

UNIVERSITE JOSEPH FOURIER

N° attribué par la bibliothèque

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE JOSEPH FOURIER

Spécialité : «signal, image, parole, télécoms»

préparée au Laboratoire des Images et des Signaux de Grenoble

dans le cadre de l'École Doctorale «**électronique, électrotechnique, automatique, télécommunications, signal**»

présentée et soutenue publiquement

par

Eric BRUNO

le 09 novembre 2001

Titre :

**De l'estimation locale à l'estimation globale de mouvement
dans les séquences d'images**

Directeur de thèse : Denis PELLERIN

JURY

Monsieur J.M. Jolion , Président
Monsieur P. Bouthemy , Rapporteur
Monsieur M. Kunt, , Rapporteur
Monsieur D. Pellerin, , Directeur de thèse
Monsieur J. Hérault , Examineur

Remerciements

Je voudrais tout d'abord remercier les personnes qui m'ont fait l'honneur de participer à mon jury, messieurs Patrick Bouthemy et Murat Kunt comme rapporteurs, monsieur Jean-Michel Jolion en tant que président ainsi que monsieur Jeanny Hérault. Je tiens à remercier tout particulièrement Patrick Bouthemy pour l'attention qu'il a porté à la lecture de ce manuscrit, me permettant d'y apporter de nombreuses corrections. Je remercie mon directeur de thèse monsieur Denis Pellerin qui m'a permis d'effectuer cette étude.

Je voudrais également remercier toutes les personnes qui m'ont écouté et conseillé durant ces trois années. C'est pour une large part grâce à elles que j'ai pu mener et achever ce travail de recherche. Merci à Olivier Chomat pour les nombreuses discussions sur les problèmes de la reconnaissance d'activités et de l'extraction du mouvement. Merci à Anisse Taleb qui par sa disponibilité et ses conseils m'a permis de démarrer ce travail de thèse et de comprendre un peu mieux le travail de chercheur. Merci également à Thomas Belli, Cyrille Hory, Mathias Pingault pour les nombreuses discussions scientifiques que nous avons pu avoir ensemble.

Ces remerciements ne seraient pas complets sans citer Guillaume, Sébastien, Mathieu, Cédric, Jun, et tous ceux que j'oublie, pour ces trois ans passés ensemble. Je garderai de cette période un souvenir marqué par la bonne humeur et l'optimisme de toutes ces personnes.

Enfin, je remercie Marie pour m'avoir apporté son soutien et son réconfort lors de cette thèse et pour la patience dont elle a fait preuve durant la phase de rédaction de ce manuscrit.

Résumé

Le thème de cette thèse est l'estimation et l'analyse du mouvement dans les séquences d'images. L'étape d'estimation consiste à déterminer le champ de vitesse entre deux images successives. La phase d'analyse permet d'extraire de cette mesure du mouvement des informations descriptives de la séquence.

Nous avons développé deux estimateurs de mouvement basés sur des modèles de mouvement de natures différentes. Les deux algorithmes s'inscrivent dans le cadre des méthodes différentielles et reposent sur la résolution d'un système d'équations du flot optique sur-contraînt par le modèle de mouvement. L'estimation des paramètres du modèle est dans les deux cas obtenue par une méthode robuste, itérative et multirésolution.

La première approche utilise un modèle de mouvement translationnel estimé localement. Une décomposition de la séquence par un banc de filtres de Gabor spatiaux et l'application de l'équation du flot optique sur chaque sous-bande permet de contourner le problème d'ouverture. La décomposition est implantée par des filtres récursifs, permettant une réduction importante du coût calculatoire de l'algorithme. Les résultats obtenus sur des séquences artificielles et réelles nous montrent que la méthode permet de mesurer de manière fiable et précise le flot optique entre deux images.

La deuxième méthode est basée sur une modélisation globale du mouvement. Le modèle, valide en tout point de l'image, doit être suffisamment riche pour approximer les mouvements complexes présent dans les séquences d'images. Nous avons défini deux types de modèle : l'un est basé sur les séries de Fourier, l'autre sur les séries d'ondelettes. Les paramètres de Fourier ou d'ondelettes sont estimés en résolvant le système d'équations du flot optique appliqué en chaque pixel de l'image. Les meilleurs résultats sont obtenus par un modèle basé sur des ondelettes B-splines, et sont comparables aux résultats de l'approche d'estimation locale.

Le dernier point de cette thèse concerne l'analyse du mouvement. En effet, les paramètres des modèles du mouvement global fournissent une signature compacte et précise du mouvement et peuvent être vus comme des descripteurs du mouvement. Une expérience de classification de vidéos et une de segmentation temporelle de vidéos montrent que les coefficients d'ondelettes B-spline du flot optique sont effectivement des descripteurs pertinents.

Mots clés

Estimation du mouvement, ondelettes de Gabor et B-splines, modèles paramétriques de mouvements, multirésolution, M-estimateur, indexation de vidéos, reconnaissance d'activités.

Abstract

The subject of this thesis deals with motion estimation and analysis in image sequences. Motion estimation consists in extracting optical flow from two image frames. From this measure, analysis stage tries to extract informations about video content and structure.

We have developed two motion estimators based on two different motion models. These two algorithms are based on the resolution of a system of velocity equations over-constrained by the motion model. Model parameters are estimated in both cases by using robust least-squares technique embended in a coarse to fine strategy.

The first method is based on a locally constant velocity field assumption. A spatial Gabor-like filter bank, based on recursive implementation, provides a fast multichannel decomposition of frame sequence. Then, applying the velocity equation on each channel allows to avoid the aperture problem. The accuracy and robustness of our algorithm are demonstrated by testing it on artificial and real image sequences.

The second approach is based on global approximation of image motion by parametric motion models. The global motion parameters are estimated over the entire image by applying on each location the velocity equation. We have defined two global motion models : one based on Fourier series expansion and the other on the wavelet series expansion. Best results are obtained from a B-spline wavelet-based motion model, and the accuracy of the algorithm are similar to the Gabor approach.

The last point of this document concerns motion analysis by using global motion model parameters as motion descriptors. The estimated motion parameters (motion wavelet coefficients or Fourier components) provide a global, robust and compact description of the motion content. Two experiences on video classification and temporal segmentation have shown that motion wavelet coefficients are efficient to characterize dynamic content of image sequences.

Key words

Motion estimation, Gabor and B-spline wavelets, parametric motion model, multiresolution, M-estimator, motion-based video indexing, activity recognition.

Table des matières

Table des matières	9
1 Introduction	13
1.1 Problème traité	13
1.2 Contribution	14
1.3 Plan du mémoire	15
I Estimation locale du mouvement	17
2 Techniques d'estimation locale du mouvement	19
2.1 Équation du flot optique et problème d'ouverture	20
2.2 Différentes méthodes d'estimation du mouvement	21
2.2.1 Méthodes de mise en correspondance	21
2.2.2 Méthodes fréquentielles	21
2.2.3 Méthodes différentielles	22
2.3 Estimation du mouvement par l'équation du flot optique au 1 ^{er} ordre	23
2.3.1 Validité de l'équation du flot optique : approches multiéchelle et multirésolution	23
2.3.2 Résolution du problème d'ouverture	23
2.4 Utilisation de bancs de filtres en vision par ordinateur	26
3 Estimation locale du mouvement par banc de filtres de Gabor récursifs	29
3.1 Introduction	29
3.2 Position du problème	29
3.3 Les bancs de filtres de Gabor et leur implantation récursive	30
3.3.1 Construction du banc de filtres	31
3.3.2 Suppression de la composante continue au moyen d'un préfiltrage passe-haut	33
3.3.3 Implantation récursive des filtres de Gabor	34
3.4 Résolution robuste du système surdéterminé	35
3.5 Estimation progressive du mouvement par l'approche multirésolution	38
3.6 Résultats expérimentaux	40

3.6.1	Affinement des réglages du banc de filtres	42
3.6.2	Coût de calcul	43
3.6.3	Comparaison quantitative avec d'autres algorithmes	44
3.6.4	Séquences réelles	46
3.7	Conclusion et perspectives	48
 II Modèles paramétriques pour l'estimation et la description du mouvement global		49
4	Les modèles paramétriques de mouvement	51
4.1	Introduction	51
4.2	Motivations pour la modélisation du mouvement	52
4.2.1	Intérêt de la représentation du mouvement par un modèle paramétrique	52
4.2.2	Description du contenu dynamique de séquence d'images	53
4.3	Différents modèles paramétriques de mouvement	54
4.3.1	Modèles polynomiaux	54
4.3.2	Modèles spécifiques	56
4.3.3	Modélisation globale du mouvement	56
4.4	Estimation des paramètres d'un modèle linéaire de mouvement	59
4.4.1	Estimation robuste et incrémentale d'un modèle paramétrique de mouvement	59
4.5	Conclusion et solutions proposées	60
5	Modélisation globale du mouvement par séries de Fourier	63
5.1	Modèle de mouvement basé sur les séries de Fourier	63
5.1.1	Rappel sur les séries de Fourier	63
5.1.2	Modélisation du mouvement	64
5.2	Estimation des coefficients de Fourier	65
5.2.1	Choix de l'estimateur robuste	65
5.2.2	Solution au sens des moindres-carrés robuste par la Décomposition en Valeurs Singulières	66
5.3	Estimation par enrichissement progressif du modèle de mouvement	68
5.4	Résultats expérimentaux	68
5.4.1	Séquences artificielles	68
5.4.2	Séquences réelles	69
5.5	Conclusion : vers une nouvelle modélisation	70
6	Modélisation globale du mouvement par bases d'ondelettes	77
6.1	Introduction	77
6.2	Définition d'un modèle de mouvement basé sur les ondelettes	78
6.2.1	Notions de base sur la transformation en ondelettes discrètes	78

6.2.2	Modélisation du mouvement	79
6.3	Estimation des paramètres du modèle de mouvement	80
6.3.1	Estimation multirésolution	80
6.3.2	Résolution d'un système creux	80
6.4	Choix d'une base d'ondelettes	83
6.4.1	Ondelettes B-splines	84
6.4.2	Précision de l'approximation en fonction du degré de Φ^N	85
6.5	Résultats expérimentaux	87
6.5.1	Résultats en fonction de l'ondelette	87
6.5.2	Détermination expérimentale du seuil ϵ appliqué au préconditionneur .	87
6.5.3	Comparaison quantitative sur des séquences artificielles	88
6.5.4	Résultats sur des séquences naturelles	91
6.6	Conclusion	95
7	Classification et segmentation temporelle de vidéos par les paramètres des modèles de mouvement global	97
7.1	Transformation des coefficients de fonctions d'échelle en coefficients d'ondelettes	98
7.2	Pertinence des paramètres de mouvement pour la classification de vidéos . . .	99
7.2.1	Description de la base de vidéos	99
7.2.2	Définition des descripteurs de mouvement	99
7.2.3	Classification hiérarchique ascendante de Ω	101
7.2.4	Résultats	102
7.3	Segmentation temporelle de séquences vidéo	105
7.3.1	Segmentation hiérarchique de la séquence	108
7.3.2	Séparation du mouvement dominant du mouvements des objets	108
7.3.3	Résultats de segmentation temporelle à partir de θ_{cam} et θ_{obj}	109
7.4	Conclusion	112
8	Conclusion et perspectives	115
	Annexes	117
A	M-estimateurs et moindres-carrés pondérés et itérés (MCPI)	119
A.1	Les M-estimateurs	119
A.2	L'algorithme MCPI	120
B	Méthode du gradient conjugué et biconjugué préconditionné	121
C	Publications	123
	Bibliographie	125

Chapitre 1

Introduction

1.1 Problème traité

La recherche en Vision par Ordinateur s’applique à définir des algorithmes permettant la perception et la compréhension du monde physique à partir d’informations visuelles (images et séquences d’images). Un système de vision complet s’articule autour de trois étapes : l’extraction d’attributs de bas-niveau (couleur, texture, forme, orientation, mouvement,...), l’analyse des attributs, fournissant des informations de plus haut-niveau sur la scène (reconnaissance, segmentation, catégorisation), et enfin l’interprétation de la scène, permettant au système de répondre à des questions du type “*qu’est ce qui se passe ?*” ou “*quels sont les objets présents dans la scène ?*”. Cette architecture, très générique, se retrouve dans tous les domaines d’applications de la vision par ordinateur, comme la robotique, la compression vidéo (MPEG-4) [55], l’indexation par le contenu d’images et de vidéos [56], les environnements intelligents [16], ou encore les applications de vidéo-surveillance. Notre travail, présenté dans cette thèse, tente d’apporter une contribution à ces recherches en proposant de nouveaux algorithmes pour la mesure et l’analyse du mouvement dans des séquences d’images.

Lorsque l’information visuelle varie au cours du temps, le mouvement, parmi tous les attributs de bas-niveau, est peut être le plus essentiel. Il offre des quantités importantes d’informations sur la structure tridimensionnelle de la scène, la trajectoire des objets, l’ego-mouvement, l’activité qui se déroule dans la scène. L’estimation de toutes ces informations nécessite une mesure précise et pertinente du mouvement dans l’image.

L’architecture générale d’un algorithme d’estimation de mouvement comporte trois niveaux distincts :

- Niveau de modélisation : il définit le modèle physique reliant le mouvement aux variations temporelles de la séquence d’images. En réalité, les variations temporelles sont dues à la projection du mouvement 3D dans le plan de l’image, ainsi qu’à des effets tels que les transparences, les ombres, les variations de luminosité, les occultations. Un modèle parfait serait très complexe et impossible à estimer. Aussi, il est toujours nécessaire d’apporter des simplifications en définissant des modèles plus ou moins complexes.

- Niveau de la mesure : afin d’estimer les paramètres du modèle physique, il est tout d’abord nécessaire d’effectuer un certain nombre de mesures sur la séquence d’images. Différents types de mesures peuvent être effectués sur la fonction de luminance, tels que le calcul des dérivées spatio-temporelles, le filtrages spatio-temporelles ou la mesure de corrélations entre deux instants successifs de la fonction.
- Niveau de l’estimation : il concerne la façon dont les mesures sont combinées pour estimer les paramètres du modèle physique. Ce niveau est le plus souvent basé sur une estimation au sens des moindres-carrés ou moindres-carrés robustes.

1.2 Contribution

Deux approches pour l’estimation du mouvement sont présentées dans ce document. Elles diffèrent l’une de l’autre essentiellement dans la définition du modèle reliant le mouvement aux variations temporelles. La première approche est basée sur un modèle très simple, permettant une estimation locale, alors que la deuxième utilise un modèle complexe fournissant une estimation globale du mouvement. Ces deux approches font l’objet chacune d’une partie de la thèse.

Première partie : estimation locale du mouvement

Cette partie concerne l’estimation locale du mouvement par banc de filtres de Gabor spatiaux. Le mouvement est estimé à partir de mesures *locales* des variations spatio-temporelles de la décomposition multi-bandes de la séquence d’images. Le résultat obtenu permet d’associer à chaque pixel un vecteur de vitesse relatif aux variations de la séquence entre deux instants différents.

Notre approche est motivée par le développement d’une architecture de vision bas-niveau dont l’élément central est composé de bancs de filtres de Gabor spatiaux. Ces filtres sont en effet des outils de perception visuelle très performants, et la possibilité d’effectuer en parallèle, par un même banc de filtres, des mesures de différentes caractéristiques de bas-niveau telles que les contours, les textures, les formes et encore le mouvement pourrait conduire à un système de vision très efficace. L’idée d’unifier les mesures d’indices visuels de bas-niveau par une architecture basée sur un banc de filtres linéaires fait depuis plusieurs années l’objet de recherches au LIS [4, 28, 65, 72, 79].

Deuxième partie : estimation globale du mouvement

Cette partie concerne l’estimation de modèles du mouvement global. La mesure du mouvement s’effectue par l’estimation d’un modèle paramétrique de mouvement appliqué à tout le support de l’image. L’estimation obtenue, qui consiste en un nombre réduit de paramètres de mouvement, est *globale* et permet de calculer en tout point de la scène un vecteur vitesse. Les modèles utilisés doivent être suffisamment riches pour approximer de façon satisfaisante les

mouvements complexes présents dans les séquences d'images. Nous avons défini deux modèles de mouvement permettant une modélisation globale : l'un est basé sur les séries de Fourier, l'autre sur les séries d'ondelettes.

L'utilisation de ces modèles permet de remonter à une mesure précise du mouvement (en particulier pour le modèle basé sur les ondelettes), et en même temps, fournit au travers des paramètres estimés des descripteurs pertinents sur le mouvement. Nous montrons en effet, par des expériences de classification et de segmentation temporelle de vidéos, que les paramètres des modèles globaux donnent une description compacte et précise du mouvement, facilitant ainsi l'analyse et l'interprétation de séquences d'images.

1.3 Plan du mémoire

Première partie : estimation locale du mouvement

Le chapitre 2 décrit la problématique de l'estimation du mouvement et présente les différentes techniques d'estimation locale de champs de vitesses, et plus particulièrement celles basées sur une décomposition multi-bandes de la séquence d'images.

Le chapitre 3 détaille notre algorithme d'estimation du mouvement par banc de filtres de Gabor spatiaux. Cette méthode combine les informations issues de bancs de filtres positionnés à différentes échelles spatiales pour estimer de manière robuste le mouvement entre deux images successives de la séquence. Afin de réduire le coût de calcul de l'algorithme, nous utilisons des filtres récursifs approximant les filtres de Gabor. Les résultats obtenus sur des séquences artificielles et réelles, ainsi qu'une comparaison des performances avec d'autres algorithmes, montrent que notre approche permet d'obtenir une mesure précise du mouvement.

Deuxième partie : estimation globale du mouvement

Le chapitre 4 établit un état de l'art concis sur le problème de la définition et de l'estimation de modèles de mouvement. Ce chapitre présente aussi l'intérêt de la modélisation du champ de vitesse à la fois pour l'estimation et la description du mouvement.

Le chapitre 5 porte sur la modélisation du mouvement global par des séries de Fourier. Les séries de Fourier permettent d'approximer par quelques termes un grand nombre de fonctions continues par morceaux. Aussi, un modèle basé sur ce développement semble adapté pour approximer le mouvement global. Les résultats obtenus sur des séquences artificielles et réelles sont satisfaisants, mais ne permettent pas une estimation précise du mouvement.

Le chapitre 6 décrit le modèle de mouvement basé sur les séries d'ondelettes. Ce nouveau modèle est introduit pour bénéficier de la supériorité des ondelettes sur les séries de Fourier pour approximer et décrire un signal complexe. Un autre avantage des ondelettes est l'allègement en coût de calcul qu'elles permettent d'obtenir en introduisant des matrices creuses dans l'algorithme d'estimation. Parmi toutes les ondelettes, nous avons choisi les B-splines pour leur régularité, leur compacité ainsi que pour leurs propriétés d'interpolation. La

méthode est à nouveau testée sur des séquences artificielles et réelles et comparée à d'autres algorithmes.

Le chapitre 7 présente deux expériences de classification et de segmentation temporelle de vidéos par les paramètres des modèles de mouvement. La première expérience consiste à classer une base de vidéos contenant différentes activités humaines à partir des paramètres estimés sur chaque séquence. Les résultats montrent que les paramètres du modèle basé sur les ondelettes sont pertinents au sens de la description du mouvement. Dans un deuxième temps, nous avons développé un algorithme de segmentation temporelle de vidéos basé sur la classification hiérarchique des paramètres de mouvement estimés sur la séquence. En utilisant les propriétés multiéchelle de la description en ondelettes, nous avons pu segmenter des séquences sportives selon le mouvement de la caméra et selon le mouvement des joueurs et des objets.

Les conclusions et perspectives sur notre travail font l'objet du chapitre 8.

Première partie

Estimation locale du mouvement

Chapitre 2

Techniques d'estimation locale du mouvement

L'estimation du mouvement est un problème fondamental pour l'analyse de séquences d'images. Il consiste à mesurer la projection 2D dans le plan de l'image d'un mouvement réel 3D, dû à la fois au mouvement des objets dans la scène et aux déplacements de la caméra. Le mouvement 2D, appelé flot optique, est une variable cachée et n'est accessible que par l'analyse des variations temporelles de la séquence d'images. Une séquence d'images peut être représentée par sa fonction de luminance $I(x, y, t)$. L'hypothèse de conservation de la luminance stipule que la luminance d'un point physique de la séquence d'image ne varie pas au cours du temps, c'est à dire :

$$I(\mathbf{p}, t) = I(\mathbf{p} + \mathbf{v}(\mathbf{p})\delta t, t + \delta t) \quad (2.1)$$

avec $\mathbf{p} = (x, y)^T$ et $\mathbf{v}(\mathbf{p}) = (u, v)^T$ le vecteur vitesse associé au point \mathbf{p} et au temps t . Les composantes u et v sont respectivement la vitesse selon x et selon y .

Cette hypothèse n'est pas respectée dans le cas d'occultations, de transparences, de réflexions spéculaires, et plus généralement de tout ce qui peut produire des variations brutales de l'illumination de la scène (flash d'appareil photo, lumières clignotantes, ...). Pour prendre en compte ces phénomènes, l'hypothèse de conservation de la luminance peut être enrichie de manière à autoriser les variations de luminances et de contrastes [57] :

$$I(\mathbf{p}, t) = c(\mathbf{p}, t)I(\mathbf{p} + \mathbf{v}(\mathbf{p})\delta t, t + \delta t) + b(\mathbf{p}, t) \quad (2.2)$$

où $c(\mathbf{p}, t)$ et $b(\mathbf{p}, t)$ représentent les variations multiplicatives et additives de la luminance et sont deux nouvelles inconnues à déterminer. Black *et. al.* [8] ont généralisé l'équation (2.2) de manière à prendre en compte les variations dues à la fois au mouvement, aux variations de luminance et de contraste et à l'apparition soudaine de nouvelles régions dans la scène (par exemple sur un visage, les paupières qui s'ouvrent font apparaître les yeux).

L'utilisation d'une modélisation complexe de la variation de luminance nécessite d'estimer un grand nombre de paramètres et conduit à des solutions peu stables. En pratique

l'hypothèse de la conservation de la luminance sous sa forme simple (2.1) est la plus usitée et permet d'obtenir les meilleurs résultats.

2.1 Équation du flot optique et problème d'ouverture

En admettant que $I(x, y, t)$ soit une fonction continue et dérivable, sa dérivée temporelle totale, selon l'hypothèse de conservation de la luminance, est nulle :

$$\frac{dI(\mathbf{p}, t)}{dt} = \nabla I(\mathbf{p}, t) \cdot \mathbf{v}(\mathbf{p}) + \frac{\partial I(\mathbf{p}, t)}{\partial t} = 0 \quad (2.3)$$

où $\nabla I = [I_x, I_y]$ est le gradient spatial de I . Cette dernière équation est appelée l'équation du flot optique et permet d'estimer la composante de vitesse "normale" :

$$v_N = -\frac{I_t}{\|\nabla I\|} \quad (2.4)$$

avec $I_t = \frac{\partial I(\mathbf{p}, t)}{\partial t}$ et v_N la projection du vecteur vitesse sur le gradient spatial d'intensité ∇I .

La relation (2.4) fait apparaître un problème intrinsèque à l'estimation du mouvement : l'hypothèse de conservation de la luminance appliquée en un point de l'image ne permet de retrouver que la composante de vitesse parallèle au gradient spatial d'intensité. Cette indétermination est connue sous le nom du *problème d'ouverture*. De plus, l'estimation est impossible dans le cas où $\nabla I = 0$.

La figure 2.1 illustre ces deux problèmes. Admettons que l'on regarde une surface en mouvement (le rectangle) au travers de petites fenêtres (symbolisées par les cercles). Trois cas sont possibles :

- cas 1 : lorsque le gradient d'intensité est nul, le mouvement n'est pas perceptible.
- cas 2 : lorsque le gradient d'intensité dans la fenêtre est orienté dans une seule direction, le mouvement est perçu comme normal au contour.
- cas 3 : la combinaison d'informations issues de gradients d'intensités de différentes orientations permet de remonter au mouvement réel de l'objet.

L'estimation locale du mouvement n'est possible que par l'addition de contraintes issues d'un voisinage spatial ou spatio-temporel (cas 3 de la figure 2.1), la solution obtenue est alors une moyenne du mouvement sur ce voisinage. Ce voisinage doit être suffisamment large pour contraindre la solution, sans recouvrir de régions contenant des mouvements différents (la solution serait alors une moyenne de tous ces mouvements).

Ainsi, en plus de la conservation de la luminance, une hypothèse de *continuité spatiale* du flot optique est nécessaire pour déterminer le mouvement. La situation la plus fréquente dans laquelle l'hypothèse de continuité spatiale du champ des vitesses n'est pas valide se rencontre à la frontière de deux régions animées de mouvements différents.

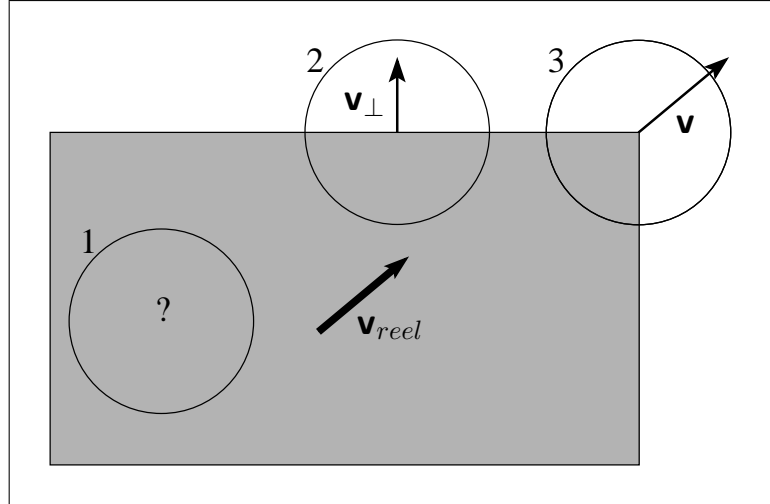


Fig. 2.1: Illustration du problème d'ouverture (voir texte)

2.2 Différentes méthodes d'estimation du mouvement

L'ensemble des techniques d'estimation du flot optique a l'obligation d'intégrer l'information sur un voisinage spatial et parfois spatio-temporel afin d'obtenir une mesure locale du mouvement. Ces techniques sont groupées en trois grandes familles, selon que la vitesse est estimée à partir de la luminance $I(\mathbf{p}_i, t)$, de la transformée de Fourier de $I(\mathbf{p}_i, t)$ ou des dérivées spatio-temporelles de $I(\mathbf{p}_i, t)$. Elles sont appelées respectivement méthodes de mise en correspondance, méthodes fréquentielles et méthodes différentielles.

2.2.1 Méthodes de mise en correspondance

L'approche de mise en correspondance ou *block matching* consiste à trouver le déplacement (r_x, r_y) qui apparie au mieux des régions ou des éléments caractéristiques (contours, coins, ...) de la scène entre deux instants consécutifs. L'appariement est en général calculé par une corrélation ou par une distance entre les régions de l'image aux instants t et $t+1$. La recherche du déplacement (r_x, r_y) s'effectue sur une gamme de valeurs discrètes.

Ces techniques, bien qu'imprécises (du fait de la discrétisation du déplacement estimé), sont robustes, simples à mettre en oeuvre et se retrouvent dans la majorité des standards de compression vidéo (MPEG1 et 2, H.261, H.263) [14, 37, 53, 54]. De plus, elles permettent de mesurer des déplacements de grande amplitude et sont très utilisées en stéréo-vision.

2.2.2 Méthodes fréquentielles

Considérons une image en espace et en temps $I(x, y, t)$ et sa transformée de Fourier $\hat{I}(f_x, f_y, f_t)$. Soit r_x et r_y respectivement les mouvements horizontaux et verticaux. La transformée de Fourier de l'image en mouvement est :

$$TF(I(x - r_x t, y - r_y t, t)) = \hat{I}(u, v, w + r_x u + r_y v) \quad (2.5)$$

Les fréquences spatiales sont inchangées, mais toutes les fréquences temporelles sont translatées par le produit de la vitesse et des fréquences spatiales. Il s'agit d'identifier dans l'espace des fréquences un plan de vitesse d'équation $w + r_x u + r_y v = 0$ pour retrouver les composantes r_x et r_y .

L'information de mouvement est en général extraite par des filtres orientés en espace et en temps (du type Gabor 3D ou filtre large-bande). La contrainte imposée pour éviter le problème d'ouverture est spatio-temporelle : le mouvement à déterminer est supposé constant à la fois sur le support spatial et temporel des filtres orientés. Le résultat obtenu est lissé en espace et en temps, ce qui peut poser des problèmes pour des séquences d'images où les actions sont rapides et saccadées.

On distingue deux approches : les méthodes basées sur l'énergie [30, 72, 80] et les méthodes exploitant la phase du signal [23]. Ces dernières permettent d'obtenir de bons résultats, la phase étant peu sensible aux variations d'illumination.

2.2.3 Méthodes différentielles

Dans le cas d'une séquence d'images numériques, la luminance $I(x, y, t)$ est temporellement échantillonnée selon une période δt . L'hypothèse de conservation de la luminance d'un point physique (x, y, t) se déplaçant de $\delta x = u\delta t$ selon x et de $\delta y = v\delta t$ selon y pendant un temps δt s'écrit :

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) \quad (2.6)$$

En développant le terme de droite de l'équation (2.6) autour du point (x, y, t) , on obtient [33] :

$$\frac{\partial I(x, y, t)}{\partial x} \delta x + \frac{\partial I(x, y, t)}{\partial y} \delta y + \frac{\partial I(x, y, t)}{\partial t} \delta t = O(\delta x^2 \delta y^2 \delta t^2) \quad (2.7)$$

Le terme de droite de l'équation (2.7) représente les termes d'ordres supérieur à 1 du développement de Taylor. Les dérivées partielles sont estimées par des filtres numériques. En posant $\delta t = 1$, cette équation peut s'écrire de façon compacte :

$$\nabla I(\mathbf{p}_i, t) \cdot \mathbf{v}(\mathbf{p}_i) + I_t(\mathbf{p}_i, t) = O(\|\mathbf{v}\|^2) \quad (2.8)$$

avec $\mathbf{p}_i = (x_i, y_i)^T$ la position des pixels de l'image et $\mathbf{v}(\mathbf{p}_i) = (u, v)^T$ le vecteur de vitesse à estimer.

Le développement (2.8) fournit une approximation au 1^{er} ordre de l'équation du flot optique (2.3). Cette approximation est valide tant que le terme $O(\|\mathbf{v}\|^2)$ est négligeable devant le terme du premier ordre. Ceci impose que les variations spatiales de la fonction de luminance dans le voisinage de (x, y, t) soient linéaires et que l'amplitude de la vitesse soit faible.

Les méthodes d'estimation de mouvement basées sur la formulation de Taylor de l'équation du flot optique (appelées méthodes différentielles) ont à résoudre deux problèmes essentiels : respecter les conditions de validité de l'équation (2.8) et résoudre le problème d'ouverture.

2.3 Estimation du mouvement par l'équation du flot optique au 1^{er} ordre

Les méthodes différentielles, basées sur la résolution de l'équation du flot optique, présentent de nombreux avantages face aux méthodes de mise en correspondance et fréquentielles. L'équation du flot optique, du fait de sa nature différentielle, permet une estimation sub-pixellique du mouvement (contrairement aux méthodes de mise en correspondance). La mesure du mouvement ne nécessite qu'un calcul local des dérivées spatio-temporelles de la séquence. Ces opérations ont un coût de calcul faible comparées aux filtrages spatio-temporels imposés par les méthodes fréquentielles. Ces deux avantages, associés au fait que l'équation du flot optique est linéaire par rapport au vecteur vitesse, expliquent le succès et le nombre très important de travaux portant sur les méthodes différentielles (par exemple [3, 5, 6, 33, 59, 71, 86]).

Néanmoins, comme nous l'avons décrit dans la section précédente, l'utilisation de l'équation du flot optique au 1^{er} ordre impose de respecter les conditions de validité liées au développement de Taylor et de résoudre le problème d'ouverture. Dans cette section, nous décrivons les techniques permettant de résoudre ces deux problèmes.

2.3.1 Validité de l'équation du flot optique : approches multiéchelle et multirésolution

Pour respecter les conditions de validité de l'équation (2.8), il est nécessaire de linéariser la fonction $I(x, y, t)$ est de réduire l'amplitude du vecteur de vitesse \mathbf{v} . Un sous-échantillonnage spatial de $I(x, y, t)$ permet à la fois de linéariser les variations de la luminance par l'opération de lissage, et de diminuer l'amplitude de la vitesse apparente par l'opération de sous-échantillonnage.

Cette constatation a motivé l'approche multirésolution. Celle-ci consiste à combiner le mouvement estimé à différentes échelles spatiales, afin d'obtenir à la fois une mesure des vitesses de grande amplitude et une bonne résolution spatiale du flot optique. L'approche multirésolution, qui consiste à travailler sur une décomposition pyramidale des images, est équivalente à l'approche multiéchelle [74] et permet de diminuer fortement le coût calculatoire de l'algorithme. La technique de raffinement de la mesure du mouvement à partir d'un niveau grossier jusqu'à un niveau fin de résolution est appelée *coarse to fine strategy*. Elle est largement utilisée [6, 5, 13] du fait de la solidité du cadre théorique dans lequel elle s'inscrit et de la précision de l'estimation qu'elle permet d'obtenir. Cette technique est discutée plus en détails dans le chapitre suivant.

2.3.2 Résolution du problème d'ouverture

Afin de résoudre le problème d'ouverture, inhérent à l'équation du flot optique, il est nécessaire d'ajouter des contraintes (spatiales ou temporelles) à cette équation pour lever

l'ambiguïté sur la mesure du mouvement. Ce dernier point fait l'objet d'une intense recherche et un grand nombre de solutions originales a été proposé.

Régularisation explicite

Une méthode de régularisation, proposée par Horn et Schunck [33], consiste à introduire une contrainte de lissage sur le champ des vitesses estimé en ajoutant un terme de régularisation à l'équation du flot optique. L'estimation du mouvement se fait alors en minimisant une fonction de coût E :

$$E = \sum_{\mathbf{p}_i \in \Omega} ((\nabla I(\mathbf{p}_i, t) \cdot \mathbf{v}(\mathbf{p}_i) + I_t(\mathbf{p}_i, t))^2 + \lambda \|\nabla \mathbf{v}(\mathbf{p}_i)\|^2) \quad (2.9)$$

où Ω est le support d'estimation et correspond en général à toute l'image. La minimisation de E , relativement au champ des vitesses \mathbf{v} , conduit à résoudre un système dont les inconnues sont les vecteurs de vitesses en chaque pixel de l'image. Le terme de régularisation $\lambda \|\nabla \mathbf{v}(\mathbf{p}_i)\|^2$ permet de réduire la dimension de l'espace des solutions en favorisant les solutions les plus régulières.

Bien que l'optimisation soit globale, le terme de régularisation est calculée sur un voisinage local (le support de $\nabla \mathbf{v}(\mathbf{p}_i)$ dans l'exemple ci-dessus). Le résultat obtenu est alors une estimation locale du flot optique.

De nombreux chercheurs ont reformulé le terme de régularisation afin de prendre en compte les discontinuités du mouvement par l'introduction par exemple de champs de Markov [7, 31] ou d'estimateurs robustes [9, 49].

Régularisation implicite

Une régularisation implicite est fondée sur l'hypothèse d'un mouvement homogène sur une région de l'image. Par exemple, l'approche proposée par Lucas et Kanade [44] consiste à supposer le flot optique constant sur une fenêtre spatiale Ω :

$$\mathbf{v}(\mathbf{p}_i) = (u_0, v_0)^T \quad \forall \mathbf{p}_i \in \Omega \quad (2.10)$$

L'estimation de mouvement au point \mathbf{p}_i consiste alors en l'identification des deux paramètres (u_0, v_0) . Cette estimation est obtenue par la minimisation d'une erreur quadratique définie sur le support Ω :

$$\mathbf{v} = \arg \min_{\mathbf{v}} \sum_{\mathbf{p}_i \in \Omega} [\nabla I(\mathbf{p}_i, t) \cdot \mathbf{v}(\mathbf{p}_i) + I_t(\mathbf{p}_i, t)]^2 \quad (2.11)$$

L'erreur quadratique peut être remplacée par une norme robuste d'erreur [6] de manière à rendre l'estimation invariante aux données aberrantes. Pratiquement, cette technique permet de résoudre en chaque pixel \mathbf{p}_i un système surdéterminé d'équations du flot optique (le nombre d'équations est égal au nombre de pixels dans le voisinage Ω) et dont la solution est le vecteur vitesse moyen de la région Ω .

Lorsque la taille du support Ω est importante, l'hypothèse de la vitesse constante n'est généralement pas vérifiée. Il est alors possible de relâcher cette contrainte par l'introduction de modèles plus complexes, tels que des transformations affines et quadratiques du plan. Ces méthodes sont plus largement discutées dans le chapitre 4.

Régularisation basée sur une décomposition multi-bandes de la séquence d'images

Une troisième approche consiste à appliquer l'équation du flot optique non pas directement sur la séquence $I(t)$, mais sur une décomposition en sous-bandes de la séquence d'images [5, 52, 71, 81, 86]. Considérons la convolution en un pixel \mathbf{p}_0 de l'équation du flot optique (2.8) par un banc de filtres $G_k(\mathbf{p}_i)$ (avec $k = 1, \dots, N$, et N le nombre de filtres)

$$\int \int (\nabla I(\mathbf{p}_i, t) \cdot \mathbf{v}(\mathbf{p}_i) + I_t(\mathbf{p}_i, t)) G_k(\mathbf{p}_i - \mathbf{p}_0) d\mathbf{p}_i = 0 \quad (2.12)$$

En faisant l'hypothèse que \mathbf{v} est constant sur le support du filtre G_k , les composantes de vitesse peuvent être extraites du produit de convolution. On obtient alors :

$$\begin{aligned} (\nabla I * G_1) \cdot \mathbf{v}(\mathbf{p}_0) + I_t * G_1 &= 0 \\ (\nabla I * G_2) \cdot \mathbf{v}(\mathbf{p}_0) + I_t * G_2 &= 0 \\ &\vdots \\ (\nabla I * G_N) \cdot \mathbf{v}(\mathbf{p}_0) + I_t * G_N &= 0 \end{aligned} \quad (2.13)$$

où $*$ est l'opérateur de convolution 2D. Cette technique fournit en chaque pixel de la séquence N équations de mouvement, et permet de remonter au flot optique en résolvant localement le système d'équations (2.13). La solution au sens des moindres-carrés est alors :

$$\mathbf{v} = \arg \min_{\mathbf{v}} \sum_{k=0}^N [(\nabla I * G_k) \cdot \mathbf{v} + I_t * G_k]^2 \quad (2.14)$$

De même que pour l'approche de Lucas et Kanade (2.11), la contrainte supplémentaire permettant de résoudre le problème d'ouverture est apportée par le voisinage spatial (ou spatio-temporel dans le cas de filtres spatio-temporels). Ainsi le système (2.13) n'est valide que si les variations du flot optique sont faibles par rapport au support du filtre G_k .

L'avantage de cette dernière approche, comme le montrent Weber & Malik [86], est d'obtenir localement un système de N équations du flot optique mieux conditionné que celui obtenu sur un voisinage spatial de N pixels. Les algorithmes basés sur cette technique sont performants, et dans le cas où l'implantation du banc de filtres est rapide, l'estimation du mouvement est une opération peu complexe. On peut citer en exemple deux algorithmes illustrant l'efficacité de cette approche :

Weber & Malik [86] ont développé un algorithme dont les performances sont comparables aux meilleures approches d'estimation du mouvement. La décomposition en sous-bandes utilisée est obtenue par des dérivées de gaussiennes d'ordre 1 et 2.

Récemment, Bernard & Mallat [5] ont proposé un algorithme utilisant une base d'ondelettes discrètes comme banc de filtres. La décomposition en sous-bandes est obtenue par un algorithme de transformée en ondelettes rapide ce qui permet de diminuer la complexité et le temps de calcul nécessaire à l'estimation du mouvement.

2.4 Utilisation de bancs de filtres en vision par ordinateur

La décomposition multi-bandes d'images ou de séquence d'images est une opération très répandue en vision par ordinateur et traitement d'images. Les bancs de filtres sont utilisés pour la reconnaissance par apparence locale (reconnaissance d'activités [16], d'objets [18], de visages [34, 45, 88, 89]), la segmentation de textures [32, 38], la classification de scènes [29, 63] ou encore la détection de points d'intérêt [48, 42].

Le développement d'un algorithme dont l'architecture centrale est basée sur un banc de filtres (figure 2.2) pourrait réaliser en parallèle toutes ces tâches, et permettrait d'obtenir une information de haut-niveau sur la scène analysée [1].

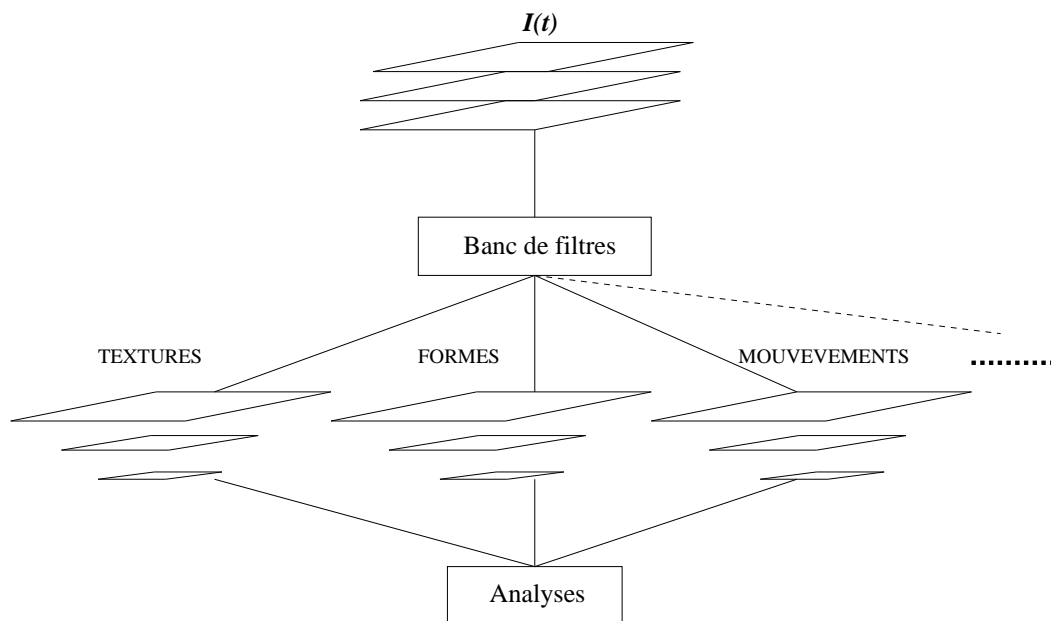


Fig. 2.2: Architecture de vision par ordinateur basée sur un banc de filtres. L'extraction de différents attributs visuels par un banc de filtres, puis leur analyse, apportent une information de haut-niveau sur la scène.

Les filtres de Gabor sont largement utilisés pour toutes ces applications. Ce succès est dû à la fois à leurs propriétés spatiales et fréquentielles (ils permettent un échantillonnage du spectre en échelles et en orientations tout en atteignant la borne d'Heisenberg) et à leur similarité avec la réponse impulsionnelle des cellules simples rencontrées dans le cortex visuel du primate [39]. Aussi, de nombreuses recherches ont été menées sur les filtres de Gabor en tant qu'outils de perception d'attributs visuels bas-niveau (forme, texture, contour, cartes

de saillances). Dans le cadre de ces travaux, et afin d'enrichir les possibilités perceptives des filtres de Gabor spatiaux, nous nous sommes intéressés au problème de la mesure du mouvement par ces filtres.

Chapitre 3

Estimation locale du mouvement par banc de filtres de Gabor récursifs

3.1 Introduction

Dans ce chapitre nous présentons une approche originale pour estimer le mouvement à partir d'une décomposition de la séquence d'images par des bancs de filtres de Gabor récursifs. Ces filtres sont une *approximation* de filtres de Gabor réels dont l'implantation est rapide. L'estimation de la vitesse est obtenue par résolution robuste du système local d'équations du flot optique (3.3). Cette estimation est incorporée dans un schéma multirésolution permettant la mesure de déplacements importants. Les performances de l'algorithme sont évaluées sur des séquences artificielles et réelles.

3.2 Position du problème

L'estimateur de mouvement développé dans cette partie est basé sur l'équation (2.13) présentée au chapitre précédent. Rappelons que cette équation s'écrit :

$$u \left(\frac{\partial I}{\partial x} * G_i \right) + v \left(\frac{\partial I}{\partial y} * G_i \right) + \frac{\partial I}{\partial t} * G_i = 0 \quad i = 1 \dots N. \quad (3.1)$$

La convolution des dérivées spatio-temporelles de la séquence d'images par un banc de N filtres G_i donne pour chaque pixel de l'image un système à N équations dont les deux inconnues sont les composantes du vecteur vitesse (u, v) . Une étape supplémentaire, qui consiste à permuter l'opérateur de convolution avec l'opérateur de dérivation, permet de réduire le nombre de filtrages :

$$u \frac{\partial(I * G_i)}{\partial x} + v \frac{\partial(I * G_i)}{\partial y} + \frac{\partial(I * G_i)}{\partial t} = 0 \quad i = 1 \dots N. \quad (3.2)$$

Dans ce cas, l'opération de filtrage porte directement sur l'image $I(t)$, et non plus sur les dérivées spatio-temporelles.

Ainsi, en chaque pixel de la séquence d'images nous obtenons le système suivant :

$$\underbrace{\begin{pmatrix} \Omega_1^x & \Omega_1^y \\ \Omega_2^x & \Omega_2^y \\ \vdots & \vdots \\ \Omega_N^x & \Omega_N^y \end{pmatrix}}_M \cdot \underbrace{\begin{pmatrix} u \\ v \end{pmatrix}}_{\mathbf{v}} = - \underbrace{\begin{pmatrix} \Omega_1^t \\ \Omega_2^t \\ \vdots \\ \Omega_N^t \end{pmatrix}}_B \quad (3.3)$$

où $\Omega_i^p = \frac{\partial(I * G_i)}{\partial p}$

Une solution stable est obtenue si le système (3.3) est sur-contraint et bien conditionné. De plus, le support spatial des filtres G_i ne doit pas être trop étendu afin de respecter l'hypothèse d'un flot optique localement constant. Ces conditions impliquent d'utiliser des filtres à support le plus étroit possible permettant d'obtenir une décomposition de l'image en sous-bandes le moins corrélées possibles, *i.e.* des filtres minimisant l'incertitude de Heisenberg.

Dans ce but, différents bancs de filtres ont été proposés : Weber & Malik [86] utilisent des dérivées de gaussiennes spatio-temporelles, la méthode de Bernard & Mallat [5] est basée sur une décomposition en ondelettes de l'image. Nous pouvons citer encore Tsao & Chen [81] (filtres de Gabor), Simoncelli [71] (filtres gaussiens) ou Mitiche *et al.* [52] (opérateurs spatiaux variés). Plusieurs enseignements peuvent être tirés de ces approches :

- Une redondance entre les filtres permet d'être moins sensible au bruit [5].
- L'utilisation de filtres complexes plutôt que réels permet d'obtenir une estimation plus stable de la vitesse [5].
- La convolution d'une image par un banc de filtres est une opération coûteuse en temps de calcul. Il est préférable d'utiliser une implémentation rapide de ces filtres.

Les filtres de Gabor, utilisés par Tsao & Chen, répondent à la majorité de ces contraintes :

- ils sont complexes,
- ils minimisent l'incertitude de Heisenberg,
- ils effectuent une analyse de l'image en échelles et en orientations, ce qui autorise une paramétrisation optimale des filtres (contrairement aux dérivées de gaussiennes, qui ne permettent qu'une analyse en échelle).

Néanmoins, le temps nécessaire à l'étape de filtrage est très important et est un frein à l'utilisation intensive d'un tel banc de filtres pour l'estimation du mouvement.

3.3 Les bancs de filtres de Gabor et leur implantation récursive

Une fonction de Gabor se définit comme une fonction gaussienne modulée par une onde sinusoïdale :

$$G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}} e^{j2\pi(xf_{x0} + yf_{y0})} \quad (3.4)$$

avec σ_x et σ_y la largeur respectivement selon x et y de la fonction gaussienne et f_{x_0} et f_{y_0} les fréquences spatiales de la modulation. Ce filtre, de type passe-bande orienté, a une réponse impulsionnelle complexe (figure 3.1.a et b). La fréquence centrale f_0 est donnée par $\sqrt{f_{x_0}^2 + f_{y_0}^2}$, l'orientation θ par $\arctan(f_{x_0}/f_{y_0})$ et la largeur de bande par σ_x et σ_y . La réponse fréquentielle du filtre de Gabor est (figure 3.1.c) :

$$TF(G)(f_x, f_y) = e^{-2\pi^2(\sigma_x^2(f_x - f_{x_0})^2 + \sigma_y^2(f_y - f_{y_0})^2)} \quad (3.5)$$

Les filtres de Gabor sont d'autant plus sélectifs en fréquence que leur support spatial est large, et inversement. Considérons un banc de N filtres de Gabor où les filtres sont isotropes ($\sigma = \sigma_x = \sigma_y$), positionnés sur une couronne définie par la fréquence centrale f_0 , chaque filtre ayant pour orientation $\theta_i = \frac{i\pi}{N}$ (figure 3.2). Ce banc de filtres est noté $G_{f_0, \sigma, N}$. Le filtre de la i^{eme} sous-bande a pour réponse impulsionnelle $G_i(x, y)$:

$$G_i(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}} e^{j2\pi f_0(x \cos \theta_i + y \sin \theta_i)} \quad (3.6)$$

Le réglage des paramètres f_0 , σ et N est déterminant pour estimer de façon précise le mouvement. Ces trois paramètres influent en particulier sur le calcul des gradients de l'image, sur le conditionnement du système d'équations du flot optique, ainsi que sur le coût calculatoire de l'opération.

3.3.1 Construction du banc de filtres

Le choix des paramètres f_0 , σ et N est un problème important, car il détermine la stabilité du système d'équations du flot optique (3.3) et impose la contrainte d'un flot optique constant sur un voisinage plus ou moins grand. Supposons deux filtres de Gabor (3.6) ayant la même largeur σ , la même fréquence centrale f_0 et positionnés aux orientations θ_1 et θ_2 . La corrélation entre ces deux sous-bandes est :

$$C(\delta\theta) = \frac{\langle G_1, G_2 \rangle}{\|G\|^2} = \exp(-2\pi^2 f_0^2 \sigma^2 (1 - \cos(\delta\theta))), \quad (3.7)$$

avec $\delta\theta = \theta_2 - \theta_1 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

Le minimum de C est atteint lorsque $\sigma \rightarrow \infty$, $f_0 \rightarrow \infty$ et $N = 2$ ($\delta\theta = \frac{\pi}{2}$). Ces bornes ne sont évidemment pas en accord avec les contraintes imposées par notre problème :

- Plus σ est grand, plus C tend vers zéro, et plus large est le domaine où le mouvement est supposé localement constant. L'utilisation de filtres à support spatial large ne permet d'estimer qu'un flot optique très lissé. A l'inverse, lorsque σ est petit, il est possible de mesurer des variations rapides du mouvement, mais la résolution du système (3.3) est très instable. Le choix de σ est un compromis entre la stabilité de l'estimation et la résolution spatiale du flot optique. Un certain nombre d'auteurs [3, 9] suggèrent qu'une fenêtre spatiale de taille 5×5 pixels est une limite inférieure pour estimer un vecteur de mouvement de façon fiable (pour une séquence d'images naturelles). En respectant cette limite, σ doit se situer dans l'intervalle [2.0, 4.0].

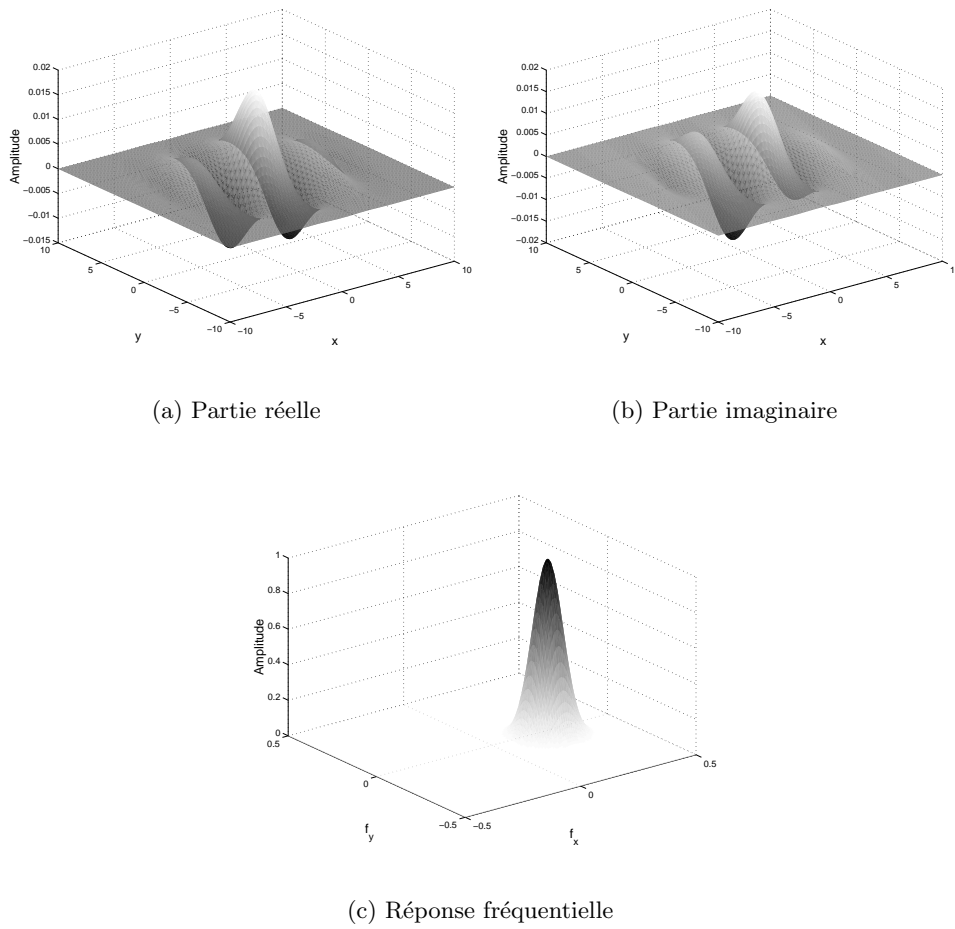


Fig. 3.1: Réponse impulsionnelle (réelle et imaginaire) et fréquentielle d'un filtre de Gabor

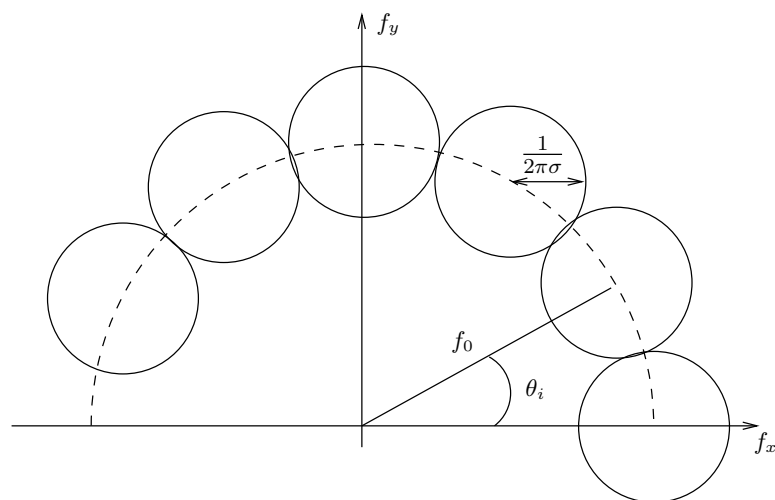


Fig. 3.2: Un banc de filtres de Gabor est décrit par la fréquence centrale f_0 , la largeur σ des filtres et le nombre de filtres N ($\theta_i = \frac{i\pi}{N}$)

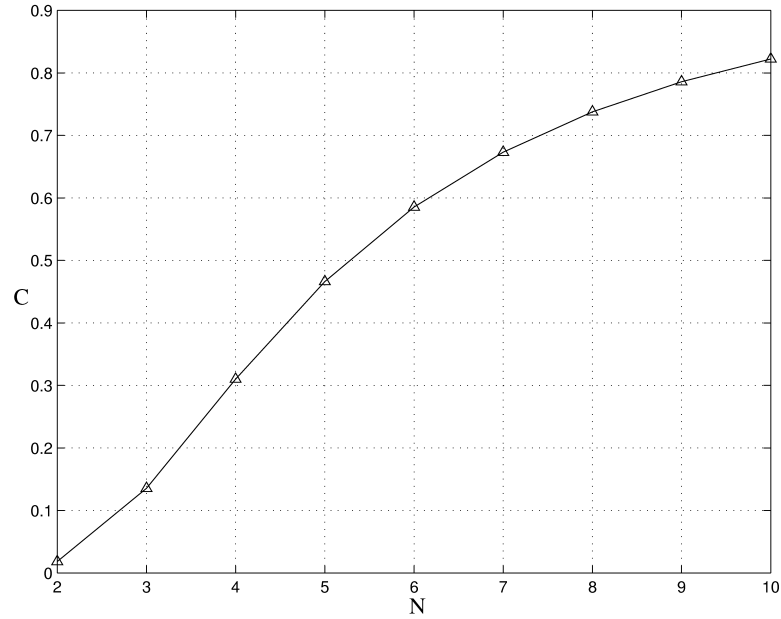


Fig. 3.3: Corrélation C en fonction du nombre de filtres N avec $f_0 = 0.15$ et $\sigma = 3.0$

- La fréquence centrale f_0 doit être inférieure à la limite imposée par l'estimation des dérivées spatiales. Le gradient de l'image est calculé par le masque $[-1 \ 8 \ 0 \ -8 \ 1]/12$. L'estimation de la dérivée est alors très bonne pour des fréquences spatiales inférieures à 0.2. En considérant la largeur fréquentielle d'un filtre de Gabor (défini par σ), la fréquence centrale f_0 doit être inférieure à 0.15
- Chaque filtre G_i apporte une contrainte supplémentaire au système (3.3) et renforce la stabilité de l'estimation. Néanmoins, car la corrélation (3.7) tend vers 1 quand N croît, l'ajout de plusieurs sous-bandes n'apporte pas plus d'informations au système (3.3) et est coûteux en temps de calcul. La figure 3.3 représente l'évolution de la corrélation en fonction du nombre de filtres N , lorsque $\sigma = 3.0$ et $f_0 = 0.15$. En considérant que si $C > 0.5$, les sous-bandes sont très corrélées, environ 5 ou 6 filtres contraignent suffisamment le système (3.3) pour en estimer la solution.

En résumé, le banc de filtres de Gabor utilisé est paramétré comme suit :

$$\sigma \in [2.0, 4.0], \quad f_0 < 0.15, \quad N \in [3, 6] \quad (3.8)$$

La figure 3.4 représente la fonction de transfert d'un filtre de Gabor dont les paramètres se trouvent dans l'intervalle défini par (3.8).

3.3.2 Suppression de la composante continue au moyen d'un préfiltrage passe-haut

La fonction de transfert des filtres de Gabor définie ci-dessus n'est pas nulle en zéro. L'ensemble des filtres d'un banc $G_{f_0, \sigma, N}$ contient ainsi la même information située dans les

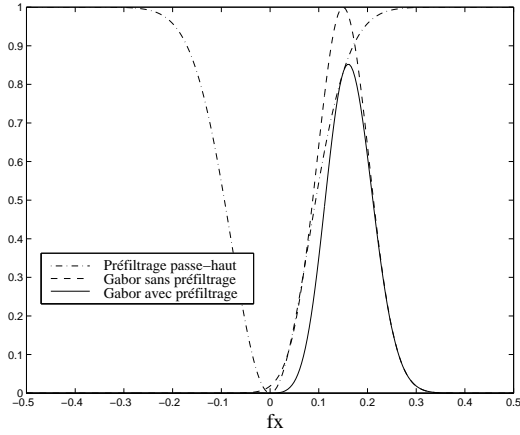


Fig. 3.4: Réponse fréquentielle d'un filtre de Gabor avec et sans préfiltrage passe haut (avec $f_0 = 0.15$, $\sigma = 3.0$ pour le filtre de Gabor et $\sigma_1 = 2.0$ pour la gaussienne inversée)

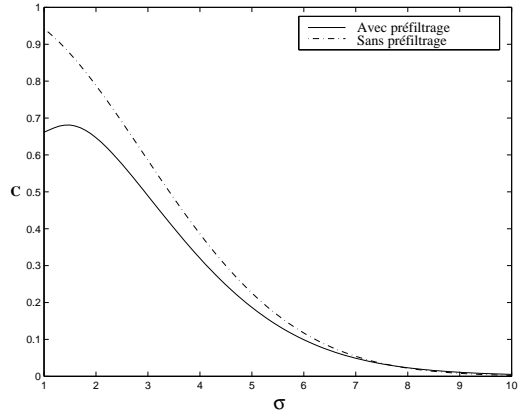


Fig. 3.5: Corrélation C en fonction de σ ($f_0 = 0.15$, $N = 6$, filtres récursifs du 3^{ème} ordre) avec et sans préfiltrage passe-haut

basses fréquences, ce qui rend le système (3.3) mal conditionné. Cet effet est renforcé par le fait que le spectre d'une image naturelle décroît en $1/f$. De plus, il est à noter que les variations globales de luminosité sont de type basses fréquences, et qu'il est nécessaire de les supprimer pour rester dans l'hypothèse de conservation de la luminosité.

Un simple préfiltrage des images par un filtre passe-haut permet de résoudre ces deux problèmes. Nous utilisons pour cela un filtrage gaussien renversé de largeur $\sigma = 2$ (figure 3.4). L'implantation récursive est utilisée pour effectuer ce préfiltrage. La figure 3.5 représente la corrélation entre deux filtres de Gabor en fonction de leur largeur σ , lorsque le signal est ou n'est pas préfiltré. Le préfiltrage permet effectivement de diminuer la corrélation entre les sous-bandes lorsque σ tend vers zéro.

3.3.3 Implantation récursive des filtres de Gabor

Classiquement, le filtrage d'une image par un tel banc de filtres peut s'effectuer par convolution spatiale ou par multiplication dans l'espace fréquentielle de l'image avec chaque filtre de Gabor. Dans les deux cas, l'opération est coûteuse en temps de calcul. Spinei [72] propose une implantation rapide basée sur un filtrage récursif permettant d'approximer les filtres de Gabor.

Soit $y(k)$ le produit de convolution d'un signal $x(k)$ par un filtre de Gabor à une dimension :

$$y(k) = x(k) * \left[K e^{-\frac{k^2}{2\sigma^2}} e^{j2\pi f_0 k} \right], \quad (3.9)$$

où K est une constante de normalisation. En développant la relation (3.9), on peut montrer

que :

$$y(k) = \left(\left[x(k)e^{-j2\pi f_0 k} \right] * K e^{-\frac{k^2}{2\sigma^2}} \right) e^{j2\pi f_0 k}. \quad (3.10)$$

Le filtre de Gabor peut être décomposé en trois étapes successives : une modulation, un filtrage passe-bas gaussien et une démodulation.

Le filtrage gaussien peut être approximé par un filtre récursif. On peut ainsi réduire la complexité de cette étape. Vliet & Young [83] proposent une technique permettant d'obtenir des filtres récursifs d'ordre N qui approximent au mieux l'enveloppe gaussienne. La fonction de transfert de ces filtres peut se décomposer en deux sous-systèmes : un causal et un anti-causal

$$H(z) = H_+(z) \cdot H_-(z). \quad (3.11)$$

L'implantation de H_+ et H_- s'effectue par deux équations aux différences respectivement causale et anti-causale :

$$\begin{aligned} v[n] &= \alpha x[n] - \sum_{i=1}^N b_i v[n-i] \\ y[n] &= \alpha v[n] - \sum_{i=1}^N b_i y[n+i], \end{aligned} \quad (3.12)$$

avec $x[n]$ le signal d'entrée et $y[n]$ de sortie. Les coefficients α et b_i sont calculés en fonction de la largeur du filtre choisie par les formules décrites dans [83].

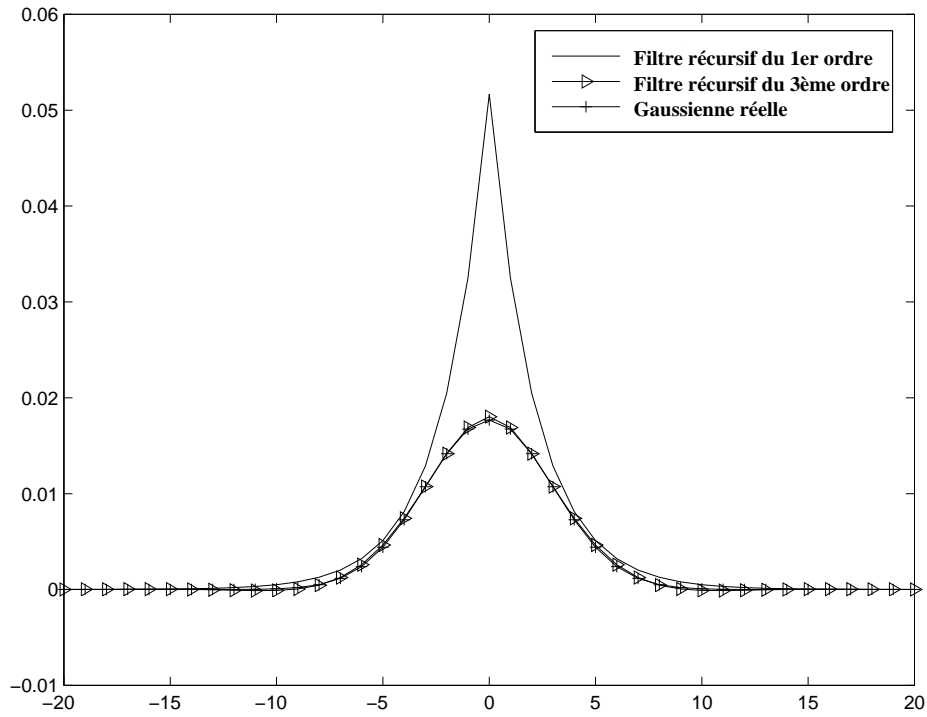
Le nombre d'opérations par pixel nécessaire au filtrage gaussien récursif est constant quelle que soit la sélectivité du filtre. Le temps de calcul dépend uniquement de l'ordre du filtre récursif.

Nous avons testé les filtres récursifs d'ordre 1 et 3 (figure 3.6). Les filtres d'ordre 3 sont une très bonne approximation du filtrage gaussien réel. A l'opposé, l'approximation à l'ordre 1 est moins bonne mais peu coûteuse en temps de calcul. On peut noter que B.E. Shi [69] a proposé une implantation analogique (circuit VLSI) des filtres de Gabor approximatés à l'ordre 1.

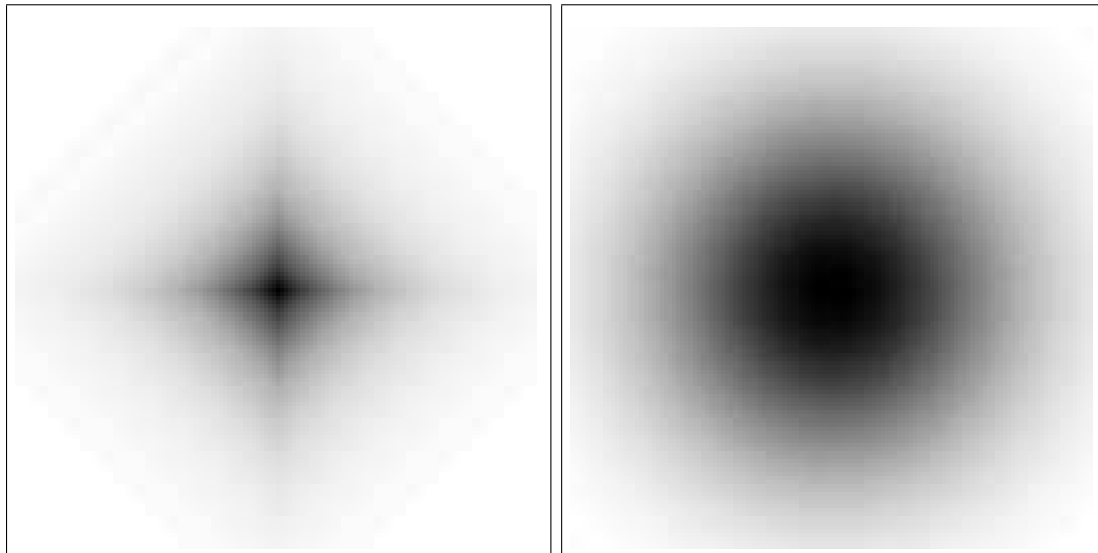
Les paramètres que nous avons déterminés en section 3.3.1 concernent un banc de filtres de Gabor réel. Ainsi, on peut considérer que si ces réglages restent corrects pour l'approximation au 3^{ème} ordre, ils le sont beaucoup moins pour un filtre récursif du 1^{er} ordre. Une étude expérimentale est menée dans la section 3.6.1 afin d'adapter les valeurs théoriques des paramètres f_0 , σ et N aux filtres de Gabor récursifs.

3.4 Résolution robuste du système surdéterminé

L'équation du flot optique (2.8) appliquée sur chaque sous-bande de Gabor de la séquence d'images, produit pour chaque pixel \mathbf{p}_i le système surdéterminé (3.3). Les réglages du banc de filtres, étudiés ci-dessus, doivent permettre au système d'être bien conditionné. Néanmoins, lorsque que la texture de l'image n'est pas assez riche, que l'hypothèse de la conservation de la



(a) Réponses impulsionnelles 1D des filtres récurrents d'ordre 1, 3 et de la gaussienne



(b) Réponses impulsionnelles 2D des filtres récurrents d'ordre 1 et 3

Fig. 3.6: Approximation récursive d'une gaussienne

luminance n'est pas respectée, ou encore que le flot optique n'est absolument pas constant sur le support des filtres, la solution du système est erronée. Le problème est donc d'estimer de façon précise et surtout robuste le champ de vitesses à partir des dérivées spatio-temporelles des sous-bandes de Gabor.

Les équations du système (3.3) sont à coefficients complexes (les filtres de Gabor étant complexes) et la solution \mathbf{v} est réelle. Le système linéaire (3.3) peut alors s'écrire comme un système réel [5] :

$$\underbrace{\begin{bmatrix} \text{Re}(M) \\ \text{Im}(M) \end{bmatrix}}_A \mathbf{v} = - \underbrace{\begin{bmatrix} \text{Re}(B) \\ \text{Im}(B) \end{bmatrix}}_Y \quad (3.13)$$

La matrice A est réelle de taille $2N \times 2$, $\mathbf{v} = (u, v)^T$ est le vecteur vitesse recherché et Y est un vecteur réel de taille $2N \times 1$. N est le nombre de filtres de Gabor utilisés. La solution au sens des moindres-carrés d'un tel système surdéterminé est obtenue par minimisation d'une erreur quadratique :

$$\mathbf{v} = \arg \min_{\mathbf{v}} \sum_i r_i^2 \quad (3.14)$$

avec $r_i = uA_{i,1} + vA_{i,2} + Y_i$ le résidu de l'équation du flot optique associé à la sous-bande i . L'erreur quadratique (3.14) prend en compte tous les résidus r_i de manière identique, ce qui ne garantit pas un comportement robuste. Il est important de diminuer le poids de données "aberrantes" dans le processus de minimisation en remplaçant la norme quadratique par une mesure robuste de l'erreur, tel qu'un M-estimateur (voir annexe A). La solution robuste du système (3.13) est alors :

$$\mathbf{v} = \arg \min_{\mathbf{v}} \sum_i \rho(r_i) \quad (3.15)$$

La minimisation (3.15) s'effectue de manière itérative par l'algorithme des moindres carrés pondérés itérés (MCPI) détaillé dans l'annexe A :

$$\mathbf{v} = \arg \min_{\mathbf{v}} \sum_i w_i r_i^2 \quad (3.16)$$

avec $w(x) = \frac{1}{x} \frac{\partial \rho(x)}{\partial x}$ le poids attribué au résidu r_i . La minimisation est menée de manière alternée : estimation des poids w_i puis estimation de \mathbf{v} (pour la première itération, l'ensemble des poids w_i est égal à 1).

Afin de garantir une estimation robuste de la vitesse, la solution doit être calculée uniquement à partir des équations valides. Pour cela, nous utilisons le M-estimateur "Biweight de Tukey" (figure 3.7) qui permet d'affecter un poids nul aux données aberrantes :

$$\rho(x) = \begin{cases} \frac{C^2}{6} (1 - (1 - (\frac{x}{C})^2)^3) & \text{si } |x| \leq C \\ \frac{C^2}{6} & \text{si } |x| > C \end{cases} \quad (3.17)$$

et :

$$w(x) = \begin{cases} (\frac{x^2 - C^2}{C^2})^2 & \text{si } |x| \leq C \\ 0 & \text{si } |x| > C \end{cases} \quad (3.18)$$

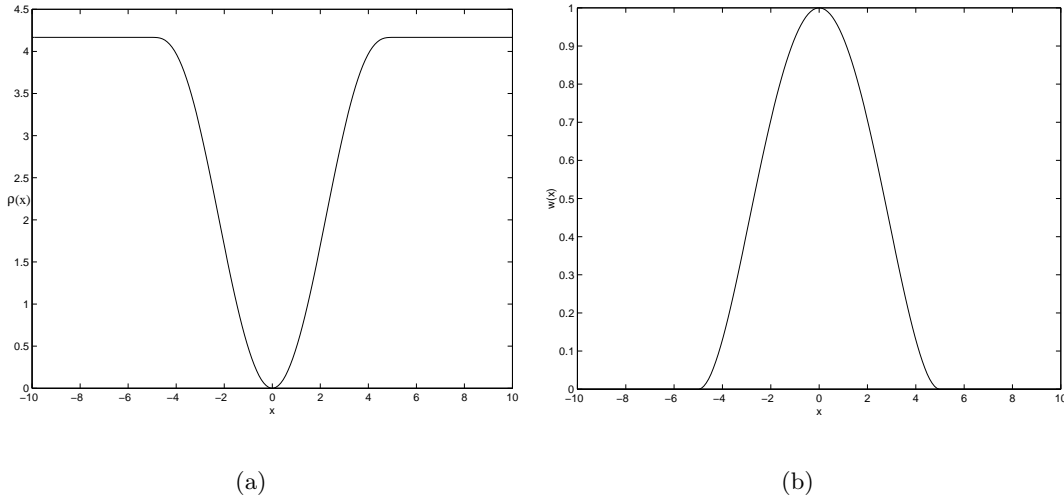


Fig. 3.7: Fonction “*Biweight de Tukey*” avec a) la fonction $\rho(x)$ et b) la fonction poids $w(x)$ associée.

Les données dont le résidu r_i est supérieur à la valeur de C , appelée point de rejet, sont supprimées du procédé d’estimation. On peut assimiler C à un écart-type et utiliser l’estimateur correspondant, mais Tukey propose un estimateur plus robuste, la médiane des écarts absolus qui vaut

$$\alpha = 1.4826 \operatorname{med}_i(|r_i - \operatorname{med}_j(r_j)|) \quad (3.19)$$

tel que $C = \lambda\alpha$ où λ est un facteur de proportionnalité qui détermine la robustesse de l’estimateur. Si $\lambda = 2.795$, l’estimateur de Tukey a le même comportement qu’un algorithme des moindres-carrés en présence de bruit gaussien.

En résumé, l’estimation robuste de la vitesse à chaque itération s’écrit (en utilisant les notations matricielles) :

$$\mathbf{v} = -(A^T W A)^{-1} A^T W Y \quad (3.20)$$

avec

$$W = \operatorname{diag}(w_i) \quad (3.21)$$

Un simple test sur les valeurs propres de la matrice $A^T W A$ permet de s’assurer que le système (3.13) est bien de rang plein. Si $\frac{\lambda_2}{\lambda_1} < \epsilon$ (où λ_2 est la plus petite valeur propre), le système n’est pas inversible, et aucune vitesse n’est estimée à cet endroit.

3.5 Estimation progressive du mouvement par l’approche multirésolution

L’équation du flot optique en tant que développement de Taylor est valide dans le cas où la fonction de luminance peut être supposée localement linéaire, et/ou lorsque le déplacement

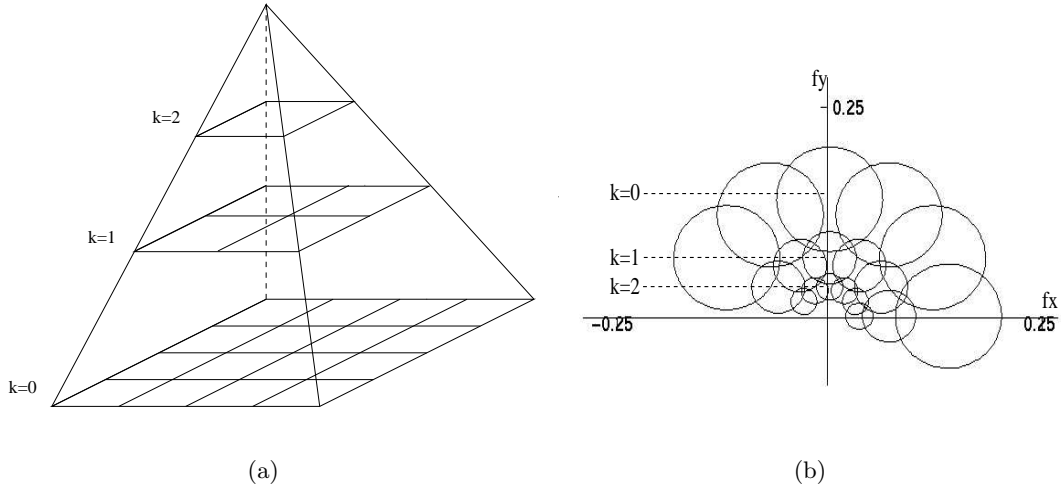


Fig. 3.8: a) Pyramide d'images passe-bas à 3 niveaux et b) le banc de filtres de Gabor équivalent

\mathbf{v} est petit. L'hypothèse de planéité de la fonction de luminance $I(\mathbf{p}_i, t)$ ne peut être respectée (il existe énormément de discontinuités, de textures plus ou moins riches dans une image naturelle), et l'on considère que l'équation du flot optique ne permet d'estimer que des déplacements de faible amplitude.

Dans la section 2.2.3 du chapitre 2, nous avons montré que la notion de petits ou grands déplacements est relative à la taille du filtre dérivateur utilisé pour approximer les dérivées spatiales de l'image. Ainsi, pour estimer des vitesses de grande amplitude, il est nécessaire d'utiliser un filtre à support très large, ou de manière équivalente, de travailler sur une version sous-échantillonnée de l'image. En contrepartie, le flot optique obtenu sera spatialement très lissé.

Une technique largement répandue en analyse de mouvement [5, 6, 13, 86] consiste à combiner l'information issue des images de la séquence à différentes résolutions spatiales : une approximation grossière du flot optique est estimée à un niveau de résolution faible. Le déplacement entre les images est alors compensé par cette estimation, afin que le mouvement résiduel puisse être mesuré à une résolution plus fine. Ce procédé est itéré jusqu'à atteindre le niveau de résolution original.

L'implantation de cette technique passe par la construction d'une pyramide d'images passe-bas à partir des images de la séquence (figure 3.8.a). Notons que l'approche multirésolution est équivalente à une approche multiéchelle où nous utiliserions un banc de filtres composé de plusieurs couronnes de filtres de Gabor $G_{2^{-k}f_0, 2^k\sigma, N}$ correspondant aux différents niveaux de pyramide k (figure 3.8.b). L'avantage de l'approche multirésolution est de diminuer le coût calculatoire de l'algorithme. Soit $\mathbf{v}_k = (u_k, v_k)^T$ le flot optique estimé au niveau k de la pyramide d'images. Le mouvement réel peut s'écrire comme la somme de \mathbf{v}_k et d'un

mouvement résiduel $\delta\mathbf{v}_k$ à déterminer :

$$\mathbf{v} = \mathbf{v}_k + \delta\mathbf{v}_k \quad (3.22)$$

Connaissant \mathbf{v}_k , l'hypothèse de conservation de la luminance s'écrit :

$$I(\mathbf{p}_i + \mathbf{v}_k + \delta\mathbf{v}_k, t + 1) - I(\mathbf{p}_i, t) = 0 \quad (3.23)$$

A nouveau, l'équation (3.23) est linéarisée par un développement de Taylor, mais cette fois-ci autour de $\mathbf{p}_i + \mathbf{v}_k$:

$$\nabla I(\mathbf{p}_i + \mathbf{v}_k, t + 1) \cdot \delta\mathbf{v}_k + I_t(\mathbf{p}_i + \mathbf{v}_k, t) = 0 \quad (3.24)$$

où $I_t(\mathbf{p}_i + \mathbf{v}_k) = I(\mathbf{p}_i + \mathbf{v}_k, t + 1) - I(\mathbf{p}_i, t)$. On retrouve l'équation du flot optique, non plus appliquée directement entre les deux images $I(t + 1)$ et $I(t)$, mais entre l'image $I(t + 1)$ compensée par le mouvement courant \mathbf{v}_k et l'image $I(t)$. Comme précédemment, le déplacement résiduel $\delta\mathbf{v}_k$ est obtenu par la résolution robuste du système

$$\nabla (I(\mathbf{p}_i + \mathbf{v}_k, t + 1) * G_i) \cdot \delta\mathbf{v}_k + I_t(\mathbf{p}_i + \mathbf{v}_k, t) * G_i = 0 \quad (3.25)$$

L'image compensée $I(\mathbf{p}_i + \mathbf{v}_k, t + 1)$ est calculée à partir de l'image originale $I(\mathbf{p}_i, t + 1)$ par interpolation bilinéaire (\mathbf{v}_k n'étant pas une grandeur entière). Cette estimation incrémentale du mouvement est effectuée du sommet à la base de la pyramide. Le flot optique estimé à un niveau de résolution l est la contribution des champs de vitesses résiduels estimés au niveau de pyramide supérieur :

$$\mathbf{v}_l = \mathbf{v}_0 + \sum_{k=N}^l \delta\mathbf{v}_k \quad (3.26)$$

où N est le nombre de niveaux de pyramide utilisés et \mathbf{v}_0 la première estimation de mouvement effectué au niveau de résolution le plus grossier.

3.6 Résultats expérimentaux

Afin de valider l'approche, nous avons testé notre algorithme sur des séquences vidéo artificielles et réelles. Trois séquences artificielles¹ dont le flot optique réel est connu sont souvent utilisées [3]. Elles permettent une analyse quantitative sur l'erreur d'estimation. La séquence *Yosemite*, de taille 252×316 , représente le survol d'une chaîne montagneuse par un avion (figure 3.9.a.b). L'amplitude du flot optique va de 0 pixel/image au point d'expansion à 5 pixels/image dans le coin inférieur gauche. Il existe une frontière de mouvement importante entre le ciel et les montagnes. Les séquences *Arbre en translation* et *Arbre en divergence* (figure 3.11.a), de taille 150×150 , ont un mouvement respectivement de translation (de 1.7 pixels/image à gauche à 2.3 pixels/image à droite) et divergent (de 0 pixels/image au centre à 2.0 pixels/image à droite) (figure 3.11.b et c). L'erreur entre le flot optique des séquences

¹téléchargeable sur ftp.csd.uwo.ca

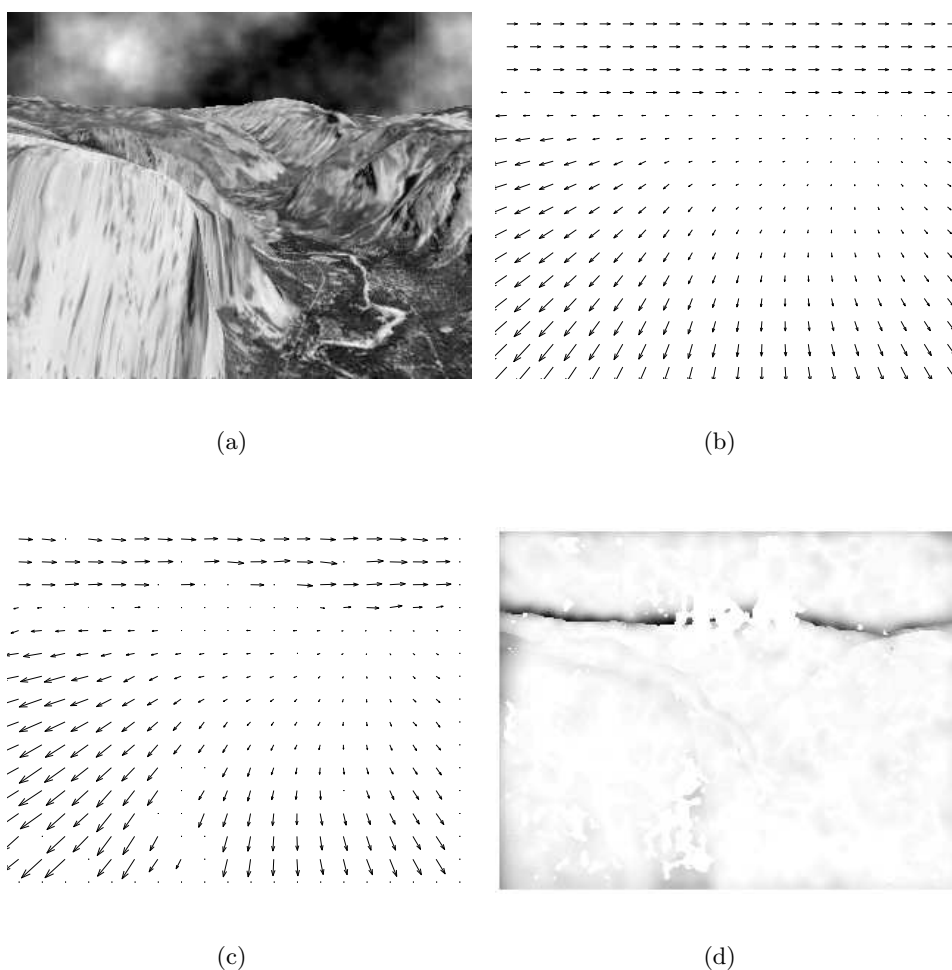


Fig. 3.9: a) Une image de la séquence Yosemite, b) le flot optique réel, c) le flot estimé (filtres récursifs au 3^{ème} ordre), et d) erreur angulaire

artificielles et le mouvement estimé est mesurée par l'erreur angulaire définie par Fleet & Jepson [23] :

$$\alpha_e = \arccos \left(\frac{uu_r + vv_r + 1}{\sqrt{u^2 + v^2 + 1}\sqrt{u_r^2 + v_r^2 + 1}} \right) \quad (3.27)$$

avec (u, v) le flot optique estimé et (u_r, v_r) le champ de déplacement vrai. Cette mesure prend en compte à la fois l'erreur commise sur l'orientation et l'amplitude des vecteurs de vitesse pour chaque pixel de l'image (figure 3.9.d). Traditionnellement, la moyenne et l'écart-type de l'erreur angulaire sur toute l'image caractérisent la précision de l'estimateur [3]. Une autre information importante est la densité du flot optique, donnant le ratio entre le nombre de vecteurs estimés et le nombre de pixels dans l'image.

Cette section présente les résultats selon un plan en quatre parties. Premièrement, à partir de la séquence artificielle *Yosemite* nous affinons expérimentalement le réglage des paramètres du banc de filtres. Puis, nous donnons le coût de calcul de l'algorithme en fonction du nombre

de filtres et de l'implantation choisie. La méthode est ensuite testée sur plusieurs séquences artificielles et comparée à différentes approches d'estimation du mouvement. Enfin, nous estimons le mouvement de séquences naturelles de qualité et de dynamique variées.

3.6.1 Affinement des réglages du banc de filtres

Dans la section 3.3 nous avons défini des intervalles de valeurs pour les trois paramètres du banc de filtres (fréquence centrale f_0 , largeur σ , nombre de filtres N). Ces paramètres doivent être affinés de façon expérimentale, en fonction de l'implantation récursive choisie (1^{er} et 3^{ème} ordre). Pour cela, nous avons testé l'algorithme sur la séquence artificielle *Yosemite* qui présente un mouvement suffisamment complexe (mouvements importants, frontières de mouvements) pour mettre en évidence l'influence des différents paramètres sur la qualité de l'estimation. Nous utilisons 3 niveaux de pyramides pour pouvoir mesurer les déplacements importants.

Les résultats présentés ci-dessous ont été obtenus en utilisant des filtres de Gabor récursifs du 3^{ème} ordre.

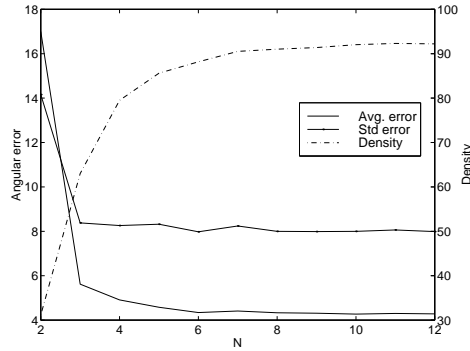
Choix du nombre de filtres N

La figure 3.10.a représente l'évolution de l'erreur angulaire (moyenne et écart-type) et de la densité de vecteurs estimés en fonction du nombre des filtres utilisés. La moyenne et l'écart-type sont quasi constants pour $N > 2$, alors que la densité de vecteurs estimés croît fortement avant d'atteindre un palier aux environs de 90% lorsque $N \geq 6$. Cette courbe illustre le problème que nous avons soulevé dans la section 3.3. Un petit nombre de filtres ne suffit pas à contraindre le système d'équations du flot optique (3.3), ce qui implique que, dans de nombreux cas, il est impossible d'obtenir une estimation de la vitesse. A l'inverse, l'ajout d'un très grand nombre de filtres supplémentaires n'améliore pas notablement les résultats, les équations étant alors très corrélées les unes aux autres. Le nombre de filtres minimum pour lequel le système est suffisamment contraint est aux alentours de 6, ce qui est conforme à ce que nous avons prévu dans la section 3.3. Néanmoins, le choix du nombre de filtres repose surtout sur le compromis que l'on se donne entre le temps de calcul et la densité de l'estimation.

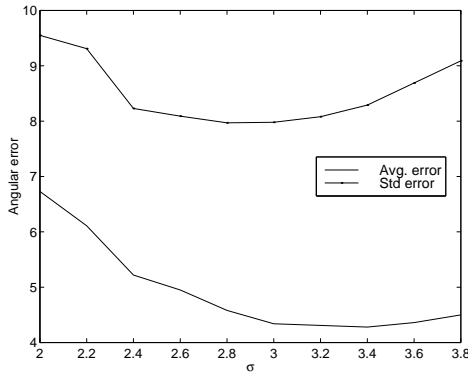
Choix de la largeur σ et de la fréquence centrale f_0

Les figures 3.10.b et 3.10.c représentent la moyenne et l'écart-type de l'erreur angulaire en fonction de la largeur des filtres σ et de la fréquence centrale f_0 , respectivement. La meilleure estimation du mouvement (sur la séquence Yosemite) est obtenue pour $f_0 \simeq 0.14$ et pour $\sigma \simeq 3.0$. Les valeurs obtenues sont conformes aux intervalles définis dans la section 3.3.

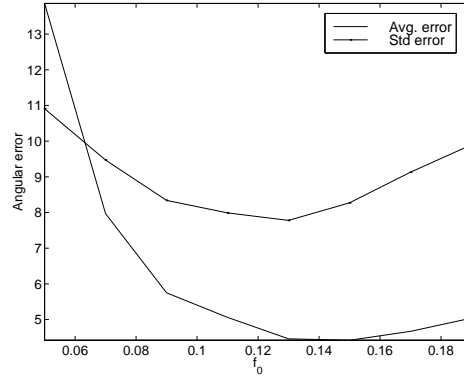
Une étude comparable a été menée en utilisant des filtres récursifs du 1^{er} ordre. La réponse impulsionnelle de ces filtres étant très différente d'une enveloppe gaussienne, le réglage des paramètres doit être a priori sensiblement différent. Les valeurs donnant la meilleure



(a) Erreur angulaire et densité de mesures en fonction du nombre de filtres N ($f_0 = 0.15$ et $\sigma = 3.0$)



(b) Erreur angulaire en fonction de la largeur des filtres de Gabor σ ($N = 6$ et $f_0 = 0.15$)



(c) Erreur angulaire en fonction de la fréquence centrale f_0 ($N = 6$ et $\sigma = 3.0$)

Fig. 3.10: Influence des paramètres f_0 , σ et N sur la qualité de l'estimation de mouvement. Ces tests ont été menés sur la séquence Yosemite en utilisant des banc de filtres récursifs au 3^{ème} ordre.

estimation du flot optique sont $f_0 = 0.14$ et $\sigma = 5.0$. Comme précédemment, le nombre de filtres N influe essentiellement sur la densité de vecteurs estimés. Celle-ci se stabilise aux alentours de 85% pour N supérieur à 8.

3.6.2 Coût de calcul

Le tableau 3.1 donne les temps de calcul nécessaire à l'estimation du flot optique de Yosemite avec une implantation au 1^{er} ordre et 3^{ème} ordre. La différence de temps de calcul entre l'implantation au 1^{er} ordre et 3^{ème} ordre n'est pas très importante. En effet l'opération de filtrage ne demande que peu de calcul comparée à la procédure MCPI.

Ordre des filtres	3 filtres	6 filtres	12 filtres
1	4.74s	8.44s	15.59s
3	5.08s	9.18s	16.63s

Tab. 3.1: Temps de calcul en seconde. Séquence *Yosemite*, 3 niveaux de pyramide, taille d'image 316×252 , PIII 700MHz, implantation C.

3.6.3 Comparaison quantitative avec d'autres algorithmes

Nous avons comparé nos résultats avec les algorithmes de Weber & Malik [86], de Bernard & Mallat [5] et de Fleet & Jepson [23]. Le tableau 3.2 donne les résultats obtenus sur les trois séquences artificielles présentées ci-dessus. L'erreur est calculée sur l'ensemble de l'image moins une bande de 16 pixels par bord.

Séquence	Auteurs	Erreur moyenne	Ecart-type	Densité
<i>Yosemite</i>	Weber & Malik	4.31°	8.66°	64.2%
	Fleet & Jepson	4.25°	11.34°	34.1%
	Bernard & Mallat	6.50°	-	96.5%
	filtres 3 ^{ème} ordre	4.33°	7.95°	89.1%
	filtres 1 ^{er} ordre	5.13°	7.98°	81,6%
<i>Arbre en divergence</i>	Weber & Malik	3.18°	2.50°	88.6%
	Fleet & Jepson	0.99°	0.78°	61.0%
	Bernard & Mallat	-	-	-
	filtres 3 ^{ème} ordre	2.81°	1.72°	97.8%
	filtres 1 ^{er} ordre	2.82°	2.56°	92.3%
<i>Arbre en translation</i>	Weber & Malik	0.49°	0.35°	96.8%
	Fleet & Jepson	0.32°	0.38°	74.5%
	Bernard & Mallat	0.78°	-	99.3%
	filtres 3 ^{ème} ordre	0.83°	0.68°	94.3%
	filtres 1 ^{er} ordre	0.92°	0.60°	88.2%

Tab. 3.2: Erreurs angulaires comparées pour les trois séquences artificielles. Nous utilisons 6 filtres de Gabor récursifs (1^{er} ordre, $f_0 = 0.14$, $\sigma = 5.0$ et 3^{ème} ordre $f_0 = 0.14$, $\sigma = 3.0$) sur 3 niveaux de pyramide. Les paramètres λ et ϵ des moindres-carrés robustes sont respectivement égaux à 3 et 10^{-2} .

Estimation avec les filtres de Gabor au 3^{ème} ordre

Les performances de notre algorithme sont comparables à celles obtenues par Weber & Malik et Fleet & Jepson. En particulier sur la séquence *Yosemite*, l'approche multirésolution permet d'obtenir une erreur faible et une densité de mesures relativement importante (contrairement à l'algorithme de Weber et à celui de Fleet).

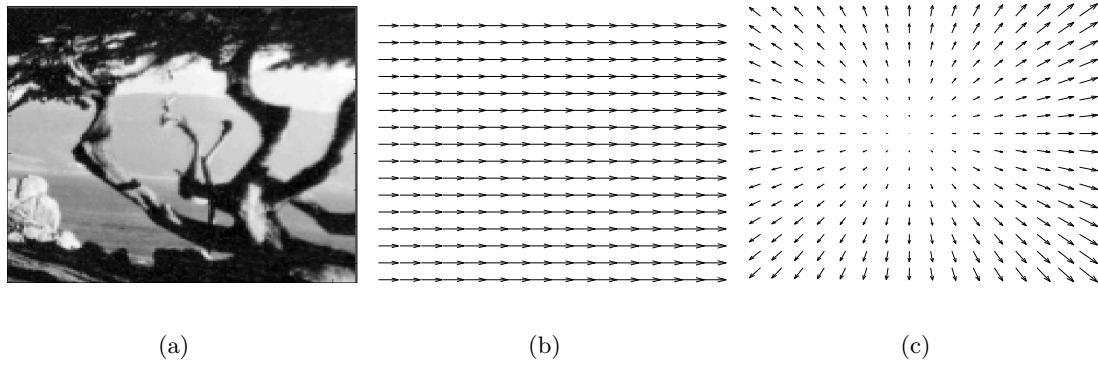


Fig. 3.11: a) Une image de la séquence *Arbre en translation* et *Arbre en divergence*, le flot optique réel de b) *Arbre en translation* et c) *Arbre en divergence*

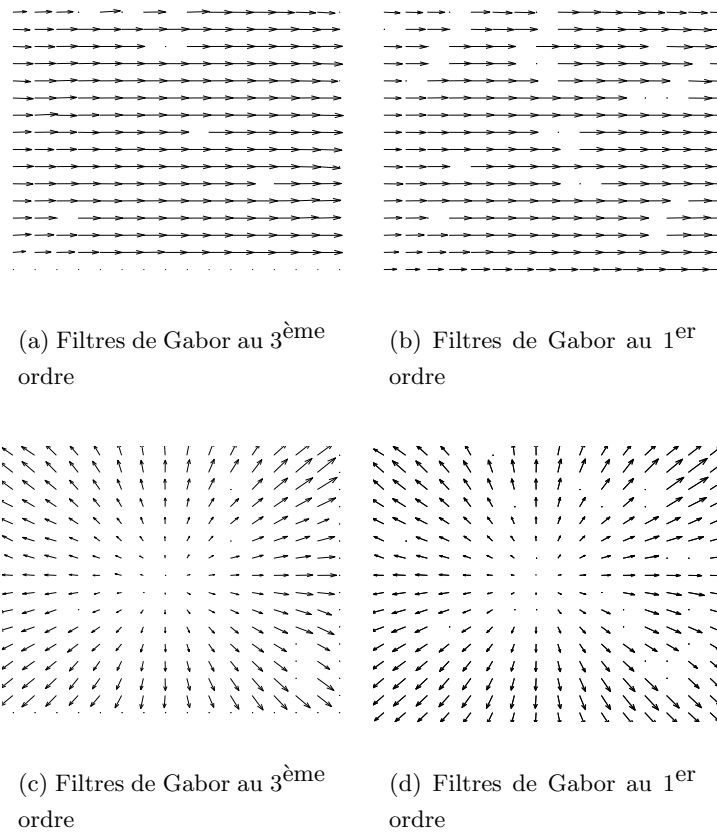


Fig. 3.12: Mouvement estimé sur a,b) *Arbre en translation* et c,d) *Arbre en divergence*

Il est à noter que l'algorithme de Weber et l'algorithme de Fleet ont besoin de respectivement 10 et 21 images pour estimer le mouvement (deux images suffisent dans notre cas). Ceci explique l'erreur très faible qu'ils obtiennent sur la séquence en *Arbre en translation*, où le mouvement est constant sur toute la durée de la séquence.

Estimation avec les filtres de Gabor au 1^{er} ordre

Les résultats restent satisfaisants. La dégradation la plus importante vient de la densité de vecteurs estimés. La décomposition issue de l'approximation au 1^{er} ordre ne permet pas d'obtenir en tout point un conditionnement suffisant du système d'équations du flot optique et donc une estimation du vecteur vitesse. Néanmoins, cette implantation permet un gain de temps de l'ordre de 30% sur les opérations de filtrage par rapport à l'approximation du 3^{ème} ordre.

3.6.4 Séquences réelles

L'algorithme a été testé sur quatre séquences réelles. Les valeurs des paramètres sont identiques à celle utilisées ci-dessus. Les deux premières séquences, *Rubik Cube* et *Mobile and calendar* (figures 3.13.a et 3.14.a), sont de très bonne qualité, alors que les deux autres, *Taxi de Hambourg* et *Tête parlante* (figure 3.15.a et 3.16.a), ont été acquises dans de moins bonnes conditions.

La séquence *Rubik Cube* contient un Rubik cube placé sur un plateau en mouvement de rotation. Les vitesses sur les bords du plateau sont d'environ 1,2 à 1,4 pixels/image, et sur le cube entre 0,2 et 0,5 pixels/image. Le flot optique estimé par les filtres du 3^{ème} ordre ou du 1^{er} ordre (figure 3.13.b et c) est consistant avec le mouvement de la séquence. Notons qu'il existe beaucoup de zones dans l'image très pauvres en textures (par exemple le dessus du plateau) où le mouvement ne peut être estimé.

La séquence *Mobile and calendar* a un mouvement global plus complexe. La caméra est en mouvement vers la gauche de l'image, avec une vitesse d'environ 0.9 pixels/image. Le calendrier, à droite de l'image, a un mouvement de translation d'environ 1,7 pixels/image vers le coin supérieur droit. Le train va vers la gauche avec une vitesse un peu inférieure à 2 pixels/image. Ce train pousse un ballon dont le mouvement de rotation est proche de 3 pixels/images. Enfin, un mobile est en mouvement (à gauche de l'image). Les deux petites billes composant ce mobile ont un mouvement très rapide. A nouveau le flot optique mesuré par les deux méthodes est consistant avec le mouvement perçu (figure 3.14.b et c). Néanmoins, le mouvement du mobile n'a pas pu être estimé.

La séquence *Taxi de Hambourg* est une séquence de rue qui contient quatre objets en mouvement : un piéton en haut à gauche, avec une vitesse d'environ 0,3 pixels/images, un taxi tournant au coin de la rue, vitesse environ 1 pixel/image, et deux autres véhicules, l'un venant de droite (partiellement caché derrière un arbre) et l'autre de gauche, se déplaçant avec des vitesses d'environ 2,5 – 3 pixels/image. La séquence est relativement bruitée, et les

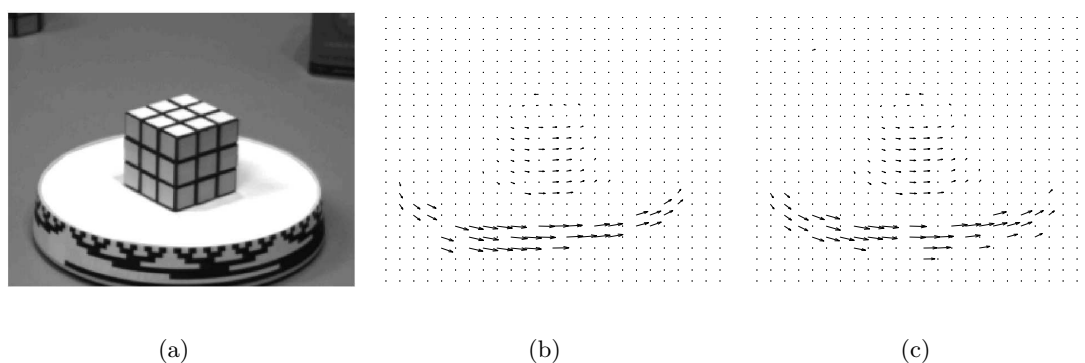


Fig. 3.13: a) Une image de la séquence *Rubik Cube* et le flot optique estimé avec les filtres de Gabor récursifs au b) 3^{ème} et c) 1^{er} ordre

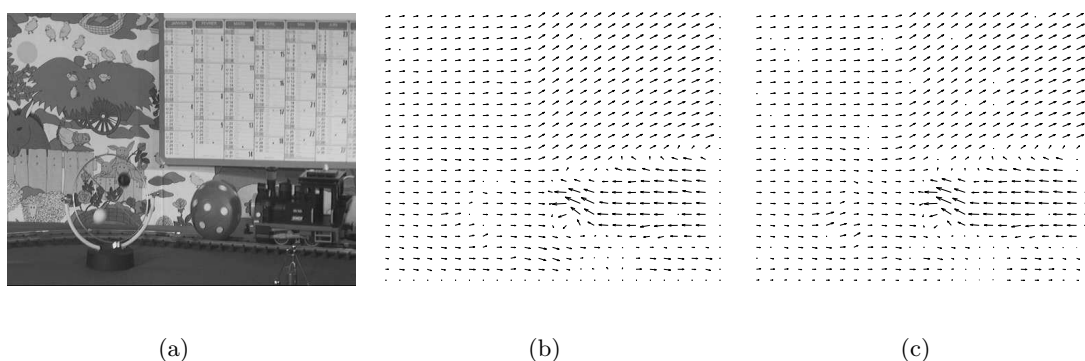


Fig. 3.14: a) Une image de la séquence *Mobile and calendar* et le flot optique estimé avec les filtres de Gabor récursifs au b) 3^{ème} et c) 1^{er} ordre

variations temporelles de la luminance dans les régions fixes sont assez importantes. Le flot optique estimé (figure 3.15.b et c) fait apparaître les trois véhicules, alors que le mouvement du piéton n'a pas été mesuré. L'estimation du mouvement par les filtres de Gabor au 1^{er} ordre est "trouée" à différents endroits en raison du bruit et des phénomènes d'occultations présents dans la séquence.

La dernière séquence, *Tête parlante* (figure 3.16.a), a été acquise grâce à une caméra fixée à un casque, lui-même posé sur la tête de la personne. Ce dispositif a été conçu pour la segmentation et le suivi des lèvres [43]. La personne étant en train de parler, les lèvres et le menton sont en mouvement vers le bas. Cette séquence présente deux difficultés essentielles : la vitesse d'ouverture de la bouche est très rapide, et l'apparition de dents rend l'hypothèse de conservation de la luminance caduque. Le mouvement estimé par les filtres du 3^{ème} ordre (figure 3.16.b) est de bonne qualité, et permet même d'observer l'ouverture asymétrique de la bouche (généralement observée chez les êtres humains).

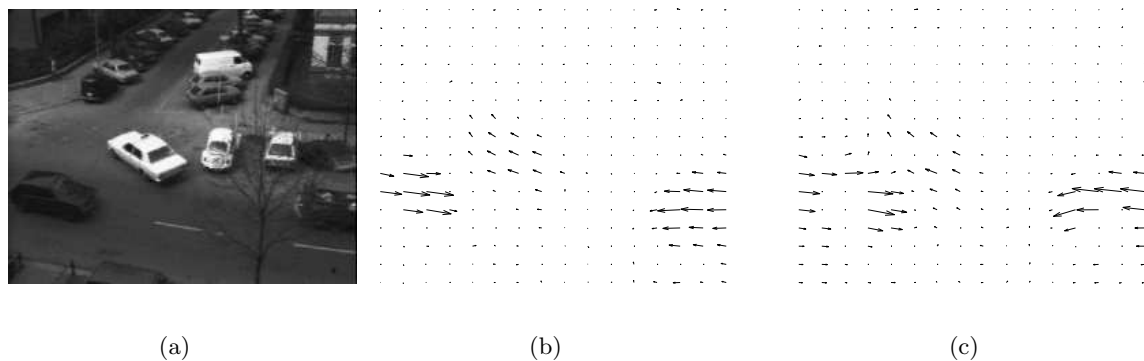


Fig. 3.15: a) Une image de la séquence *Taxi de Hambourg* et le flot optique estimé avec les filtres de Gabor récursifs au b) 3^{ème} et c) 1^{er} ordre

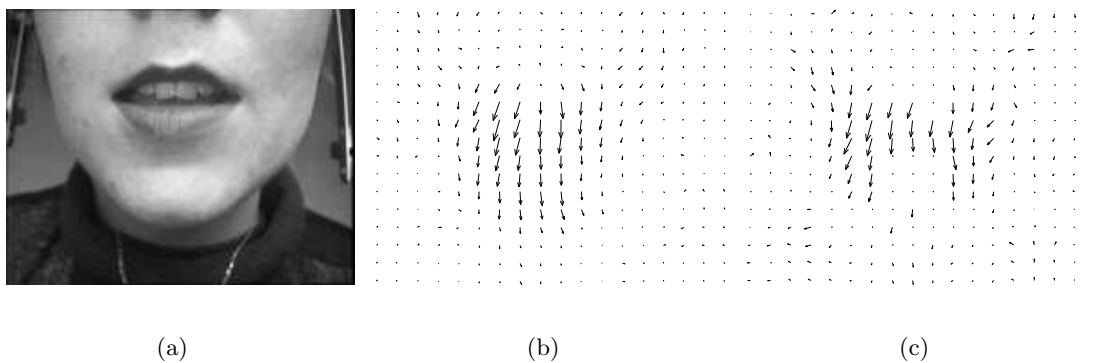


Fig. 3.16: a) Une image de la séquence *Tête parlante* et le flot optique estimé avec les filtres de Gabor récursifs au b) 3^{ème} et c) 1^{er} ordre

3.7 Conclusion et perspectives

Nous avons décrit une technique robuste d'estimation du mouvement dans les séquences d'images. Le problème d'ouverture est résolu par l'utilisation d'une décomposition de Gabor de la séquence d'images : l'équation du flot optique est appliquée sur chaque sous-bande, permettant d'obtenir un système sur-contraint d'équations en chaque point de l'image. Afin de diminuer le temps de calcul nécessaire aux filtrages, nous avons implémenté les filtres de Gabor par des filtres récursifs. Nous avons proposé une approximation au 1^{er} ordre et au 3^{ème} ordre des filtres de Gabor. La résolution du système sur-contraint est effectuée à l'aide d'une méthode robuste (M-estimateurs), et l'algorithme estime itérativement, à différentes résolutions spatiales, le flot optique entre deux images d'une séquence. La qualité de l'estimation dépend de l'implantation des bancs de filtres (1^{er} ou 3^{ème} ordre), mais elle est dans les deux cas comparable à des techniques réputées pour la précision de l'estimation, telles que l'algorithme de Fleet & Jepson et de Weber & Malik.

Deuxième partie

Modèles paramétriques pour l'estimation et la description du mouvement global

Chapitre 4

Les modèles paramétriques de mouvement

4.1 Introduction

Les méthodes d'estimation locale du mouvement, présentées aux chapitres 2 et 3, sont basées sur une modélisation localement translationnelle du flot optique. Ce modèle simple permet de contourner le problème d'ouverture et d'estimer un vecteur vitesse en chaque pixel. Le nombre de paramètres à estimer est très important, ce qui pose des problèmes à la fois sur la stabilité de l'estimation et sur l'analyse du mouvement. Dans ce dernier cas, il s'agit d'extraire, d'un volume très important de données, des caractéristiques pertinentes sur le champ de vitesses.

De ce point de vue, il nous semble intéressant de déterminer un modèle spatial et linéaire du mouvement global, défini par un nombre limité de paramètres. Ce modèle, en renforçant la contrainte sur les variations spatiales du mouvement, contribue à la stabilité de l'estimation et fournit, par l'intermédiaire des paramètres estimés, une description compacte du flot optique.

La première partie de ce chapitre présente les avantages à utiliser des modèles paramétriques globaux, tant au niveau de l'estimation que de la description du mouvement. En particulier, nous décrivons l'utilité et les limites des modèles paramétriques dans les problèmes d'indexation de vidéos et de reconnaissance d'activités. Dans un deuxième temps, nous présentons les différents modèles de mouvement utilisés ainsi que les représentations mathématiques conduisant à une description globale d'un signal. Nous détaillons ensuite un algorithme itératif et robuste d'estimation des paramètres d'un modèle générique qui sera utilisé dans les chapitres 5 et 6. Nous présentons enfin les solutions retenues pour modéliser globalement le flot optique entre deux images.

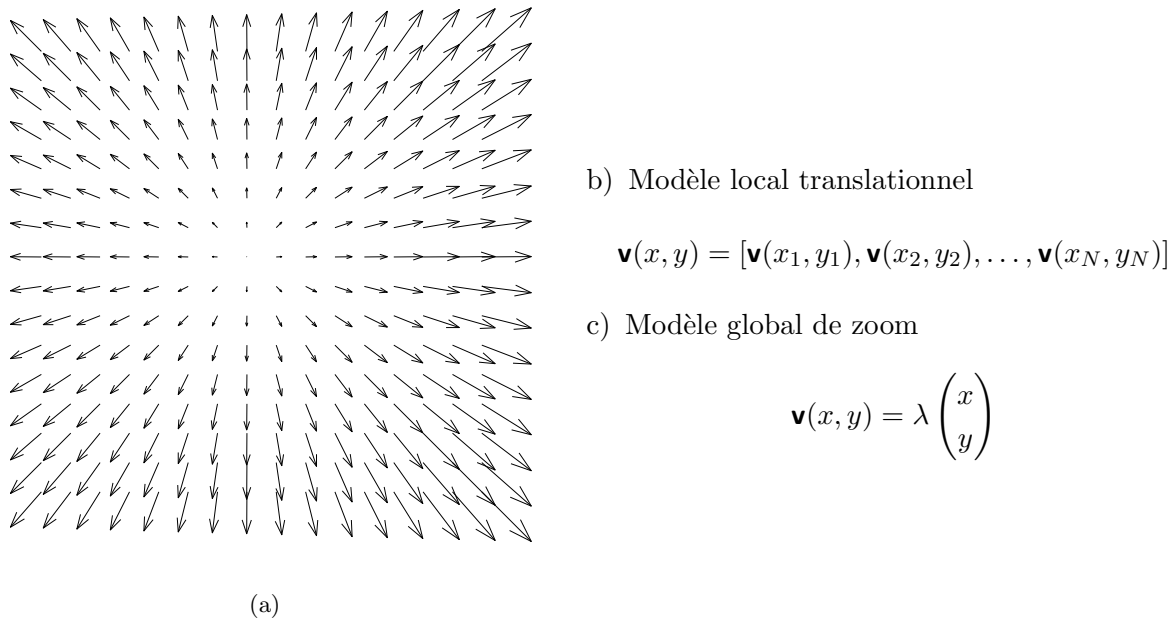


Fig. 4.1: Le flot optique (a) peut être représenté par b) un modèle localement constant et c) un modèle de zoom.

4.2 Motivations pour la modélisation du mouvement

4.2.1 Intérêt de la représentation du mouvement par un modèle paramétrique

L'approche locale part de l'hypothèse que le mouvement est constant sur une petite région de l'image. Cette modélisation, translationnelle et locale, peut être enrichie et étendue à une surface plus importante de l'image. Considérons par exemple le flot optique de la séquence "Arbre en divergence" (Figure 4.1). L'estimation du mouvement avec un modèle localement constant mesure pour chaque pixel un vecteur de déplacement. A l'inverse, la modélisation du mouvement sur toute l'image par un zoom pur ne nécessite que d'estimer le facteur de zoom λ .

Les trois avantages de ce dernier modèle sont :

- L'hypothèse de mouvement localement constant n'est dans ce cas jamais vérifiée, alors que le modèle de zoom est parfaitement adapté, et permet une meilleure estimation du flot optique. Cet exemple est évidemment un cas idéal, mais dans de nombreuses configurations (zoom, rotation, frontières de mouvement,...) l'utilisation d'un modèle plus complexe que la translation permet de mieux approximer la structure spatiale du mouvement.
- L'estimation d'un seul paramètre sur toute l'image est plus robuste que l'estimation de chaque vecteur vitesse sur un petit voisinage spatial. La contrainte sur le problème d'ouverture est très forte, supprimant toute ambiguïté sur la mesure du mouvement. Là

encore le cas est idéal, puisque le modèle n'est décrit que par un paramètre. Mais quel que soit le modèle utilisé, le nombre de paramètres est toujours largement inférieur au nombre de pixels de l'image.

- Les paramètres du modèle fournissent une description pertinente du mouvement. Dans notre cas, le paramètre λ nous renseigne directement sur le zoom effectué par la caméra. Ce dernier point est fondamental puisqu'il permet d'obtenir directement au niveau de l'étape de l'estimation une information de plus haut niveau que le flot optique. Cette information sur le mouvement peut être très utile pour la reconnaissance d'activité ou l'indexation de vidéos.

La modélisation du mouvement sur une large région de la scène nécessite néanmoins de s'assurer que le modèle choisi est conforme avec le mouvement réel sur le support considéré. Pour cela, il s'agit soit de déterminer un modèle de mouvement globalement pertinent en tout point de la scène, soit de déterminer les régions où le mouvement peut être approximé par un modèle initialement défini. Dans la section ci-dessous nous discutons du choix de l'une ou l'autre stratégie (modèle global ou segmentation au sens du mouvement) du point de vue de la description du contenu d'une séquence d'images.

4.2.2 Description du contenu dynamique de séquence d'images

Le problème de l'indexation de vidéos ou de la reconnaissance d'activité consiste à décrire le contenu de la séquence dans le but de comprendre "ce qu'il se passe" dans le document vidéo. Cette information de haut niveau permet par la suite d'identifier une vidéo dans une base de données, de localiser un passage précis dans une vidéo, de créer un résumé vidéo, de reconnaître une action particulière. Une étude approfondie de ces problèmes est exposée dans [26] et [16].

Bien que n'étant pas le seul indice visuel nécessaire à la description d'une séquence d'images, le mouvement est fondamental pour accéder au contenu dynamique d'une vidéo. La problématique consiste à extraire au cours du temps des caractéristiques relatives au mouvement, appelées descripteurs de mouvement. Par exemple, la norme des descripteurs associés au contenu des documents audiovisuels, élaborée par le groupe MPEG-7 [56], prévoit entre autre d'associer à une vidéo des descripteurs sur le mouvement de la caméra, le mouvement et la trajectoire des objets ainsi qu'une mesure de l'activité (faible ou forte).

La plupart des approches exploite une segmentation au sens du mouvement qui repose sur des modèles paramétriques 2D [26, 36, 84]. Les paramètres du modèle de mouvement sous-jacent à la segmentation sont alors une partie des descripteurs caractérisant le contenu dynamique de la vidéo. Si l'on se réfère toujours à la norme MPEG-7, les descripteurs associés à un objet en mouvement correspondent aux paramètres d'un modèle affine. Ceux associés aux mouvements de la caméra sont obtenus par l'estimation d'un modèle tridimensionnel des déplacements possibles de la caméra. Dans certains cas où le contenu de la séquence est

contrôlé (vidéo conférence, environnement routier), on peut procéder à la reconnaissance de modèles spécifiques de mouvement (voir la section 4.3.2).

L'intérêt de cette description basée sur une segmentation spatio-temporelle de la scène est de fournir une information quasi sémantique sur le contenu de la séquence (mouvement de la caméra et des objets, suivi et trajectoire des objets, paramètres physiques de modèles spécifiques). Ainsi, ces descripteurs sont pertinents pour répondre à des requêtes du type "objet allant vers la droite + zoom avant de la caméra". Malheureusement, la segmentation d'une scène en zones significatives au sens du mouvement reste limitée aux séquences possédant un contenu spatio-temporel simple. Notamment, la complexité et la diversité des contenus que l'on retrouve dans les vidéos issues de la production cinématographique ou télévisuelle ne permettent pas d'assurer le succès de tels algorithmes.

Afin de dépasser cette limitation, des solutions visant à caractériser *globalement* le mouvement par des attributs statistiques sur la texture temporelle ont été proposées dans [22, 58, 64, 78]. Ces descripteurs du mouvement global apportent une information qualitative sur le contenu dynamique de la séquence, mais ne fournissent pas une mesure directe du mouvement. De fait, il est difficile de donner une signification explicite à ces descripteurs, telle que la nature du mouvement de la caméra ou d'un objet dans la scène.

L'approche proposée dans cette thèse tente de réunir les avantages de la modélisation paramétrique du mouvement et de la description globale du mouvement. Ainsi, sans segmentation préalable au sens du mouvement, nous cherchons à définir des descripteurs globaux permettant de remonter de façon précise au mouvement entre les images.

4.3 Différents modèles paramétriques de mouvement

Un grand nombre de modèles paramétriques de mouvement ont été développés. Ils ont été utilisés à des fins variées, telles que l'estimation du flot optique, la segmentation au sens du mouvement, la construction d'images mosaïques ou encore la reconnaissance d'activités. Dans cette partie, nous étudions les propriétés des différentes familles de modèles paramétriques, et en particulier nous recherchons une classe de modèles susceptible de représenter le mouvement global sans segmentation *a priori* de la scène.

4.3.1 Modèles polynomiaux

Les modèles polynomiaux de degré N s'écrivent

$$\mathbf{v}(x, y) = \sum_{(i,j) \in \Delta} \begin{pmatrix} a_{ij} \\ b_{ij} \end{pmatrix} x^i y^j, \text{ avec } \Delta = \{(i, j) \in \mathbb{N}^2 / (i + j) \leq N\} \quad (4.1)$$

et correspondent à une transformation géométrique dont la complexité, ou le nombre de degrés de liberté, dépend du degré N choisi (Figure 4.2). On peut voir aussi le modèle polynomial comme un moyen d'approximer le flot optique sur Ω par son développement de Taylor limité aux N premiers termes, N étant le degré du polynôme (4.1).

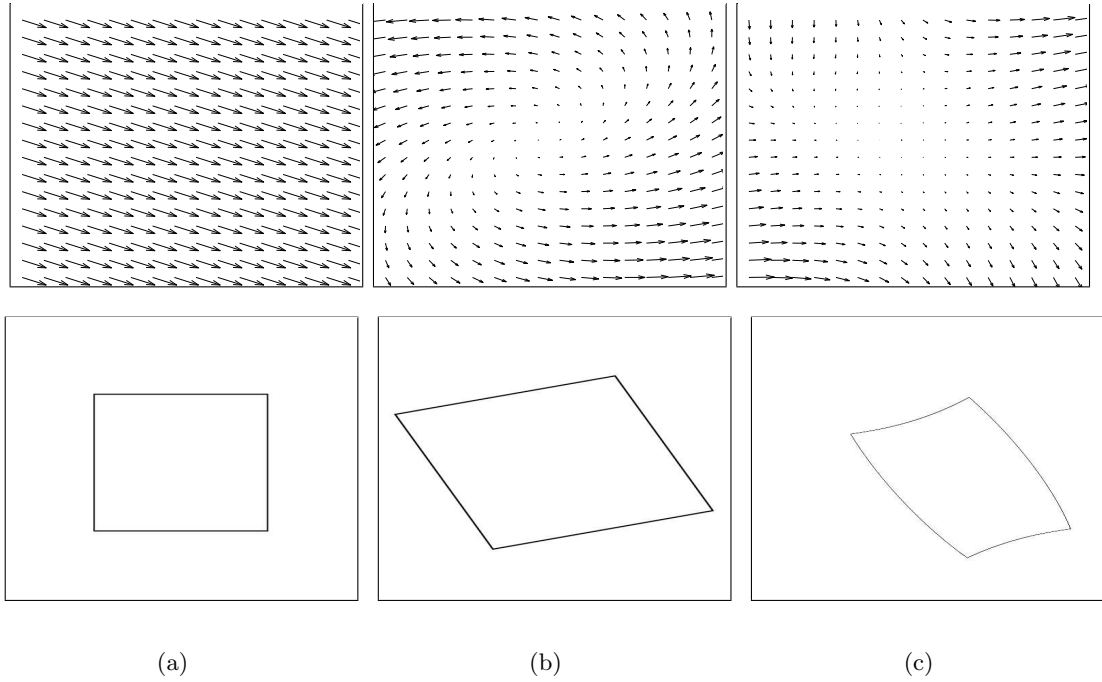


Fig. 4.2: Exemples de modèles polynomiaux de mouvement et des transformations correspondantes appliquées à un carré centré pour : a) un modèle translationnel, b) un modèle affine, et c) un modèle quadratique

L'estimation des termes élevés est peu stable car très sensible au bruit. En pratique, la limite est le modèle quadratique ($N = 2$), et le modèle le plus usité est le modèle affine, de degré $N = 1$. Ce dernier autorise une variation linéaire de la vitesse sur le support Ω , tout en étant peu sensible au bruit.

Le modèle affine permet de modéliser toutes les transformations affines du plan. En particulier, les paramètres $(a_{i,j}, b_{i,j})$ sont directement liés au déplacement translationnel ainsi qu'aux facteurs de zoom et de rotation du support d'estimation Ω . Pour ces propriétés, le modèle affine (et éventuellement le modèle quadratique) est largement utilisé pour l'estimation du flot optique [9], l'estimation des paramètres du mouvement dominant (induit par le déplacement de la caméra) [60], la segmentation du mouvement [61, 15, 85], la création de mosaïque d'images [36] ou l'indexation de vidéos [26].

Néanmoins, le mouvement global dans une image ne se résume que très rarement à une transformation affine ou éventuellement quadratique. Parallèlement à l'étape d'estimation des paramètres, il est nécessaire de déterminer les régions où le modèle de mouvement est valide. Cette étape de partitionnement peut être effectuée par différentes techniques : segmentation de l'image en niveaux de gris [9], estimation hiérarchique des mouvements dominants successifs [61] ou encore estimation de mélanges de modèles [2]. Ces opérations sont complexes et ne garantissent pas un résultat fiable lorsque la qualité de la séquence d'images est faible ou que le mouvement n'est pas structuré (déplacement d'une foule, activités sportives).

4.3.2 Modèles spécifiques

Dans les applications où les caractéristiques spatiales du champ de déplacement sont a priori connues, l'utilisation de modèles spécifiques au flot optique réel apporte une contrainte parfaitement adaptée au problème d'estimation de mouvement. Ce type d'approche est largement employé en reconnaissance et suivi de visages [46, 91], de bouches [10], de personnes [12, 70, 41].

L'a priori sur le contenu de la séquence (visage parlant, marche d'un piéton, ...) permet par exemple de définir un modèle physique pour le mouvement. Dans [12], Bregler & Malik utilisent un modèle physique du corps humain permettant de mesurer des paramètres cinématiques, tel que la rotation du genou, conformes aux mesures obtenues par des psychomotriciens. Dans d'autre cas, le modèle est obtenu à partir d'une base d'apprentissage : une base $\{\phi_k\}_k$ est calculée par analyse en composantes principales (ACP) [10] d'une collection de flots optiques. Le modèle peut être aussi défini par rapport à une structure particulière du mouvement, tel que les discontinuités du mouvement [24], ou l'estimation de déplacements purement rotationnels et divergents [20].

Ce type de modèles nécessite évidemment de segmenter dans la scène les régions dont le mouvement répond au caractère spécifique du modèle, et de ce point de vue ne peut pas être adapté à notre problème.

4.3.3 Modélisation globale du mouvement

La modélisation paramétrique du mouvement global entre deux images nécessite d'utiliser un modèle suffisamment riche pour approximer correctement le flot optique en tout point de l'image. Dans ce qui suit, nous considérons le flot optique comme une fonction *continue par morceaux*, hypothèse qui est en général acceptable. De fait, notre exposé n'aborde pas le problème de l'estimation de "textures" de mouvement, tel que le mouvement d'une flamme ou des frondaisons sous le vent [22].

Les paramètres du modèle étant des inconnues à estimer par la relation 4.11, la modélisation doit reposer sur le moins de paramètres ou degrés de liberté possible, afin de contraindre au maximum le problème d'ouverture, tout en approximant le mieux possible le champ des vitesses $\mathbf{v}(x, y)$.

Cette section présente brièvement les approches existantes concernant la modélisation globale du mouvement. Puis, nous nous attacherons à trouver une décomposition permettant la meilleure reconstruction d'un signal avec un minimum de fonctions de base, dont les coefficients fournissent une description globale du signal.

Modèles basés sur des fonctions interpolatrices

Un certain nombre d'auteurs ont formulé ce problème comme un problème d'interpolation [68, 73]. Le champ des vitesses \mathbf{v} est modélisé par une combinaison linéaire de fonctions d'interpolation :

$$\mathbf{v}(\mathbf{p}_i) = \sum_{k,l} \mathbf{c}_{k,l} \phi_{k,l}(\mathbf{p}_i) \quad (4.2)$$

où $\phi_{k,l}(\mathbf{p}_i) = \phi(x - k, y - l)$ est la fonction d'interpolation localisée autour du point (k, l) . Le choix pour la fonction ϕ est assez vaste, tous les noyaux interpolants étant possibles. Dans [68], Rakshit et Anderson testent un certain nombre de fonctions (de l'interpolation linéaire à la fonction *sinus cardinal sinc*).

Dans ces deux approches, les paramètres de mouvement $\mathbf{c}_{k,l}$ sont estimés en résolvant un système d'équations du flot optique (2.8) appliqué en chaque point de l'image Ω :

$$\nabla I(\mathbf{p}_i, t) \cdot \sum_{k,l} \mathbf{c}_{k,l} \phi_{k,l}(\mathbf{p}_i) + I_t(\mathbf{p}_i, t) = 0, \quad \forall \mathbf{p}_i \in \Omega \quad (4.3)$$

Si N est le nombre de fonctions $\phi_{k,l}$ utilisées pour modéliser le flot optique, il s'agit de résoudre un système de $\text{card}(\Omega)$ équations à $2N$ inconnues. Le nombre N de fonctions de base ϕ définit la précision de la modélisation : en augmentant le nombre de degrés de liberté, la solution obtenue peut suivre au mieux les variations spatiales du flot optique. En contre partie, plus il y a de paramètres à estimer, plus l'inversion du système (4.3) est numériquement instable.

Une approche relativement similaire, basée sur une grille de contrôle spline, a été proposée par Szeliski et Coughlan [76]. Dans ce formalisme, les paramètres du modèle $\mathbf{c}_{k,l}$ sont des points de contrôle sur la grille, et correspondent aux vecteurs de vitesses $\mathbf{v}(k, l)$ associés aux pixels de coordonnées (k, l) . Entre les points de contrôle, les vecteurs de vitesse sont calculés par interpolation spline.

Ces approches, en particulier les algorithmes de Szeliski *et al.* [76, 77] et Srinivasan *et al.* [73], sont relativement précises en terme d'estimation du flot optique.

Bien que le modèle soit défini globalement, l'influence de chaque fonction de base est très locale (le support spatial des noyaux interpolants). Les paramètres du modèle $\mathbf{c}_{k,l}$ sont une version sous-échantillonnée du flot optique réel et ne sont donc pas directement des descripteurs du mouvement global.

Modèles basés sur des ondelettes

Wu *et al.* [90] ont récemment proposé un modèle de mouvement global défini par une base d'ondelettes de Cai-Wang. Ces ondelettes sont construites à partir de fonction B-splines de degré 3, et en ce sens cette approche a des similarités avec les techniques d'interpolation décrites ci-dessus. Dans ce cas, le champ de mouvement est modélisé par un développement

en séries d'ondelettes 2D :

$$\begin{aligned}
 \mathbf{v}(\mathbf{p}_i) &= \sum_{k_1, k_2=0}^{2^L-1} \mathbf{c}_{L, k_1, k_2} \Phi_{L, k_1, k_2}(\mathbf{p}_i) \\
 &+ \sum_{j \geq L}^l \sum_{k_1, k_2=0}^{2^j-1} [\mathbf{d}_{j, k_1, k_2}^H \Psi_{j, k_1, k_2}^H(\mathbf{p}_i) \\
 &+ \mathbf{d}_{j, k_1, k_2}^D \Psi_{j, k_1, k_2}^D(\mathbf{p}_i) + \mathbf{d}_{j, k_1, k_2}^V \Psi_{j, k_1, k_2}^V(\mathbf{p}_i)]
 \end{aligned} \tag{4.4}$$

où $\Phi_{j, k_1, k_2}(\mathbf{p}_i)$ est la *fonction d'échelle*, et $\Psi_{j, k_1, k_2}^{H, D, V}(\mathbf{p}_i)$ sont les *ondelettes* représentant respectivement les variations horizontales, diagonales et verticales du flot optique. Le modèle est multiéchelle, et s'applique du niveau d'échelle le plus large L au niveau d'échelle le plus fin l .

La méthode d'estimation des paramètres de mouvement proposée par Wu *et. al.* consiste à minimiser, par rapport au modèle de mouvement (4.4), l'erreur quadratique suivante :

$$E(\mathbf{v}) = \sum_{\mathbf{p}_i \in \Omega} (I(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i), t + 1) - I(\mathbf{p}_i, t))^2 \tag{4.5}$$

Cette minimisation est menée par l'algorithme itératif de Levenberg-Marquardt.

Le point fort de la base d'ondelettes est de fournir naturellement un cadre multirésolution au problème de l'estimation du mouvement.

Description globale d'un signal

Il existe un certain nombre de techniques de décomposition du signal par projection sur une base de fonctions. A titre d'exemple, on peut citer les décompositions les plus utilisées :

- *L'Analyse en Composantes Principales (ACP)* : cette technique permet d'obtenir une base optimale pour la représentation du signal et d'en extraire les caractéristiques globales. Les fonctions de base sont obtenues à partir d'un ensemble d'apprentissage.
- *Le développement de Taylor* : ce développement fournit une description globale, composée de la moyenne et des dérivées successives du signal. Comme il a été montré dans la section 4.3.1, cette décomposition est très utilisée pour modéliser le mouvement.
- *La décomposition en séries de Fourier* : les séries de Fourier présentent l'avantage de représenter n'importe quelle fonction continue par morceaux appartenant à L^2 , l'espace des fonctions de carré intégrable. La description du signal est fréquentielle, donc globale.
- *La décomposition en séries d'ondelettes* : les ondelettes sont parfaitement adaptées à la représentation de fonctions continues par morceaux [75]. Cette décomposition fournit une représentation très compacte, un signal étant en général décrit par relativement peu de coefficients d'ondelettes. La décomposition en séries d'ondelettes opère une analyse multirésolution du signal. La description est alors à la fois globale et locale, et permettrait éventuellement une description du mouvement à ces deux échelles spatiales.

Un modèle global reposant sur une base de fonctions propres (ACP) n'est évidemment pas envisageable, les fonctions de base étant déterminées par le mouvement qui nous est inconnu.

De même le développement de Taylor ne permet pas de modéliser une fonction continue par morceaux et nécessite une segmentation au sens du mouvement de la scène (voir la section 4.3.1). Par contre, l'utilisation de modèles basés sur la décomposition en séries de Fourier ou en séries d'ondelettes semble être adaptée à notre problème.

4.4 Estimation des paramètres d'un modèle linéaire de mouvement

D'une manière générale la modélisation linéaire consiste à écrire le flot optique comme une somme pondérée de fonctions de base $\phi(\mathbf{p}_i)$ avec $\mathbf{p}_i = (x_i, y_i)^T \in \Omega$:

$$\mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}) = \sum_{k=0}^N \mathbf{c}_k \phi_k(\mathbf{p}_i) \quad \text{avec } \mathbf{c}_k = (c_x^k, c_y^k)^T \quad (4.6)$$

Le domaine Ω peut être tout ou partie de l'image et correspond à une région où l'on applique le modèle défini par les fonctions de base $\{\phi_k(\mathbf{p}_i)\}_k$. L'ensemble des coefficients (c_x^k, c_y^k) forme un vecteur de paramètres de mouvement $\boldsymbol{\theta} = [\mathbf{c}_0 \dots \mathbf{c}_N]^T$ qui caractérise complètement le mouvement sur Ω . En appliquant l'hypothèse de conservation de la luminance (2.1), une estimation de $\boldsymbol{\theta}$ peut être obtenue en minimisant une fonction de coût sur le support Ω :

$$\begin{aligned} E(\boldsymbol{\theta}) &= \sum_{\mathbf{p}_i \in \Omega} \rho(I(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}), t+1) - I(\mathbf{p}_i, t), \sigma) \\ &= \sum_{\mathbf{p}_i \in \Omega} \rho(r(\mathbf{p}_i; \boldsymbol{\theta}), \sigma) \end{aligned} \quad (4.7)$$

où $\rho(\cdot, \sigma)$ est un M-estimateur appliquée à la *différence d'images déplacées (DFD)* :

$$r(\mathbf{p}_i; \boldsymbol{\theta}) = I(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}), t+1) - I(\mathbf{p}_i, t). \quad (4.8)$$

L'utilisation d'un M-estimateur comme norme d'erreur robuste est nécessaire afin de tenir compte des erreurs dues à la violation de l'hypothèse de la conservation de la luminance ou à la non validité en tout point de Ω du modèle choisi [6, 12, 24, 60]. L'estimation de $\boldsymbol{\theta}$ est alors robuste aux données aberrantes (voir l'annexe A). Le paramètre σ détermine la robustesse du M-estimateur.

4.4.1 Estimation robuste et incrémentale d'un modèle paramétrique de mouvement

La fonction (4.7) peut être minimisée de manière itérative [59]. Pour cela, il faut tout d'abord réécrire le vecteur de paramètres de mouvement comme la somme d'une estimation initiale $\boldsymbol{\theta}$ et d'un résidu $\delta\boldsymbol{\theta}$:

$$E(\boldsymbol{\theta} + \delta\boldsymbol{\theta}) = \sum_{\mathbf{p}_i \in \Omega} \rho(r(\mathbf{p}_i; \boldsymbol{\theta} + \delta\boldsymbol{\theta}), \sigma) \quad (4.9)$$

- INITIALISATION : $k = 0, \boldsymbol{\theta}_0 = 0, \sigma = \infty$
- FAIRE
 - $\delta\boldsymbol{\theta}_k \leftarrow \text{MCPI}(\tilde{E}(\boldsymbol{\theta}_k + \delta\boldsymbol{\theta}_k), \sigma)$
 - $\boldsymbol{\theta}_{k+1} \leftarrow \boldsymbol{\theta}_k + \delta\boldsymbol{\theta}_k$
 - $\sigma \leftarrow 1.4826 \text{ med}_i(|r(\mathbf{p}_i; \boldsymbol{\theta}_k) - \text{med}_j(r(\mathbf{p}_j; \boldsymbol{\theta}_k))|)$
 - $k \leftarrow k + 1$
- TANT QUE ($\tilde{E}(\boldsymbol{\theta}_k) > \epsilon_1$ OU $\text{NORME}(\delta\boldsymbol{\theta}_k) > \epsilon_2$ ET $k < \text{MAX_ITER}$)

Fig. 4.3: Algorithme d'Estimation Robuste Itérative du Mouvement (ERIM). Les seuils prédéterminés ϵ_1 , ϵ_2 et MAX_ITER sont les critères d'arrêt de l'algorithme de minimisation

Étant donné une estimation $\boldsymbol{\theta}_k$ (initialement égale à 0) le problème est d'estimer le résidu $\delta\boldsymbol{\theta}_k$. L'estimation courante de $\boldsymbol{\theta}_{k+1}$ devient alors $\boldsymbol{\theta}_k + \delta\boldsymbol{\theta}_k$. La minimisation de (4.9) s'effectue en linéarisant $r(\mathbf{p}_i; \boldsymbol{\theta} + \delta\boldsymbol{\theta})$ par rapport à $\delta\boldsymbol{\theta}_k$:

$$\tilde{E}(\boldsymbol{\theta}_k + \delta\boldsymbol{\theta}_k) = \sum_{\mathbf{p}_i \in \Omega} \rho(r(\mathbf{p}_i; \boldsymbol{\theta}_k) + \nabla I(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}_k), t + 1)\mathbf{v}(\mathbf{p}_i; \delta\boldsymbol{\theta}_k), \sigma) \quad (4.10)$$

La minimisation de la fonction de coût \tilde{E} consiste à résoudre un système non-linéaire d'équations du flot optique. Ce système est similaire à celui obtenu par un banc de filtres de Gabor (cf. équation (3.15), chapitre 3) et peut être résolu par la technique des moindres-carrés pondérés itérés (MCPI) présentée à l'annexe A :

$$\tilde{E}(\boldsymbol{\theta}_k + \delta\boldsymbol{\theta}_k) = \sum_{\mathbf{p}_i \in \Omega} w(r(\mathbf{p}_i; \boldsymbol{\theta}_k + \delta\boldsymbol{\theta}_k)) (r(\mathbf{p}_i; \boldsymbol{\theta}_k) + \nabla I(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}_k), t + 1)\mathbf{v}(\mathbf{p}_i; \delta\boldsymbol{\theta}_k))^2 \quad (4.11)$$

avec $w(x) = \frac{1}{x} \frac{\partial \rho(x)}{\partial x}$. Rappelons qu'un poids fort indique une donnée prise en compte par le modèle et qu'un poids proche de zéro indique que la donnée est rejetée par l'estimateur.

La figure 4.3 résume la procédure itérative décrite ci-dessus, que nous appelons Estimation Robuste Itérative du Mouvement (ERIM).

4.5 Conclusion et solutions proposées

Dans ce chapitre, nous avons exposé l'intérêt que représente l'utilisation de modèles paramétriques de mouvement, à la fois pour l'estimation et la description du mouvement. Dans les deux cas, les modèles employés nécessitent généralement une étape de segmentation au sens du mouvement, ceci afin d'assurer la validité du modèle par rapport au mouvement réel. Cette étape n'est pas fiable dans le cas de séquences d'images au contenu spatio-temporel complexe, et il serait intéressant de définir un modèle de mouvement ne nécessitant aucune pré-segmentation ni connaissance *a priori*.

Dans les chapitres suivants, nous proposons deux approches pour l'estimation du mouvement global basées sur des modèles de mouvement différents. La première, qui fait l'objet du chapitre 5, modélise le mouvement par une série de Fourier. Les conclusions obtenues sur cette approche nous ont poussés à développer un nouveau modèle global, basé sur la description en ondelettes du mouvement (chapitre 6). Différents types d'ondelettes ont été testés pour modéliser le flot optique, et les meilleurs résultats ont été obtenus avec les ondelettes B-splines.

Le chapitre 7 décrit enfin quelques expériences montrant que les paramètres des modèles permettent de classer et de segmenter temporellement des vidéos en fonction de leur contenu dynamique.

Chapitre 5

Modélisation globale du mouvement par séries de Fourier

Le développement en séries de Fourier consiste à exprimer une fonction par une combinaison linéaire d'exponentielles complexes. Cette représentation présente l'avantage d'approximer un signal continu par morceaux par un nombre limité d'harmoniques. Cette propriété est notamment utilisée pour la compression d'images (JPEG) [67]. L'utilisation d'un modèle paramétrique basé sur les séries de Fourier permet ainsi théoriquement d'estimer des mouvements complexes par un nombre réduit de paramètres.

Dans ce chapitre, nous détaillons le modèle de mouvement global proposé ainsi que l'algorithme d'estimation des paramètres du mouvement. Puis nous présentons les résultats obtenus sur la capacité du modèle à approximer des mouvements complexes.

5.1 Modèle de mouvement basé sur les séries de Fourier

5.1.1 Rappel sur les séries de Fourier

Soit $f(x)$ une fonction périodique de période $T_0 = \frac{1}{\nu_0}$. Si $f(x)$ est continue par morceaux dans $[-T_0/2, T_0/2]$, la série de Fourier de $f(x)$ existe et converge :

$$f(x) = \sum_{k=-\infty}^{\infty} a_k e^{i2\pi k\nu_0 x}$$

avec

$$a_k = \frac{1}{T_0} \int_{T_0} f(x) e^{-i2\pi k\nu_0 x} dx \tag{5.1}$$

Les coefficients a_k pondèrent chaque harmonique $e^{i2\pi k\nu_0 x}$ et caractérisent la fonction $f(x)$. Évidemment, en pratique, le développement n'est effectué que sur un nombre limité de termes de manière à obtenir une approximation de $f(x)$ par un polynôme trigonométrique. Plus les harmoniques utilisées pour approximer une fonction sont d'ordre élevé (ou de fréquence élevée), plus l'approximation est fine et proche de la fonction réelle. Pour des fonctions dont

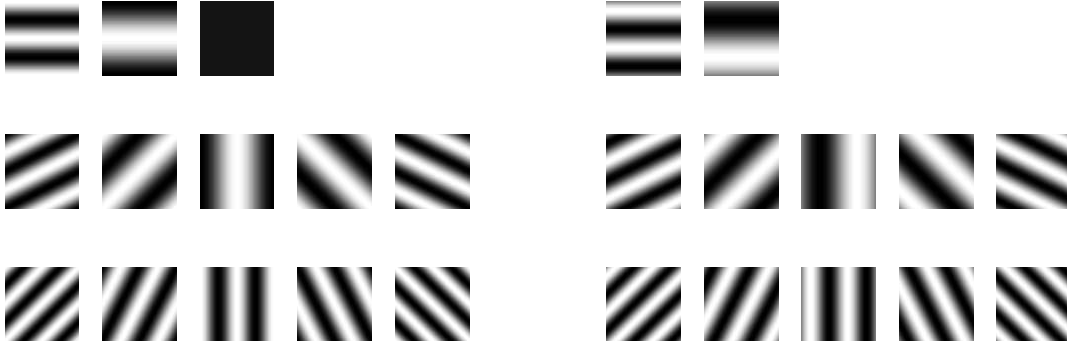


Fig. 5.1: Les 25 premières harmoniques paires et impaires ($N = 4$) utilisées pour modéliser $\mathbf{v}(x, y)$.

le comportement est régulier (non aléatoire), l'amplitude des coefficients a_k décroît rapidement lorsque la fréquence augmente. L'information sur la fonction est contenue dans les termes de "basses fréquences" et l'estimation de ces seuls termes permet alors une très bonne approximation de la fonction originale.

5.1.2 Modélisation du mouvement

Considérons $\mathbf{v}(x, y)$ le mouvement 2D entre deux images définies sur le support Ω . Le champ des vitesses $\mathbf{v}(x, y)$ peut être dupliqué sur \mathbb{R}^2 de manière à obtenir une fonction périodique 2D, où les périodes $T_{x_0} = \frac{1}{\nu_{x_0}}$ et $T_{y_0} = \frac{1}{\nu_{y_0}}$ sont respectivement la largeur et la hauteur de la fenêtre Ω . Si nous faisons l'hypothèse que le mouvement est une fonction continue par morceaux, alors la série de Fourier de $\mathbf{v}(\mathbf{p}_i)$, $\mathbf{p}_i = (x, y)^T$, existe et converge, et nous pouvons approximer le mouvement global par un nombre limité de termes :

$$\mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}) = \sum_{k,l=-N/2}^{N/2} \mathbf{c}_{k,l} \phi_{k,l}(\mathbf{p}_i) \text{ avec } \phi_{k,l}(\mathbf{p}_i) = e^{i2\pi(k\nu_{x_0}x + l\nu_{y_0}y)}$$

$$\text{et } \boldsymbol{\theta} = [\mathbf{c}_{-N/2,-N/2} \dots \mathbf{c}_{N/2,N/2}] \quad (5.2)$$

La figure 5.1 représente les harmoniques paires et impaires utilisées pour un développement avec $N = 4$. Le modèle paramétrique de mouvement correspondant contient $p = 2(N + 1)^2$ paramètres. Les harmoniques de plus hautes fréquences utilisées déterminent le niveau de détail du flot optique estimé. Un modèle défini par peu de paramètres fournira une approximation lissée du mouvement. A l'opposé, un flot optique précis (en particulier sur les contours du mouvement) ne peut être obtenu que par un modèle défini par un grand nombre de paramètres. Ce modèle est implanté en utilisant le développement en séries de Fourier réel (décomposition sur une base de cosinus et sinus) plutôt que par une base d'exponentielles complexes. Étant donné que le flot optique que l'on cherche à déterminer est réel, ces deux écritures sont totalement équivalentes.

5.2 Estimation des coefficients de Fourier

Rappelons que l'estimation des paramètres d'un modèle de mouvement tel que nous venons de le définir consiste à minimiser une fonction de coût robuste (éq. 4.7), basée sur un M-estimateur $\rho(x, \sigma)$, appliquée à la *différence d'images déplacées* $r(\mathbf{p}_i; \boldsymbol{\theta})$ (éq. 4.8). Nous avons choisi de réaliser cette minimisation de manière incrémentale, en utilisant la technique des moindres-carrés pondérés itérés (MCPI). Dans ce cas, l'expression à minimiser, comme nous l'avons déjà indiquée, est :

$$\tilde{E}(\boldsymbol{\theta}_j + \delta\boldsymbol{\theta}_j) = \sum_{\mathbf{p}_i \in \Omega} w(r(\mathbf{p}_i; \boldsymbol{\theta}_j)) (r(\mathbf{p}_i; \boldsymbol{\theta}_j) + \mathbf{v}(\mathbf{p}_i; \delta\boldsymbol{\theta}_j) \nabla I(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}_j), t + 1))^2 \quad (5.3)$$

Le procédé est itéré selon l'algorithme ERIM (figure 4.3), jusqu'à atteindre un minimum de \tilde{E} par rapport à l'incrément $\delta\boldsymbol{\theta}_j$. La solution finale de $\boldsymbol{\theta}$ est alors égale à :

$$\boldsymbol{\theta} = \sum_j \delta\boldsymbol{\theta}_j \quad (5.4)$$

5.2.1 Choix de l'estimateur robuste

Nous désirons que le modèle de Fourier (5.2) donne une approximation du mouvement en tout point de l'image. Pour cela, la fonction robuste (5.3) ne doit pas définitivement supprimer des régions entières de l'image lors du processus d'estimation. Le M-estimateur de *Geman-McClure* [27] pondère les données en fonction de leur erreur par rapport au modèle, mais à la différence de l'estimateur Biweight de Tukey (3.17), il ne supprime pas les données dont l'erreur est trop importante. Ainsi le modèle est contraint de s'adapter à l'ensemble des données, avec néanmoins une pondération permettant d'être robuste aux données réellement aberrantes. Le M-estimateur de Geman-McClure a été utilisé avec succès dans de nombreux algorithmes d'estimation de modèles de mouvement [9, 25]. Son expression est :

$$\rho(x, \sigma) = \frac{x^2}{\sigma^2 + x^2} \quad (5.5)$$

La fonction de pondération $w(x, \sigma)$, définie par rapport à $\rho(x, \sigma)$ est égale à :

$$w(x, \sigma) = \frac{1}{x} \frac{\partial \rho(x, \sigma)}{\partial x} = 2 \frac{\sigma^2}{(\sigma^2 + x^2)^2} \quad (5.6)$$

La figure 5.2 représente le M-estimateur de Geman-McClure ainsi que la fonction de pondération associée.

Le paramètre σ , qui influe sur la robustesse de l'estimateur est déterminé par la médiane des écarts absolus (3.19) définie au chapitre 3.

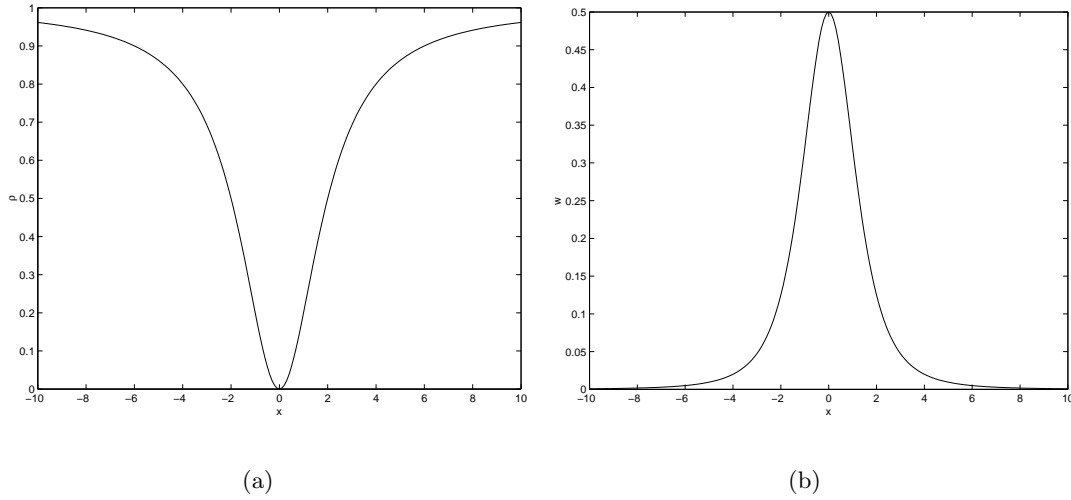


Fig. 5.2: M-estimateur de Geman-McClure : a) fonction ρ et b) pondération w

5.2.2 Solution au sens des moindres-carrés robuste par la Décomposition en Valeurs Singulières

En notation matricielle, la fonction d'erreur robuste (5.3) s'écrit :

$$\tilde{E}(\boldsymbol{\theta}_j + \delta\boldsymbol{\theta}_j) = W \cdot (M\delta\boldsymbol{\theta}_j + B)^T \cdot (M\delta\boldsymbol{\theta}_j + B) \quad (5.7)$$

où, pour

- $k \times l = [-N/2 \dots N/2]^2$, $p = 2(N+1)^2$ est le nombre de paramètres du modèle.
- $\forall \mathbf{p}_i = (x_i, y_i) \in \Omega$, Ω étant le support d'estimation, $m = \text{card}(\Omega)$ est le nombre d'équations du système :

$$\begin{aligned} M &= [I_x(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}_j))\phi_{k,l}(\mathbf{p}_i), I_y(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}_j))\phi_{k,l}(\mathbf{p}_i)], \quad M \in \mathbb{R}^{m \times p} \\ \delta\boldsymbol{\theta}_j &= [\delta\mathbf{c}_{k,l}]^T, \quad \delta\boldsymbol{\theta}_j \in \mathbb{R}^p \\ W &= [\text{diag}(w(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}_j)))], \quad W \in \mathbb{R}^{m \times m} \\ B &= r(\mathbf{p}_i; \boldsymbol{\theta}_j), \quad B \in \mathbb{R}^m \end{aligned} \quad (5.8)$$

Les colonnes de la matrice M contiennent les gradients spatiaux de l'image (successivement selon x et y), multipliés par les fonctions de bases $\phi_{k,l}$ définissant le modèle. La matrice diagonale W contient les poids $w(\mathbf{p}_i)$ associés à chaque point de l'image. Enfin le vecteur B représente la différence d'images déplacées selon le mouvement global $\boldsymbol{\theta}_j$ en chaque pixel.

Le minimum de (5.7) par rapport à $\delta\boldsymbol{\theta}_j$ conduit au système linéaire :

$$M^T W M \delta\boldsymbol{\theta}_j = M^T W B \quad (5.9)$$

Sa solution au sens des moindres-carrés est :

$$\delta\boldsymbol{\theta}_j = (M^T W M)^{-1} M^T W B \quad (5.10)$$

L'inversion de la matrice M^TWM est évidemment le coeur du problème. La stabilité du résultat dépend du conditionnement de cette matrice. De manière générale, le système (5.9) est largement sur-contraint. Le modèle de Fourier est défini par un nombre limité de paramètres, alors qu'une petite image, de taille 128×128 , permet de contraindre le problème par plus de 16000 équations.

Néanmoins, lorsque la scène est peu texturée, que les gradients spatiaux sont faibles ou orientés selon une direction unique, le problème d'ouverture se pose à nouveau. Dans ce cas la matrice M^TWM est mal conditionnée et son inversion est numériquement instable. La *Décomposition en Valeur Singulière (Singular Value Decomposition (SVD))* permet d'obtenir une solution stable définie dans un sous-espace du système [66]. Posons L tel que $W = L^2$. Soit A une matrice telle que $A = LM$ et Y un vecteur tel que $Y = LB$. Alors, l'équation

$$\delta\theta_j = (A^T A)^{-1} A^T Y \quad (5.11)$$

est équivalente à la relation (5.10).

La matrice A peut alors se décomposer comme le produit de trois matrices :

$$A = U \begin{pmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_p \end{pmatrix} V^T \quad (5.12)$$

U est une matrice de taille $m \times p$ dont les colonnes sont orthogonales entre elles. La matrice diagonale de taille $p \times p$ contient les *valeurs singulières* s_j de A et V est une matrice orthogonale de taille $p \times p$. En remplaçant A par sa décomposition en valeurs singulières dans (5.11) on obtient :

$$\delta\theta_j = -V \begin{pmatrix} 1/s_1 & & & \\ & 1/s_2 & & \\ & & \ddots & \\ & & & 1/s_p \end{pmatrix} U^T Y \quad (5.13)$$

Si le système (5.7) est mal conditionné, A n'est pas de rang plein et un certain nombre de valeurs singulières s_i sont égales ou proches de zéros. Lors de la résolution du système (5.13), l'inversion de ces valeurs singulières est numériquement faux et amène à un résultat erroné.

La solution à ce problème est de *mettre à zéro* l'inverse des valeurs singulières inférieures à un seuil prédéterminé. La solution $\delta\theta_j$ appartient alors uniquement à l'espace image de A et correspond à la meilleure solution au sens des moindres-carrés [66]. Le seuil en dessous duquel l'inverse des valeurs singulières est mis à zéro, est déterminé en fonction de la valeur singulière maximale et d'un facteur de proportionnalité k :

$$\text{Si } s_j < k \times \max(s_j) \text{ alors } 1/s_j = 0 \quad (5.14)$$

Pour les calculs en virgule flottante, il est conseillé de prendre $k = 10^{-6}$.

Pratiquement, cette technique revient à mettre à zéro les paramètres du modèle les moins bien conditionnés, et à estimer de façon fiable les paramètres restants. Le mouvement global est alors décrit par une base de Fourier incomplète, ce qui veut dire que l'estimation n'est pas parfaite, mais qu'elle est la meilleure possible vu le mauvais conditionnement du système.

5.3 Estimation par enrichissement progressif du modèle de mouvement

L'estimation directe d'un modèle de Fourier contenant un grand nombre de paramètres (donc d'harmoniques de fréquences élevées) n'est pas stable dans le cas où la séquence contient des déplacements de grande amplitude.

En effet, les harmoniques de fréquences élevées mesurent les variations spatiales rapides, et donc localisées, du mouvement. Nous avons montré dans le chapitre 2 que l'estimation de mouvements importants ne peut se faire que sur un large support, la solution obtenue étant alors spatialement très lissée. Ainsi, malgré l'utilisation de fonctions de bases définies sur l'ensemble de l'image, les harmoniques de fréquences élevées, du fait du caractère très local des informations qu'elles mesurent, ne permettent pas d'estimer des déplacements importants.

La solution à ce problème réside à nouveau dans l'estimation multirésolution du mouvement. Contrairement à l'approche présentée dans le chapitre 3, nous utilisons ici un schéma où le mouvement est estimé par *enrichissement progressif du modèle de Fourier*, et non par augmentation de la résolution de l'image.

Soit θ^N le vecteur de coefficients de Fourier correspondant à une base d'harmoniques de fréquence maximale $\nu_{max} = \frac{N\nu_0}{2}$. Le schéma débute par l'estimation de θ^l , associé à un modèle basse fréquence du mouvement défini par peu de paramètres. Le vecteur θ^l , estimé par l'algorithme ERIM, initialise l'étape suivante, où le vecteur θ^{l+1} est à son tour estimé. Ce procédé est répété jusqu'à atteindre le modèle le plus riche que l'on s'est fixé, correspondant à θ^L . Ce dernier vecteur de paramètres de mouvement donne l'estimation finale du mouvement global, approximé par un modèle à $2(L+1)^2$ paramètres.

5.4 Résultats expérimentaux

Dans cette partie, nous illustrons les capacités du modèle à approximer différents types de mouvement. Dans ce but, nous avons testé le modèle de Fourier sur des séquences artificielles et réelles.

5.4.1 Séquences artificielles

Nous avons testé l'approche sur les 3 séquences artificielles *Arbre en Translation*, *Arbre en Divergence* et *Yosemite*. Le mouvement global est estimé en utilisant successivement un modèle à 50 paramètres ($N = 4$), à 98 paramètres ($N = 6$) et à 162 paramètres ($N = 8$).

Les figures 5.3, 5.4 et 5.5 représentent le flot optique estimé pour chacune de ces séquences et pour les trois modèles de Fourier.

Le tableau 5.1 donne les résultats en terme d'erreur angulaire, obtenus sur ces trois séquences et pour les trois modèles utilisés.

Séquence	N	Erreur moyenne	Ecart-type
<i>Yosemite</i>	4	26.4°	21.8°
	6	9.0°	8.6°
	8	7.4°	8.4°
<i>Arbre en divergence</i>	4	14.3°	11.3°
	6	9.9°	11.4°
	8	8.3°	10.5°
<i>Arbre en translation</i>	4	33.5°	16.8°
	6	8.6°	7.7°
	8	7.0°	7.3°

Tab. 5.1: Erreurs angulaires comparées pour les trois séquences artificielles et aux niveaux $N = 4$, $N = 6$ et $N = 8$.

L'estimation de mouvements affines purs (sur les séquences *Arbre en Translation* et *Arbre en Divergence*) n'est pas performante. En particulier l'estimation d'une translation pure (Figure 5.3) semble poser beaucoup de problèmes avec le modèle de Fourier. Un élément d'explication peut être que le modèle comporte beaucoup trop de paramètres pour estimer un mouvement global simple.

Par contre la modélisation semble mieux convenir pour approximer un mouvement complexe, tel que celui de la séquence *Yosemite* (Figure 5.5) où le modèle donne une approximation à la fois du mouvement translationnel dans le ciel et du mouvement divergent sur les collines.

5.4.2 Séquences réelles

Dans cette dernière partie, nous testons le modèle de Fourier sur des séquences réelles. Trois séquences sont présentées : *Mobile and calendar*, *Rubik Cube* et *Taxi de Hambourg*. Ces séquences possèdent un mouvement global complexe que l'on peut retrouver dans de nombreuses autres séquences.

Le mouvement global de la séquence *Mobile and calendar* (Figure 5.6.a) est assez intéressant, car il est composé de régions en mouvement de taille très différente (cf. le flot optique de la figure 5.6.b estimé localement par bancs de filtres de Gabor). Les figures 5.6.c, d et e, représentent les flots optiques estimés aux trois niveaux de résolutions habituels.

Pour finir, les figures 5.7 et 5.8 présentent le flot optique estimé par modèle de Fourier, et pour comparaison, le flot optique localement estimé par bancs de filtres de Gabor.

5.5 Conclusion : vers une nouvelle modélisation

Les différentes expérimentations menées dans cette section nous montrent qu'un modèle de Fourier permet d'approximer globalement des mouvements de différentes natures. L'approximation reste relativement grossière, mais permet d'extraire les structures spatiales principales du flot optique, et ainsi de réaliser une description satisfaisante du mouvement.

Néanmoins, cette modélisation ne permet pas d'estimer précisément le mouvement entre deux images, notamment dans le cas de mouvements affines simples. Ce problème est très sûrement lié au fait que toutes les fonctions de base ont comme support d'estimation l'image entière. Ainsi, si un paramètre est mal estimé, l'erreur s'applique à la totalité du support, entraînant une estimation erronée du flot optique. La modélisation du mouvement par une base d'ondelettes, dont le support spatial est localisé, semble de ce point de vue plus adaptée. Le chapitre suivant présente l'algorithme d'estimation d'un modèle de mouvement global basé sur les ondelettes ainsi que les résultats obtenus.

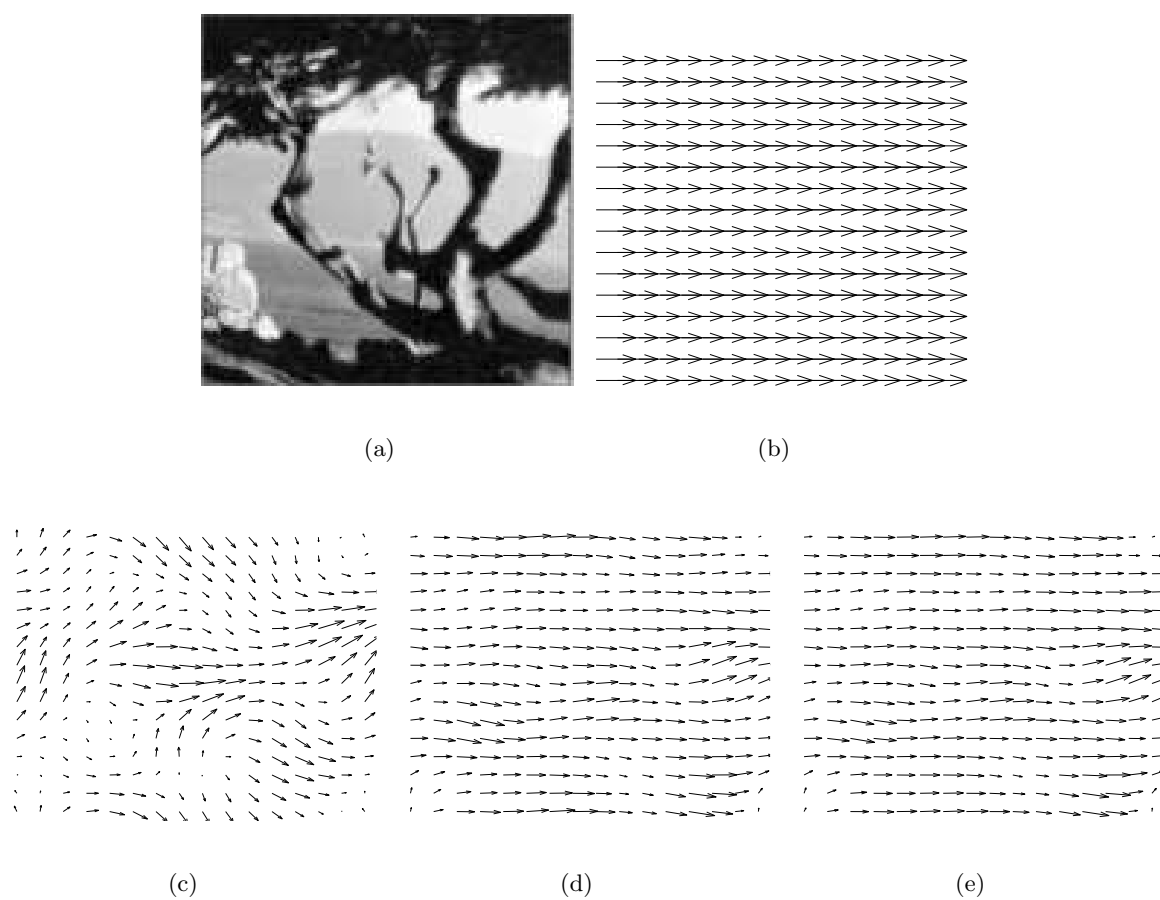


Fig. 5.3: a) Image de la séquence *Arbre en Translation*, b) flot optique réel, flots optiques estimés avec les modèles de Fourier définis pour c) $N = 4$, d) $N = 6$ et e) $N = 8$

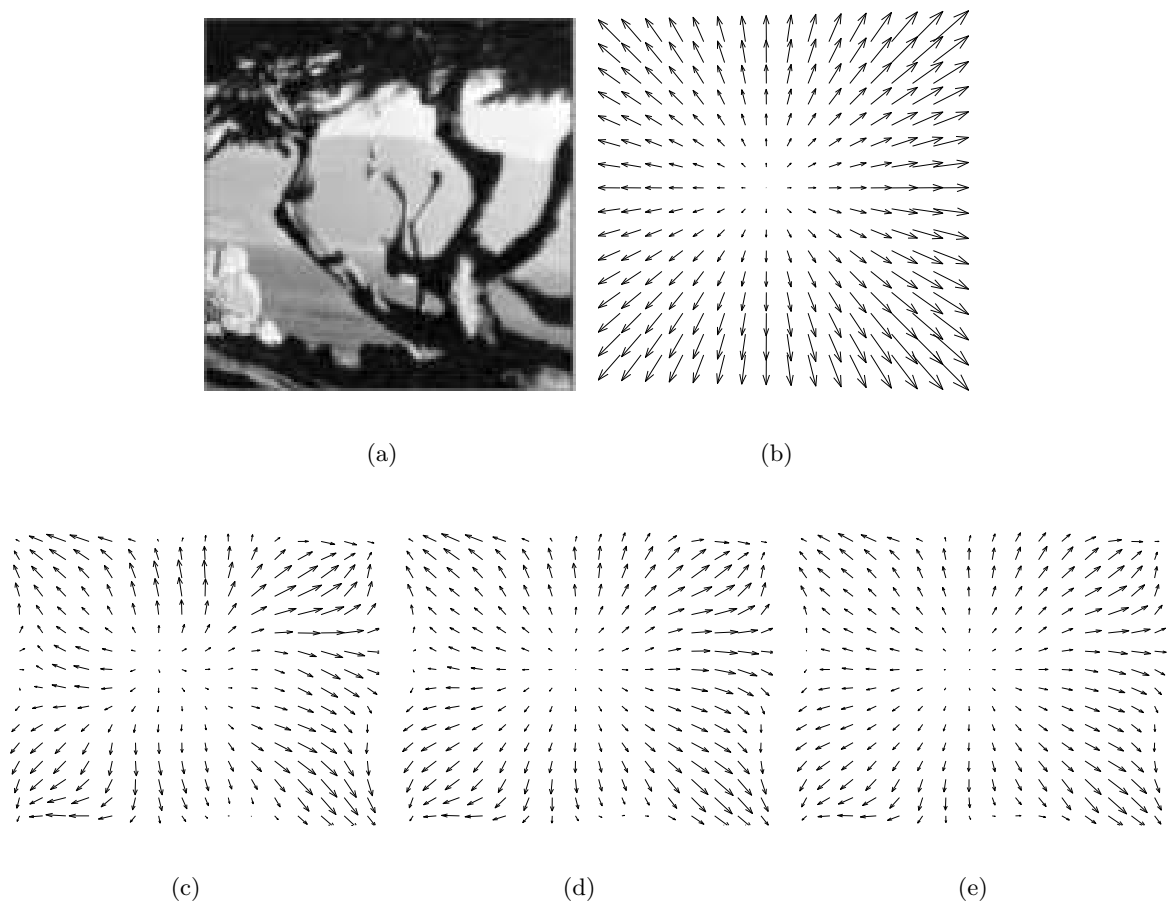


Fig. 5.4: a) Image de la séquence *Arbre en Divergence*, b) flot optique réel, flots optiques estimés avec les modèles de Fourier définis pour c) $N = 4$, d) $N = 6$ et e) $N = 8$

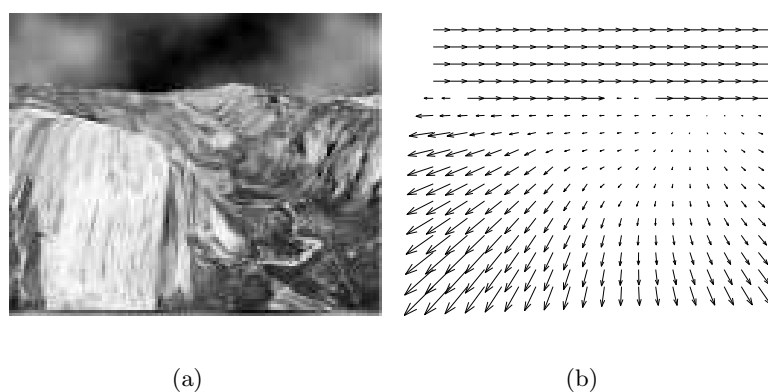


Fig. 5.5: a) Image de la séquence *Yosemite*, b) flot optique réel, flots optiques estimés avec les modèles de Fourier définis pour c) $N = 4$, d) $N = 6$ et e) $N = 8$

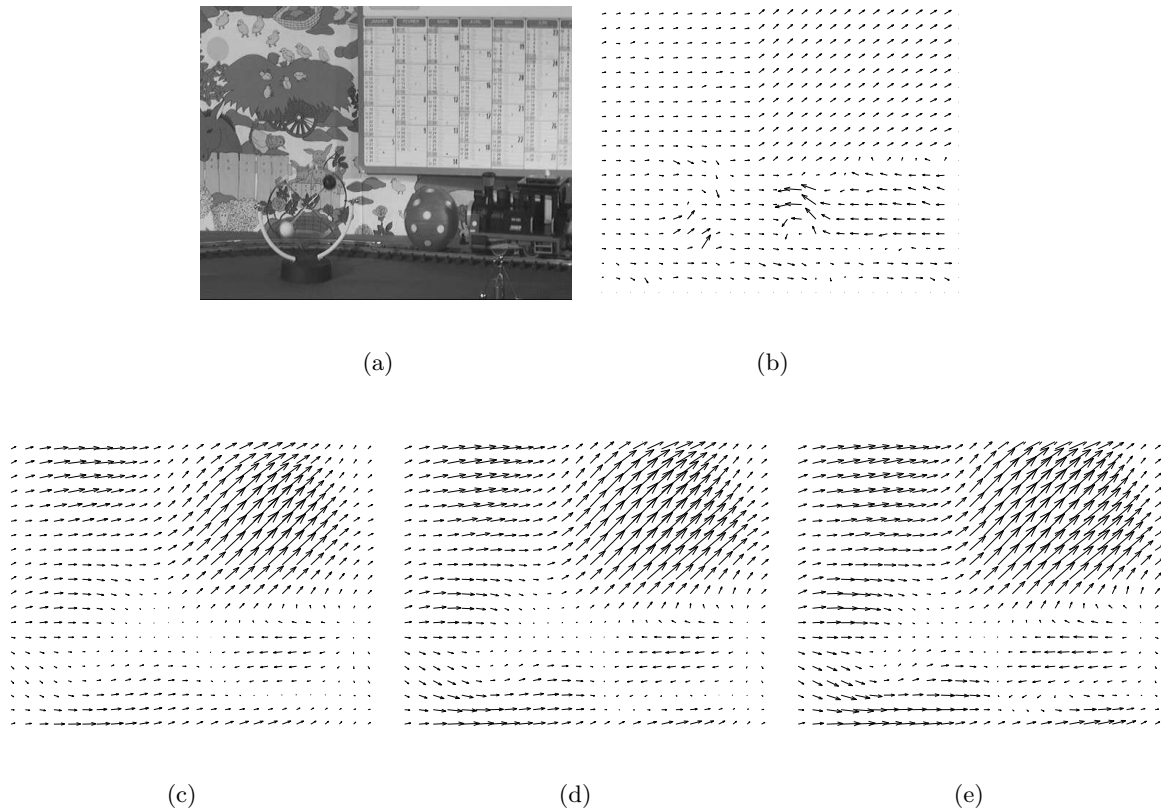


Fig. 5.6: Séquence *Mobile and calendar* avec a) image de la séquence, b) estimation locale du flot optique par filtres de Gabor spatiaux, estimation globale du flot optique au niveau c) $N = 4$, d) $N = 6$ et e) $N = 8$.

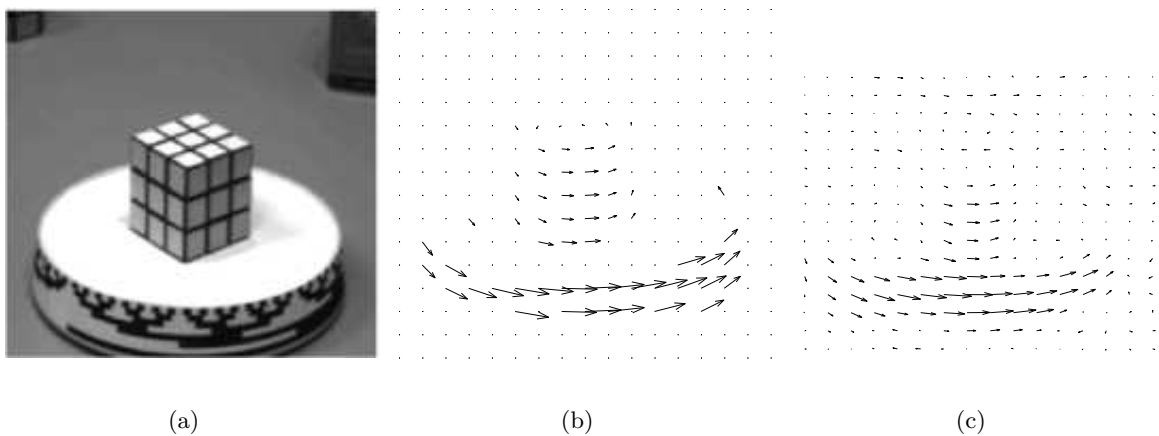


Fig. 5.7: Séquence *Rubik Cube* avec a) image de la séquence, b) estimation locale du flot optique par filtres de Gabor spatiaux et c) estimation globale du flot optique au niveau $N = 8$

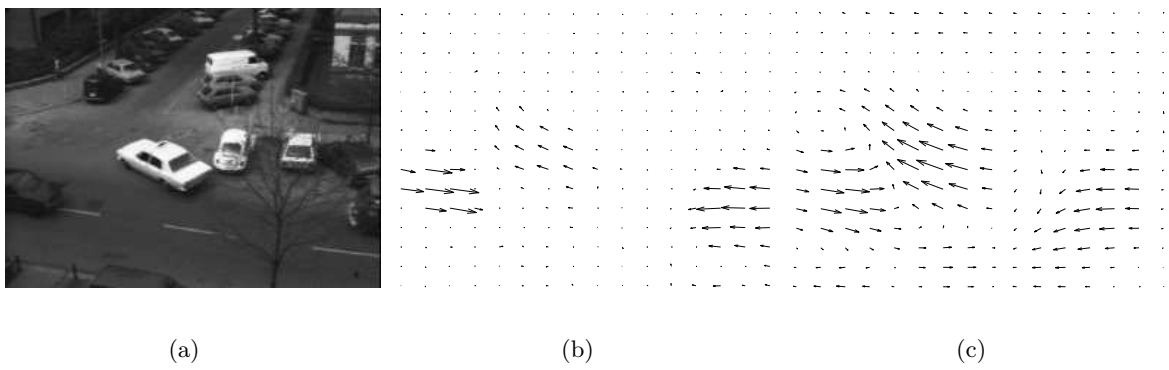


Fig. 5.8: Séquence *Taxi de Hambourg* avec a) image de la séquence, b) estimation locale du flot optique par filtres de Gabor spatiaux et c) estimation globale du flot optique au niveau $N = 8$

Chapitre 6

Modélisation globale du mouvement par bases d'ondelettes

6.1 Introduction

Tout comme la représentation en séries de Fourier, les séries d'ondelettes permettent d'approximer une fonction continue par morceaux par un nombre fini de termes. Dans la majorité des cas cette approximation est bien meilleure que par la description de Fourier, et cela avec beaucoup moins de fonctions de base et donc de coefficients utiles à la reconstruction. C'est pourquoi par exemple le nouveau format de compression d'images JPEG 2000 [40] utilise une décomposition en ondelettes plutôt qu'en cosinus discrets (comme cela est le cas pour la norme JPEG [67]) pour coder l'information. L'utilisation de modèles de mouvement basés sur les ondelettes doit en théorie nous conduire à une meilleure approximation du mouvement global.

Parallèlement à nos travaux, Wu *et. al.* [90] ont récemment proposé un modèle de mouvement global défini par une base d'ondelettes de Cai-Wang (voir la sous-section 4.3.3). Il est intéressant de constater que notre approche présente des similarités avec leurs travaux alors que nos motivations respectives étaient au départ opposées : partant des travaux de Szeliski et Coughlan [76], Wu *et. al.* ont voulu donner, par l'introduction de base d'ondelettes, un caractère plus global au modèle de mouvement. A l'inverse, partant d'un modèle de Fourier purement global, nous avons introduit la représentation en ondelettes afin de mieux modéliser les variations locales du flot optique. D'un point de vue pratique, nos approches diffèrent néanmoins sur les aspects suivants : la nature de la base d'ondelettes, l'algorithme d'estimation des paramètres de mouvement, et enfin la stratégie multirésolution utilisée.

Dans ce chapitre, après un bref rappel sur la théorie des ondelettes discrètes, nous présentons le modèle de mouvement global basé sur les séries d'ondelettes. Nous décrivons ensuite les modifications apportées à l'algorithme d'estimation des paramètres, afin de tenir compte de la localisation spatiale des ondelettes. Le choix de la base d'ondelettes est ensuite discuté en fonction des critères de compacité, de régularité et de précision d'approximation des fonctions de base. Enfin, des résultats d'estimation du mouvement global sont présentés,

permettant de valider l'approche.

6.2 Définition d'un modèle de mouvement basé sur les ondelettes

6.2.1 Notions de base sur la transformation en ondelettes discrètes

La transformation en ondelettes discrètes repose sur l'Analyse Multirésolution (A.M.) de l'espace de fonctions de carré intégrable $L^2(\mathbb{R})$ [47, 50].

Soit V_j , avec $j \in \mathbb{Z}$, une famille de sous-espaces vectoriels fermés de $L^2(\mathbb{R})$. V_j contient l'approximation \tilde{f}_j de toutes fonctions $f(x) \in L^2(\mathbb{R})$ au niveau de résolution j . L'approximation de $f(x)$ au niveau $j + 1$ contient toute l'information nécessaire pour l'approcher au niveau j , c'est à dire :

$$\forall j \in \mathbb{Z}, \quad V_j \subset V_{j+1} \quad (6.1)$$

Lorsque la résolution croît, l'approximation \tilde{f}_j tend vers $f(x)$ et $\bigcup_{j \in \mathbb{Z}} V_j$ est dense dans $L^2(\mathbb{R})$.

Le facteur d'échelle de la résolution est égal à 2 :

$$\forall j \in \mathbb{Z}, \quad (f(x) \in V_j) \iff (f(2x) \in V_{j+1}) \quad (6.2)$$

Il existe une fonction $\phi \in L^2(\mathbb{R})$ telle que la famille $\{\phi_{j,k}(x) = \phi(2^j x - k)\}_{k \in \mathbb{Z}}$ soit une base de V_j . Alors :

$$\tilde{f}_j(x) = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j,k}(x) \quad (6.3)$$

$\phi_{j,k}$ est appelée *fonction d'échelle*.

De ce qui précède on déduit qu'il existe un sous-espace W_j qui est le supplémentaire orthogonal de V_j dans V_{j+1} , c-à-d :

$$\forall j \in \mathbb{Z}, \quad V_{j+1} = V_j \oplus W_j \quad (6.4)$$

De même que pour V_j , il existe une famille de fonctions, appelées *ondelettes*, $\{\psi_{j,k}(x) = \psi(2^j x - k)\}_{k \in \mathbb{Z}}$ formant une base de W_j . La décomposition en séries d'ondelettes discrètes de $f(x) \in L^2(\mathbb{R})$ au niveau de résolution j , est alors définie comme une combinaison linéaire de fonctions d'échelles et d'ondelettes :

$$\tilde{f}_j(x) = \sum_{k \in \mathbb{Z}} c_{l,k} \phi_{l,k}(x) + \sum_{n \geq l} \sum_{k \in \mathbb{Z}} d_{n,k} \psi_{n,k}(x) \quad (6.5)$$

L'indice $l < j$ correspond au niveau de résolution le plus bas de la décomposition, et peut être choisi arbitrairement.

Ainsi, pour une résolution j donnée, la relation (6.3) est équivalente à la relation (6.5). Il en résulte que connaissant les coefficients $\{c_{j+1,k}\}_k$ à une échelle $j + 1$, on déduit tous les

coefficients d'ondelettes et de fonctions d'échelle pour l'ensemble des niveaux de résolution inférieurs :

$$\sum_k c_{j+1,k} \phi_{j+1,k}(x) = \sum_k c_{j,k} \phi_{j,k}(x) + \sum_k d_{j,k} \psi_{j,k}(x) \quad (6.6)$$

La fonction d'échelle et l'ondelette de résolution j sont obtenues par combinaison linéaire de fonctions d'échelle de niveau de résolution $j + 1$. Cette relation entre deux échelles successives est appelée l'*équation de dilatation* :

$$\begin{aligned} \phi_j(x) &= \sum_k h(k) \phi_{j+1,k}(x) \\ \psi_j(x) &= \sum_k g(k) \phi_{j+1,k}(x) \end{aligned} \quad (6.7)$$

Les coefficients de pondération $h(k)$ et $g(k)$ dépendent de la base d'ondelettes utilisée, et correspondent respectivement à un filtre passe-bas et passe-haut.

6.2.2 Modélisation du mouvement

Nous désirons modéliser le flot optique $\mathbf{v}(x, y)$ par des fonctions d'échelle 2D. Généralement, le passage de la transformation en ondelettes discrètes de dimension 1 en dimension 2 s'effectue par le produit tensoriel suivant :

$$\begin{aligned} \Phi_{j,k_1,k_2}(x, y) &= \phi(2^j x - k_1) \phi(2^j y - k_2) \\ \Psi_{j,k_1,k_2}^H(x, y) &= \psi(2^j x - k_1) \phi(2^j y - k_2) \\ \Psi_{j,k_1,k_2}^D(x, y) &= \psi(2^j x - k_1) \psi(2^j y - k_2) \\ \Psi_{j,k_1,k_2}^V(x, y) &= \phi(2^j x - k_1) \psi(2^j y - k_2) \end{aligned} \quad (6.8)$$

Les indices j , k_1 , k_2 représentent respectivement l'échelle et le décalage horizontal et vertical. Les ondelettes Ψ^H , Ψ^D et Ψ^V représentent respectivement les variations horizontales, diagonales et verticales du signal 2D analysé.

Le mouvement global peut être approximé au niveau de résolution j par une décomposition en séries d'ondelettes. La séquence d'images étant un signal discret, il est nécessaire d'imposer que la taille de son support spatial soit dyadique ($n = 2^m$, où m est quelconque), afin de pouvoir lui associer une A.M. $\{V_j\}_{j=0,\dots,m}$.

Alors, $\forall \mathbf{p}_i = (x_i, y_i)^T \in [n \times n]^2$, l'approximation au niveau j de $\mathbf{v}(\mathbf{p}_i)$ (selon les relations (6.3) et (6.5)) est :

$$\mathbf{v}_j(\mathbf{p}_i) = \sum_{k_1, k_2=0}^{2^j-1} \mathbf{c}_{j,k_1,k_2} \Phi_{j,k_1,k_2}(\mathbf{p}_i) \quad (6.9)$$

ou encore,

$$\begin{aligned}
 \mathbf{v}_j(\mathbf{p}_i) = & \sum_{k_1, k_2=0}^{2^l-1} \mathbf{c}_{l, k_1, k_2} \Phi_{l, k_1, k_2}(\mathbf{p}_i) \\
 & + \sum_{n \geq l}^{j-1} \sum_{k_1, k_2=0}^{2^n-1} [\mathbf{d}_{n, k_1, k_2}^H \Psi_{n, k_1, k_2}^H(\mathbf{p}_i) + \mathbf{d}_{n, k_1, k_2}^D \Psi_{n, k_1, k_2}^D(\mathbf{p}_i) + \mathbf{d}_{n, k_1, k_2}^V \Psi_{n, k_1, k_2}^V(\mathbf{p}_i)]
 \end{aligned} \tag{6.10}$$

L'approximation (6.9), équivalente à l'expression (6.10), présente l'avantage d'être basée uniquement sur la fonction d'échelle $\Phi(\mathbf{p}_i)$, translatée sur toute l'image et dilatée à différentes échelles. Par la suite, pour simplifier l'implantation de l'algorithme, nous utiliserons la relation (6.9) comme modèle de mouvement. Le vecteur de paramètres du mouvement global $\boldsymbol{\theta}_j = [\mathbf{c}_{j, 0, 0}, \mathbf{c}_{j, 1, 0} \dots \mathbf{c}_{j, 2^{j-1}, 2^{j-1}}]$, contenant les coefficients des fonctions d'échelle de résolution j , permet de retrouver tous les vecteurs $\boldsymbol{\theta}_l$ ainsi que les coefficients d'ondelette $\mathbf{d}_{l, k_1, k_2}^H$, $\mathbf{d}_{l, k_1, k_2}^D$ et $\mathbf{d}_{l, k_1, k_2}^V$ pour $l \in [0, j]$.

6.3 Estimation des paramètres du modèle de mouvement

L'estimation du vecteur de paramètres du mouvement $\boldsymbol{\theta}_j$ s'effectue selon un schéma quasi identique à celui présenté dans la section 5.2, chapitre 5. La fonction robuste est le M-estimateur de Geman-McClure (5.5). La minimisation de la fonction de coût est obtenue par l'algorithme ERIM (figure 4.3).

6.3.1 Estimation multirésolution

Le mouvement global est estimé de manière progressive, par enrichissement successif du modèle. En effet, de même que pour le modèle de Fourier, les fonctions d'échelle de résolutions fines ne permettent pas d'estimer des déplacements importants.

Le cadre mathématique de la transformation en ondelettes étant l'Analyse Multirésolution, notre modèle de mouvement s'adapte tout naturellement à l'estimation multirésolution. Une première estimation basée sur un modèle de faible résolution, appartenant au sous-espace V_0 (l'indice 0 est arbitraire et désigne uniquement le modèle le plus grossier utilisé), permet de mesurer les mouvements de grande amplitude. Cette estimation basse-résolution initialise l'étape suivante, où un modèle de résolution plus fine (appartenant à V_1) est utilisé afin d'affiner le flot optique. Ainsi, d'étape en étape, des détails du mouvement de plus en plus fins sont mesurés.

6.3.2 Résolution d'un système creux

Il existe néanmoins une différence notable entre l'estimation des paramètres du modèle de Fourier et du modèle d'ondelettes. Celle-ci concerne la résolution du système linéaire (5.9) conduisant à la minimisation incrémentale de la fonction de coût.

Rappelons que la solution $\delta\boldsymbol{\theta}$ de ce système s'écrit en notation matricielle (pour une résolution j fixée) :

$$\delta\boldsymbol{\theta} = (M^T W M)^{-1} M^T W B \quad (6.11)$$

avec, pour

$k_1 \times k_2 = [0 \dots 2^j - 1]^2$, $p = 2^{2j+1}$ le nombre de paramètres.

$\forall \mathbf{p}_i = (x_i, y_i) \in \Omega$, Ω étant le support d'estimation, $m = \mathbf{card}(\Omega)$ le nombre d'équations :

$$\begin{aligned} M &= [I_x(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}))\Phi_{j,k_1,k_2}(\mathbf{p}_i), I_y(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta}))\Phi_{j,k_1,k_2}(\mathbf{p}_i)], \quad M \in \mathbb{R}^{m \times p} \\ \delta\boldsymbol{\theta} &= [\delta\mathbf{c}_{k_1,k_2}]^T, \quad \delta\boldsymbol{\theta} \in \mathbb{R}^p \\ W &= [\text{diag}(w(\mathbf{p}_i + \mathbf{v}(\mathbf{p}_i; \boldsymbol{\theta})))], \quad W \in \mathbb{R}^{m \times m} \\ B &= r(\mathbf{p}_i; \boldsymbol{\theta}), \quad B \in \mathbb{R}^m \end{aligned} \quad (6.12)$$

L'utilisation de fonctions d'échelle spatialement localisées comme fonctions de base conduit à ce que la matrice $M^T W M$ soit creuse, c'est à dire que la majorité des éléments qui la composent est nulle. Cette propriété est très intéressante car elle permet un gain de temps de calculs et d'espace mémoire lors de la résolution du système, et de fait, autorise l'utilisation de modèles définis par un plus grand nombre de paramètres.

Les éléments de la matrice $M^T W M$ sont de la forme :

$$(M^T W M)_{m,n} = \sum_{\mathbf{p}_i} \lambda(\mathbf{p}_i) \Phi_{(m)}(\mathbf{p}_i) \Phi_{(n)}(\mathbf{p}_i) \quad (6.13)$$

où $\Phi_{(n)} = \Phi_{j,k_1,k_2}$, k_2 et k_1 étant respectivement le reste et le quotient de la division euclidienne de n par 2^j (j est fixé). $\lambda(\mathbf{p}_i)$ est une grandeur dépendante de la pondération $w(\mathbf{p}_i)$ et des gradients spatiaux $I_x(\mathbf{p}_i)$ et $I_y(\mathbf{p}_i)$.

Supposons que la fonction d'échelle Φ soit définie sur un support borné $[0 \dots N] \times [0 \dots N]$, c'est à dire $\Phi(\mathbf{p}_i) = 0$, $\forall \mathbf{p}_i \notin [0 \dots N] \times [0 \dots N]$. Alors l'expression (6.13) est différente de zéro si et seulement si les deux fonctions d'échelle $\Phi_{(m=2^j k_1+k_2)}$ et $\Phi_{(n=2^j l_1+l_2)}$ se recouvrent au moins partiellement :

$$\begin{aligned} |k_1 - l_1| &\leq N \\ |k_2 - l_2| &\leq N \end{aligned} \quad (6.14)$$

Ainsi, quelle que soit la nature de la fonction d'échelle utilisée, le système (6.11) est d'autant plus creux que le support de Φ est petit. La borne minimale est atteinte pour l'ondelette de Haar, définie sur $[0, 1]^2$. Dans ce cas la matrice $M^T W M$ est diagonale par blocs, chaque bloc définissant un sous-système indépendant de tous les autres. La résolution de ce sous-système donne une estimation locale de la vitesse à l'endroit où la fonction est non-nulle. La figure 6.1 représente la matrice $M^T W M$ pour des fonctions d'échelles définies sur $[0, 1]^2$, $[0, 2]^2$, $[0, 3]^2$ et $[0, 4]^2$.

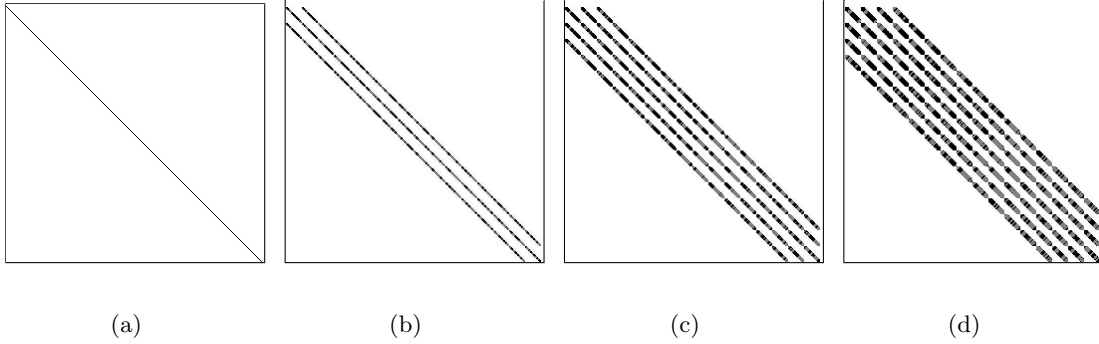


Fig. 6.1: Matrice $M_j^T W M_j$ pour des fonctions d'échelle définies sur a) $[0, 1]^2$, b) $[0, 2]^2$, c) $[0, 3]^2$ et d) $[0, 4]^2$ au niveau $j = 4$. Les régions blanches et noires représentent respectivement les entrées nulles et non-nulles de la matrice.

La méthode du gradient conjugué

La méthode du *gradient conjugué* est généralement employée pour résoudre des systèmes creux de la forme [66] :

$$A\mathbf{x} = \mathbf{b} \quad (6.15)$$

où A est une matrice symétrique définie positive (ce qui est le cas de $M^T W M$ lorsque le système est bien conditionné). L'intérêt de la résolution par le gradient conjugué est que les calculs nécessaires à l'estimation de \mathbf{x} ne sont que des multiplications de A et de sa transposée avec des vecteurs. Les éléments de A étant majoritairement égaux à zéro, ces opérations s'effectuent très rapidement.

L'idée de cette technique est d'estimer la solution du système par un algorithme itératif minimisant la fonction

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - \mathbf{x}^T \mathbf{b} \quad (6.16)$$

Cette fonction est minimale lorsque son gradient

$$\nabla f = A\mathbf{x} - \mathbf{b} \quad (6.17)$$

est égal à zéro, ce qui est équivalent à (6.15).

La convergence de l'algorithme peut être accélérée en l'appliquant à une forme *préconditionnée* du système (6.15) :

$$(\tilde{A}^{-1} A)\mathbf{x} = \tilde{A}^{-1} \mathbf{b} \quad (6.18)$$

Si \tilde{A} est une matrice proche de A , telle que $\tilde{A}^{-1} A \approx \mathbf{1}$, l'algorithme converge en quelques étapes. La matrice \tilde{A} est appelée un *préconditionneur*. Tout le problème est évidemment de déterminer \tilde{A} sans calculer A^{-1} . Une solution consiste à définir \tilde{A} comme une matrice diagonale dont les termes diagonaux sont égaux à ceux de A [66].

Dans notre cas, où $A = M^T W M$, les éléments de $\tilde{A} = \text{diag}(A)$ sont successivement de la forme

$$\tilde{A} = \begin{pmatrix} \sum_{\mathbf{p}_i} w I_x^2 \Phi_{j,0,0}^2 & & & & & \\ & \sum_{\mathbf{p}_i} w I_y^2 \Phi_{j,0,0}^2 & & & & \\ & & \ddots & & & \\ & & & \sum_{\mathbf{p}_i} w I_x^2 \Phi_{j,N,N}^2 & & \\ & & & & \sum_{\mathbf{p}_i} w I_y^2 \Phi_{j,N,N}^2 & \end{pmatrix} \quad (6.19)$$

et représentent une mesure de la texture sur le support de la fonction d'échelle. Si une fonction d'échelle $\Phi_{(n)}$ recouvre une zone uniforme de l'image, l'élément correspondant dans \tilde{A} est proche de zéro. Le coefficient \mathbf{c}_{j,k_1,k_2} , associé à $\Phi_{(n)}$, est indéterminé et l'estimation de l'ensemble des coefficients en est perturbée (on rencontre à nouveau le problème d'ouverture).

Cette indétermination peut être évitée en mettant à zéro l'inverse des éléments du préconditionneur \tilde{A} dont l'amplitude est inférieure à un seuil donné :

$$\text{Si } \tilde{A}_{n,n} < \epsilon \text{ alors } 1/\tilde{A}_{n,n} = 0 \quad (6.20)$$

L'estimation \mathbf{c}_{j,k_1,k_2} est alors égale à zéro et l'instabilité est atténuée. La solution n'appartient pas strictement à l'espace image de l'application associée à A et, contrairement à la résolution par *SVD*, n'est pas optimale au sens des moindres-carrés. Néanmoins, comme nous le verrons dans la partie des résultats, le seuillage de l'inverse du préconditionneur \tilde{A} (6.20) permet de contrôler de façon satisfaisante le mauvais conditionnement des coefficients.

Si le système n'est pas de rang plein, et que des valeurs de \tilde{A}^{-1} sont mises à zéro, la matrice A n'est plus définie positive. Dans ce cas nous utilisons la généralisation du gradient conjugué, appelé *gradient biconjugué*, qui est valable dans le cas où le système n'est pas symétrique ou défini positif. Les détails de l'algorithme du gradient conjugué et biconjugué préconditionné sont donnés en annexe B.

6.4 Choix d'une base d'ondelettes

Il existe une infinité de bases d'ondelettes discrètes, et il nous est évidemment impossible de toutes les tester pour déterminer la plus adaptée à la modélisation du mouvement. La base doit être définie en fonction des deux contraintes suivantes : $\{\Phi_{j,k_1,k_2}\}_{k_1,k_2}$ doit permettre la meilleure approximation possible du mouvement et la fonction Φ doit être définie sur le support le plus compact possible, afin de réduire le temps de calcul.

On peut montrer [75] que toute fonction $f(x) \in C^p$ peut être approximée à l'ordre 2^{-jp} par sa projection $f_j(x)$ dans V_j par une base $\{\phi_{j,k}\}_k$:

$$\|f(x) - f_j(x)\| \leq C 2^{-jp} \|f^{(p)}(x)\| \quad (6.21)$$

La constante C et l'exposant p dépendent de la fonction d'échelle ϕ , $f^{(p)}(x)$ désigne la dérivée $p^{\text{ième}}$ de $f(x)$. Le nombre p , appelé ordre d'approximation de ϕ , est lié au degré des polynômes

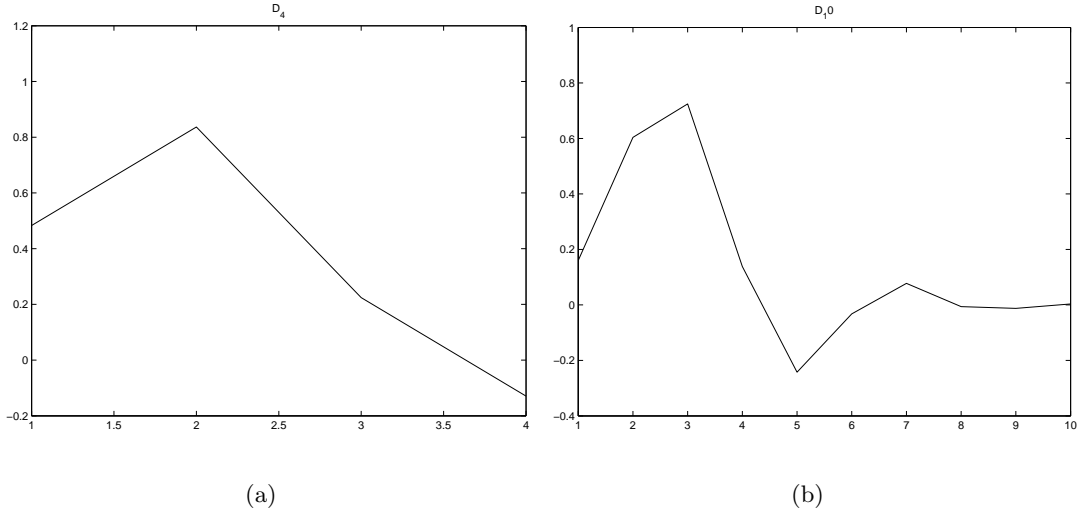


Fig. 6.2: Fonctions d'échelle de Daubechies a) D4 et b) D10

parfaitement reconstituable par $\{\phi_{j,k}\}_k$:

$$\left\| \sum_{i=0}^{p-1} a_i x^i - \sum_k c_k \phi_{j,k}(x) \right\| = 0, \quad \forall a_i \in \mathbb{R} \quad (6.22)$$

Ainsi l'utilisation de fonctions d'échelle ayant un ordre d'approximation $p = 2$ permet de modéliser les mouvements affines. Si $p = 3$, la base est capable de modéliser toute transformation quadratique du plan.

En général, plus p est grand, plus le support de ϕ est important. Afin d'obtenir le meilleur compromis entre la précision et le temps de calcul, nous devons choisir une famille de fonctions d'échelle ayant un support minimum pour un ordre p maximum. Les ondelettes de Daubechies [17], noté D_{2p} et dont le support est défini sur $[0, 2p]$, sont, parmi toutes les ondelettes orthogonales, celles qui offrent un support minimum pour un ordre p maximum. Elles sont néanmoins très irrégulières et non symétriques (Figure 6.2), ce qui pose des problèmes à la fois de bords et de régularité du flot optique.

La propriété d'orthogonalité des fonctions ϕ n'est jamais utilisée et n'est donc pas une condition nécessaire à la résolution du problème. Aussi, nous pouvons utiliser des bases biorthogonales d'ondelettes B-splines qui semblent bien adaptées à la modélisation du mouvement.

6.4.1 Ondelettes B-splines

Les *B-splines* sont des fonctions ayant une régularité et une symétrie maximum. Elles maximisent l'ordre d'approximation p par rapport à la taille du support compact.

De plus, il a été montré que la constante C de l'inégalité (6.21) est minimum pour les fonctions d'échelle B-splines [75].

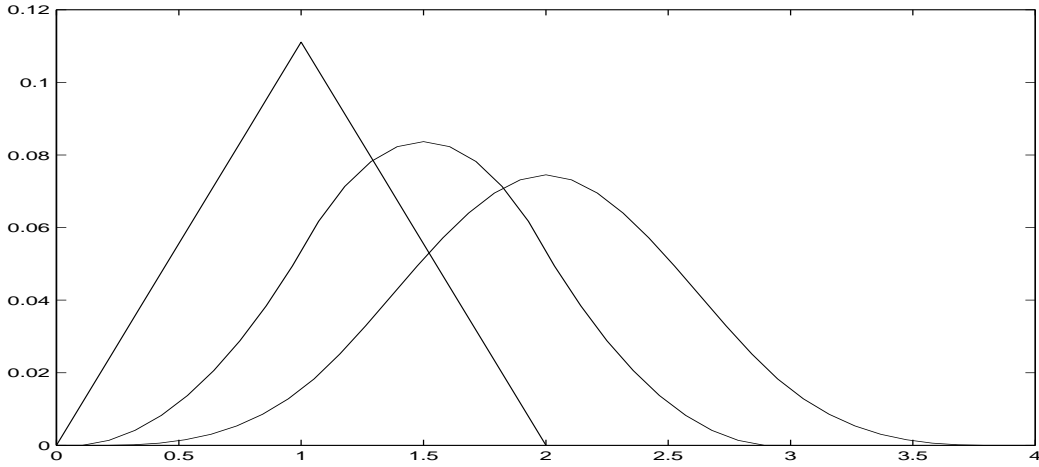


Fig. 6.3: Fonctions d'échelle B-splines de degré 1, 2 et 3 pour $j = 0$

La fonction B-spline de degré N s'écrit [82] :

$$\phi^N(x) = \frac{1}{(N)!} \sum_{k=0}^{N+1} \binom{N+1}{k} (-1)^k (x-k)_+^N \quad (6.23)$$

où, quel que soit n

$$x_+^n = \begin{cases} x^n & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases} \quad (6.24)$$

$\phi^N(x)$ a pour support $[0, N+1]$ et pour ordre d'approximation $p = N$. La figure 6.3 représente les fonctions B-spline 1D de degré 1, 2 et 3 que nous utiliserons par la suite.

ϕ^N définit la fonction d'échelle au niveau de résolution 0, c-à-d $\phi_0^N(x) = \phi^N(x)$. Comme nous travaillons sur un signal à support fini, $\phi_0^N(x)$ est dilatée afin d'être définie sur tout le signal. Alors, les fonctions d'échelle au niveau de résolution j , translatées d'un facteur k , sont définies par :

$$\phi_{j,k}^N(x) = \phi_0^N(2^j x - k) \quad (6.25)$$

Les fonctions de base $\Phi_{j,k_1,k_2}^N(\mathbf{p}_i)$ sont obtenues par la relation (6.8), formant des bases de B-splines à différentes résolutions (Figure 6.4).

6.4.2 Précision de l'approximation en fonction du degré de Φ^N

Le degré N des fonctions B-splines modélisant le mouvement a une influence très importante sur le flot optique estimé.

Nous savons que plus le degré N est élevé, plus le modèle peut approximer des variations d'ordre élevé d'une fonction modélisée (cf. inégalité 6.21). Néanmoins, lorsque la fonction contient des variations d'ordre inférieur au degré des B-splines, la solution obtenue au sens des moindres-carrés a tendance à osciller autour de la solution réelle.

Les figures 6.5 et 6.6 illustrent ces problèmes. L'approximation au sens des moindres-carrés d'une fonction linéaire oscille lorsque le degré N est supérieur à 1 (Figure 6.5.b et c),

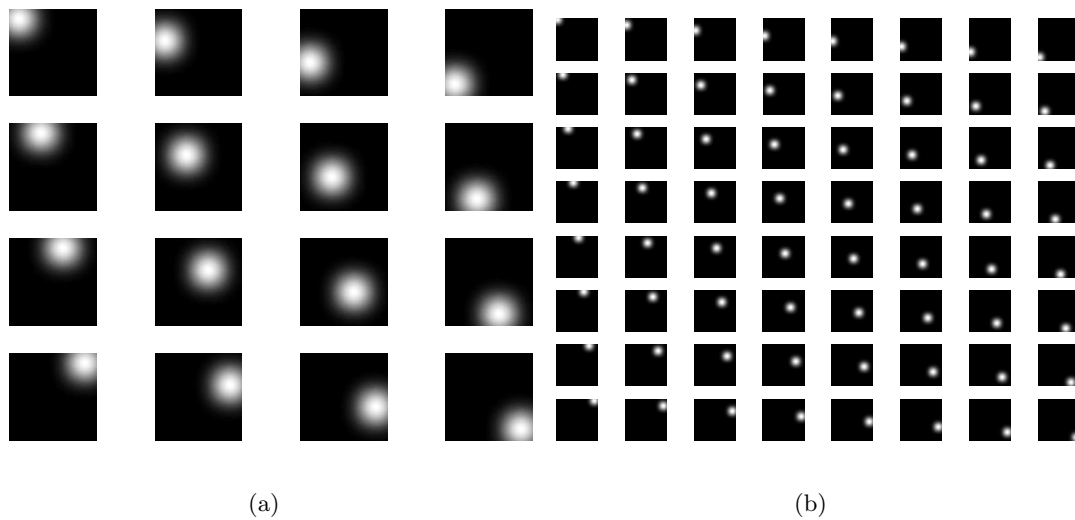


Fig. 6.4: Bases de fonctions d'échelles B-splines cubiques pour a) $j = 2$ (4×4 fonctions de base et b) $j = 3$ (8×8 fonctions de base)

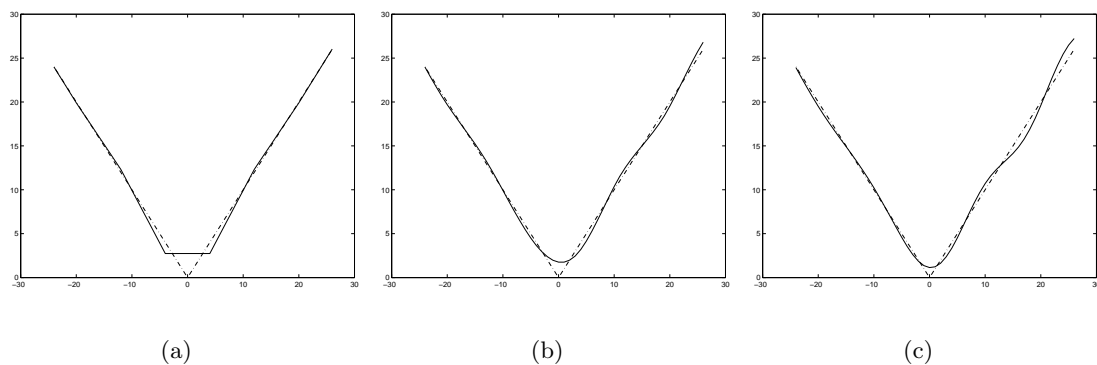


Fig. 6.5: Approximation au sens des moindres-carrés de la fonction $f(x) = |x|$ (en pointillé) avec une base de B-splines de degré a) 1, b) 3 et c) 5. Plus le degré est élevé, plus la solution oscille autour de $f(x)$.

alors qu'elle est très proche de la solution réelle lorsque $N = 1$ (Figure 6.5.a). En revanche, lorsque la fonction est indéfiniment dérivable, l'estimation converge vers la solution réelle lorsque N augmente (Figure 6.6.a, b et c).

Ce résultat nous indique que l'utilisation de B-splines de degré élevé n'est pas forcément le plus adapté pour modéliser le mouvement, où, les frontières mises à part, les variations sont souvent d'un ordre faible.

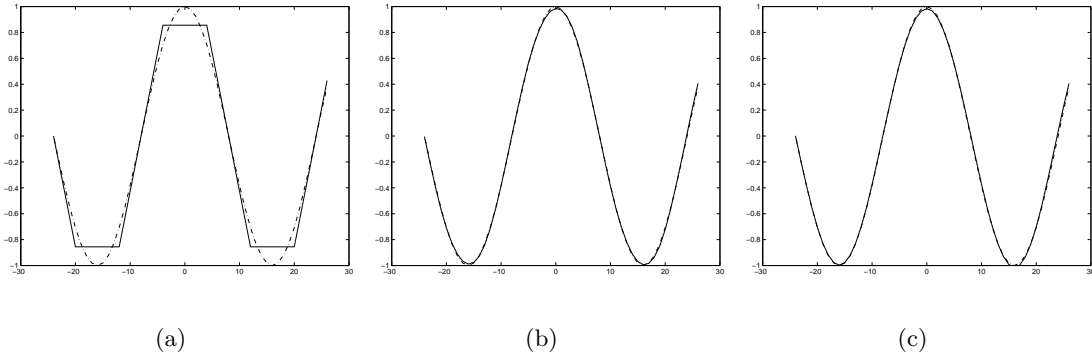


Fig. 6.6: Approximation au sens des moindres-carrés de la fonction $f(x) = \cos(x)$ (en pointillé) avec une base de B-splines de degré a) 1, b) 3 et c) 5. Plus le degré est élevé, plus la solution converge vers la fonction infiniment dérivable $\cos(x)$

6.5 Résultats expérimentaux

Dans cette partie nous présentons les résultats obtenus par le modèle d'ondelettes, en fonction des différents paramètres de l'algorithme et sur plusieurs types de mouvements. Une comparaison entre les performances de notre algorithme et celles obtenues par différents auteurs est aussi donnée. Les séquences d'images sont spatialement ré-échantillonnées afin d'être de dimensions dyadiques.

6.5.1 Résultats en fonction de l'ondelette

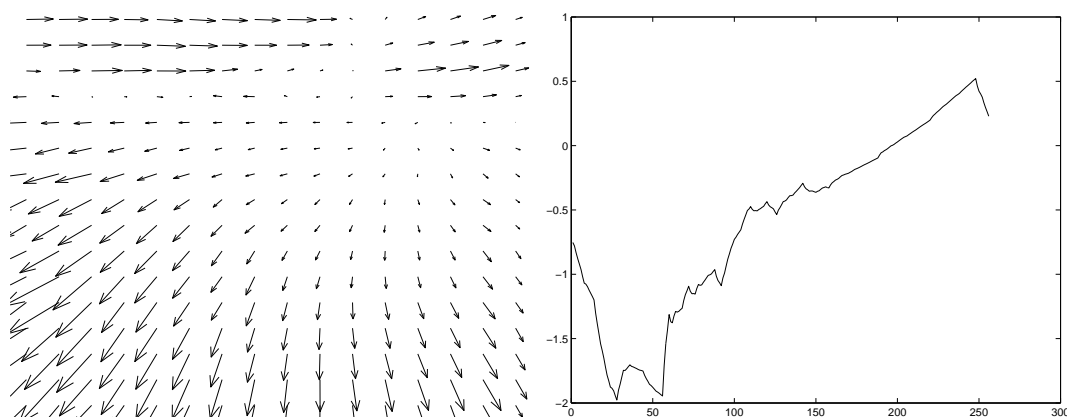
Nous avons testé les fonctions d'échelle de Daubechies pour modéliser le mouvement. Pour cela, nous avons estimé le mouvement global sur la séquence artificielle *Yosemite*. La figure 6.7 présente le flot optique estimé par une base de Daubechies d'ordre 2 et par une base de B-splines de degré 1 (et d'ordre 2), ainsi qu'une coupe de la composante horizontale de \mathbf{v} .

L'irrégularité de la fonction d'échelle de Daubechies se retrouve dans l'approximation du mouvement (Figure 6.7.a), alors que l'utilisation de B-splines permet d'estimer un flot optique régulier.

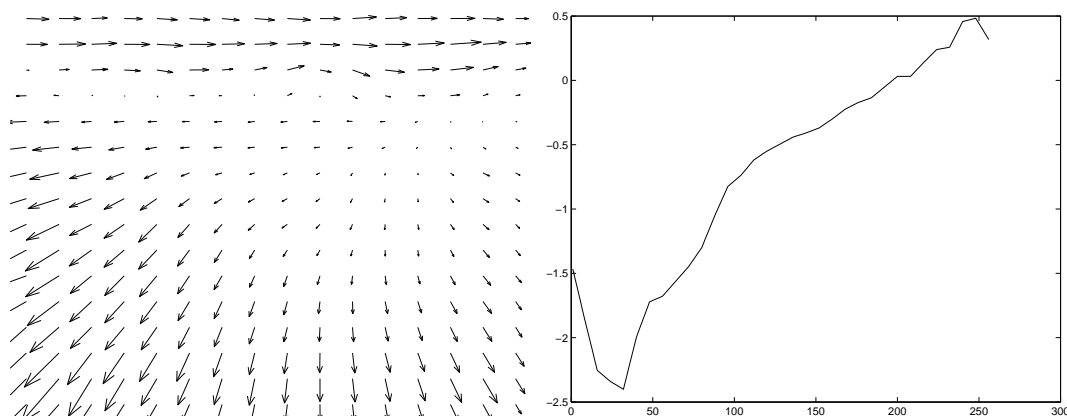
Nous avons calculé l'erreur angulaire (équation 3.27) entre le flot optique réel de la séquence *Yosemite* et le mouvement estimé par les deux types de modèles. La moyenne de l'erreur pour ondelettes de Daubechies est de 9.0° et pour les B-splines de 5.6° .

6.5.2 Détermination expérimentale du seuil ϵ appliqué au préconditionneur

Le seuil ϵ , appliqué à la matrice \tilde{A}^{-1} dans l'équation (6.18), permet d'éliminer les coefficients insuffisamment contraints par les gradients de l'image. Dans l'exemple du *Rubik Cube* (figure 6.8), l'influence du seuillage est déterminante : le fond de la scène est uniforme et rend le système (6.11) singulier. Les coefficients estimés sur le fond uniforme sont erronés (figure 6.8.a) et il en résulte une très mauvaise mesure du flot optique dans cette zone. Le



(a)



(b)

Fig. 6.7: Flot optique et coupe de la composante de vitesse horizontale v_x estimé sur la séquence *Yosemite*, dans le cas où le mouvement est estimé par une base de a) D_4 et b) B-splines de degré 1 (dans les deux cas estimation sur 3 résolutions du modèle $j = 2, 3, 4$).

seuillage du préconditionneur permet de mettre à zéro ces coefficients mal conditionnés et ainsi d'obtenir une estimation valide du flot optique (figure 6.8.b).

Nous avons déterminé expérimentalement le seuil ϵ en dessous duquel l'estimation est instable. Par la suite, pour tous les résultats donnés, ϵ est égal à 2.5.

6.5.3 Comparaison quantitative sur des séquences artificielles

Nous avons testé l'algorithme sur les trois séquences d'images dont nous connaissons le flot optique réel (*Arbre en translation*, *Arbre en divergence* et *Yosemite*) afin de mesurer la précision de l'algorithme.

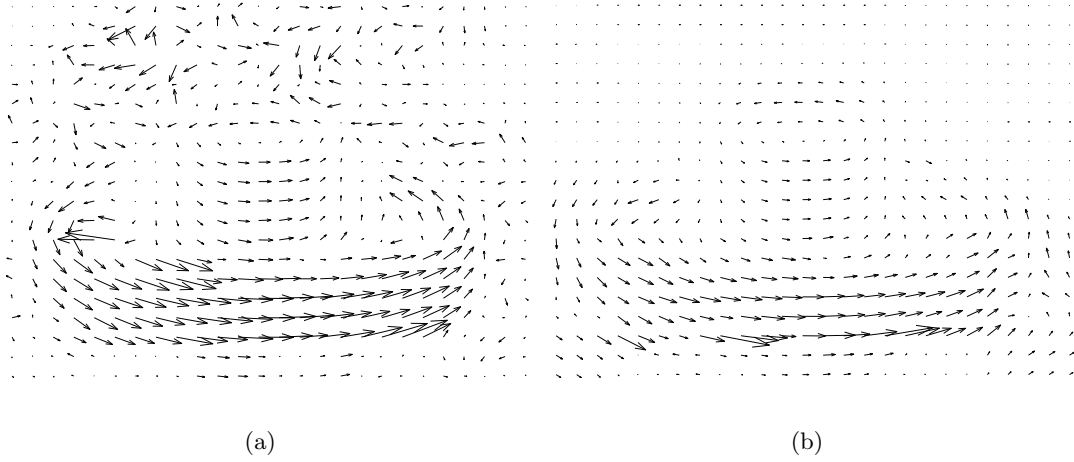


Fig. 6.8: Flot optique estimé sur la séquence *Rubik Cube* : a) sans seuillage et b) avec seuillage de \tilde{A}^{-1} ($\epsilon = 2.5$).

Séquence Yosemite				
B-spline	Niveau	Moyenne	Ecart-type	Densité
Φ^1	$j = 2$	9.1°	9.6°	100%
	$j = 3$	6.4°	8.5°	100%
	$j = 4$	5.6°	9.7°	100%
Φ^2	$j = 2$	9.2°	9.6°	100%
	$j = 3$	5.0°	7.5°	100%
	$j = 4$	4.6°	9.0°	100%
Φ^3	$j = 2$	10.0°	10.7°	100%
	$j = 3$	5.1°	8.0°	100%
	$j = 4$	4.4°	8.7°	100%
Srinivasan [73]		8.9°	10.6°	100%
Szeliski [76]		3.1°	7.6°	39.6%
Wu [90]		4.48°	8.66°	100%

Tab. 6.1: Erreurs angulaires (moyenne et écart-type) obtenues sur la séquence *Yosemite*, avec des B-splines de degré 1, 2 et 3 et aux niveaux de résolution 2, 3 et 4 (estimation progressive du mouvement sur ces 3 niveaux de résolution).

Sur la séquence *Yosemite*, nous avons calculé la moyenne et l'écart-type de l'erreur angulaire (3.27) aux niveaux de résolution $j = 2, 3$ et 4 et pour les B-splines de degré 1, 2 et 3 (tableau 6.1). Pour comparaison, les résultats obtenus par Srinivasan & Chellappa [73], Szeliski & Coughlan [76] et de Wu *et al.* [90] (voir chapitre 4) sont donnés en bas du tableau. Les flots optiques obtenus à ces trois niveaux, ainsi que les coefficients θ_j correspondant à la composante horizontale du mouvement sont représentés figure 6.9.

Le tableau 6.3 présente les résultats obtenus sur les séquences *Arbre en divergence* et

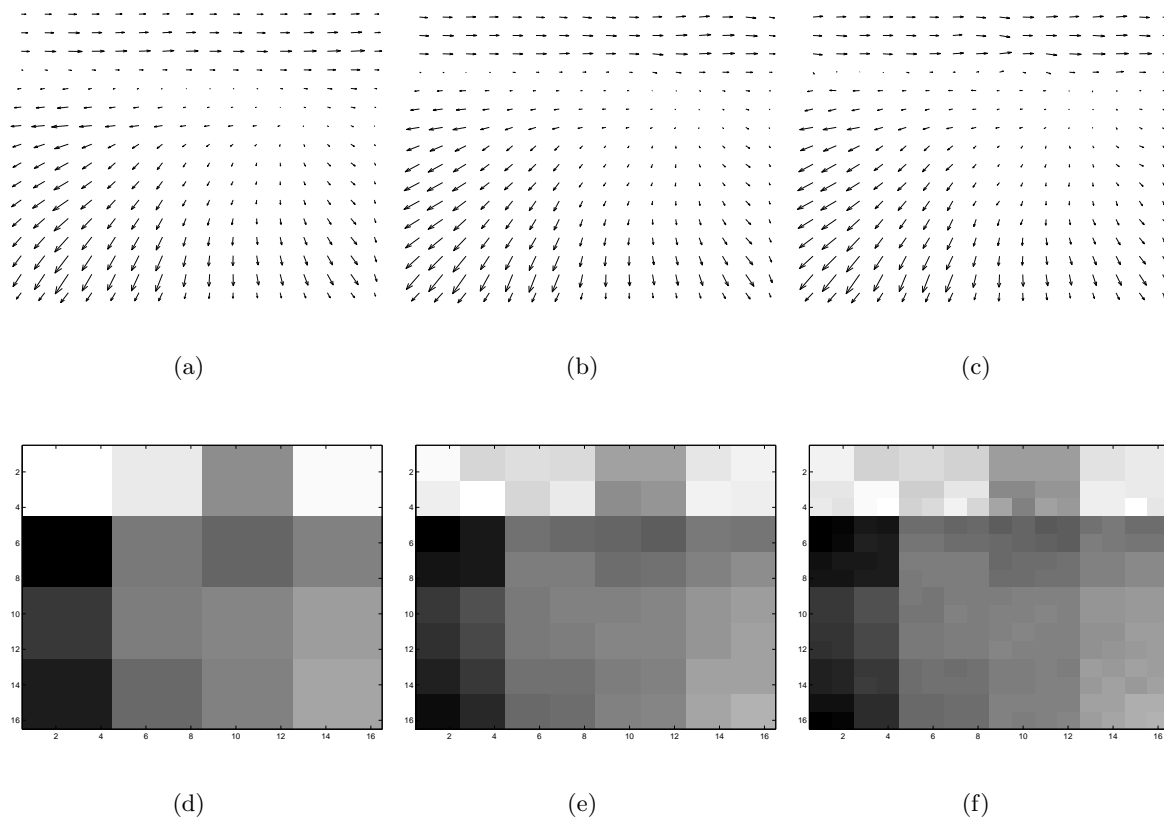


Fig. 6.9: Flots optiques estimés avec Φ^3 sur la séquence *Yosemite* aux niveaux a) 2, b) 3, c) 4 et d,e,f) coefficients θ_j associés (seuls les coefficients correspondant à la composante horizontale du mouvement sont représentés).

Arbre en translation.

Fonction de base	Temps de calcul
Φ^1	15,47s
Φ^2	30,05s
Φ^3	55,20s

Tab. 6.2: Temps de calcul en seconde. Séquence *Yosemite*, 3 niveaux de résolution, taille d'image 256×256 , PIII 700MHz, implantation MATLAB

Les performances obtenues au niveau de résolution 4 sur ces trois séquences sont comparables aux différentes approches d'estimation du mouvement global. En particulier, l'algorithme de Wu *et. al.* fournit des résultats équivalents aux nôtres alors qu'il est basé sur des B-splines de degré 4 (susceptibles de mieux modéliser le mouvement pour une complexité algorithmique supérieure).

Il apparaît que la modélisation du mouvement par des B-splines de degré élevé est mieux

Séquence <i>Arbre en Translation</i>			
B-spline	Moyenne	Ecart-type	Densité
Φ^1	0.54°	0.87°	100%
Φ^2	0.58°	0.74°	100%
Φ^3	0.63°	0.8°	100%
Srinivasan [73]	0.61°	0.26°	100%
Szeliski [76]	0.35°	0.34°	100%
Wu [90]	0.21°	0.26°	100%

Séquence <i>Arbre en Divergence</i>			
B-spline	Moyenne	Ecart-type	Densité
Φ^1	2.9°	1.7°	100%
Φ^2	3.5°	1.9°	100%
Φ^3	3.7°	1.8°	100%
Srinivasan [73]	2.94°	1.64°	100%
Szeliski [76]	0.98°	0.74°	100%
Wu [90]	0.54°	0.62°	100%

Tab. 6.3: Erreurs angulaires (moyenne et écart-type) obtenues sur les séquences *Arbre en translation* et *Arbre en divergence*, avec des B-splines de degrés 1, 2 et 3 au niveau de résolution 3 (estimation multirésolution du mouvement sur 2 niveaux, $j = 2$ et 3).

adaptée pour les mouvements complexes (*Yosemite*) que pour les mouvements simples (*Arbre en divergence* et *Arbre en translation*). Ce résultat est cohérent avec les remarques exposées dans la section 6.4.2. Néanmoins, les différences d'erreurs entre les trois B-splines sont faibles (variation de 1,2° sur *Yosemite* entre Φ^1 et Φ^3) alors que le temps de calcul augmente fortement lorsque le degré N augmente (tableau 6.2).

6.5.4 Résultats sur des séquences naturelles

Nous présentons les résultats obtenus sur des séquences d'images naturelles afin de tester la validité du modèle sur des mouvements complexes.

Les séquences réelles sont *Mobile and calendar* (figure 6.10), *Rubik Cube* (figure 6.11), *Taxi de Hamburg* (figure 6.12), présentées dans les chapitres 3 et 5, ainsi que les séquences *Papillon* (figure 6.13) et *Trial* (figure 6.14). La séquence *Papillon*, représentant un papillon ouvrant ses ailes, est de bonne qualité. Au contraire la séquence *Trial* est une séquence MPEG très compressée, et donc de mauvaise qualité. Le mouvement global de cette séquence est dû principalement à trois facteurs : le mouvement de la caméra vers la gauche, la personne qui se lève au premier plan, et le vélo qui tente de franchir l'obstacle dans le fond de la scène.

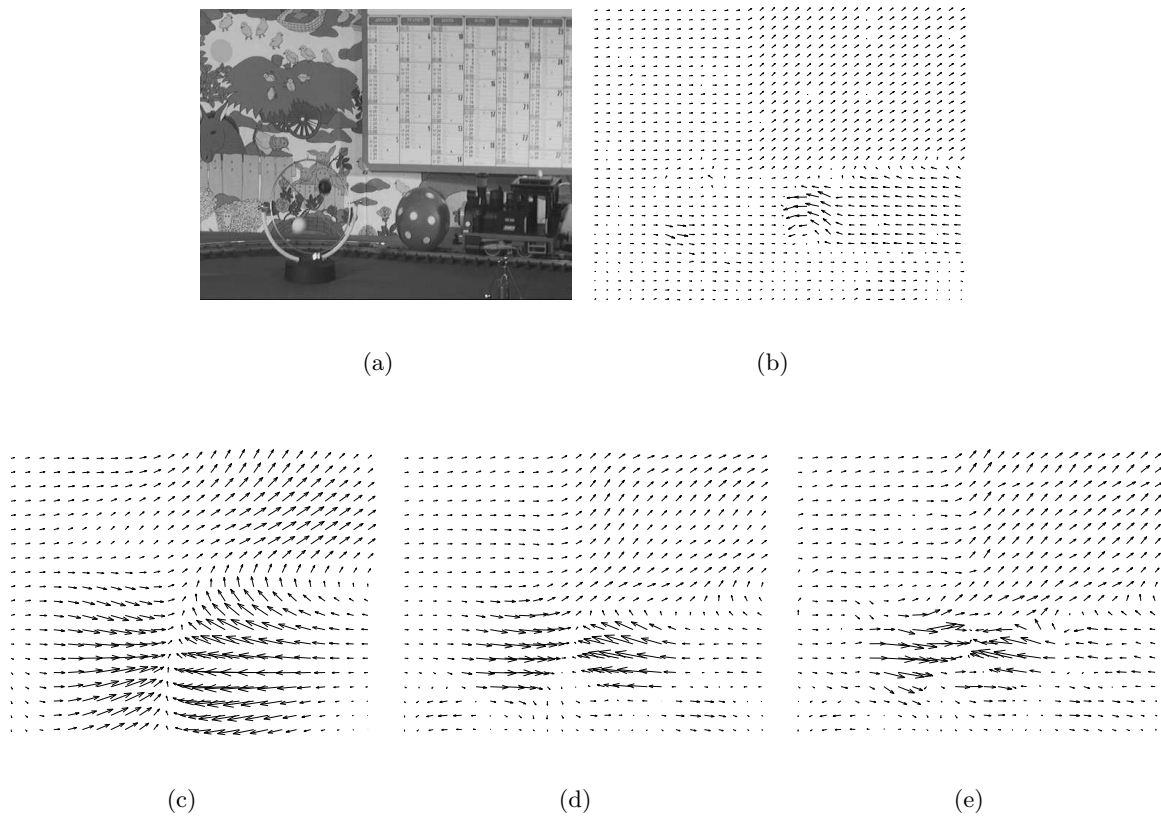


Fig. 6.10: Séquence *Mobile and calendar* avec a) image de la séquence, b) estimation locale du flot optique par filtres de Gabor spatiaux sur 3 niveaux de pyramide, estimation globale du flot optique avec Φ^3 au niveau c) 2, d) 3 et e) 4.

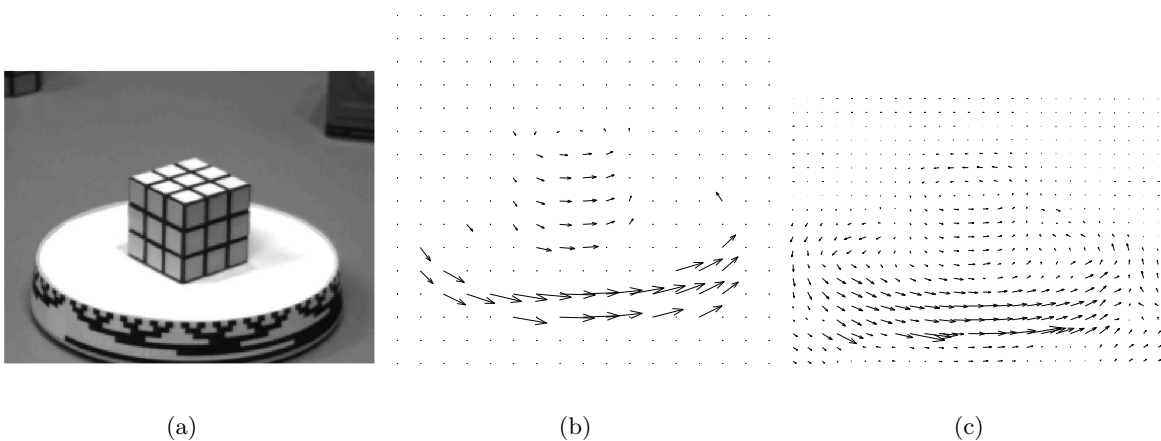


Fig. 6.11: Séquence *Rubik Cube* avec a) image de la séquence, b) estimation locale du flot optique par filtres de Gabor spatiaux sur 3 niveaux de pyramide et c) estimation globale du flot optique avec Φ^3 sur 3 niveaux de résolution

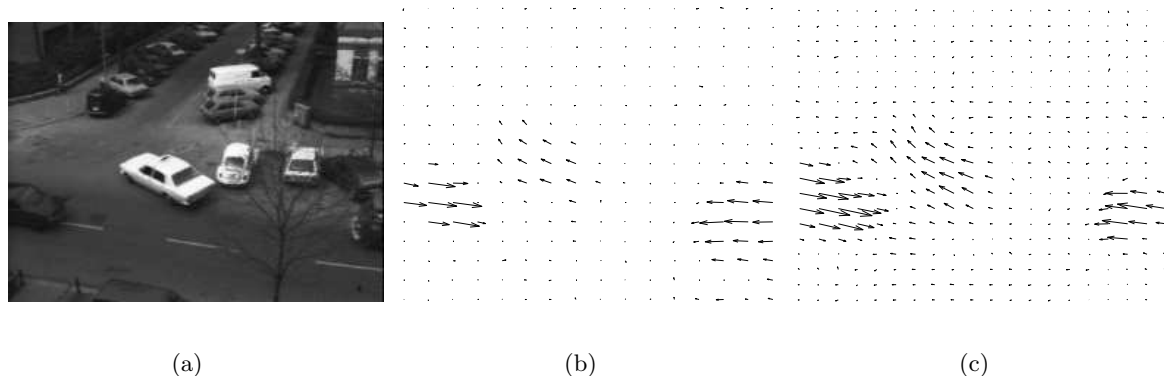


Fig. 6.12: Séquence *Taxi de Hambourg* avec a) image de la séquence, b) estimation locale du flot optique par filtres de Gabor spatiaux sur 3 niveaux de pyramide et c) estimation globale du flot optique avec Φ^3 sur 3 niveaux de résolution

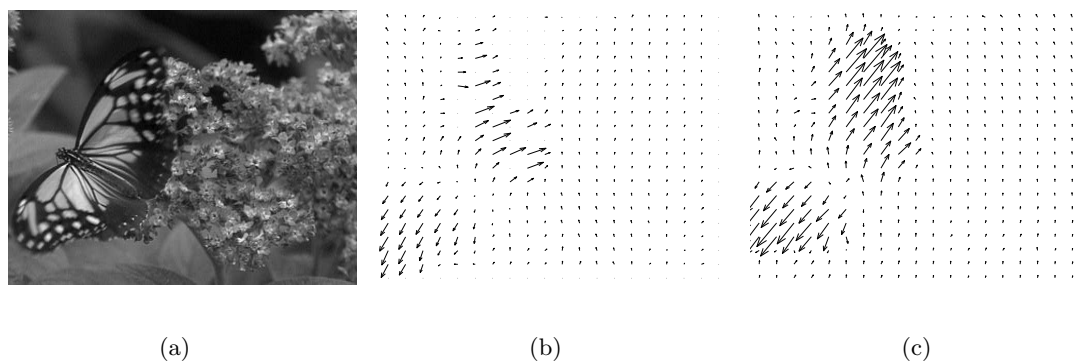


Fig. 6.13: a) Séquence *Papillon*, b) estimation locale du flot optique par filtres de Gabor spatiaux sur 3 niveaux de pyramide et c) estimation globale du flot optique avec Φ^3 sur 3 niveaux de résolution.

Les flots optiques estimés sur ces différentes séquences sont proches des champs denses estimés par l'algorithme basé sur les filtres de Gabor (chapitre 3). D'une manière générale, nous observons que la qualité de l'estimation est intrinsèquement liée au modèle de mouvement utilisé. Plus la résolution du modèle est importante, plus le flot optique estimé gagne en détails (variations spatiales de plus en plus fines). Malheureusement, du fait du problème d'ouverture, cette progression en résolution s'accompagne d'une augmentation de l'erreur sur l'estimation, et se traduit par un champ de mouvement irrégulier. L'exemple de la séquence *Mobile and calendar* est caractéristique. On peut observer à la fois un accroissement de la résolution du champ des vitesses, obtenu à chaque niveau supérieur du modèle de mouvement, et en même temps des erreurs de plus en plus importantes dans les zones relativement uniformes de la scène (le calendrier, le mur derrière le mobile, . . .) (figure 6.10.c, d et e).

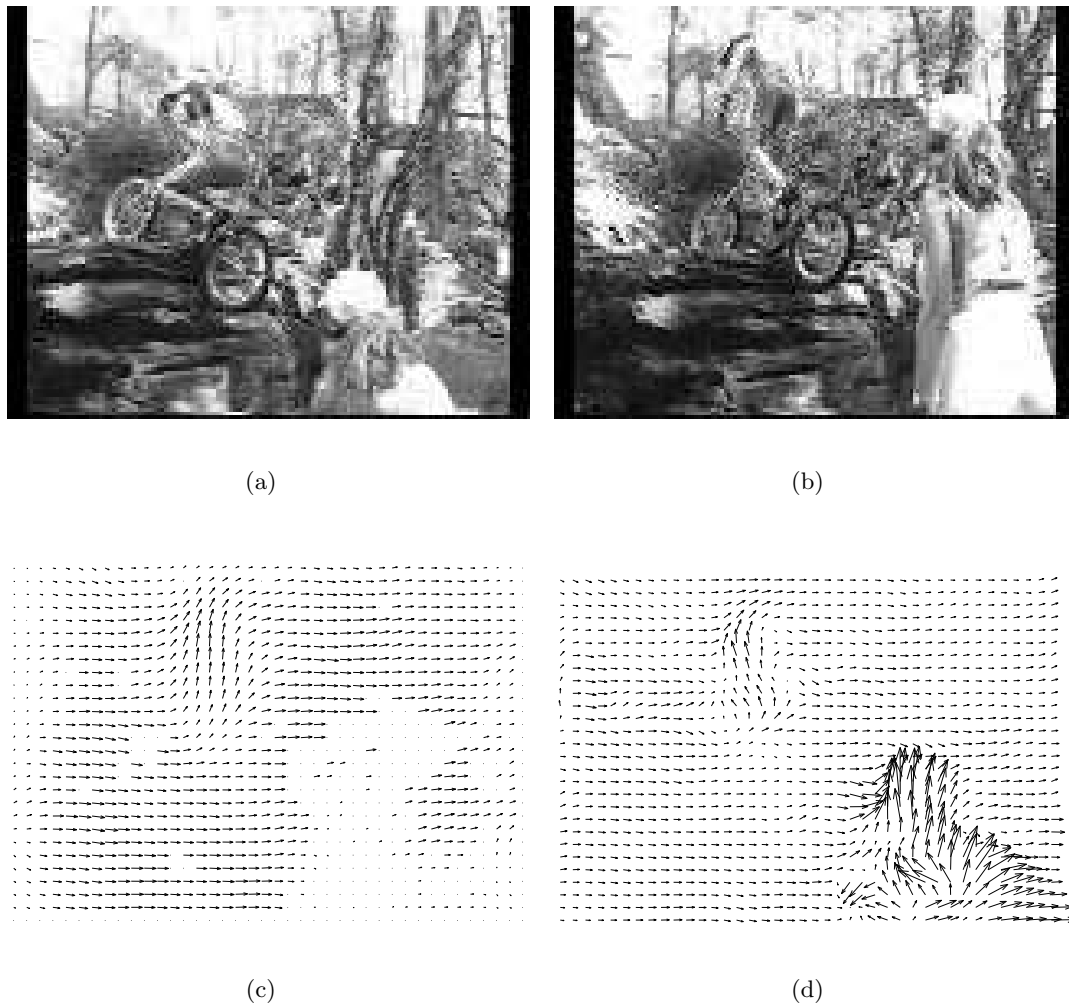


Fig. 6.14: a) Image 60 et b) image 80 de la séquence *Trial*. Flot optique estimé entre les images 60 et 61 par : c) les filtres de Gabor spatiaux sur 3 niveaux de pyramide et d) par le modèle de mouvement global à 3 niveaux de résolution, basé sur Φ^3 .

Les résultats obtenus sur la séquence *Trial* sont intéressants. Malgré la faible qualité de la séquence et la complexité du mouvement, l'estimation est cohérente avec le mouvement perçu, et met en évidence le mouvement de la caméra, de la personne au premier plan et du vélo dans le fond (figure 6.14.d). De plus, comparé au champ dense des vitesses estimé par les filtres de Gabor spatiaux (figure 6.14.c), l'estimation du modèle global de mouvement permet de mesurer le déplacement rapide du personnage au premier plan. Il en est de même pour la séquence *Papillon*, où le mouvement des ailes est estimé par le modèle global et non par les filtres de Gabor spatiaux (figures 6.13.b et c).

6.6 Conclusion

Nous avons développé un nouveau modèle de mouvement global basé sur la décomposition en ondelettes. Le choix de la base d'ondelettes a été guidé par un certain nombre d'impératifs sur la fonction d'échelle : le support compact doit être de taille minimal pour un ordre d'approximation maximum et la fonction doit être symétrique et régulière. Les B-splines satisfaisant le mieux à ces trois impératifs, ont été choisies comme fonctions de base. Des expérimentations nous ont montré qu'elles permettaient d'obtenir de meilleurs résultats que par l'utilisation d'ondelettes de Daubechies (ondelettes qui ont aussi un support minimal pour un ordre maximal).

Le modèle de mouvement global basé sur une décomposition en ondelettes B-splines permet d'approximer des mouvements complexes. Les résultats obtenus sur des séquences artificielles et naturelles nous montrent que cette approximation est précise et fiable. L'utilisation des paramètres du modèle pour l'indexation et la classification de vidéos nous semble tout à fait envisageable.

Chapitre 7

Classification et segmentation temporelle de vidéos par les paramètres des modèles de mouvement global

La variabilité du mouvement dans les séquences d'images est très importante du fait du nombre élevé de degrés de liberté qu'il comporte. Aussi, pour comparer, classer ou indexer des vidéos à partir de cette information, il est nécessaire d'en réduire le volume en ne sélectionnant que les descripteurs de mouvement pertinents. Le caractère pertinent est tout d'abord lié à la nature de la description que l'on se propose d'effectuer. Par exemple, si la finalité de l'application est de découper une séquence en fonction du déplacement de la caméra, seul les descripteurs liés au mouvement dominant seront pertinents, alors que toutes les mesures relatives au mouvement des objets sont inutiles et peuvent induire des erreurs dans la segmentation temporelle. Dans un deuxième temps, les descripteurs sont jugés pertinents si le résultat obtenu correspond à une réalité psycho-visuelle, c'est à dire similaire à un découpage opéré par un opérateur humain.

Les paramètres des modèles de Fourier et d'ondelettes sont une signature compacte du mouvement global de la scène. L'utilisation de ces paramètres comme descripteurs de mouvement permet d'analyser le contenu dynamique de vidéos. Plus le modèle est complexe, plus la description est précise, et à l'inverse, un modèle, défini par peu de paramètres, décrit uniquement les structures principales du flot optique. Dans le cadre de l'indexation de vidéos, il est nécessaire de décrire le contenu par ses caractéristiques principales de manière à ce qu'elles soient communes à plusieurs séquences. Cela permet ainsi de créer des classes d'activités dans lesquelles un utilisateur peut trouver plusieurs séquences correspondant à sa requête (par exemple). Si la description est trop précise, chaque séquence définit sa propre classe.

Ce chapitre présente plusieurs expériences de description du mouvement par les coefficients de Fourier et d'ondelettes. La première est un exemple de classification hiérarchique d'une base

de vidéos contenant différentes activités humaines bien distinctes. Cette expérience permet de comparer la pertinence des descripteurs de mouvement selon qu'ils soient des coefficients de Fourier ou d'ondelettes. Puis, nous nous intéressons à la segmentation temporelle de séquences par les paramètres de mouvement. Les propriétés multiéchelle des ondelettes permettant de définir des descripteurs sur le mouvement dominant et sur le mouvement des objets, il est alors possible de segmenter une vidéo en fonction du déplacement de la caméra ou du mouvement des objets.

7.1 Transformation des coefficients de fonctions d'échelle en coefficients d'ondelettes

Dans le chapitre 6, nous avons vu que le développement en séries d'ondelettes (6.5) à une résolution donnée est équivalent au développement en séries de fonctions d'échelle (6.3) au niveau de résolution supérieur. Pour simplifier l'algorithme, nous avons modélisé le mouvement global par des fonctions d'échelle, les coefficients estimés apportant ainsi une information très locale sur le mouvement (correspondant au support de la fonction d'échelle). La transformation de ces coefficients d'échelle en coefficients d'ondelettes permet de décrire le mouvement depuis l'échelle la plus large (toute l'image) à l'échelle la plus fine (fixée par la résolution initiale du modèle) (figure 7.1). Ainsi, les caractéristiques globales du mouvement sont décrites par quelques coefficients d'ondelettes correspondant aux échelles larges et sont donc plus facilement identifiables.

Les coefficients d'ondelettes $d_{i,m,n}$ s'obtiennent à partir des coefficients d'échelle $c_j(k_1, k_2)$ par le produit scalaire

$$d_{i,m,n} = \left\langle c_j(k_1, k_2), 2^{-i/2} \tilde{\Psi}(2^{-i}k_1 - m, 2^{-i}k_2 - n) \right\rangle \quad i = 0, \dots, j-1 \quad (7.1)$$

où $\tilde{\Psi}$ est l'ondelette duale satisfaisant la condition de biorthogonalité

$$\left\langle \Psi_{j,m,n}, \tilde{\Psi}_{l,p,q} \right\rangle = \delta_{j,l} \cdot \delta_{m,p} \cdot \delta_{n,q} \quad (7.2)$$

L'ondelette duale s'obtient par un jeu de filtrages et de sous-échantillonnages de la fonction B-spline (6.23) [82]. Le passage des coefficients d'échelle aux coefficients d'ondelettes s'effectue par un algorithme de transformée en ondelettes rapides. La séquence de coefficients obtenue est alors de la forme $\{c_0, \mathbf{d}_0^H, \mathbf{d}_0^V, \mathbf{d}_0^D, \dots, \mathbf{d}_{i,k_1,k_2}^H, \mathbf{d}_{i,k_1,k_2}^V, \mathbf{d}_{i,k_1,k_2}^D, \dots\}$, $i = 1 \dots j-1$ et $(k_1, k_2) \in [0, 2^i - 1]^2$. Les vecteurs c_0 , \mathbf{d}_{i,k_1,k_2}^H , \mathbf{d}_{i,k_1,k_2}^V et \mathbf{d}_{i,k_1,k_2}^D représentent respectivement la moyenne et les variations horizontales, verticales et diagonales à différentes échelles du champ des vitesses. Cette description est similaire à celle obtenue par les séries de Fourier où les termes du développement correspondent à la moyenne et aux variations selon différentes directions et fréquences du mouvement. Dorénavant, chaque fois que nous parlerons des paramètres du modèle d'ondelettes, cela correspondra aux coefficients d'ondelettes et non aux coefficients d'échelle.

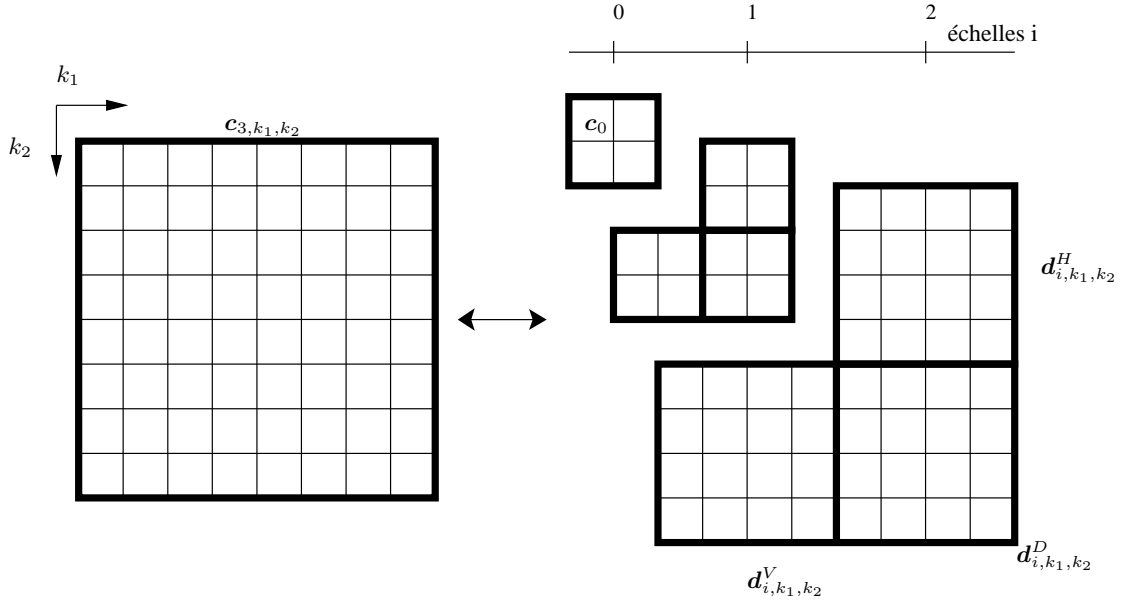


Fig. 7.1: Transformation des coefficients d'échelle $\{c_{j,k_1,k_2}\}$ en coefficients d'ondelettes $\{c_0, d_{i,k_1,k_2}^H, d_{i,k_1,k_2}^V, d_{i,k_1,k_2}^D\}$.

7.2 Pertinence des paramètres de mouvement pour la classification de vidéos

Les coefficients de Fourier et d'ondelettes portent une information sur le mouvement global entre images successives d'une séquence. Pour évaluer la pertinence de cette information, nous classons une base de vidéos dont le contenu dynamique permet de déterminer sans ambiguïté une classification a priori. La comparaison entre les résultats obtenus par les coefficients de Fourier, d'ondelettes et la classification a priori donne ainsi un élément de réponse sur la pertinence des paramètres du mouvement considérés.

7.2.1 Description de la base de vidéos

La base contient 30 vidéos représentant six types d'activités humaines [16] (figure 7.2). Ces séquences ont été acquises par une caméra fixe placée à l'intersection de trois couloirs et en face d'un escalier. Les personnes passant à cette endroit ont six possibilités : monter et descendre par l'escalier, traverser la scène par la droite ou par la gauche et enfin s'approcher ou s'éloigner de la caméra. Chaque séquence est assignée *a priori* à une classe en fonction de l'activité qu'elle contient. Ces classes sont *up*, *down*, *left*, *right*, *come* et *go* (voir la figure 7.2). Les séquences contiennent toutes dix images et la fréquence d'acquisition est de 10 Hz.

7.2.2 Définition des descripteurs de mouvement

Les coefficients estimés à partir de chaque paire d'images $(t, t + 1)$ d'une séquence S forment un vecteur de paramètres de mouvement $\theta(t) = [c_1 \dots c_N](t)$.

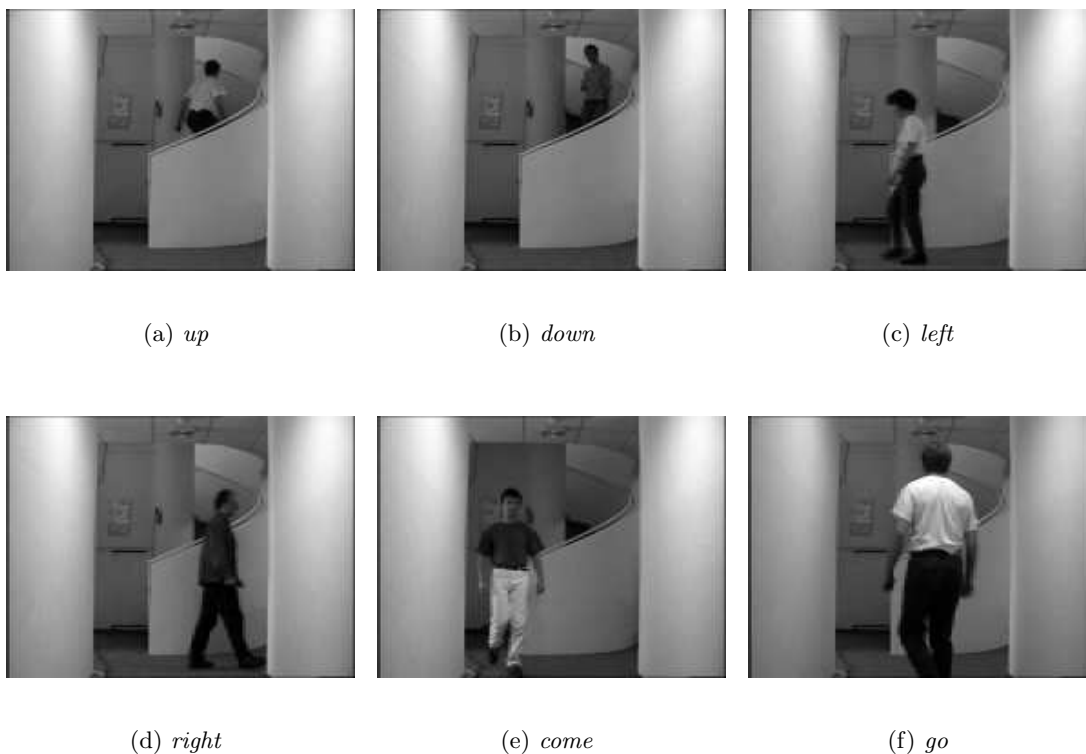


Fig. 7.2: Base de trente vidéos représentant six activités exécutées par cinq personnes différentes

Nous définissons un descripteur de mouvement associé à une séquence S contenant M images par la relation suivante :

$$\Theta_S = \frac{\sigma^{-1}}{M} \sum_{t=1}^M \theta(t) \quad (7.3)$$

avec σ la matrice diagonale contenant les écarts-types des composantes de $\theta(t)$ suivant t :

$$\sigma_{i,i} = \sqrt{\frac{1}{M} \sum_{t=1}^M (\theta_i(t) - \bar{\theta}_i(t))^2} \quad (7.4)$$

La normalisation de $\theta(t)$ par σ minimise l'influence des composantes variant fortement au cours du temps et donc peu fiable pour caractériser la séquence. L'ensemble des descripteurs Θ_S pour K vidéos décrivent un espace de descripteurs du mouvement :

$$\Omega = \{\Theta_{S_i}\}, \quad i = 1 \dots K \quad (7.5)$$

La dimension de Ω est égale au nombre de fonctions de base utilisées dans le modèle de mouvement (typiquement 32, 128, 512 pour les ondelettes et 50, 98, 162 pour Fourier). Il est nécessaire d'utiliser un algorithme de classification afin d'obtenir une représentation compréhensible de cet espace.

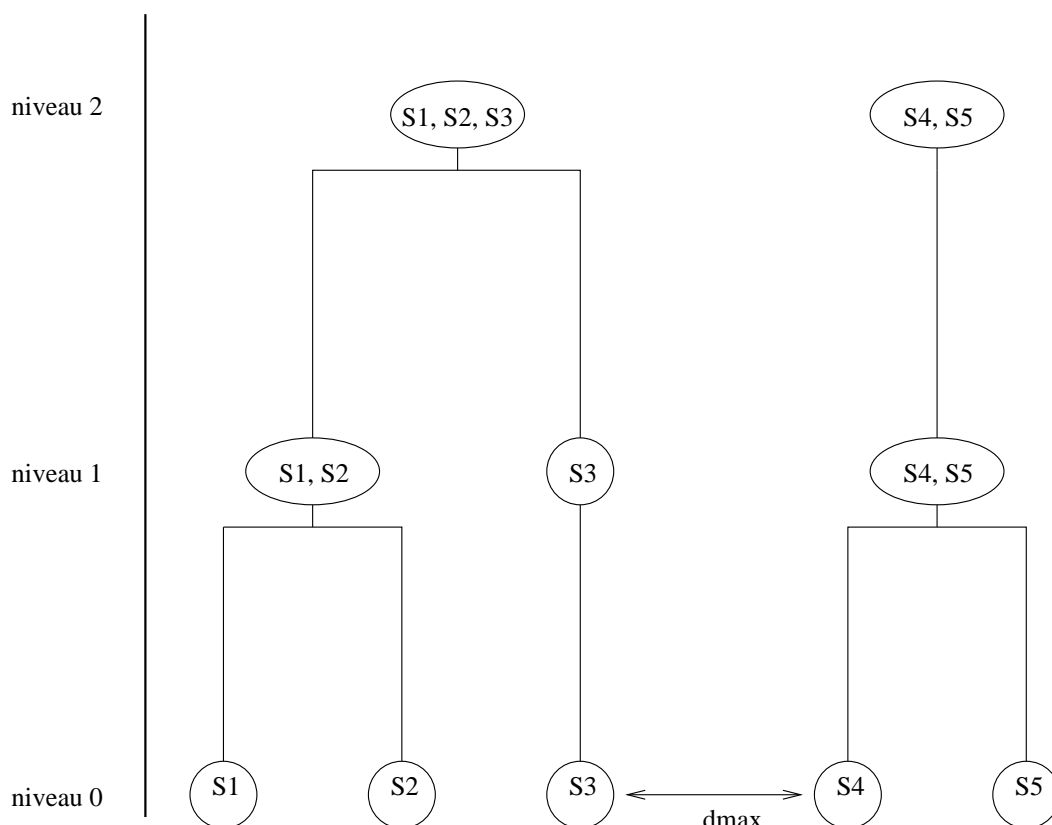


Fig. 7.3: Classification Hiérarchique Ascendante. Exemple sur un espace à 1 dimension contenant 5 éléments. Les éléments les plus proches au sens de la distance euclidienne sont successivement fusionnés jusqu'à ce que la distance entre les noeuds de l'arborescence soit supérieure à une distance d_{max} prédéfinie.

7.2.3 Classification hiérarchique ascendante de Ω

La *classification hiérarchique ascendante (CHA)* [19] est une méthode utilisée avec succès pour l'indexation de bases d'images et de vidéos [21, 51]. Elle consiste à organiser la base selon une arborescence exprimant les similarités relatives à une distance $d(\cdot, \cdot)$ entre les vidéos dans l'espace des descripteurs Ω contenant K éléments (figure 7.3). La distance choisie est la distance euclidienne dans l'espace Ω .

La procédure est initialisée en définissant chaque vecteur de descripteurs Θ_{S_i} associé à une séquence S_i comme un noeud $\mathcal{N}_{0,i}$ du niveau 0 de l'arborescence. Le niveau supérieur est construit en fusionnant les deux noeuds $\mathcal{N}_{0,n}$ et $\mathcal{N}_{0,m}$ les plus proches au sens de la distance $d(\cdot, \cdot)$. Après cette première fusion, on fusionne la paire $(\mathcal{N}_{0,k}, \mathcal{N}_{0,l})$ dont la distance est minimale pour l'ensemble des noeuds privés de $(\mathcal{N}_{0,n}, \mathcal{N}_{0,m})$. Cette procédure est itérée jusqu'à ce que la distance minimale sur l'ensemble restant soit supérieure ou égale à une distance d_{max} fixée. Ce seuil assure une homogénéité du mouvement entre les séquences associées à un même noeud. Les noeuds isolés sont alors projetés au niveau supérieur de

l'arborescence.

$$\mathcal{N}_{1,i} = \begin{cases} \arg \min_{(\mathcal{N}_{0,n}, \mathcal{N}_{0,m}) \in \Gamma_i^2} d(\mathcal{N}_{0,n}, \mathcal{N}_{0,m}) & \text{si } d(\mathcal{N}_{0,n}, \mathcal{N}_{0,m}) < d_{max} \\ \mathcal{N}_{0,i} & \text{sinon} \end{cases} \quad (7.6)$$

avec $\Gamma_i = \{\mathcal{N}_{0,j}, j = 1, \dots, K \setminus \mathcal{N}_{0,i} \in \mathcal{N}_{1,k}, k < i\}$ l'ensemble des noeuds de niveau 0 non encore fusionnés. Pour les niveaux supérieurs à 0, il s'agit de fusionner des noeuds contenant plusieurs éléments. La distance $\tilde{d}(\mathcal{N}_{j,n}, \mathcal{N}_{j,m})$ au niveau j est définie comme la distance euclidienne entre le centre d'inertie $\mathcal{G}_{j,n}$ des éléments de $\mathcal{N}_{j,n}$ et le centre d'inertie $\mathcal{G}_{j,m}$ des éléments de $\mathcal{N}_{j,m}$:

$$\tilde{d}(\mathcal{N}_{j,n}, \mathcal{N}_{j,m}) = d(\mathcal{G}_{j,n}, \mathcal{G}_{j,m}) \quad (7.7)$$

Les niveaux supérieurs de l'arborescence sont obtenus par la procédure décrite en (7.6) en remplaçant la distance $d(\cdot, \cdot)$ par $\tilde{d}(\cdot, \cdot)$. La procédure s'arrête lorsque la distance minimum entre les noeuds du dernier niveau est supérieure ou égale au seuil d_{max} .

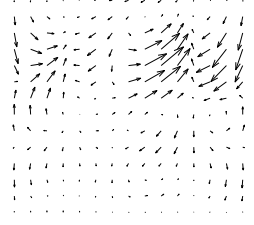
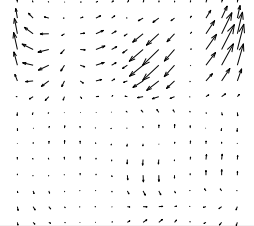
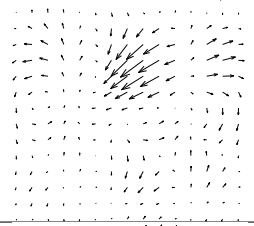
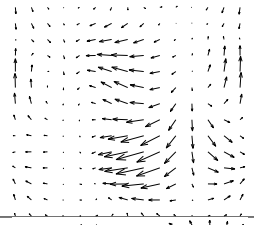
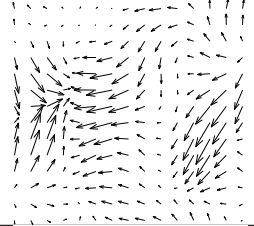
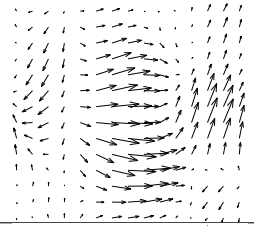
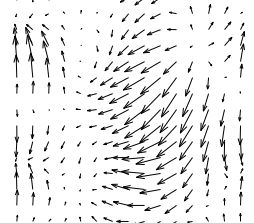
Le choix de d_{max} a une influence cruciale sur la qualité de la classification obtenue. Il détermine la distance minimale entre classes qui nous est a priori inconnue. La valeur optimale de d_{max} dépend à la fois de la nature des séquences à classer et des descripteurs de mouvement utilisés. Afin de rendre d_{max} peu sensible à ces variations, nous normalisons les distances calculées entre les différents noeuds à chaque niveau de l'arborescence. Par cette normalisation, à un niveau donné, la distance entre deux noeuds est relative à la distance maximale du niveau. La valeur de d_{max} est fixée à 0.2 pour tous les résultats présentés ci-dessous.

7.2.4 Résultats

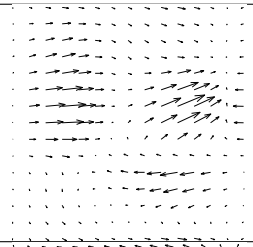
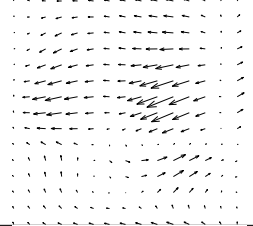
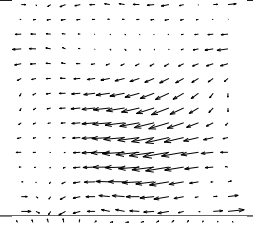
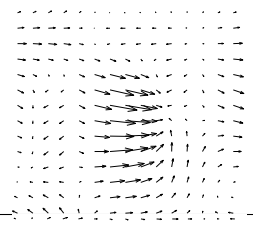
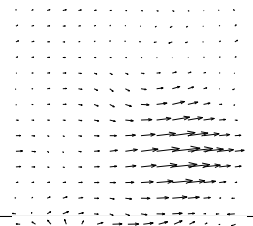
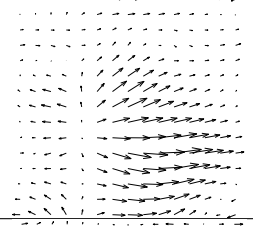
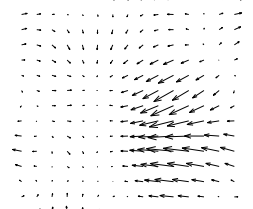
L'algorithme de classification est appliqué aux espaces de descripteurs de Fourier et d'ondelettes estimés sur la base de vidéos. Les modèles de Fourier et d'ondelettes utilisés contiennent respectivement 50 et 32 fonctions de bases, fournissant une description grossière du mouvement. Pour le modèle d'ondelettes, des fonctions B-splines de degré 1 sont utilisées.

Les tableaux 7.1 et 7.2 représentent la classification de la base de vidéos obtenue au dernier niveau de l'arborescence de l'algorithme CHA. La première colonne indique le numéro i du noeud $\mathcal{N}_{h,i}$ situé sur le niveau le plus haut de l'arborescence. La deuxième colonne indique le nom des séquences contenues dans le noeud $\mathcal{N}_{h,i}$ et la troisième colonne représente sous forme de flot optique le centre d'inertie $\mathcal{G}_{h,i}$ associé à $\mathcal{N}_{h,i}$.

Les descripteurs de Fourier et ondelettes permettent une catégorisation des vidéos relative aux mouvements qu'elles contiennent. Les résultats montrent en effet que la majorité des classes obtenues contiennent des vidéos représentant les mêmes activités. Dans le cas où une classe contient des séquences d'activités différentes (classes \mathcal{N}_5 et \mathcal{N}_6 du tableau 7.1 et \mathcal{N}_3 et \mathcal{N}_5 du tableau 7.2), on peut remarquer que le mouvement est relativement similaire entre ces séquences. Ainsi, les activités *go* et *left*, caractérisées par un mouvement dominant vers la gauche, et les activités *come* et *right*, caractérisées par un mouvement dominant vers la droite,

i	$\mathcal{N}_{h,i}$	$\mathcal{G}_{h,i}$
1	up1, up2, up3, up4, up5	
2	down2, down5	
3	down3, down4	
4	left1, left2, left3	
5	left4, left5, go1, down1	
6	come5, right2, come3, right1, come2, come1, come4, right4, right3	
7	go2, go3, go4, go5	

Tab. 7.1: Classification de la base de vidéos obtenue par les descripteurs de Fourier.

i	$\mathcal{N}_{h,i}$	$\mathcal{G}_{h,i}$
1	up1, up2, up3, up4, up5	
2	down1, down2, down3, down4, down5	
3	left1, left2, left3, left4, left5	
4	right1, right2	
5	right3, right4, right5, come4	
6	come1, come2, come3, come5	
7	go1, go2, go3, go4, go5	

Tab. 7.2: Classification de la base de vidéos obtenue par les coefficients d'ondelettes B-splines de degré 1.

sont très similaires au sens du mouvement, ce qui entraîne des erreurs dans la catégorisation de certaines séquences.

Comparés au descripteurs de Fourier, les descripteurs d'ondelettes permettent une meilleure classification de la base de vidéos (tableau 7.2). Les séquences de type *up*, *down*, *left* et *go* sont parfaitement classées et seules les séquences de type *right* posent réellement des problèmes en créant deux classes pour cette activité. Ce résultat confirme la supériorité des ondelettes sur les séries de Fourier pour résumer par quelques coefficients l'information essentielle contenue dans le mouvement global.

Les classifications obtenues avec les coefficients d'ondelettes B-splines de degré 2 et 3 (tableaux 7.3 et 7.4) sont moins pertinentes que celle obtenues avec les B-splines de degré 1, et cela pour un temps de calcul évidemment plus important. De même, la description du mouvement par des modèles de résolutions plus fines ($j = 3$ et $j = 4$) ne permet pas d'améliorer la classification : cette description trop précise et souvent bruitée conduit à la création d'un grand nombre de classes ne contenant qu'une ou deux séquences.

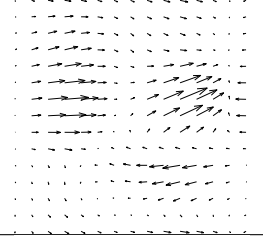
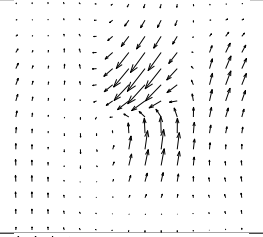
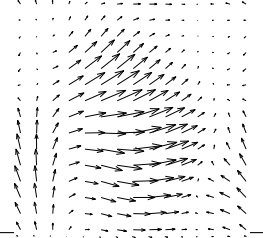
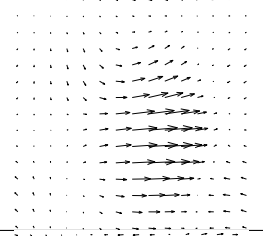
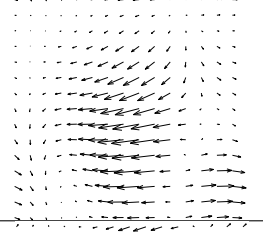
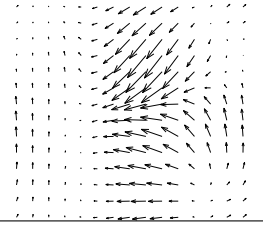
En conclusion de cette expérience, les descripteurs de mouvement basés sur les coefficients d'ondelettes B-splines de degré 1 au niveau de résolution 2 semblent être le meilleur choix pour caractériser le contenu de vidéos. Ils présentent l'avantage à la fois d'être pertinents et de minimiser le temps de calcul nécessaire à leur estimation. Aussi, les expériences décrites dans la suite de ce chapitre seront toutes basées sur ces descripteurs.

7.3 Segmentation temporelle de séquences vidéo

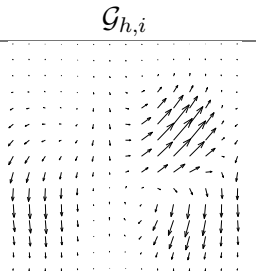
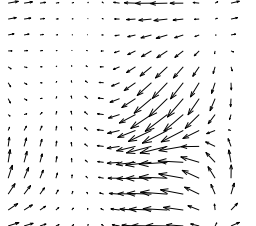
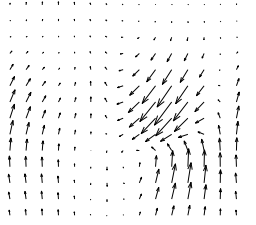
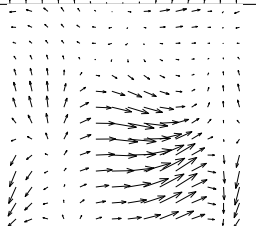
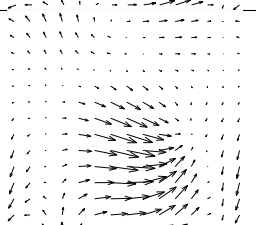
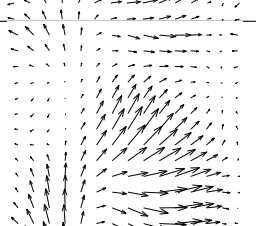
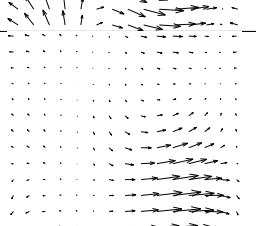
La segmentation temporelle d'une vidéo consiste à découper la séquence selon certains critères tels que la présence de changement de plan ou le changement de type de mouvement de caméra [11]. La structuration de la vidéo facilite par la suite son indexation par le contenu.

Les coefficients d'ondelettes $\theta(t)$ estimés pour chaque paire d'images d'une séquence S forment un espace de descripteurs du mouvement de S . Appelons Ω_S cette espace. Une classification de Ω_S , basée sur une version modifiée de la procédure CHA (pour conserver les relations temporelles existant entre les descripteurs), permet de partitionner la séquence S selon un critère de similarité de mouvement.

L'application directe de cette technique conduit à une segmentation dépendant à la fois du mouvement dominant (souvent induit par le déplacement de la caméra) et du mouvement des objets présents dans la scène. La description multiéchelle des ondelettes nous permet de séparer de façon approximative le mouvement dominant du mouvement des objets. En utilisant cette propriété, nous pouvons construire pour une séquence deux segmentations temporelles : l'une basée sur le déplacement de la caméra et l'autre sur le mouvement des objets.

i	$\mathcal{N}_{h,i}$	$\mathcal{G}_{h,i}$
1	up1, up2, up3, up4, up5	
2	down1, down2, down3, down4, down5	
3	come1, come2, come3, come5, right1, right2	
4	right3, right4, right5, come4	
5	left1, left2, left3, left4, left5, go3	
6	go1, go2, go4, go5	

Tab. 7.3: Classification de la base de vidéos obtenue par les coefficients d'ondelettes B-splines de degré 2.

i	$\mathcal{N}_{h,i}$	$\mathcal{G}_{h,i}$
1	up1, up2, up3, up4, up5	
2	down1, go2, go1, go4, go5, left5, go3, left4, left3, left1, left2	
3	down2, down3, down4, down5	
4	right1	
5	right2	
6	come1, come2, come3, come5	
7	come4, right5, right4, right3	

Tab. 7.4: Classification de la base de vidéos obtenue par les coefficients d'ondelettes B-splines de degré 3.

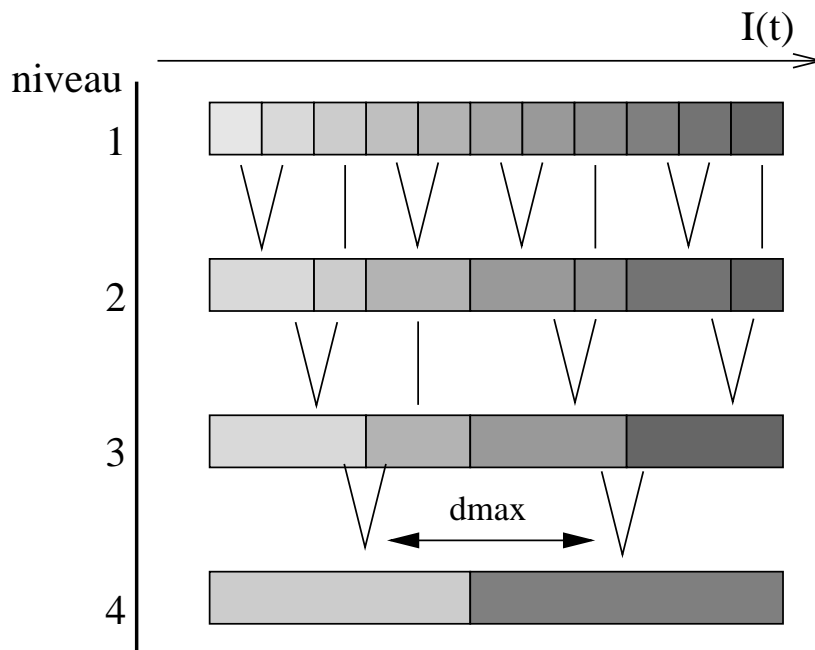


Fig. 7.4: Segmentation temporelle d'une séquence par classification hiérarchique. Chaque image I de la séquence S est caractérisée par un vecteur de coefficients d'ondelettes. Ces vecteurs sont classés hiérarchiquement avec comme contrainte de ne pouvoir être associés qu'à leurs plus proches voisins temporels.

7.3.1 Segmentation hiérarchique de la séquence

La segmentation temporelle est effectuée à l'aide de l'algorithme CHA. Celui-ci est modifié de manière à ne fusionner que les noeuds connectés temporellement. Cette modification consiste à remplacer dans le schéma de fusion (7.6) à un niveau j quelconque, la distance \tilde{d} par une nouvelle mesure \tilde{d}_t [22] :

$$\tilde{d}_t(\mathcal{N}_{j,k}, \mathcal{N}_{j,l}) = \begin{cases} \tilde{d}(\mathcal{N}_{j,k}, \mathcal{N}_{j,l}) & \text{si } \mathcal{N}_{j,k} \text{ et } \mathcal{N}_{j,l} \text{ sont connectés temporellement} \\ d_{max} & \text{sinon} \end{cases} \quad (7.8)$$

Les distances sont normalisées et $d_{max} = 0.2$ comme précédemment. L'algorithme CHA modifié s'effectue sur l'espace Ω_S contenant l'ensemble des descripteurs de mouvement estimés sur chaque paire d'images de la séquence S . La classification obtenue fournit alors un découpage temporelle de la séquence (figure 7.4).

7.3.2 Séparation du mouvement dominant du mouvements des objets

La description multiéchelle du mouvement (cf. transformation 7.1) permet de séparer l'information globale, relative au mouvement dominant de l'information locale, relative aux déplacements des objets dans la scène. En séparant simplement les coefficients d'ondelettes au niveau de résolution 0 des autres, il est possible de définir des descripteurs associés au

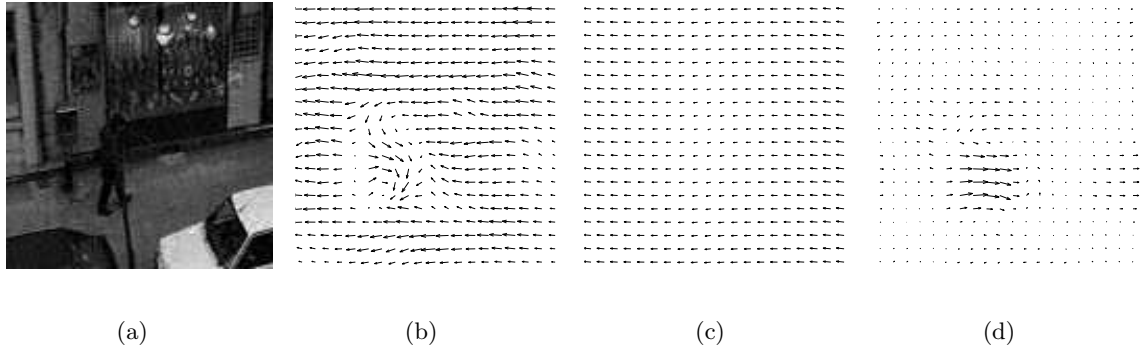


Fig. 7.5: Exemple de séparation du mouvement de la caméra du mouvement des objets avec :
a) image de la séquence, b) mouvement global estimé avec une base de B-splines de degré 1 sur 2 niveaux de résolution (2 et 3), c) flot optique reconstruit à partir de θ_{cam} et d) flot optique reconstruit à partir de θ_{obj} .

mouvement dominant et au mouvement des objets. Nous définissons le descripteur du mouvement dominant induit par les déplacements de la caméra comme l'ensemble des coefficients d'ondelettes de niveau zéro :

$$\theta_{cam} = [\mathbf{c}_0, \mathbf{d}_0^H, \mathbf{d}_0^V, \mathbf{d}_0^D] \quad (7.9)$$

L'utilisation de ces quatre coefficients permet d'exprimer exactement toutes fonctions linéaires de L^2 (cf. chapitre 6), et sont donc bien adaptés pour décrire les mouvements affines.


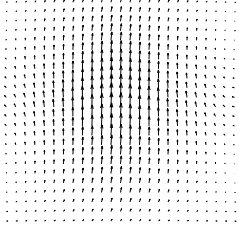


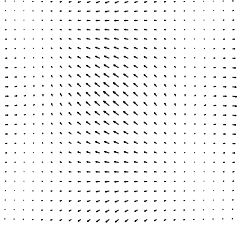


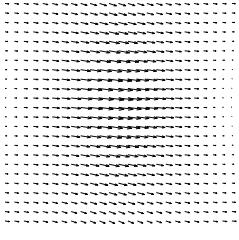

Le descripteur associé au mouvement des objets est défini comme l'ensemble des coefficients d'ondelettes de résolution supérieure à zéro :

$$\theta_{obj} = [\mathbf{d}_{i,k_1,k_2}^H, \mathbf{d}_{i,k_1,k_2}^V, \mathbf{d}_{i,k_1,k_2}^D], \quad i > 0 \text{ et } (k_1, k_2) \in [0, 2^i - 1]^2 \quad (7.10)$$

Le descripteur θ_{cam} n'est pas totalement invariant aux mouvements des objets (d'autant moins que les objets occupent une surface importante de la scène). De même, lorsque le mouvement dominant est complexe (quadratique ou plus), des termes liés à ce mouvement se trouvent dans θ_{obj} . Néanmoins, dans la majorité des cas, les informations issues de θ_{cam} et θ_{obj} correspondent relativement bien aux déplacements de la caméra et aux mouvements des objets. La figure 7.5 illustre le type d'informations sur le mouvement fournit par ces deux descripteurs. La séquence représente un piéton marchant vers la droite de la scène. La caméra pivote sur la droite de la scène induisant un mouvement dominant vers la gauche (figure 7.5.b). Les flots optiques reconstruits à partir de θ_{cam} et θ_{obj} mettent en évidence le mouvement dominant vers la gauche (figure 7.5.c) et le mouvement du piéton (figure 7.5.d).

7.3.3 Résultats de segmentation temporelle à partir de θ_{cam} et θ_{obj}

Dans cette partie, nous présentons des résultats de segmentation temporelle obtenus sur deux séquences réelles. Ces deux vidéos représentent des activités sportives (football et trial)

i	Première image	\mathcal{G}_i	Dernière image
1			
2			
3			

Tab. 7.5: Segmentation de la séquence *foot* selon θ_{cam}


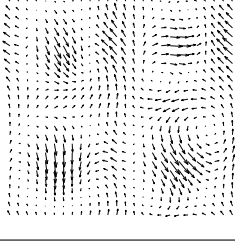


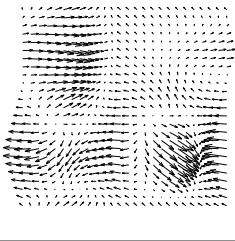


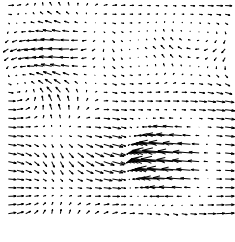

et sont caractérisées par des mouvements de caméra importants ainsi qu'une grande mobilité des personnages présents dans la scène. Chacune d'elle est constituée d'un seul plan (pas de coupure entre deux scènes).

La première séquence, que nous appellerons *foot*, est un extrait d'une vidéo d'un match de football. Elle est composée de 139 images. La caméra suit le ballon tout au long de la séquence. L'action commence au centre droit du terrain, où trois joueurs (et le ballon) courent vers le fond droit du terrain. Après 20 images le joueur centre et le ballon atterit sur la gauche du terrain (image 85 environ). Un autre joueur, venant du milieu de terrain, court vers le ballon et marque le but.

La deuxième séquence, appelée *trial*, contient 98 images et représente une épreuve de trial. Un cycliste part de la gauche de la scène et se dirige vers un obstacle qu'il escalade (de l'image 45 à l'image 80 environ). Arrivé au sommet, le cycliste exécute un rebond pour retrouver l'équilibre. La caméra suit le cycliste durant toute la séquence.

Les tableaux 7.5, 7.6, 7.7 et 7.8 présentent les résultats obtenus au dernier niveau de la classification hiérarchique. La première colonne indique le numéro i du noeud \mathcal{N}_i . La deuxième et la quatrième colonne montrent respectivement la première et la dernière image de \mathcal{N}_i . Dans la troisième colonne se trouve, sous forme de flot optique, le centre d'inertie \mathcal{G}_i associé à \mathcal{N}_i .

La segmentation de la séquence *foot* à partir des descripteurs du mouvement dominant

i	Première image	\mathcal{G}_i	Dernière image
1	$n^\circ 1$ 		$n^\circ 48$ 
	$n^\circ 49$ 		$n^\circ 86$ 
3	$n^\circ 87$ 		$n^\circ 139$ 


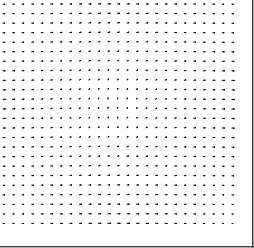


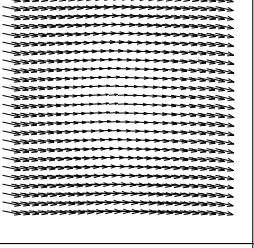


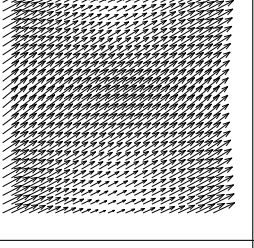


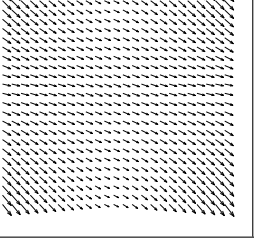

Tab. 7.6: Segmentation de la séquence *foot* selon θ_{obj}

θ_{cam} met en évidence les trois phases de jeu que nous avons identifiées précédemment (tableau 7.5). Le premier groupe correspond à la course des trois joueurs le long de la ligne, le deuxième au centre du ballon et le troisième au tir dans le but. Les centres d'inertie associés à chaque groupe donnent une indication sur le mouvement de la caméra.

La segmentation obtenue à partir des descripteurs de mouvement des objets θ_{obj} (tableau 7.6) est moins facile à analyser, les joueurs ayant des mouvements indépendants les uns des autres. Néanmoins, on peut reconnaître dans le flot optique correspondant au centre d'inertie \mathcal{G}_2 , le mouvement de l'attaquant qui va marquer le but (en haut à gauche du flot optique). De même, le mouvement du joueur se situant à droite du but dans la dernière image de la séquence apparaît nettement dans le flot optique du troisième groupe.

La segmentation de séquence *trial* selon θ_{cam} (tableau 7.7) permet de découper la séquence selon les déplacements de la caméra qui suit le vélo. Le centre d'inertie de chaque groupe donne une indication sur ces déplacements : successivement immobilité, translation vers la droite, déplacement vers le haut, puis vers le bas.

Avec les descripteurs θ_{obj} , la segmentation obtenue (tableau 7.8) correspond bien aux différentes phases du mouvement du vélo. Il est quasiment immobile de l'image 1 à 11. Puis de l'image 12 à 31, il s'avance vers l'obstacle. De l'image 32 à 76, le vélo gravit le monticule. Les deux derniers groupes correspondent respectivement à la phase ascendante et descendante

i	Première image	\mathcal{G}_i	Dernière image
1	 $n^{\circ} 1$		 $n^{\circ} 11$
2	 $n^{\circ} 12$		 $n^{\circ} 51$
3	 $n^{\circ} 52$		 $n^{\circ} 95$
4	 $n^{\circ} 96$		 $n^{\circ} 98$


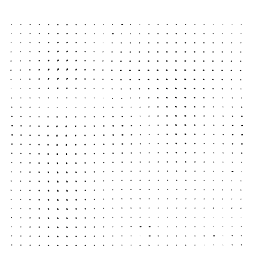


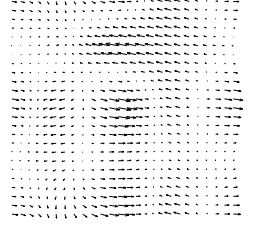


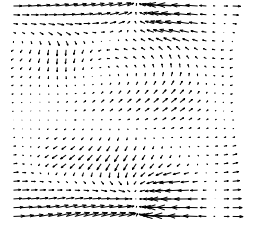


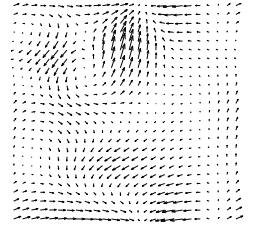


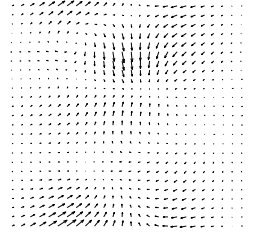

Tab. 7.7: Segmentation de la séquence *trial* selon θ_{cam}

du rebond qu'effectue le cycliste en fin de parcours.

7.4 Conclusion

Dans ce chapitre, nous avons montré que les paramètres des modèles du mouvement global sont utilisables pour la classification et la segmentation temporelle de vidéos. Une première expérience, consistant à classer différentes vidéos par un algorithme CHA, a montré que les paramètres du modèle basé sur les B-splines de degré 1 sont les plus satisfaisants pour ce type d'application. Ils fournissent une description du contenu permettant une bonne classification de la base et leur estimation nécessite un faible temps de calcul.

En modifiant l'algorithme CHA de manière à tenir compte de la relation temporelle entre

i	Première image	\mathcal{G}_i	Dernière image
1	$n^{\circ} 1$ 		$n^{\circ} 11$ 
2	$n^{\circ} 12$ 		$n^{\circ} 30$ 
3	$n^{\circ} 31$ 		$n^{\circ} 76$ 
4	$n^{\circ} 77$ 		$n^{\circ} 85$ 
5	$n^{\circ} 86$ 		$n^{\circ} 98$ 

Tab. 7.8: Segmentation de la séquence *trial* selon θ_{obj}

les images, nous avons développé une technique de segmentation temporelle de séquences d'images. Cette segmentation peut être obtenue en fonction du mouvement dominant ou du mouvement des objets. Les descripteurs du mouvement dominant correspondent aux coefficients des ondelettes de résolution égale à zéro. Les descripteurs associés aux mouvements des objets sont les coefficients de niveaux supérieurs à zéro. Les résultats obtenus montrent que ces descripteurs sont aussi pertinents pour segmenter une séquence selon le mouvement dominant et le mouvement des objets.

Ces résultats sur la classification et la segmentation temporelle ouvrent des perspectives intéressantes vers l'indexation vidéo et la reconnaissance d'activités.

Chapitre 8

Conclusion et perspectives

Cette thèse a exploré le thème de l'estimation et l'analyse du mouvement dans des séquences d'images. Nous avons développé deux approches permettant la mesure du flot optique. Toutes deux inscrites dans le cadre des méthodes différentielles, elles diffèrent par la complexité du modèle de mouvement utilisé ainsi que par la taille du support d'estimation. Ces choix de modèles permettent dans les deux cas de contourner le problème d'ouverture.

La méthode basée sur les filtres de Gabor, décrite au chapitre 3, fait l'hypothèse d'un mouvement localement translationnel. En projetant l'équation du flot optique en un pixel de l'image sur un banc de N filtres de Gabor, nous obtenons un système de N équations d'inconnues les deux composantes de vitesse du pixel considéré. Le système est surdéterminé car chaque filtre intègre localement l'information spatiale selon des propriétés fréquentielles différentes (orientations, échelles). La résolution de ce système local par une technique d'estimation robuste conduit à une mesure locale de la vitesse. Pour pouvoir estimer des mouvements de grande amplitude tout en préservant la résolution du flot optique, nous combinons, dans un cadre multirésolution, les estimations obtenues à différentes échelles spatiales. Afin de réduire la complexité algorithmique, nous avons utilisé une approximation récursive des filtres de Gabor. L'ordre du filtrage récursif (1 ou 3) détermine la qualité du flot optique obtenu mais aussi le coût de calcul de l'algorithme. Ainsi, à partir d'expérimentations sur des séquences artificielles et naturelles, nous avons montré qu'il était possible d'obtenir une estimation fiable et robuste du flot optique .

A l'inverse, la deuxième approche, décrite aux chapitres 5 et 6, consiste à estimer une modélisation globale du mouvement. Nous supposons que le modèle défini est valide en tout point de l'image. En appliquant l'équation du flot optique en chaque pixel, nous obtenons à nouveau un système dont les inconnues sont les paramètres du modèle de mouvement. Le nombre de paramètres étant très inférieur au nombre d'équations (égal au nombre de pixels), le système est, dans la majorité des cas, surdéterminé et est résolu par une technique d'estimation robuste. Le résultat obtenu est alors un jeu de paramètres de mouvement caractérisant de façon compacte et précise le flot optique sur tout le support de la scène. Nous avons testé deux types de modèles globaux. L'un est basé sur les séries de Fourier (chapitre 5), l'autre sur les séries d'ondelettes (chapitre 6). Ces modèles sont définis par un certain nombre de

fonctions de bases, qui correspond au nombre de termes de la série. Ainsi, plus ce nombre est élevé, plus la résolution spatiale du modèle est fine, mais plus le nombre de paramètres à estimer est important. La limite théorique est atteinte lorsque le modèle comporte autant de paramètres que de pixels dans l'image. Les mouvements de grande amplitude peuvent être estimés par un modèle de résolution faible, et les détails par un modèle de résolution haute. En reprenant l'idée de la multirésolution, l'estimation du mouvement est effectuée en combinant les estimations obtenues sur des modèles de différentes résolutions. Les résultats les plus probants ont été obtenus avec des modèles définis par des ondelettes B-splines. Selon la résolution du modèle utilisé, il est possible d'obtenir une estimation plus ou moins précise du flot optique. Ainsi, un modèle basé sur un pavage de 16×16 B-splines permet d'obtenir des performances de qualité équivalente aux meilleurs algorithmes d'estimation de mouvement.

Nous avons montré au chapitre 7 que les paramètres d'un modèle de basse résolution sont des descripteurs pertinents du mouvement pour la segmentation temporelle et la classification de vidéos. Une première expérience, consistant à classer une base de vidéos contenant différentes activités humaines par la classification hiérarchique des paramètres de mouvement estimés sur chaque séquence, a montré que les coefficients d'ondelettes B-splines sont adaptés pour la description du mouvement. Puis, nous avons développé une technique de segmentation temporelle de séquence, basée sur une version modifiée de la classification hiérarchique des paramètres estimés entre chaque image de la séquence. Les propriétés multi-échelles de la description en ondelettes nous ont permis de définir de manière très simple des descripteurs sur le mouvement dominant et sur le mouvement des objets. Ces deux nouveaux descripteurs sont relativement indépendants l'un de l'autre, et ont permis de segmenter temporellement des séquences sportives selon les déplacements de la caméra et le mouvement des personnages.

Les perspectives de notre travail se situent bien évidemment au niveau des deux algorithmes proposés. Chaque méthode est en fait un outil de vision bas-niveau qui n'a d'intérêt qu'en étant intégré dans un système complet de perception-réaction : perception de l'environnement, et réaction relative à l'interprétation de l'information perçue.

En ce qui concerne notre première approche, les filtres de Gabor étant un point de départ commun à de nombreux algorithmes de vision bas-niveau, il est logique de penser qu'un système de vision artificielle pourra combiner de manière naturelle des informations de vitesses et de textures, de formes ou de contours. Pour cela, il serait intéressant d'étudier par la suite les possibilités d'interaction entre le mouvement et d'autres attributs bas-niveau pour la segmentation [87], la reconnaissance d'objets ou d'activités [62], ...

Les perspectives les plus intéressantes des modèles du mouvement global concerne la description du contenu de vidéos. Basée sur une modélisation grossière du mouvement, la méthode est peu complexe. De même, les descripteurs sur le déplacement de la caméra et le mouvement des objets s'obtiennent sans calcul à partir des coefficients d'ondelettes. Le développement de systèmes d'indexation de vidéos ou de reconnaissance d'activités basés sur les descripteurs globaux de mouvement serait une suite intéressante à ce travail.

Annexes

Annexe A

M-estimateurs et moindres-carrés pondérés et itérés (MCPI)

Nous décrivons la technique des M-estimateurs ainsi que l'algorithme MCPI. Cette technique permet d'estimer de manière robuste un modèle à partir de données bruitées. Aussi, nous l'utilisons dans différents chapitres pour estimer des modèles de mouvement à partir des dérivés spatio-temporelles de la séquence d'images. La présentation qui suit est générale, et ne présuppose rien sur le modèle et les données utilisées.

A.1 Les M-estimateurs

Soit r_i le *résidu* associé à la i^{eme} donnée d_i . Il correspond à la différence entre la i^{eme} observation et sa valeur obtenue par le modèle $M(i; \boldsymbol{\theta})$ ($\boldsymbol{\theta} = [\theta_1, \dots, \theta_m]$ est le vecteur de paramètres à déterminer) :

$$r_i = d_i - M(i, \boldsymbol{\theta}). \quad (\text{A.1})$$

La méthode classique des moindres-carrés tente, pour ajuster le modèle aux données, de minimiser le carré des résidus,

$$\min_{\boldsymbol{\theta}} \sum_i r_i^2. \quad (\text{A.2})$$

Cette opération est instable si certaines données sont aberrantes vis-à-vis du modèle utilisé. En effet, ces données aberrantes ont une influence très importante du fait de la minimisation quadratique.

Un M-estimateur est une fonction qui pondère l'importance allouée à chaque donnée en fonction de sa conformité au modèle. Cette conformité est évaluée par le résidu r_i . Le terme quadratique de la minimisation au sens des moindres-carrés est alors remplacé par un M-estimateur ρ appliqué aux résidus r_i [35] :

$$\min_{\boldsymbol{\theta}} \sum_i \rho(r_i). \quad (\text{A.3})$$

ρ est une fonction symétrique, définie positive et ne possédant qu'un seul minimum en zéro. La croissance de ρ doit être inférieure à la fonction x^2 . Il existe un grand nombre de fonctions

possibles, et, dans le cas général, il est impossible d'en choisir une sans être arbitraire. Une revue sur les différents M-estimateurs utilisés en vision par ordinateur se trouve dans la thèse de Black [6].

La minimisation de (A.3) est obtenue en résolvant les m équations

$$\sum_i \psi(r_i) \frac{\partial r_i}{\partial \theta_j} = 0, \text{ pour } j = 1, \dots, m, \quad (\text{A.4})$$

où la dérivée $\psi(x) = \frac{d\rho(x)}{dx}$ est appelée *fonction d'influence*.

A.2 L'algorithme MCPI

Soit $w(x)$ la fonction de pondération définie par

$$w(x) = \frac{\psi(x)}{x}. \quad (\text{A.5})$$

L'équation (A.4) devient

$$\sum_i w(r_i) r_i \frac{\partial r_i}{\partial \theta_j} = 0, \text{ pour } j = 1, \dots, m. \quad (\text{A.6})$$

C'est exactement le système d'équations que l'on obtient en résolvant le problème des moindres-carrés pondérés et itérés suivant :

$$\min_{\theta} \sum_i w(r_i^{(k-1)}) r_i^2, \quad (\text{A.7})$$

où l'exposant k indique le numéro de l'itération en cours. Le poids $w(r_i^{(k-1)})$, associé à d_i , est recalculé après chaque itération en fonction de la nouvelle estimation de $\theta^{(k)}$, et est utilisé à l'itération suivante. A la première itération, tous les poids sont initialisés à 1.

On peut résumer le procédé par l'algorithme suivant :

1. Première estimation θ^0 =moindres-carrés (A.2)
2. Calcul des résidus r_i^0 =relation (A.1)
3. Calcul des poids $w_i(r_i^0)$ =relation (A.5)
4. Nouvelle estimation θ^1 =moindres-carrés-pondérés (A.7)
5. Test de convergence : si $(\frac{|\theta^1 - \theta^0|}{\theta^0} > \epsilon$ et $k < k_{max}$) alors retour au pas 2.

Les critères d'arrêt ϵ et k_{max} définissent le nombre d'itérations de l'algorithme. En pratique, on utilise peu d'itérations car le procédé a tendance à rejeter de nouveaux points à chaque itération. Typiquement, dans notre cas, $k_{max} = 3$ et $\epsilon = 10^{-2}$.

Annexe B

Méthode du gradient conjugué et biconjugué préconditionné

La méthode du gradient conjugué est un algorithme itératif permettant de résoudre des systèmes linéaires de la forme $A\mathbf{x} = \mathbf{b}$, $A \in \mathbb{R}^{N \times N}$ par minimisation de la fonctionnelle quadratique $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{x}^T \mathbf{b}$. La minimisation est obtenue en générant une succession de directions de recherches \mathbf{p}_k et de vecteurs \mathbf{x}_k minimisant $f(\mathbf{x})$. A chaque étape, α_k est estimé tel qu'il minimise $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)$, et $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$. Les vecteurs \mathbf{p}_k et \mathbf{x}_k sont construits de telle sorte que \mathbf{x}_{k+1} minimise f sur le sous-espace vectoriel défini par les directions de recherches précédentes $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k\}$. Après N itérations, nous trouvons le vecteur \mathbf{x}_N qui minimise f sur l'ensemble de l'espace vectoriel, c'est à dire la solution du système. Lorsque A n'est pas symétrique ou définie positive la généralisation du gradient conjugué, appelée gradient biconjugué, est utilisée. Sous la forme préconditionnée du système, c-à-d $(\tilde{A}^{-1}A)\mathbf{x} = \tilde{A}^{-1}\mathbf{b}$, les équations itératives sont de la forme :

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{p}_k \\ \alpha_k &= \frac{\bar{\mathbf{r}}_k^T \tilde{A}^{-1} \mathbf{r}_k}{\bar{\mathbf{p}}_k^T A \mathbf{p}_k} \\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k A \mathbf{p}_k \\ \bar{\mathbf{r}}_{k+1} &= \bar{\mathbf{r}}_k - \alpha_k A^T \mathbf{p}_k \\ \beta_k &= \frac{\bar{\mathbf{r}}_{k+1}^T \tilde{A}^{-1} \mathbf{r}_{k+1}}{\bar{\mathbf{r}}_k^T \tilde{A}^{-1} \mathbf{r}_k} \\ \mathbf{p}_{k+1} &= \tilde{A}^{-1} \mathbf{r}_{k+1} + \beta_k \mathbf{p}_k \\ \bar{\mathbf{p}}_{k+1} &= \tilde{A}^{-1} \bar{\mathbf{r}}_k + \beta_k \mathbf{p}_k \\ \mathbf{x}_0 &= \tilde{A}^{-1} A \mathbf{b} \\ \mathbf{r}_0 &= \bar{\mathbf{r}}_0 = \mathbf{b} - A \mathbf{x}_0 \\ \mathbf{p}_0 &= \bar{\mathbf{p}}_0 = \mathbf{r}_0\end{aligned}$$

La méthode du gradient conjugué est un cas particulier du gradient biconjugué, où $\bar{\mathbf{r}}_k = \mathbf{r}_k$ et $\bar{\mathbf{p}}_k = \mathbf{p}_k$.

Annexe C

Publications

Ce travail a donné lieu aux communications et publications suivantes :

E. Bruno et D. Pellerin. *Robust motion estimation using spatial Gabor-like filters*. Signal Processing, Volume 82/2, pp. 175-187, Mars 2002.

E. Bruno et D. Pellerin. *Modélisation du mouvement global sur une base d'ondelettes : application à l'indexation de séquences vidéo*. GRETSI'01, pp. 153-156, Toulouse, France, Septembre 2001.

E. Bruno et D. Pellerin. *Global motion model based on B-spline wavelets : application to motion estimation and video indexing*. Proc. of the 2nd IEEE region 8 and Eurasip International Symposium on Image and Signal Processing and Analysis, ISPA'01, pp. 289-294, Pula, Croatie, Juin 2001.

E. Bruno et D. Pellerin. *Global motion Fourier series expansion for video indexing and retrieval*. Advances in Visual Information System, VISUAL'00, LNCS 1929, pp. 327-337, Lyon, France, Novembre 2000.

E. Bruno et D. Pellerin. *Robust motion estimation using spatial Gabor filters*. X^{ème} European Signal Processing Conference, EUSIPCO'00, pp. 1465-1468, Tampere, Finlande, Septembre 2000.

E. Bruno et D. Pellerin. *Estimateur de mouvement par Bancs de Filtres de Gabor spatiaux*. ORASIS'99, pp. 25-33, Aussois, France, Avril 1999.

Bibliographie

- [1] E.H. Adelson and J.R. Bergen. *The Plenoptic Function and the Element of Early Vision*. M.S. Landy and J.A. Movshon, Massachusetts Institute of Technology, 1991.
- [2] S. Ayer and H.S. Sawhney. Layered representation of motion video using robust maximum-likelihood estimation of mixture model and MDL encoding. In *Proceedings of IEEE International Conference on Computer Vision*, pages 777–784, Cambridge, MA, USA, June 1995.
- [3] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 1(12) :43–77, 1994.
- [4] W. Beaudot. *Modèle biologique de la rétine des vertébrés, application au traitement d'images*. PhD thesis, Institut National Polytechnique de Grenoble, Laboratory LIS, Grenoble, France, 1995.
- [5] Ch. Bernard, S. Mallat, and J-J. Slotine. Discrete wavelet analysis : a new framework for fast optic flow computation. In *European Conference on Computer Vision (ECCV'98)*, Freiburg, Germany, June 1998.
- [6] M.J. Black. *Robust Incremental Optical Flow*. PhD thesis, Yale University, Department of Computer Science, 1992.
- [7] M.J. Black and P. Anandan. Robust dynamic motion estimation over time of optical flow. In *Proc. Computer Vision and Pattern Recognition, CVPR'91*, pages 231–236, June 1991.
- [8] M.J. Black, D.J. Fleet, and Y. Yacoob. A framework for modeling appearance change in image sequences. In *IEEE International Conference on Computer Vision, ICCV'98*, Mumbai, India, January 1998.
- [9] M.J. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1) :57–92, July 1996.
- [10] M.J. Black, Y. Yacoob, A.D. Jepson, and D.J. Fleet. Learning parametrized models of image motion. In *Computer Vision and Pattern Recognition CVPR'97*, pages 561–567, Puerto Rico, June 1997.
- [11] P. Bouthemy, M. Gelgon, and F. Ganansia. A unified approach to shot detection and camera motion characterization. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(7) :1030–1044, October 1999.

- [12] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Proceedings of IEEE International Conference Computer Vision and Pattern Recognition*, pages 8–15, Santa Barbara, 1998.
- [13] P.J. Burt, C. Yen, and X. Xu. Multiresolution flow-through motion analysis. In *Proc. Conference Computer Vision and Pattern Recognition*, pages 246–252, Washington, 1983.
- [14] CCITT. Recommendation H.261 : Video codec for audio-visual services at $p \times 64$ kbits/s. *COM XV-R 37-E*, 1989.
- [15] M.M. Chang, A. Murat Tekalp, and M. Ibrahim Sezan. Simultaneous motion estimation and segmentation. *IEEE Transactions on Image Processing*, 6(9) :1326–1333, September 1997.
- [16] Olivier Chomat. *Caractérisation d'éléments d'activités par la statistique conjointe de champs réceptifs*. PhD thesis, Institut National Polytechnique de Grenoble, 2000.
- [17] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.
- [18] V. Colin de Verdière and J.L. Crowley. Reconnaissance d'objets par apparence locale. In *11ème Congrès de Reconnaissance de Forme et d'Intelligence Artificielle*, Clermont-Ferrand, Janvier 1998.
- [19] E. Diday, G. Govaert, Y. Lechevallier, and J. Sidi. Clustering in pattern recognition. *Digital Image Processing*, pages 19–58, 1981.
- [20] C. Dimou and I. Pitas. Parameter estimation of non - translational motion fields. In *Proceedings of European Signal Processing Conference*, Trieste, Italy, September 1996.
- [21] R. Fablet and P. Bouthemy. Motion-based feature extraction and ascendant hierarchical classification for video indexing and retrieval. In *Proc. of the 3rd Int. Conf. on Visual Information Systems, VISual99*, volume 1614, pages 221–228, June 1999.
- [22] R. Fablet and P. Bouthemy. Statistical motion-based retrieval with partial query. In *Proc. of the 4th Int. Conf. on Visual Information Systems, VISUAL'00*, volume 1929, pages 96–107, November 2000.
- [23] D. Fleet and A. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5 :77–104, 1990.
- [24] D. J. Fleet, M. J. Black, and A. D. Jepson. Motion feature detection using steerable flow fields. In *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR-98*, pages 274–281, 1998.
- [25] D.J. Fleet, M.J. Black, Y. Yacoob, and A.D. Jepson. Design and use of linear models for image motion analysis. *International Journal of Computer Vision*, 36(3) :171–193, September 2000.
- [26] Marc Gelgon. *Segmentation spatio-temporelle et suivi dans une séquence d'image : application à la structuration et à l'indexation de vidéo*. PhD thesis, Université de Rennes I, 1998.

-
- [27] S. Geman and D.E. McClure. Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, LII(4) :5–21, 1987.
- [28] A. Guerin and M. Elghadi. Shape from texture by local frequencies estimation. In *SCIA '99*, Kangerlussuaq, Greenland, June 1999.
- [29] N. Guyader and J. Hérault. Représentation espace-fréquence pour la catégorisation d'images. In *GRETSI 2001*, Toulouse, France, September 2001.
- [30] D.J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, pages 297–302, 1988.
- [31] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuities optical flow using markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12), 1993.
- [32] T. Hofmann, J. Puzicha, and J. Buhmann. Unsupervised texture segmentation in deterministic annealing framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8) :803–818, August 1998.
- [33] B.K.P. Horn and B.G. Schunk. Determining optical flow. *Artificial Intelligence*, 17 :185–204, 1981.
- [34] K. Hotta, T. Mishima, T. Kurita, and S. Umeyama. Face matching through information theoretical attention points and its applications to face detection and classification. In *IEEE Conf. on Automatic Face and Gesture Recognition, FG'00*, pages 34–39, Grenoble, France, April 2000.
- [35] P.J. Huber. *Robust statistics*. Wiley, Erewhon, NC, 1981.
- [36] M. Irani and P. Anandan. Video indexing based on mosaic representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 86(5) :905–921, May 1998.
- [37] ITU-T. Recommendation H.263 : Video coding of narrow telecommunication channel < 64kbits/s. 1995.
- [38] A.K. Jain and K. Karu. Learning texture discrimination masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2) :195–205, 1996.
- [39] J.P. Jones and L.A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58 :1233–1258, 1987.
- [40] ISO/IEC JTC1/SC29/WG1 N1646R JPEG2000. Jpeg2000 part 1 final committee draft version 1.0. 2000.
- [41] S. Ju, M.J. Black, and Y. Yacoob. Cardboard people : A parametrized model of articulated image motion. In *Proceedings of IEEE Conference on Automatic Face and Gesture Recognition*, pages 38–44, Killington, Vermont, October 1996.
- [42] P. Kalocsai, H. Neven, and J. Steffens. Statistical analysis of Gabor filter representation. In *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 360–365, 1998.

- [43] M. Lievin, P. Delmas, P.Y. Coulon, F. Luthon, and V. Fristot. Automatic lip tracking : Bayesian segmentation and active contours in a cooperative scheme. In *IEEE Conf. on Multimedia, Computing and Systems, ICMCS'99*, Fiorenza, Italy, June 1999.
- [44] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. of the DARPA IU Workshop*, pages 121–130, 1981.
- [45] M.J. Lyons, J. Budynek, A. Plante, and S. Akamatsu. Classifying facial attributes using a 2-D Gabor wavelet representation and discriminant analysis. In *IEEE Conf. on Automatic Face and Gesture Recognition, FG'00*, pages 202–207, Grenoble, France, April 2000.
- [46] M. Malcius and F. Prêteux. A robust model-based approach for 3D head tracking in video sequences. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pages 169–174, Grenoble, March 2000.
- [47] S. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 :674–693, July 1989.
- [48] S. McKenna, S. Gong, R.P. Wurtz, J. Tanner, and D. Banin. Tracking facial feature points with gabor wavelets and shape models. In Josef Bigün, Gerard Chollet, and Gunilla Borgefors, editors, *1st International Conference on Audio- and Video-based Biometrics Person Authentication, LNCS 1206*, pages 35–42. Springer Verlag, March 1997.
- [49] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5) :703–719, May 1998.
- [50] Y. Meyer. *Ondelettes et Opérateurs I*. Hermann, 1996.
- [51] R. Milanese, D. Squire, and T. Pun. Correspondence analysis and hierarchical indexing for content-based image retrieval. In *ICIP'96*, Lausanne, Switzerland, September 1996.
- [52] A. Mitiche, Y. F. Wang, and J. K. Aggarwal. Experiment in computing optical flow with the gradient-based, multiconstraint method. *Pattern Recognition*, 20(2) :173–179, 1987.
- [53] ISO/IEC JTC1 IS 11172-2 (MPEG-1). Information technology - coding and moving pictures and associated audio for digital storage media at up to about 1.5Mbits/s. 1993.
- [54] ISO/IEC JTC1 IS 13818-2 (MPEG-2). Information technology - generic coding and moving pictures and associated audio information. 1996.
- [55] ISO/IEC JTC1/SC29/WG11 N4030 (MPEG-4). Mpeg-4 overview (V.18). Singapore, 2001.
- [56] ISO/IEC JTC1/SC29/WG11 N4031 (MPEG-7). Overview of the Mpeg-7 standard (version 5.0). Singapore, 2001.
- [57] S. Negahdaripour and C. Yu. A generalized brightness change model for computing optical flow. In *IEEE International Conference on Computer Vision, ICCV'93*, pages 2–11, January 1998.

-
- [58] R. Nelson and P. Polana. Qualitative recognition of motions using temporal texture. In *CVGIP : Image Understanding*, volume 1, pages 78–89, July 1992.
- [59] J. M. Odobez. *Estimation, détection et segmentation : une approche robuste et markovienne*. PhD thesis, Université de Rennes 1, Irisa No 1304, 1994.
- [60] J.-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4) :348–365, December 1995.
- [61] J.-M. Odobez and P. Bouthemy. Direct incremental model-based image motion segmentation for video analysis. *Signal Processing*, 66(2) :143–155, April 1998.
- [62] R. Okada, Y. Shirai, and J. Miura. Tracking a person with 3-D motion by integrating optical flow and depth. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 336–341, Grenoble, France, March 2000.
- [63] A. Oliva and A. Torralba. Scene classification from low-level features. In *ARVO'99*, Fort Lauderdale, Florida, USA, 1999.
- [64] K. Otsuka, T. Horikoshi, S. Suzuki, and M. Fujii. Feature extraction of temporal texture based on spatio-temporal motion trajectory. In *Proc. Int. Conf. on Pattern Recognition, ICPR'98*, August 1998.
- [65] P. Palagi. *Modélisation des premiers étages du cortex visuel par filtrages orientés et cartes topologiques. Application à la segmentation de texture*. PhD thesis, Institut National Polytechnique de Grenoble, Laboratory LIS, Grenoble, France, 1995.
- [66] W. H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C, Second Edition*. Cambridge University Press, 1992.
- [67] ISO/IEC JTC1/SC29/WG1 FCD 14495 public draft. Lossless and near-lossless coding of continuous tone still images (jpeg-ls). 1997.
- [68] S. Rakshit and C.H. Anderson. Computation of optical flow using basis functions. *IEEE Transactions on Image Processing*, 6(9) :1246–1254, September 1997.
- [69] B. E. Shi. Gabor-type filtering in space and time with cellular neural networks. *IEEE Transactions on Circuits and Systems-I : Fundamental Theory and Applications*, 45(2) :121–132, February 1998.
- [70] H. Sidenbladh, M.J. Black, and D.J. Fleet. Stochastic tracking of 3D human figures using 2D image motion. In *Proceedings of European Conference on Computer Vision*, Dublin, June 2000.
- [71] E.P. Simoncelli. *Distributed Representation and Analysis of Visual Motion*. PhD thesis, Massachusetts Institute of Technology, Dpt of Electrical Engineering and Computer Science, 1993.
- [72] A. Spinei, D. Pellerin, and J. Héroult. Spatiotemporal energy-based method for velocity estimation. *Signal Processing*, 65(6) :347–362, 1998.

- [73] S. Srinivasan and R. Chellappa. Noise-resilient estimation of optical flow by use of overlapped basis functions. *Journal of the Optical Society of America A*, 16(3) :493–507, March 1999.
- [74] C. Stiller and J. Konrad. Estimating motion in image sequences. a tutorial on modeling and computation of 2D motion. *IEEE Signal Processing Magazine*, July 1999.
- [75] G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.
- [76] R. Szeliski and J. Coughlan. Spline-based image registration. *International Journal of Computer Vision*, 22(3) :199–218, 1997.
- [77] R. Szeliski and H. Shum. Motion estimation with quadtree splines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12) :1199–1210, December 1996.
- [78] M. Szummer and R.W. Picard. Temporal texture modeling. In *ICIP'96*, September 1996.
- [79] A.B. Torralba. *Architectures analogiques pour la vision et circuits neuromorphiques*. PhD thesis, Institut National Polytechnique de Grenoble, Laboratory LIS, Grenoble, France, 1999.
- [80] A.B. Torralba and J. Héroult. An efficient neuromorphic analog network for motion estimation. *IEEE Transactions on Circuits and Systems-I*, 46(2), February 1999.
- [81] T.R. Tsao and V.Chen. A neural sheme for optical flow computation based on gabor filters and generalized gradient method. *Neurocomputing*, 6(3) :305–325, 1994.
- [82] M. Unser, A. Aldroubi, and M. Eden. Fast B-spline transforms for continuous image representation and interpolation. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 13 :277–285, March 1991.
- [83] L. J. van Vliet, I. T. Young, and P. W. Verbeek. Recursive gaussian derivative filters. In *International Conference on Pattern Recognition, ICPR'98*, volume 1, pages 509–514, Brisbane, Australia, August 1998.
- [84] N. Vasconcelos and A. Lippman. Spatiotemporal motion model for video summarization. In *Proceedings of Computer Vision and Pattern Recognition, CVPR'97*, Santa Barbara, CA, 1997.
- [85] J.Y.A. Wang and E.H. Adelson. Representing moving images with layers. *IEEE Transactions on Image Processing*, 3(5) :625–638, September 1994.
- [86] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. *International Journal of Computer Vision*, 14(1) :67–81, 1995.
- [87] Y. Weiss and H. Adelson. A unified mixture framework for motion segmentation : Incorporating spatial coherence and estimating the number of models. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, CVPR'96*, pages 321–326, San Fransisco, 1996.
- [88] X. Wu and B Bhanu. Gabor wavelet representation for 3-D object recognition. *IEEE Transactions on Image Processing*, 6(1) :47–64, 1995.

- [89] Y. Wu and K. Toyama. Wide-range, person- and illumination-insensitive head orientation estimation. In *IEEE Conf. on Automatic Face and Gesture Recognition, FG'00*, pages 183–188, Grenoble, France, March 2000.
- [90] Y.T. Wu, T. Kanade, C.C. Li, and J. Cohn. Image registration using wavelet-based motion model. *International Journal of Computer Vision*, 38(2) :129–152, July 2000.
- [91] Y. Zhang and C. Kambhamettu. Robust 3d head tracking under partial occlusion. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pages 176–182, Grenoble, March 2000.