \hat{\mathbf{z}}_2)$	Modèle constant					Modèle bilinéaire				
	CCN <sub>2D</sub>	$\Delta_{\text{moy}}$ $\times 10^{-2}$	$\Delta_{\text{ET}}$ $\times 10^{-2}$	$\Delta_2$ $\times 10^{-2}$	P <sub>Err Z</sub> %	CCN <sub>2D</sub>	$\Delta_{\text{moy}}$ $\times 10^{-2}$	$\Delta_{\text{ET}}$ $\times 10^{-2}$	$\Delta_2$ $\times 10^{-2}$	P <sub>Err Z</sub> %
$M = 16$	0,9136	3,14	8,15	2,18	15,0	0,9135	3,13	8,16	2,38	17,7
$M = 8$ , opt.	0,9066	2,80	8,46	2,30	15,1	0,9117	2,93	8,24	2,20	20,2
$M = 8$ , 2 lig.	0,9015	3,64	8,71	2,50	15,7	0,8990	3,61	8,83	2,56	20,0
$M = 8$ , 2 col.	0,8995	2,69	8,78	2,45	15,1	0,9025	2,79	8,66	2,39	19,9
$M = 4$ , opt.	0,9013	2,48	8,68	2,39	15,6	0,9086	2,86	8,37	2,26	23,8
$M = 4$ , 1 lig.	0,8875	3,22	9,28	2,75	16,0	0,8838	3,33	9,47	2,85	24,2
$M = 2$ , opt.	0,8985	3,66	8,81	2,56	16,8	0,9028	3,98	8,66	2,52	26,1

TAB. 5.4: Évaluation des performances en fonction du nombre et de la disposition des images.

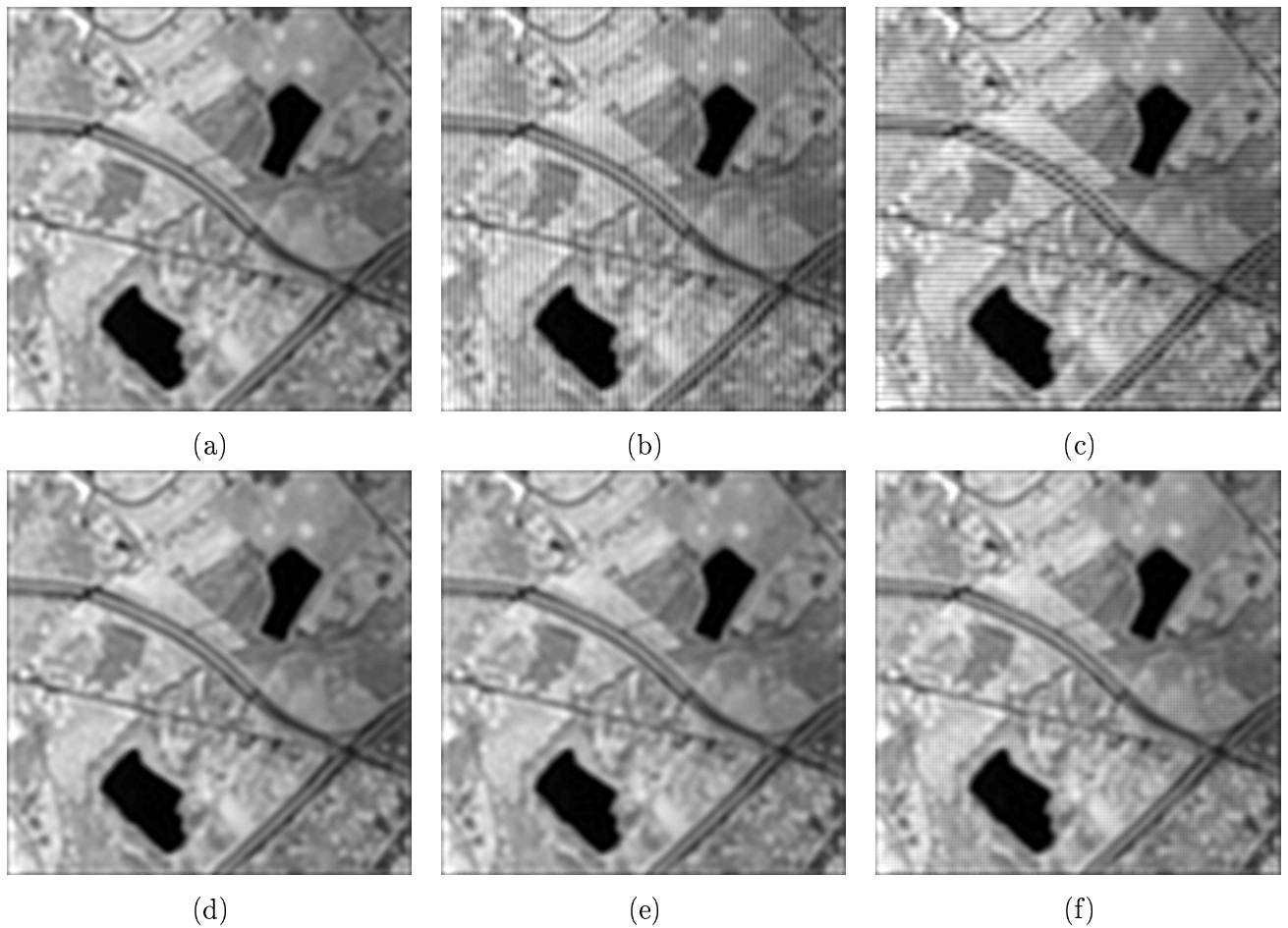


FIG. 5.10: Résultats de SR en fonction du nombre des images utilisées, RSB = 30 dB.  
 (a)  $\hat{\mathbf{f}}_2$ , bil.,  $M = 16$  (b)  $\hat{\mathbf{f}}_2$ , bil.,  $M = 8$ , 2 col. (c)  $\hat{\mathbf{f}}_2$ , bil.,  $M = 4$ , 1 lig. (d)  $\hat{\mathbf{f}}_2$ , bil.,  $M = 8$ , opt.  
 (e)  $\hat{\mathbf{f}}_2$ , bil.,  $M = 4$ , opt. (f)  $\hat{\mathbf{f}}_2$ , bil.,  $M = 2$ , opt.

On note également qu'une sélection « intelligente » des images BR, qui évite des répartitions de mouvement inter-images de type « ligne » FIG. 5.10 (c) ou « colonne » (b), mais qui propose des configurations dispersées, peut donner de très bons résultats de reconstruction. Et ceci, même dans des cas apparemment critiques tels que  $M = 2$  (f),  $M = 4$  (e) ou  $M = 8$  (d).

#### 5.7.4 Résultats obtenus à partir de données réelles

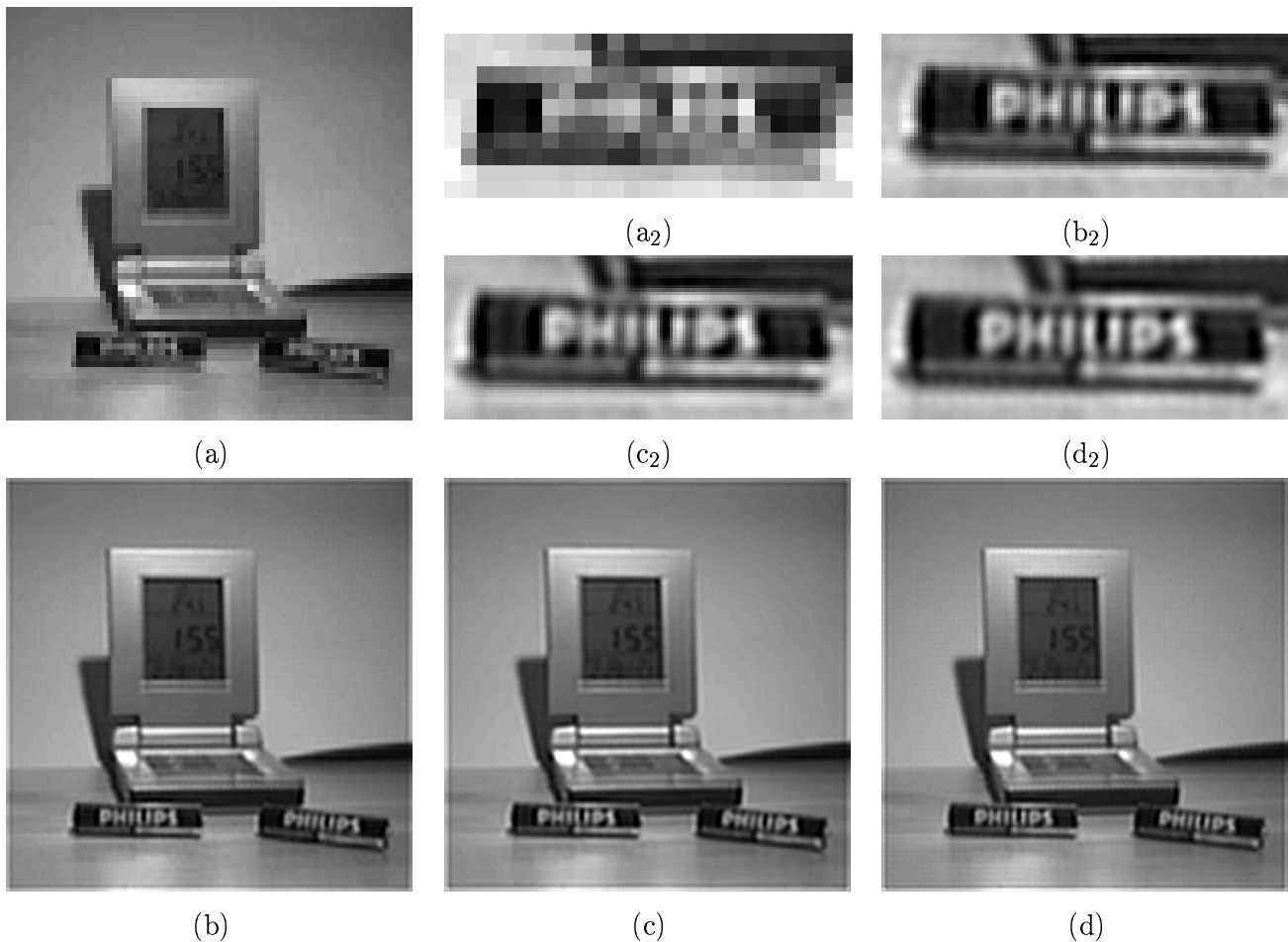


FIG. 5.11: Résultats de SR sur la séquence réelle « réveil » et agrandissement d'une région.

(a) une image BR de taille  $70 \times 69$  pixels (b)  $\hat{f}_1$ ,  $k_0 = 4$ ,  $M = 16$  (c)  $\hat{f}_1$ ,  $k_0 = 5$ ,  $M = 25$

(d)  $\hat{f}_1$ ,  $k_0 = 6$ ,  $M = 35$  (a<sub>2</sub>) agrandissement de taille  $10 \times 25$  pixels sur la pile « PHILIPS » à gauche dans l'image (a) (b<sub>2</sub>), (c<sub>2</sub>), (d<sub>2</sub>) agrandissements de la même région dans (b), (c) et (d).

Une séquence réelle montrant un réveil et deux piles « PHILIPS » a été obtenue à partir d'une webcam. Des résultats obtenus à partir de cette séquence baptisée « réveil » ont servi d'illustration dans [Humblot 05c, Humblot 05b]. Les images de FIG. 5.11 montrent un aspect qui n'a pas été abordé dans les articles précédemment cités. Il s'agit d'une comparaison des résultats de SR obtenus à partir de différents facteurs  $k_0$ .

Les images extraites de la séquence vidéo ont une taille de  $70 \times 69$  pixels. L'une d'entre elles peut

être observée sur (a). Les images (b), (c) et (d) montrent les résultats de SR obtenus respectivement pour  $k_0 = 4, 5$  et  $6$ . Leurs dimensions sont donc  $70 k_0 \times 69 k_0$  pixels. Toutefois, ces images de tailles réelles différentes sont montrées à la même échelle pour comparaison.

Dans chacun de ces trois cas, le nombre d'images utilisées a été de  $M = k_0^2$ , sauf pour  $k_0 = 6$ , où 35 images représentatives de 35 déplacements différents ont été utilisées au lieu de 36. Le nombre de déplacements détectés est directement lié au mouvement présent dans la séquence vidéo utilisée. Pour « réveil », le mouvement a été obtenu en remuant légèrement la caméra lors de la prise.

Le nombre idéal d'images (ou presque) a pu être utilisé dans ces différents cas. Il en résulte que la qualité des images semble augmenter avec la valeur de  $k_0$ . Cette impression est renforcée lorsqu'on regarde des agrandissements de la région où se situe la pile de gauche. Les images (a<sub>2</sub>), (b<sub>2</sub>), (c<sub>2</sub>) et (d<sub>2</sub>) de FIG. 5.11 sont des agrandissements de cette même région dans (a), (b), (c) et (d).

Lorsqu'on s'intéresse plus particulièrement à l'agrandissement (d<sub>2</sub>) tiré de l'image HR (d) obtenue avec  $k_0 = 6$ , ses détails sont plus précis que les agrandissements obtenus avec  $k_0 = \{4, 5\}$ .

### 5.7.5 Bilan de l'évaluation et interface Matlab de simulation

Un rappel synthétique des différents résultats obtenus au cours de cette évaluation est proposé dans la dernière partie de ce chapitre.

- Si l'on possède une image HR de référence, le **critère d'évaluation CCN<sub>2D</sub>** permet une comparaison qualitative pertinente des images HR obtenues. Il semble envisageable de définir un seuil permettant de rejeter de manière certaine un mauvais résultat de SR. Les expérimentations menées suggèrent que ce seuil se situe entre 0,89 et 0,90.
- En se basant sur ce critère, le **modèle bilinéaire permet d'obtenir des images HR sensiblement meilleures** que celles obtenues avec le modèle constant. Augmenter le nombre  $K$  de régions homogènes à utiliser dans la segmentation jointe diminue très légèrement la qualité de l'image reconstruite. Le choix  $K = 1$  donne de bons résultats sur l'image HR, tout en la privant d'un résultat de segmentation joint. Dans ce cas particulier les deux modèles sont identiques, et donc leurs résultats ne peuvent pas être comparés.
- Les résultats de segmentation sont d'autant plus riches que le nombre de classe  $K$  est important. Concernant la qualité de ce résultat, le **modèle constant donne de meilleurs résultats de classification** (plus proche de la réalité) que le modèle bilinéaire.
- Une certaine **robustesse** des méthodes présentées **en présence d'un bruit de 15 dB** a été montré.
- Une **dégradation importante des résultats** a été constatée **lorsqu'il y a plus de 50 % d'erreur dans l'estimation du mouvement inter-images**. Ainsi, l'importance de la **qualité du recalage** a été mise en évidence. Il est bon de rappeler qu'une méthode précise de recalage est primordiale à une bonne reconstruction. Cette limite pourrait être ajustée de manière plus précise

en affinant les simulations proposées.

- Enfin, le nombre d'images utilisées dans la reconstruction par SR ne constitue pas un élément critique de la qualité de l'image obtenue. Une répartition dispersée et bien distribuée des mouvements inter-images sur  $Q$  semble être d'une importance beaucoup plus grande que le nombre d'images utilisées pour la reconstruction.

Un système « intelligent » de sélection et de réduction des images BR pourrait être envisagé par la suite.

L'un ou l'autre des deux modèles est à privilégier en fonction de l'application visée.

L'utilisation du **modèle constant** permet à l'algorithme de s'exécuter **plus rapidement**, et il donne de **meilleurs résultats** de segmentation.

Le **modèle bilinéaire**, plus long à l'exécution de la méthode, permet de construire des **images HR de meilleure qualité**.

Une plate forme de test à partir du logiciel Matlab a été réalisée pendant cette thèse. L'image FIG. 5.12 illustre cette interface de simulation.

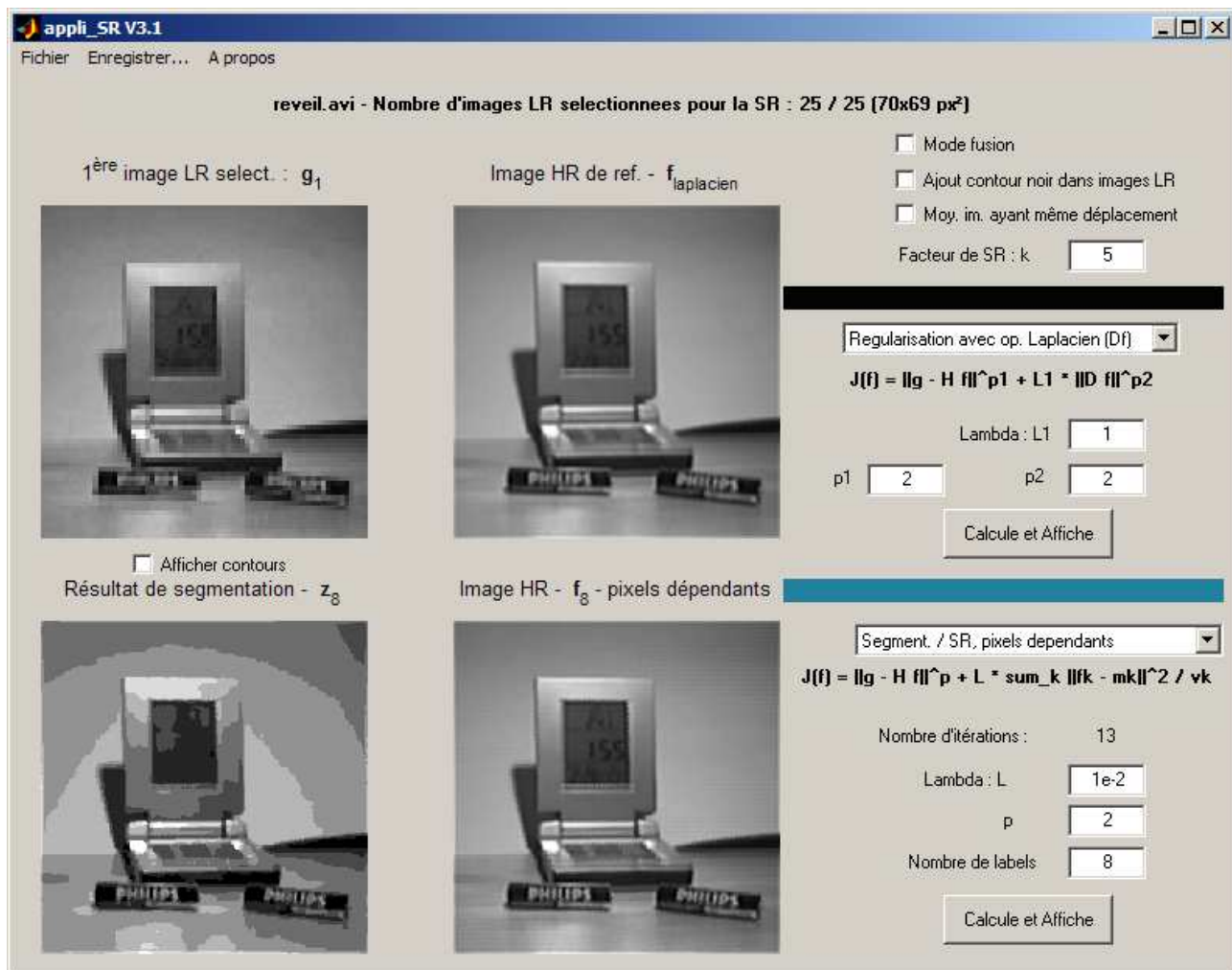


FIG. 5.12: Interface Matlab de l'outil de simulation de SR.

## Chapitre 6

# Conclusions et perspectives

La présentation du sujet de thèse et la définition de concepts et notations utilisés font l'objet du **Chapitre 1**. Le contexte applicatif est décrit dans la partie **1.1** ; pour le **recalage des images**, ce contexte a permis de limiter le champs des transformations possibles entre deux images au mouvement de translation global.

Dans le **Chapitre 2**, une méthode de **détection de petits objets** de la taille d'un pixel permet de définir un **critère d'évaluation**. Celui-ci utilise le **contraste** d'un pixel défini comme son niveau de gris moins la moyenne des niveaux de gris de ses quatre voisins les plus proches (dans le cas d'une modélisation de Markov d'ordre 1). Un pixel est détecté comme « cible » lorsque son contraste est d'au moins 55 en valeur absolue sur une échelle allant de 0 à 255. Ceci correspond à un contraste d'**au moins 21,5 % de la dynamique de l'image**. Ce **critère d'évaluation** pourra être ré-utilisé **pour comparer** ce résultat avec ceux obtenus avec **d'autres méthodes de détection**.

Deux méthodes utilisant le principe de corrélation de phase dans le domaine fréquentiel sont décrites dans le **Chapitre 3**. Elles permettent d'estimer précisément, et de manière sous-pixellique, un mouvement global de translation entre deux images. Une étude réalisée à partir d'une fonction analytique connue comme image de référence permet de conclure que la méthode utilisant le **gradient des images** donne de **meilleurs résultats** que celle qui utilise **directement les images**. Et dans un contexte non bruité, l'erreur commise sur le résultat est **inférieur à 2 % du déplacement réel**.

D'autre part, la méthode utilisant le **gradient** des images donne également de **bons résultats** sur des images comportant des **régions faiblement ou mal tracées**, avec des **contours peu discernables** et pour lesquelles l'autre méthode ne fonctionne pas. Ces traitements s'effectuent rapidement grâce à l'utilisation de la transformée de Fourier rapide (TFR).

L'**équation de recalage** dans le cas d'une approximation bi-linéaire est présentée dans ce chapitre (cf. (3.19), p. 45). Et une meilleure manière de gérer les **régions inconnues** dans un processus de recalage d'images est proposée : **cette région doit être prise à l'identique dans l'image de référence**.

Les deux méthodes présentées pourraient facilement s'appliquer dans un but d'**estimation des mouvements locaux** dans l'image, sur des petites sections d'images dont les dimensions sont des puissances de 2, de manière à être compatibles avec l'utilisation de la TFR.

Toutefois, ceci implique une étude non menée dans ce travail par manque de temps, et qui aborderait le problème du raccord des différents blocs de l'image entre eux. En particulier, lorsque de grandes diffé-

rences sont constatées dans l'estimation des différents mouvements des régions dans l'image. Cet aspect est une des **perspectives** de ce travail.

Deux **états de l'art** sont présentés :

- des **méthodes de recalage** d'images dans la section **3.2** du **Chapitre 3**,
- des **méthodes de super-résolution (SR)** dans le **Chapitre 4**.

Les premiers travaux qui ont appréhendé, formulé et proposé des solutions à ces deux problématiques y sont analysés succinctement. Pour la SR, des évolutions récentes développées par les différentes équipes mondiales spécialistes du domaine sont également présentées.

Enfin, une **nouvelle approche** de la SR est présentée dans le **Chapitre 5**. Il s'agit de la partie **originale** et **novatrice** de cette thèse. Elle s'inscrit dans le **cadre de l'estimation bayésienne**, et permet d'obtenir à la fois une **image HR**, et un **résultat de segmentation corrélé à l'image super-résolue**.

Dans cette approche, l'image HR est modélisée par un champ de Markov composite « **intensités-labels** ». Les variables cachées représentant les **labels** des régions sont modélisées par un **champ de Potts**. Les **intensités** des pixels dans chaque classe sont modélisées par un **champ de Gauss-Markov**. Bien que ce type de modélisation ait déjà été proposé et utilisé en segmentation d'images, leur utilisation dans les problèmes inverses est récente. En particulier, l'utilisation de cette modélisation en SR est nouvelle. Tout comme la mise en œuvre des algorithmes appropriés d'optimisation et d'échantillonnage, utilisés pour le calcul des estimateurs du problème inverse de SR.

L'image HR est calculée avec un critère d'optimisation utilisant des informations obtenues à partir du résultat de segmentation. Il s'agit d'estimer conjointement l'image HR, sa segmentation en un nombre fini de classes, puis de segments, ainsi que tous les hyper-paramètres du problème en utilisant un **schéma global hybride d'échantillonnage de Gibbs**.

Le **résultat de segmentation** peut s'avérer **très utile pour un post-traitement** de détection, ou d'extraction de caractéristiques géométriques et de leurs suivis. Cependant, cet aspect devrait être encore approfondi.

## Perspectives

Quelques pistes d'études futures sont évoquées ci-après :

- Utilisation de **méthodes de recalage considérant des mouvements plus généraux** tels que des transformations de type rotation et homothétie.
- **Généralisation** de cette méthode pour les **images en couleurs**.
- **Évaluation objective** des méthodes de SR reste à définir, même si les simulations réalisées ont montré que le **coefficient de corrélation croisée** est un **bon estimateur de la qualité visuelle lorsqu'existe une image de référence** représentant la scène HR (cf. (5.12), p. 128).  
En effet, l'estimation d'une méthode de SR peut être envisagée lorsqu'on utilise des séquences de synthèse obtenues à partir d'une image HR volontairement filtrée (afin de considérer le rôle intégrateur et l'optique du capteur), translatée, décimée et corrompue par un bruit. Il est alors possible de comparer les images HR reconstruites à l'image HR initiale.  
Néanmoins, il reste difficile de trouver une métrique précise permettant de caractériser une « **bonne** » image super-résolue **en l'absence d'une image de référence** (cf. 5.6). Pour l'instant les critères d'évaluation restent très **subjectifs**, essentiellement basés sur une **interprétation visuelle** des résultats.
- Extension et implémentation de la méthode de SR proposée pour la **création d'une séquence vidéo HR**.
- Le modèle de Potts utilisé pour la modélisation des labels peut être remplacé par des **modèles plus riches de segmentation** tels que la variation inhomogène du modèle de Potts, ou le modèle Chien [Descombes 99]. Cette évolution permettrait une meilleure adaptation de la méthode de SR aux scènes texturées.
- L'utilisation d'autres **modèles, qui lient les pixels de l'image HR à ceux de la variable cachée de segmentation**, sont à envisager. Il s'agit par exemple des modèles « couple » et « triplet ». L'article [Pieczynski 00] présente les champs couples. Ce modèle permet d'obtenir des lois *a posteriori* qui suivent des distributions markoviennes, lorsque la variable cachée n'est pas nécessairement de Markov, et il fait disparaître le problème de la constante de normalisation.  
Plus récemment, les champs triplets ont été proposés. Ils sont présentés de manière générale dans [Benboudjema 05b, Benboudjema 05a]. Les illustrations présentées montrent de bons résultats de segmentation obtenus à partir d'une image aérienne texturée.

La présente étude pourrait être prolongée pour déterminer si on peut obtenir des résultats équivalents sur des images obtenues à l'aide d'une caméra de « haute résolution », plutôt qu'avec une caméra de moins bonne résolution complétée par un traitement de SR. Il conviendrait de combiner le traitement de SR à celui de détection, sur deux séquences vidéo montrant une même scène à partir de deux caméras de résolutions différentes.





## Annexe A

### Calculs de lois marginales *a posteriori*

1.

$$\begin{aligned} p(\mathbf{z} | \mathbf{f}, \boldsymbol{\theta}, \mathbf{g}) &\propto p(\mathbf{g} | \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z}) \\ &\propto \mathcal{N}(\mathbf{g} | \mathbf{H}\mathbf{f}, \sigma_\epsilon^2 \mathbf{I}) p(\mathbf{z}) \end{aligned}$$

2.

$$\begin{aligned} p(\mathbf{z} | \boldsymbol{\theta}, \mathbf{g}) &\propto p(\mathbf{g} | \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z}) \\ &\propto \mathcal{N}(\mathbf{g} | \mathbf{H}\mathbf{m}_z, \mathbf{H}\boldsymbol{\Sigma}_z\mathbf{H}^t + \boldsymbol{\Sigma}_\epsilon) p(\mathbf{z}) \end{aligned}$$

$$3. p(\boldsymbol{\theta} | \mathbf{f}, \mathbf{z}, \mathbf{g}) = \prod_{i=1}^M p(\sigma_{\epsilon_i}^2 | \mathbf{f}, \mathbf{g}_i) \prod_{k=1}^K p(m_k | \sigma_k^2, \mathbf{f}, \mathbf{z}) p(\sigma_k^2 | m_k, \mathbf{f}, \mathbf{z})$$

(a)

$$\begin{aligned} p(m_k | \sigma_k^2, \mathbf{f}, \mathbf{z}) &\propto p(\mathbf{f} | m_k, \sigma_k^2, \mathbf{z}) p(m_k | \sigma_k^2, \mathbf{z}) \\ &\propto \mathcal{N}(\mathbf{f}_k | m_k \mathbf{1}, \sigma_k^2 \mathbf{I}) \times \mathcal{N}(m_k | m_{k0}, \sigma_{k0}^2) \sim \mathcal{N}(m_k | m_{k0}^{apost}, \sigma_{k0}^{2\,apost}) \\ &\propto \left( \frac{1}{2\pi \sigma_k^2} \right)^{\frac{n_k}{2}} \exp\left( -\frac{1}{2\sigma_k^2} \|\mathbf{f}_k - m_k \mathbf{1}\|^2 \right) \\ &\quad \times \frac{1}{\sqrt{2\pi \sigma_{k0}^2}} \exp\left( -\frac{1}{2\sigma_{k0}^2} |m_k - m_{k0}|^2 \right) \end{aligned}$$

$$\frac{\delta p(m_k | \sigma_k^2, \mathbf{f}, \mathbf{z})}{\delta m_k} \propto \left( \sum_{\mathbf{r} \in \mathcal{R}_k} \frac{1}{\sigma_k^2} (f(\mathbf{r}) - m(\mathbf{r})) - \frac{1}{\sigma_{k0}^2} (m_k - m_{k0}) \right) \exp(\dots)$$

$$\frac{\delta p(m_k | \sigma_k^2, \mathbf{f}, \mathbf{z})}{\delta m_k} = 0 \implies \frac{1}{\sigma_k^2} \sum_{\mathbf{r} \in \mathcal{R}_k} f(\mathbf{r}) + \frac{m_{k0}}{\sigma_{k0}^2} = m_k \left( \frac{n_k}{\sigma_k^2} + \frac{1}{\sigma_{k0}^2} \right)$$

$$\sigma_{k0}^{2\,apost} = \left( \frac{n_k}{\sigma_k^2} + \frac{1}{\sigma_{k0}^2} \right)^{-1} \quad \text{et} \quad m_{k0}^{apost} = \sigma_{k0}^{2\,apost} \left( \frac{1}{\sigma_k^2} \sum_{\mathbf{r} \in \mathcal{R}_k} f(\mathbf{r}) + \frac{m_{k0}}{\sigma_{k0}^2} \right)$$

(b)

$$\begin{aligned}
p(\sigma_k^2 | m_k, \mathbf{f}, \mathbf{z}) &\propto p(\mathbf{f} | \sigma_k^2, m_k, \mathbf{z}) p(\sigma_k^2 | m_k, \mathbf{z}) \\
&\propto \mathcal{N}(\mathbf{f}_k | m_k \mathbf{1}, \sigma_k^2 \mathbf{I}) \times \mathcal{IG}(\sigma_k^2 | \alpha_{k0}, \beta_{k0}) \sim \mathcal{IG}(\sigma_k^{2\text{apost}} | \alpha_{k0}^{\text{apost}}, \beta_{k0}^{\text{apost}}) \\
&\propto \left( \frac{1}{2\pi \sigma_k^2} \right)^{\frac{n_k}{2}} \exp\left( -\frac{1}{2\sigma_k^2} \|\mathbf{f}_k - m_k \mathbf{1}\|^2 \right) \left( \frac{1}{\sigma_k^2} \right)^{\alpha_{k0}-1} \exp\left( -\beta_{k0} \frac{1}{\sigma_k^2} \right) \\
&\propto \left( \frac{1}{\sigma_k^2} \right)^{\alpha_{k0} + \frac{n_k}{2} - 1} \exp\left( -\frac{1}{\sigma_k^2} \left( \beta_{k0} + \frac{\|\mathbf{f}_k - m_k \mathbf{1}\|^2}{2} \right) \right) \\
\alpha_{k0}^{\text{apost}} &= \alpha_{k0} + \frac{n_k}{2} \quad \text{et} \quad \beta_{k0}^{\text{apost}} = \beta_{k0} + \frac{1}{2} \sum_{\mathbf{r} \in \mathcal{R}_k} |f(\mathbf{r}) - m_k(\mathbf{r})|^2
\end{aligned}$$

(c)

$$\begin{aligned}
p(\sigma_{\epsilon_i}^2 | \mathbf{f}, \mathbf{g}_i) &\propto p(\mathbf{g}_i | \sigma_{\epsilon_i}^2, \mathbf{f}) p(\sigma_{\epsilon_i}^2 | \mathbf{f}) \\
&\propto \mathcal{N}(\mathbf{g}_i | \mathbf{H}_i \mathbf{f}, \sigma_{\epsilon_i}^2 \mathbf{I}) \times \mathcal{IG}(\sigma_{\epsilon_i}^2 | \alpha_0^{\epsilon_i}, \beta_0^{\epsilon_i}) \sim \mathcal{IG}(\sigma_{\epsilon_i}^2 | \alpha_0^{\epsilon_i \text{apost}}, \beta_0^{\epsilon_i \text{apost}}) \\
&\propto \left( \frac{1}{2\pi \sigma_{\epsilon_i}^2} \right)^{\frac{R}{2}} \exp\left( -\frac{1}{2\sigma_{\epsilon_i}^2} \|\mathbf{g}_i - \mathbf{H}_i \mathbf{f}\|^2 \right) \left( \frac{1}{\sigma_{\epsilon_i}^2} \right)^{\alpha_0^{\epsilon_i}-1} \exp\left( -\beta_0^{\epsilon_i} \frac{1}{\sigma_{\epsilon_i}^2} \right) \\
&\propto \left( \frac{1}{\sigma_{\epsilon_i}^2} \right)^{\alpha_0^{\epsilon_i} + \frac{R}{2} - 1} \exp\left( -\frac{1}{\sigma_{\epsilon_i}^2} \left( \beta_0^{\epsilon_i} + \frac{\|\mathbf{g}_i - \mathbf{H}_i \mathbf{f}\|^2}{2} \right) \right) \\
\alpha_0^{\epsilon_i \text{apost}} &= \alpha_0^{\epsilon_i} + \frac{R}{2} \quad \text{et} \quad \beta_0^{\epsilon_i \text{apost}} = \beta_0^{\epsilon_i} + \frac{1}{2} \sum_{\mathbf{r} \in \mathcal{R}} |g_i(\mathbf{r}) - [\mathcal{H}_i f](\mathbf{r})|^2
\end{aligned}$$

## Annexe B

# Acronymes et glossaire

<i>aliasing</i>	recouvrement du spectre fréquentiel d'un signal.
bit	<i>binary digit</i> , nombre binaire. Par définition vaut 0 ou 1.
BR	basse résolution, LR en anglais.
CCD	<i>charge-coupled device</i> , composant à couplage de charge, se dit d'une matrice 2-D de capteurs carrés sensibles à la lumière. Type de capteurs numériques présents dans les appareil d'acquisition d'images ou de vidéo.
CCN <sub>2D</sub>	coefficient de corrélation normalisé 2-D, permet de mesurer la similarité entre deux fonctions discrètes 2-D. Peut se calculer avec l'équation (5.12), p. 128.
CP	corrélation de phase.
DRI	détection, reconnaissance et identification (familles d'applications militaires).
FBP	<i>filtered back-projection</i> , méthode algébrique de rétro-projection filtrée.
FMM	<i>finite mixture model</i> , modèle de mélange fini.
FTO	fonction de transfert optique, elle décrit « l'image d'un objet ponctuel isolé situé sur un fond uniformément noir ».
i.i.d.	indépendant et identiquement distribué, se dit d'une v.a.
HD	haute définition.
HMM	<i>hidden Markov model</i> , modèle de Markov caché.
HMRF	Huber MRF.
HR	haute résolution.
IBP	<i>iterative back-projection</i> , méthode itératives de rétro-projection.
IR	infrarouge.
LMMSE	<i>linear minimum mean square error</i> , erreur quadratique moyenne du minimum linéaire.
LR	<i>low resolution</i> , BR en français.
MAP	se dit d'un estimateur au sens du maximum <i>a posteriori</i> de probabilité.

MC	moindres carrés.
MCMC	<i>Markov chain Monte-Carlo</i> , se dit d'un échantillonneur de Gibbs par chaîne de Markov de Monte-Carlo.
MCR	moindres carrés récursifs.
ML	<i>maximum likelihood</i> , MV en français.
MRF	<i>Markov random field</i> , champs de Markov aléatoires.
MV	maximum de vraisemblance, ML en anglais.
NTSC	<i>national television systems committee</i> , système de codage sur 525 lignes, mais $480 \times 720$ pixels sont affichés car le reste est utilisé pour la synchronisation. Affiche 59,94 trames par seconde ou 29,97 trames entrelacées.
PAL	<i>phase-alternating line</i> qui signifie changement de phase à chaque ligne, mode de codage des vidéos sur 625 lignes, mais $576 \times 720$ pixels seulement sont affichés car 8 % des lignes servent à la synchronisation. Affiche 25 images par seconde ou 50 champs entrelacés par seconde.
pel, pixel	<i>picture element</i> , élément d'image. Point représentant le plus petit élément de teinte homogène d'une image pouvant être transmis sous forme numérique.
PMM	<i>Potts Markov model</i> , modèle de Potts Markov.
POCS	<i>projection onto convex sets</i> , méthodes de projection sur des ensembles convexes.
PSF	<i>point spread function</i> , fonction d'étalement de point (cf. FTO).
RSB	rapport signal à bruit. Peut se calculer avec l'équation (4.2), p. 65.
RTLS	<i>recursive TLS</i> , méthode totale des MCR.
SAR	<i>synthetic aperture radar</i> , radar à ouverture synthétique.
SD	<i>standard definition</i> , définition standard.
SPC	spectre de puissance croisée, <i>cross-power spectrum</i> en anglais.
SR	super-résolution, <i>super-resolution</i> ou <i>superresolution</i> en anglais.
TF	transformée de Fourier.
TFC	transformée de Fourier continue.
TFD	transformée de Fourier discrète.
TFI	transformée de Fourier inverse.
TFR	transformée de Fourier rapide.
TLS	<i>total least squares</i> , méthode totale des MC.
TNT	télévision numérique terrestre.
v.a.	variable aléatoire.

# Bibliographie

- [Aggarwal 88] J. Aggarwal & N. Nandhakumar. *On the computation of motion from sequences of images – A review*. Proc. of the IEEE, vol. 76, no. 8, pages 917–935, août 1988.
- [Aizawa 91] K. Aizawa, T. Komatsu & T. Saito. *Acquisition of very high resolution images using stereo cameras*. Proc. SPIE Visual Communications and Image Processing, vol. 1605, pages 318–328, 1991.
- [Aizawa 92] K. Aizawa, T. Komatsu & T. Saito. *A scheme for acquiring very high resolution images using multiples cameras*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), volume 3, pages 289–292, mars 1992.
- [Alam 00] M. Alam, J. Bogner, R. Hardie & B. Yasuda. *Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames*. IEEE Transactions on Instrumentation and Measurement, vol. 49, no. 5, pages 915–923, oct. 2000.
- [Alvarez 04] L.D. Alvarez, J. Mateos, R. Molina & A.K. Katsaggelos. *High resolution images from compressed low resolution video : Motion estimation and observable pixels*. Inter. Journal of Imaging Systems and Technology, vol. 14, no. 2, pages 58–66, août 2004.
- [Anandan 87] P. Anandan. *A unified perspective on computational techniques for the measurement of visual motion*. In Proc. IEEE Conf. on Computer Vision, pages 219–230, 1987.
- [Argyriou 03] V. Argyriou & T. Vlachos. *Sub-pixel motion estimation using gradient cross-correlation*. In The 7<sup>th</sup> International Symposium on Signal Processing and its Applications (ISSPA), Paris, juillet 2003.
- [Baker 02] S. Baker & T. Kanade. *Limits on super-resolution and how to break them*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, pages 1167–1183, sept. 2002.
- [Barron 94] J. Barron, D. Fleet & S. Beauchemin. *Performance of optical flow techniques*. Inter. Journal of Computer Vision, vol. 12, no. 1, pages 43–77, 1994.
- [Bazaraa 93] M. Bazaraa, H. Sherali & C. Shetty. *Nonlinear programming : Theory and algorithms*. John Wiley & Sons, 2nd edition, 1993.
- [Benboudjema 05a] D. Benboudjema & W. Pieczynski. *Segmenting non stationary images with triplet Markov fields*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 1, pages 317–320, sept. 2005.

- [Benboudjema 05b] D. Benboudjema & W. Pieczynski. *Unsupervised image segmentation using triplet Markov fields*. Computer Vision and Image Understanding, vol. 99, no. 3, pages 476–498, 2005.
- [Bertero 98] M. Bertero & P. Boccacci. Introduction to inverse problems in imaging. Institute of Physics Publishing, 1998.
- [Bertsekas 95] D. Bertsekas. Nonlinear programming. Athena Scientific, Belmont, Massachusetts, 1995.
- [Besag 74] J. Besag. *Spatial interaction and the statistical analysis of lattice systems*. Journal of Royal Statistical Society, vol. 2, pages 192–236, 1974.
- [Bierling 88] M. Bierling. *Displacement estimation by hierarchical block matching*. Proc. SPIE Visual Communications and Image Processing, vol. 1001, pages 942–951, nov. 1988.
- [Borman 04] S. Borman. *Topics in Multiframe Superresolution Restoration*. Thèse, University of Notre Dame, Indiana, États-Unis, avril 2004.
- [Bose 93a] N. Bose, H. Kim & H. Valenzuela. *Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), volume 5, pages 269–272, avril 1993.
- [Bose 93b] N. Bose, H. Kim & H. Valenzuela. *Recursive total least squares algorithm for image reconstruction from noisy, undersampled multiframes*. Multidimensional Systems and Signal Processing, vol. 4, no. 3, pages 253–268, juill. 1993.
- [Bose 94] N. Bose, H. Kim & B. Zhou. *Performance analysis of the TLS algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 3, pages 571–575, nov. 1994.
- [Bouman 93] C. Bouman & K. Sauer. *A generalized Gaussian image model for edge-preserving MAP estimation*. IEEE Transactions on Image Processing, vol. 2, no. 3, pages 296–310, juill. 1993.
- [Brown 81] J. Brown. *Multi-channel sampling of low-pass signals*. IEEE Transactions on Circuits and Systems, vol. 28, no. 2, pages 101–106, 1981.
- [Brown 92] L. Brown. *A survey of image registration techniques*. ACM Computing Survey, vol. 24, no. 4, pages 325–376, déc. 1992.
- [Cafforio 83] C. Cafforio & F. Rocca. *The differential method for motion estimation*. In Image Sequence Processing and Dynamic Scene Analysis, pages 104–124. Springer-Verlag, T. S. Huang edition, 1983.
- [Cao 01] F. Cao & B. Collin. *Region tracking using level set registration*. In SPIE Aerosense, Acquisition, Tracking, and Pointing XV, avril 2001.
- [Capel 00] D. Capel & A. Zisserman. *Super-resolution enhancement of text image sequences*. In Proc. of the 15<sup>th</sup> Inter. Conf. on Pattern Recognition, volume 1, pages 600–605, sept. 2000.

- [Carlotto 82] M. Carlotto & V. Tom. *Interpolation of two-dimensional surfaces using the Gerchberg algorithm*. SPIE, vol. 359, pages 226–232, 1982.
- [Champagnat 96] F. Champagnat, Y. Goussard & J. Idier. *Unsupervised deconvolution of sparse spike trains using stochastic approximation*. IEEE Transactions on Signal Processing, vol. 44, no. 12, pages 2988–2998, déc. 1996.
- [Cheeseman 94] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz & R. Hanson. *Super-resolved surface reconstruction from multiple images*. Rapport technique FIA-94-12, NASA Ames Research Center, déc. 1994.
- [Chen 94] Q. Chen, M. Defrise & F. Deconinck. *Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 12, pages 1156–1168, déc. 1994.
- [Chen 98] D. Chen & R. Schultz. *Extraction of high-resolution video stills from MPEG image sequences*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 2, pages 465–469, oct. 1998.
- [Clark 85] J. Clark, M. Palmer & P. Lawrence. *A transformation method for the reconstruction of functions from nonuniformly spaced samples*. IEEE Transactions on Acoustics Speech and Signal Processing, vol. 33, no. 4, pages 1151–1165, oct. 1985.
- [Combettes 93] P. Combettes. *The foundations of set theoretic estimation*. Proc. of the IEEE, vol. 82, no. 2, pages 182–208, fév. 1993.
- [Corral 02] E. Corral, G. Thomas & B. Flores. *Effects of using superresolution techniques in ISAR imagery*. In Proc. of the IEEE Canadian Conf. on Electrical and Computer Engineering, volume 1, pages 353–357, mai 2002.
- [Cortijo 03] F. Cortijo, S. Villena, R. Molina & A. Katsaggelos. *Bayesian super-resolution of text image sequences from low-resolution observations*. In Proc. of 7<sup>th</sup> Inter. Symp. on Signal Proc. and Its Appli., volume 1, pages 421–424, juill. 2003.
- [Davila 91] C. Davila. *Recursive total least squares algorithms for adaptive filtering*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 1853–1856, fév. 1991.
- [Davila 94] C. Davila. *Efficient recursive total least squares algorithms for FIR adaptive filtering*. IEEE Transactions on Signal Processing, vol. 42, no. 2, pages 268–280, fév. 1994.
- [Dekeyser 00] F. Dekeyser, P. Bouthemy, P. Pérez & E. Payot. *Super-resolution from noisy image sequences exploiting a 2D parametric motion model*. In Proc. IEEE Inter. Conf. on Pattern Recognition (ICPR), volume 3, pages 350–353, sept. 2000.
- [Dekeyser 01] F. Dekeyser. *Restauration de séquence d'image par des approches spatio-temporelles : filtrage et super-résolution par le mouvement*. Thèse, VISTA (IRISA), Université de Rennes 1, France, nov. 2001.
- [Dempster 77] A. Dempster, N. Laird & D. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society : Series B, vol. 39, no. 1, pages 1–38, nov. 1977.



- [Descombes 99] X. Descombes, R. Morris, J. Zerubia & M. Berthod. *Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood*. IEEE Transactions on Image Processing, vol. 8, no. 7, pages 954–963, juill. 1999.
- [Dublanche 97] F. Dublanche, J. Idier & P. Duvaut. *Direction-of-arrival and frequency estimation using Poisson-Gaussian modeling*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 3501–3504, avril 1997.
- [Dufaux 94] F. Dufaux. *Multigrid block matching motion estimation for generic video coding*. Thèse, Ecole Polytechnique Fédérale de Lausanne, Suisse, 1994.
- [Dufaux 95] F. Dufaux & F. Moscheni. *Motion estimation techniques for digital TV : A review and a new contribution*. Proceedings of the IEEE, vol. 83, no. 6, pages 858–876, juin 1995.
- [Elad 95] M. Elad & A. Feuer. *Super-resolution restoration of continuous image sequence using the LMS algorithm*. In Proc. of the 18<sup>th</sup> IEEE Convention in Israel, mars 1995.
- [Elad 96] M. Elad & A. Feuer. *Super-resolution reconstruction of an image*. In Proc. of the 19<sup>th</sup> IEEE Convention in Israel, pages 391–394, nov. 1996.
- [Elad 97] M. Elad & A. Feuer. *Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images*. IEEE Transactions on Image Processing, vol. 6, no. 12, pages 1646–1658, déc. 1997.
- [Elad 99] M. Elad & A. Feuer. *Superresolution restoration of an image sequence : Adaptive filtering approach*. IEEE Transactions on Image Processing, vol. 8, no. 3, pages 387–395, mars 1999.
- [Enkelmann 88] W. Enkelmann. *Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences*. Computer Graphics and Image Processing, vol. 43, pages 150–177, août 1988.
- [Erdem 92] A. Erdem, M. Sezan & M. Ozkan. *Motion-compensated multiframe Wiener restoration of blurred and noisy image sequences*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), volume 3, pages 293–296, mars 1992.
- [Eren 97] P. Eren, M. Sezan & A. Tekalp. *Robust, object-based high-resolution image reconstruction from low-resolution video*. IEEE Transactions on Image Processing, vol. 6, no. 10, pages 1446–1451, oct. 1997.
- [Farsiu 04] S. Farsiu, D. Robinson, M. Elad & P. Milanfar. *Fast and robust multiframe super resolution*. IEEE Transactions on Image Processing, vol. 13, no. 10, pages 1327–1344, oct. 2004.
- [Fleet 90] D. Fleet & A. Jepson. *Computation of component image velocity from local phase information*. Inter. Journal of Computer Vision, vol. 5, pages 77–104, 1990.
- [Fleet 91] D. Fleet & A. Jepson. *Stability of phase information*. In Proceedings of the IEEE, pages 52–60, oct. 1991.

- [Frooosh 02] H. Frooosh, J. Zerubia & M. Berthod. *Extension of phase correlation to subpixel registration*. IEEE Transactions on Image Processing, vol. 11, no. 3, pages 188–200, 2002.
- [Franke 82a] R. Franke. *Scattered data interpolation : Tests of some methods*. Mathematics of Computation, vol. 38, no. 157, pages 181–200, jan. 1982.
- [Franke 82b] R. Franke. *Smooth interpolation of scattered data by local thin-plate splines*. Computers and Mathematics with Applications, vol. 8, no. 4, pages 273–281, 1982.
- [Frieden 87] B. Frieden & H. Aumann. *Image reconstruction from multiple 1-D scans using filtered localized projection*. Applied Optics, vol. 26, no. 17, pages 3615–3621, sept. 1987.
- [Fuh 91] C. Fuh & P. Maragos. *Affine models for image matching and motion detection*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), volume 4, pages 2409–2412, mai 1991.
- [Féron 05] O. Féron & A. Mohammad-Djafari. *Image fusion and unsupervised joint segmentation using a HMM and MCMC algorithm*. Journal of Electronic Imaging, vol. 14, no. 2, pages 1–12, avril 2005.
- [Geman 84] S. Geman & D. Geman. *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 6, pages 721–741, nov. 1984.
- [Geman 92] D. Geman & G. Reynolds. *Constrained restoration and the recovery of discontinuities*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 3, pages 367 – 383, mars 1992.
- [Geman 95] D. Geman & C. Yang. *Nonlinear image recovery with half-quadratic regularization*. IEEE Transactions on Image Processing, vol. 4, no. 7, pages 932 – 946, juillet 1995.
- [Glazer 84] F. Glazer. *Multilevel relaxation in low-level computer vision*. In Multiresolution Image Processing Analysis. Springer-Verlag, A. Rosenfeld edition, 1984.
- [Golub 80] G. Golub & C. Loan. *An analysis of the total least squares problems*. SIAM Journal on Numerical Analysis, vol. 17, pages 883–893, 1980.
- [Golub 96] G. Golub & C. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, Maryland, 3rd edition, 1996.
- [Gonzalez 03] R. Gonzalez & R. Woods. *Digital image processing*. Prentice-Hall, 2nd edition, 2003.
- [Greenspan 02] H. Greenspan, G. Oz, N. Kiryati & S. Peled. *MRI inter-slice reconstruction using super resolution*. Magnetic Resonance Imaging, vol. 20, pages 437 – 446, 2002.
- [Guichard 97] F. Guichard & J.-M. Morel. *The state of the art in numerical analysis, chapitre Partial differential equations and image iterative filtering*, pages 525–562. Clarendon Press, Oxford, I. S. Duff and G. A. Watson edition, 1997.

- [Gunturk 02] B. Gunturk, Y. Altunbasak & R. Mersereau. *Multiframe resolution-enhancement methods for compressed video*. IEEE Signal Processing Letters, vol. 9, no. 6, pages 170–174, juin 2002.
- [Gunturk 03] B. Gunturk, A. Batur, Y. Altunbasak, M. Hayes & R. Mersereau. *Eigenface-domain super-resolution for face recognition*. IEEE Transactions on Image Processing, vol. 12, no. 5, pages 597–606, mai 2003.
- [Gunturk 04] B. Gunturk, Y. Altunbasak & R. Mersereau. *Super-resolution reconstruction of compressed video using transform-domain statistics*. IEEE Transactions on Image Processing, vol. 13, no. 1, pages 33–43, jan. 2004.
- [Hadamard 23] J. Hadamard. Lectures on the cauchy problem in linear partial differential equations. Yale University Press, New Haven, Connecticut, 1923.
- [Hardie 97a] R. Hardie, K. Barnard & E. Armstrong. *Joint MAP registration and high-resolution image estimation using a sequence of undersampled images*. IEEE Transactions on Image Processing, vol. 6, no. 12, pages 1621–1633, déc. 1997.
- [Hardie 97b] R. Hardie, T. Tuinstra, J. Bognar, K. Barnard & E. Armstrong. *High resolution image reconstruction from digital video with global and non-global scene motion*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 1, pages 153–156, oct. 1997.
- [Hardie 98] R. Hardie, K. Barnard, J. Bognar, E. Armstrong & E. Watson. *High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system*. Optical Engineering, vol. 37, no. 1, pages 247–260, jan. 1998.
- [Herman 80] G. Herman. Image reconstruction from projections : The fundamentals of computerized tomography. Academic Press, New York, 1980.
- [Hong 97a] M. Hong, M. Kang & A. Katsaggelos. *An iterative weighted regularized algorithm for improving the resolution of video sequences*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 2, pages 474–477, oct. 1997.
- [Hong 97b] M. Hong, M. Kang & A. Katsaggelos. *A regularized multichannel restoration approach for globally optimal high-resolution video sequence*. In Proc. SPIE Visual Communications and Image Processing, volume 3024, pages 1306–1316, fév. 1997.
- [Horn 81] B. Horn & B. Schunck. *Determining optical flow*. Artificial Intelligence, vol. 17, pages 185–203, fév. 1981.
- [Humblot 05a] F. Humblot, B. Collin & A. Mohammad-Djafari. *Evaluation and practical issues of subpixel image registration using phase correlation methods*. In Inter. Conf. on Physics in Signal and Image Processing (PSIP '05), pages 115–120, Toulouse, France, jan. 2005.
- [Humblot 05b] F. Humblot & A. Mohammad-Djafari. *Super-resolution and joint segmentation in Bayesian framework*. In 25th Inter. Workshop on Bayesian Inference and Maximum Entropy Methods (MaxEnt '05). AIP Conference Proceedings, volume 803, pages 207–214, San José, CA, États-Unis, août 2005.

- [Humblot 05c] F. Humblot & A. Mohammad-Djafari. *Super-resolution using hidden markov model and Bayesian detection estimation framework*. EURASIP Journal on Applied Signal Processing, special issue on Super-Resolution Imaging Analysis, Algorithms, and Applications, article accepté, à paraître fin 2005.
- [Hunt 91] B. Hunt. *Image super-resolution : Emerging prospects*. In Proc. of the SPIE : Applications of Digital Image Processing XIV, volume 1567, pages 600–608, juil. 1991.
- [Hunt 92] B. Hunt. *Quantifying the super-resolution capabilities of the CLEAN image processing algorithm*. In Proc. of the SPIE : Applications of Digital Image Processing XV, pages 300–306, juil. 1992.
- [Hunt 95] B. Hunt. *Super-resolution of images : Algorithms, principles, performances*. Inter. Journal of Imaging Systems and Technology, vol. 6, no. 4, pages 297–308, 1995.
- [Irani 90] M. Irani & S. Peleg. *Super resolution from image sequences*. In Proc. IEEE Inter. Conf. on Pattern Recognition (ICPR), volume 2, pages 115–120, juin 1990.
- [Irani 91] M. Irani & S. Peleg. *Improving resolution by image registration*. CVGIP : Graphical Models and Image Processing, vol. 53, no. 3, pages 231–239, mai 1991.
- [Irani 93] M. Irani & S. Peleg. *Motion analysis for image enhancement : Resolution, occlusion and transparency*. Journal of Visual Communications and Image Representation, vol. 4, no. 4, pages 324–335, déc. 1993.
- [Irani 94] M. Irani, B. Rousso & S. Peleg. *Computing occluding and transparent motions*. Inter. Journal of Computer Vision, vol. 12, no. 1, pages 5–16, fév. 1994.
- [Jacquemod 92] G. Jacquemod, C. Odet & R. Goutte. *Image resolution enhancement using subpixel camera displacement*. Signal Processing, vol. 26, no. 1, pages 139–146, 1992.
- [Jain 81] J. Jain & A. Jain. *Displacement measurement and its application in interframe image coding*. In IEEE Transactions on Communications, volume 29, pages 1799–1808, déc. 1981.
- [Kaltenbacher 96] E. Kaltenbacher & R. Hardie. *High resolution infrared image reconstruction using multiple, low resolution, aliased frames*. In Proc. IEEE National Aerospace and Electronics Conf., volume 2, pages 702–709, mai 1996.
- [Katsaggelos 90] A. Katsaggelos. *A multiple input image restoration approach*. Journal of Visual Communication and Image Representation, vol. 1, no. 1, pages 93–103, sept. 1990.
- [Keren 88] D. Keren, S. Peleg & R. Brada. *Image sequence enhancement using sub-pixel displacements*. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 742–746, juin 1988.
- [Keys 81] R. Keys. *Cubic convolution interpolation for digital image processing*. IEEE Transactions on Acoustics Speech and Signal Processing, vol. 29, no. 6, pages 1153–1160, déc. 1981.

- [Kim 90a] S. Kim & N. Bose. *Reconstruction of 2-D bandlimited discrete signals from nonuniform samples*. IEE Proceedings-F Radar and Signal Processing, vol. 137, no. 3, pages 197–204, juin 1990.
- [Kim 90b] S. Kim, N. Bose & H. Valenzuela. *Recursive reconstruction of high resolution image from noisy undersampled multiframe*. IEEE Transactions on Acoustics Speech and Signal Processing, vol. 38, no. 6, pages 1013–1027, juin 1990.
- [Kim 91] S. Kim & W. Su. *Recursive high-resolution reconstruction of blurred multiframe images*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 2977–2980, mai 1991.
- [Kim 93a] S. Kim & W. Su. *Recursive high-resolution reconstruction of blurred multiframe images*. IEEE Transactions on Image Processing, vol. 2, no. 4, pages 534–539, oct. 1993.
- [Kim 93b] S. Kim & W. Su. *Subpixel accuracy image registration by spectrum cancellation*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), volume 5, pages 153–156, avril 1993.
- [Kim 94] H. Kim. *High Resolution Image Reconstruction from Undersampled Multiframe*. Thèse, Pennsylvania State University, États-Unis, mai 1994.
- [Kirkpatrick 83] S. Kirkpatrick, C. Gellat, Jr. & M. Vecchi. *Optimization by simulated annealing*. Science, vol. 220, no. 4598, pages 671–680, mai 1983.
- [Komatsu 93a] T. Komatsu, K. Aizawa, T. Igarashi & T. Saito. *Signal-processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis*. IEE Proceedings-I Communications, Speech and Vision, vol. 140, no. 1, pages 19–25, fév. 1993.
- [Komatsu 93b] T. Komatsu, T. Igarashi, K. Aizawa & T. Saito. *Very high resolution imaging scheme with multiple different aperture cameras*. Signal Processing : Image Communication, vol. 5, pages 511–526, déc. 1993.
- [Kuglin 75] C. D. Kuglin & D. C. Hines. *The phase correlation image alignment method*. In Proc. IEEE 1975 Int. Conf. Cybernetics and Society, pages 163–165, sept. 1975.
- [Landweber 51] L. Landweber. *An iterative formula for Fredholm integral equations of the first kind*. American Journal of Mathematics, vol. 73, pages 615–624, 1951.
- [Lin 01] Z. Lin & H.-Y. Shum. *On the fundamental limits of reconstruction-based super-resolution algorithms*. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), volume 1, pages 1171–1176, 2001.
- [Lin 04] Z. Lin & H.-Y. Shum. *Fundamental limits of reconstruction-based superresolution algorithms under local translation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 1, pages 83–97, jan. 2004.
- [Liu 93] B. Liu & A. Zaccarin. *New fast algorithms for the estimation of block motion vectors*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 3, no. 2, pages 148–157, avril 1993.

- [Lorette 97] A. Lorette, H. Shekarforoush & J. Zerubia. *Super-resolution with adaptive regularization*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 1, pages 169–172, oct. 1997.
- [Lucas 85] B. Lucas & T. Kanade. *An iterative image registration technique with an application to stereo vision*. IEEE Transactions on Acoustics Speech and Signal Processing, vol. 33, no. 4, pages 1151–1165, oct. 1985.
- [Mann 94] S. Mann & R. Picard. *Virtual bellows : Constructing high quality stills from video*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), pages 363–367, 1994.
- [Mohammad-Djafari 01] A. Mohammad-Djafari. Problèmes inverses (polycopié de cours du DEA ATS), chapitre 9 (Approche bayésienne), pages 141–168. LSS, Supélec, mai 2001.
- [Molina 03a] R. Molina, J. Mateos, A. Katsaggelos & M. Vega. *Bayesian multichannel image restoration using compound Gauss-Markov random fields*. IEEE Transactions on Image Processing, vol. 12, no. 12, pages 1642–1654, déc. 2003.
- [Molina 03b] R. Molina, M. Vega, J. Abad & A. Katsaggelos. *Parameter estimation in bayesian high-resolution image reconstruction with multisensors*. IEEE Transactions on Image Processing, vol. 12, no. 12, pages 1655–1667, déc. 2003.
- [Monasse 00a] P. Monasse. *Morphological Representation of Digital Images and Application to Registration*. Thèse, Université Paris IX Dauphine, France, 2000.
- [Monasse 00b] P. Monasse & F. Guichard. *Fast computation of a contrast invariant representation*. IEEE Transactions on Image Processing, vol. 9, no. 5, pages 860–872, mai 2000.
- [Morris 96] R. Morris, X. Descombes & J. Zerubia. *The Ising/Potts model in not well suited to segmentation tasks*. In Proc. of the IEEE Digital Signal Processing Workshop, pages 263–266, sept. 1996.
- [Musmann 85] H. Musmann, P. Pirsch & H. Grallert. *Advances in picture coding*. In Proc. of the IEEE, volume 73, pages 523–548, avril 1985.
- [Nagel 83] H. Nagel. *Displacement vectors derived from second-order intensity variations in image sequences*. Computer Graphics and Image Processing, vol. 21, pages 85–117, 1983.
- [Nagel 86] H. Nagel. *Image sequence – ten (octal) years – from phenomenology toward a theoretical foundation*. In Proc. Inter. Joint Conf. on Pattern Recognition, pages 1174–1185, oct. 1986.
- [Nakazawa 95] Y. Nakazawa, T. Komatsu & T. Saito. *High-resolution image acquisition based on temporal integration with hierarchical estimation of image warping*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 3, pages 244–247, oct. 1995.
- [Netravali 79] A. Netravali & J. Robbins. *Motion compensated television coding part I*. Bell Systems Tech. J., vol. 58, no. 3, pages 629–668, avril 1979.
- [Nguyen 00] N. Nguyen. *Numerical Algorithms For Image Superresolution*. Thèse, Stanford University, Californie, États-Unis, juillet 2000.

- [Nikolova 01] M. Nikolova. *Image restoration by minimizing objective functions with non-smooth datafidelity terms*. In IEEE Workshop on Variational and Level Set Methods (ICCV), pages 11–18, juill. 2001.
- [Omrane 04] N. Omrane & P. Palmer. *Super-resolution for translated satellite images using the walsh functions*. In Proc. of SPIE, Image and Signal Processing for Remote Sensing IX, volume 5238, pages 62–71, fév. 2004.
- [Oskoui-Fard 88] P. Oskoui-Fard & H. Stark. *Tomographic image reconstruction using the theory of convex projections*. IEEE Transactions on Medical Imaging, vol. 7, no. 1, pages 45–58, mars 1988.
- [Papoulis 77] A. Papoulis. *Generalized sampling expansion*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 11, pages 652–654, nov. 1977.
- [Patti 94] A. Patti, M. Sezan & A. Tekalp. *High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), pages 343–347, nov. 1994.
- [Patti 95] A. Patti, M. Sezan & A. Tekalp. *High resolution standards conversion of low resolution video*. In IEEE Transactions on Acoustics Speech and Signal Processing, volume 4, pages 2197–2200, 1995.
- [Patti 97a] A. Patti, M. Sezan & A. Tekalp. *Robust methods for high-quality stills from interlaced video in the presence of dominant motion*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no. 2, pages 328–342, avril 1997.
- [Patti 97b] A. Patti, M. Sezan & A. Tekalp. *Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time*. IEEE Transactions on Image Processing, vol. 6, no. 8, pages 1064–1076, août 1997.
- [Patti 98] A. Patti, A. Tekalp & M. Sezan. *A new motion compensated reduced order model Kalman filter for space-varying restoration of progressive and interlaced video*. IEEE Transactions on Image Processing, vol. 7, no. 4, pages 543–554, avril 1998.
- [Patti 99] A. Patti & Y. Altunbasak. *Super-resolution image estimation for transform coded video with application to MPEG*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 3, pages 179–183, oct. 1999.
- [Peleg 87] S. Peleg, D. Keren & L. Schweitzer. *Improving image resolution using subpixel motion*. Pattern Recognition Letters, vol. 5, no. 3, pages 223–226, mars 1987.
- [Pieczynski 00] W. Pieczynski & A.-N. Tebbache. *Pairwise Markov random fields and segmentation of textured images*. Machine Graphics & Vision, vol. 9, no. 3, pages 705–718, 2000.
- [Reddy 96] B. Reddy & B. Chatterji. *An FFT-based technique for translation, rotation, and scale-invariant image registration*. IEEE Transactions on Image Processing, vol. 5, no. 8, pages 1266–1271, août 1996.
- [Reeves 93] S. Reeves & S. Eddins. *Comments on “Iterative procedures for reduction of blocking effects in transform image coding”*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 3, no. 6, pages 439–440, déc. 1993.

- [Roberts 87] R. Roberts & C. Mullis. *Digital signal processing*. Addison-Wesley, 1987.
- [Rocheftort 05] G. Rocheftort. *Imagerie haute-résolution à partir de séquences vidéo*. Thèse, Université de Paris-sud XI, Orsay, France, mai 2005.
- [Roullot 01] E. Roullot. *Analyse d'acquisition multiples anisotropes en Angiographie par Résonance Magnétique 3D : modélisation et reconstruction régularisée pour l'amélioration de la résolution spatiale*. Thèse, ENST, Paris, France, déc. 2001.
- [Saito 94] T. Saito, T. Komatsu & K. Aizawa. *An image processing algorithm for a super high definition imaging scheme with multiple different-aperture cameras*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), volume 5, pages 45–48, avril 1994.
- [Samson 02] V. Samson. *Approche régularisée pour la détection d'objets ponctuels dans une séquence d'images*. Thèse, Université de Paris-sud XI, Orsay, France, déc. 2002.
- [Samson 04] V. Samson, F. Champagnat & J.-F. Giovanelli. *Point target detection and subpixel position estimation in optical imagery*. Applied Optics, vol. 43, no. 2, pages 257–263, jan. 2004.
- [Sauer 87] K. D. Sauer & J. P. Allebach. *Iterative reconstruction of band-limited images from nonuniformly spaced samples*. IEEE Transactions on Circuits and Systems, vol. 34, no. 12, pages 1497–1506, déc. 1987.
- [Sawchuk 72] A. Sawchuk. *Space-variant image motion degradation and restoration*. Proceedings of the IEEE, vol. 60, no. 7, pages 854–861, juill. 1972.
- [Schultz 92] R. Schultz & R. Stevenson. *Improved definition image expansion*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), volume 3, pages 173–176, mars 1992.
- [Schultz 94] R. Schultz & R. Stevenson. *A Bayesian approach to image expansion for improved definition*. IEEE Transactions on Image Processing, vol. 3, no. 3, pages 233–241, mai 1994.
- [Schultz 95a] R. Schultz & R. Stevenson. *Extraction of high-resolution frames from video sequences*. In Proc. of the IEEE Workshop on Nonlinear Image Processing, juin 1995.
- [Schultz 95b] R. Schultz & R. Stevenson. *Improved definition video frame enhancement*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), volume 4, pages 2169–2172, mai 1995.
- [Schultz 95c] R. Schultz & R. Stevenson. *Video resolution enhancement*. In Proc. SPIE Image and Video Processing, volume 3, pages 23–34, fév. 1995.
- [Schultz 96] R. Schultz & R. Stevenson. *Extraction of high-resolution frames from video sequences*. IEEE Transactions on Image Processing, vol. 5, no. 6, pages 996–1011, juin 1996.
- [Schultz 97a] R. Schultz, L. Meng & R. Stevenson. *Subpixel motion estimation for multi-frame resolution enhancement*. In Proc. SPIE Visual Communications and Image Processing, volume 3024, pages 1317–1328, fév. 1997.



- [Schultz 97b] R. Schultz & R. Stevenson. *Bayesian estimation of subpixel-resolution motion fields and high-resolution video stills*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 3, pages 62–65, oct. 1997.
- [Schultz 98] R. Schultz, L. Meng & R. Stevenson. *Subpixel motion estimation for super-resolution image sequence enhancement*. Journal of Visual Communication and Image Representation, vol. 9, no. 1, pages 38–50, mars 1998.
- [Seferidis 93] V. Seferidis & M. Ghanbari. *General approach to block-matching motion estimation*. Optical Engineering, vol. 32, no. 7, pages 1464–1474, juill. 1993.
- [Segall 04] C. Segall, A. Katsaggelos, R. Molina & J. Mateos. *Bayesian resolution enhancement of compressed video*. IEEE Transactions on Image Processing, vol. 13, no. 7, pages 898–911, juill. 2004.
- [Sementilli 93] P. Sementilli, M. Nadar & B. Hunt. *Poisson MAP super-resolution estimator with smoothness constraint*. In Proc. SPIE Neural and Stochastic Methods in Image and Signal Processing II, volume 2032, pages 2–13, juill. 1993.
- [Sementilli 94] P. Sementilli, M. Nadar & B. Hunt. *Empirical evaluation of a bound on image superresolution performance*. In Proc. SPIE Image Reconstruction and Restoration, volume 2302, pages 178–187, juill. 1994.
- [Sezan 84] M. Sezan & H. Stark. *Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion*. IEEE Transactions on Medical Imaging, vol. MI-3, no. 2, pages 91–98, juin 1984.
- [Shah 96] N. Shah & A. Zakhor. *Multiframe spatial resolution enhancement of color video*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 1, pages 985–988, sept. 1996.
- [Shah 99] N. Shah & A. Zakhor. *Resolution enhancement of color video sequences*. IEEE Transactions on Image Processing, vol. 8, no. 6, pages 879–885, juin 1999.
- [Shannon 49] Claude Shannon. *Communication in the presence of noise*. Proc. of the IRE, vol. 37, no. 1, pages 10–21, jan. 1949. Réimprimé dans Proc. of the IEEE, vol. 86 (2) p. 447–457, fév. 1998.
- [Shechtman 05] E. Shechtman, Y. Caspi & M. Irani. *Space-time super-resolution*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 4, pages 531–545, avril 2005.
- [Shekarforoush 95] H. Shekarforoush, M. Berthod & J. Zerubia. *3D super-resolution using generalized sampling expansion*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 2, pages 300–303, oct. 1995.
- [Sheppard 98] D. Sheppard, B. Hunt & M. Marcellin. *Iterative multiframe super-resolution algorithms for atmospheric turbulence degraded imagery*. Journal of Optical Society of America A, vol. 15, no. 4, pages 422–434, avril 1998.
- [Snoussi 03] H. Snoussi. *Approche bayésienne en séparation de sources. Applications en imagerie*. Thèse, Université de Paris-sud XI, Orsay, France, sept. 2003.
- [Snoussi 04] H. Snoussi & A. Mohammad-Djafari. *Fast joint separation and segmentation of mixed images*. Journal of Electronic Imaging, vol. 13, no. 2, pages 349–361, avril 2004.

- [Srinivas 90] C. Srinivas & M. Srinath. *A stochastic model-based approach for simultaneous restoration of multiple misregistered images*. In Proc. SPIE Visual Communications and Image Processing, pages 1416–1427, oct. 1990.
- [Srinivasan 85] R. Srinivasan & K. Rao. *Predictive coding based on efficient motion estimation*. IEEE Transactions on Communications, vol. 33, no. 8, pages 888–896, août 1985.
- [Stark 89] H. Stark & P. Oskoui. *High-resolution image recovery from image-plane arrays, using convex projections*. Journal of Optical Society of America, vol. 6, no. 11, pages 1715–1726, nov. 1989.
- [Su 94] W. Su & S. Kim. *High-resolution restoration of dynamic image sequences*. Inter. Journal of Imaging Systems and Technology, vol. 5, no. 4, pages 330–339, 1994.
- [Tekalp 92] A. Tekalp, M. Ozkan & M. Sezan. *High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration*. In Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 169–172, mars 1992.
- [Tekalp 95] A. Tekalp. Digital video processing. Prentice Hall, Upper Saddle River, New Jersey, 1995.
- [Tikhonov 77] A. Tikhonov & V. Arsenin. Solutions of ill-posed problems. V. H. Winston & Sons, Washington, District of Columbia, 1977.
- [Tom 94a] B. Tom & A. Katsaggelos. *Multi-channel image identification and restoration using the expectation maximization algorithm*. In Proc. SPIE Applications of Digital Image Processing XVII, volume 2298, pages 316–331, juill. 1994.
- [Tom 94b] B. Tom & A. Katsaggelos. *Reconstruction of a high resolution image from multiple degraded mis-registered low resolution images*. In Proc. SPIE Visual Communications and Image Processing, pages 971–981, sept. 1994.
- [Tom 94c] B. Tom, A. Katsaggelos & N. Galatsanos. *Reconstruction of a high resolution image from registration and restoration of low resolution images*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 3, pages 553–557, nov. 1994.
- [Tom 95] B. Tom & A. Katsaggelos. *Reconstruction of a high-resolution image by simultaneous registration, restoration and interpolation of low-resolution images*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 2, pages 539–542, oct. 1995.
- [Tom 96a] B. Tom & A. Katsaggelos. *An iterative algorithm for improving the resolution of video sequences*. In Proc. SPIE Visual Communications and Image Processing, pages 1430–1438, mars 1996.
- [Tom 96b] B. Tom & A. Katsaggelos. *Resolution enhancement of video sequences using motion compensation*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 1, pages 713–715, sept. 1996.
- [Tsai 84] R. Tsai & T. Huang. Advances in computer vision and image processing, volume 1, chapitre Multiframe image restoration and registration, pages 317–339. JAI Press Inc., 1984.

- [Tupin 02] F. Tupin, M. Sigelle, I. Bloch & H. Maître. Le traitement des images (polycopié du cours ANIM), volume 1, chapitre 7 (Définition et simulation d'un champ de Markov), pages 115–162. Département TSI, Télécom-Paris, déc. 2002.
- [Ur 92] H. Ur & D. Gross. *Improved resolution from subpixel shifted pictures*. CVGIP : Graphical Models and Image Processing, vol. 54, no. 2, pages 181–186, mars 1992.
- [Vetterli 95] M Vetterli & J. Kovacevic. Wavelets and subband coding. Prentice-Hall, Englewood Cliffs, NJ, 1995.
- [Walker 84] D. Walker & K. Rao. *Improved pel-recursive motion compensation*. IEEE Transactions on Communications, vol. COM-32, pages 1128–1134, oct. 1984.
- [Wang 04] Z. Wang & F. Qi. *On ambiguities in super-resolution modeling*. IEEE Signal Processing Letters, vol. 11, no. 8, pages 678–681, août 2004.
- [Winkler 95] G. Winkler. Image analysis, random fields and dynamic monte carlo methods. A mathematical introduction. Springer-Verlag, 1995. Applications of mathematics.
- [Wolberg 00] G. Wolberg & S. Zokai. *Robust image registration using log-polar transform*. In Proc. IEEE Inter. Conf. on Image Processing (ICIP), volume 1, sept. 2000.
- [Xu 05] X. Xu & R. Narayanan. *Enhanced resolution in SAR/ISAR imaging using iterative sidelobe apodization*. IEEE Transactions on Image Processing, vol. 14, no. 4, pages 537–547, avril 2005.
- [Yeh 90] S. Yeh & H. Stark. *Iterative and one-step reconstruction from nonuniform samples by convex projections*. Journal of Optical Society of America, vol. 7, no. 3, pages 491–499, mars 1990.
- [Youla 78] D. Youla. *Generalized image restoration by the method of alternating orthogonal projections*. IEEE Transactions on Circuits and Systems, vol. 25, no. 9, pages 694–702, sept. 1978.
- [Youla 82] D. Youla & H. Webb. *Image restoration by the method of convex projections : Part 1 – Theory*. IEEE Transactions on Medical Imaging, vol. MI-1, no. 2, pages 81–94, oct. 1982.
- [Yuen 94] P. Yuen & B. Hunt. *Constraining information for improvements to MAP image super-resolution*. In Proc. SPIE Image Reconstruction and Restoration, volume 2302, pages 144–145, juill. 1994.
- [Zitová 03] B. Zitová & J. Flusser. *Image registration methods : a survey*. Image and Vision Computation, vol. 21, no. 11, pages 977–1000, oct. 2003.
- [Zomet 00] A. Zomet & S. Peleg. *Efficient super-resolution and applications to mosaics*. In Proc. of the 15<sup>th</sup> Inter. Conf. on Pattern Recognition, volume 2, pages 579–583, sept. 2000.

# Table des figures

1.1	Comparaison de deux approches pour la détection. . . . .	10
1.2	Convention utilisée pour représenter les coordonnées des images numériques. . . . .	12
2.1	Pixels tests et leurs 4 voisins (dans le même ordre que dans TAB. 2.1). . . . .	22
2.2	Localisation des 5 pixels dans l'image IR utilisée pour les simulations. . . . .	24
2.3	Comparaison résultats cas entier/réel - Pixel (128, 54), zone chaude. . . . .	24
2.4	Comparaison résultats cas entier/réel - Pixel (218, 292), 3 voisins chauds. . . . .	25
2.5	Comparaison résultats cas entier/réel - Pixel (30, 260), zone froide. . . . .	25
2.6	Superposition des 5 cas entiers. . . . .	26
2.7	Superposition des 5 cas réels. . . . .	26
3.1	Exemple d'un déplacement sous-pixelique sur une image de 9 pixels. . . . .	45
3.2	Image numérique utilisée pour l'évaluation des méthodes de recalage. . . . .	48
3.3	Différence entre les images obtenues en prenant $q = 1$ et $q = 11$ . . . . .	48
3.4	Résultats d'estimation des déplacements sous la forme de constellation. . . . .	49
3.5	Effet du recalage observé à partir d'images de différence. . . . .	49
3.6	Exemple de déplacement sous-pixelique connu. . . . .	52
3.7	Image translatée corrigée par une translation de (-10,17, -12,70). . . . .	52
3.8	(a) Différences entre images réelle et translatée, (b) et entre images réelle et recalée. . .	52
3.9	Deux images IR réelles montrant un déplacement de nuages. . . . .	53
3.10	Image recalée et normes du gradient des deux images de la FIG. 3.9. . . . .	53
3.11	(a) Différences entre 1 <sup>ere</sup> et 350 <sup>e</sup> image, (b) et entre 1 <sup>ere</sup> et 350 <sup>e</sup> image recalée. . . . .	53
3.12	Zoom sur le point de recalage et valeur des niveaux de gris des pixels voisins. . . . .	54
3.13	1 <sup>ere</sup> et 21 <sup>e</sup> image d'une séquence montrant un mouvement vertical. . . . .	56
3.14	Image recalée et normes du gradient des deux images de la FIG. 3.13. . . . .	56
3.15	(a) Différences entre 1 <sup>ere</sup> et 21 <sup>e</sup> image, (b) et entre 1 <sup>ere</sup> et 21 <sup>e</sup> image recalée. . . . .	56
4.1	Génération d'une séquence HR à partir d'une séquence BR. . . . .	62
4.2	Illustration au niveau pixelique des opérateurs $\mathbf{D}$ , $\mathbf{M}$ , $\mathbf{D}^\top$ et $\mathbf{M}^\top$ dans le processus de SR. . . . .	64
4.3	Illustration du problème direct en images. . . . .	65
5.1	Les cliques associées à deux systèmes de voisinage en dimension 2. . . . .	112
5.2	Comparaison des deux modèles d'homogénéité de régions : constant et bilinéaire. . . . .	121
5.3	Comparaison d'images HR et BR dans un exemple synthétique. . . . .	129
5.4	Résultats de SR obtenus dans un cas faiblement bruité, RSB = 30 dB. . . . .	132

5.5	Résultats de SR obtenus dans un cas bruité, $\text{RSB} = 15$ dB. . . . .	132
5.6	Résultats de segmentation pour $K$ allant de 2 à 6 avec modèle constant, $\text{RSB} = 30$ dB. .	133
5.7	Résultats de segmentation pour $K$ allant de 4 à 6 avec modèle constant, $\text{RSB} = 15$ dB. .	133
5.8	Résultats de segmentation pour $K$ allant de 4 à 6 avec modèle bilinéaire, $\text{RSB} = 30$ dB.	134
5.9	Résultats de SR avec mauvaises données de recalage $\hat{\mathbf{s}}_0$ , $\text{RSB} = 30$ dB. . . . .	135
5.10	Résultats de SR en fonction du nombre des images utilisées, $\text{RSB} = 30$ dB. . . . .	137
5.11	Résultats de SR sur la séquence réelle « réveil » et agrandissement d'une région. . . . .	138
5.12	Interface Matlab de l'outil de simulation de SR. . . . .	140

# Liste des tableaux

2.1	Niveaux de gris des pixels voisins aux pixels tests et leur moyenne. . . . .	22
2.2	Temps de calcul pour 256 réalisations de l'algorithme (cas entier et réel). . . . .	23
3.1	Moyenne et maximum des erreurs (réalisés à partir de 400 estimations, exprimés en %). . . . .	48
4.1	Résumé et comparaisons des méthodes de SR dans les domaines fréquentiel et spatial. . . . .	104
4.2	Comparaisons qualitatives de deux méthodes de SR : MAP et POCS. . . . .	105
5.1	Évaluation des performances en milieu bruité. Modèle constant d'homogénéité. . . . .	130
5.2	Évaluation des performances en milieu bruité. Modèle bilinéaire d'homogénéité. . . . .	130
5.3	Évaluation des performances en fonction des données de recalage. . . . .	135
5.4	Évaluation des performances en fonction du nombre et de la disposition des images. . . . .	137







# Détection de petits objets dans une image en utilisant les techniques de super-résolution

## Résumé

Cette thèse concerne la détection de petits objets dans une image en utilisant les techniques de super-résolution (SR). La reconstruction d'une image par une méthode de SR consiste à produire une image de haute résolution (HR), à partir de plusieurs images de basse résolution (BR) obtenues par l'intermédiaire de différentes caméras, ou à partir d'une séquence vidéo acquise avec une seule caméra. Obtenir une image HR nécessite deux étapes : le recalage des images BR dans un repère commun, et la construction de l'image HR par leur fusion.

Ce mémoire présente donc deux parties. La première est consacrée à la détection et aux méthodes de recalage d'images, et la seconde aux techniques de restauration d'images par SR. Concernant la première partie, plusieurs méthodes ont été évaluées : une méthode fréquentielle de recalage utilisant le principe de corrélation de phase d'une part, et une méthode de détection de petits points basée sur un estimateur MAP dans le formalisme bayésien d'autre part. Dans la seconde partie, une nouvelle méthode de SR utilisant une modélisation markovienne hiérarchique de l'image HR dans le cadre de l'estimation bayésienne est proposée. Cette nouvelle approche, qui est basée sur l'idée que l'image HR est constituée de zones homogènes, permet d'obtenir non seulement une image de bonne qualité, mais également un résultat de segmentation de la scène HR.

**Mots clés.** Super-résolution, estimation et détection bayésienne, classification, segmentation, recalage sous-pixellique d'image, méthodes de corrélation de phase.

---

## Detection of Small Objects Using Super-Resolution Image Processing Techniques

### Abstract

This thesis deals with the detection of small objects in an image using super-resolution (SR) techniques. Image reconstruction using a SR method consists in producing a high-resolution (HR) image, from several low-resolution (LR) images obtained either from several cameras, or from a video sequence shot with one single video camera. Obtaining a HR image requires two steps: registration of the LR images on a common grid, and the estimation of the HR image using a data fusion approach.

Therefore, this manuscript presents two parts. The first main section deals with detection and registration methods, and the second one with multi-frame SR restoration techniques. About the first part, several methods have been evaluated: a frequency-domain registration method using phase correlation principle on the one hand, and a speck detection method based on a MAP estimator in a Bayesian framework on the other hand. In the second part, a new SR method using a hierarchical Markov model for the HR image in the Bayesian framework is proposed. This new approach, which is based on the idea that the HR image is made of homogeneous regions, provides not only an image of good quality, but also a segmentation result of the HR scene.

**Keywords.** Super-resolution, Bayesian detection and estimation, classification, segmentation, subpixel image registration, phase correlation methods.