Université de Montréal

Estimation de mouvement sans restriction par filtres en quadrature localisés

par

Gaspard Petit

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en informatique

décembre, 2006

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé :

Estimation de mouvement sans restriction par filtres en quadrature localisés

présenté par

Gaspard Petit

a été évalué par un jury composé des personnes suivantes:

Max Mignotte
(Président-rapporteur)

Sébastien Roy
(Directeur de recherche)

Victor Ostromoukhov
(Membre du jury)

Mémoire accepté le 7 avril 2006

RÉSUMÉ

Nous présentons une nouvelle méthode pour l'estimation de mouvement entre deux images où les déplacement ne sont pas limités par la longueur d'onde des filtres utilisés. Cette approche utilise des signatures de phase calculées pour chacun des pixels à l'aide d'une banque de filtres localisés.

Dans un premier temps, la corrélation de ces signatures permet d'estimer un mouvement entier entre deux pixels. Puisque nous utilisons des filtres très minces et ayant un support limité à une seule longueur d'onde, la méthode détecte efficacement les bordures de discontinuités de mouvement. De plus, nous expliquons comment la comparaison des filtres peut être robuste aux occlusions et invariante au changement de contraste et d'orientation.

Une étape subséquente raffine le mouvement à une précision de sous-pixel en utilisant le gradient des signatures. Plutôt que d'utiliser une minimisation de distance pour chacune des réponses de la signature, nous proposons une nouvelle méthode de résolution par vote. Cette méthode est plus robuste au bruit et détecte les mouvements multiples, tel qu'observés en zones de discontinuité de mouvement.

La méthode, quoique sans terme de régularisation, se compare de façon très favorable aux autres méthodes récentes avec régularisation. Des résultats sont présentés pour des séquences dont le mouvement est connu et sur des séquences présentant de grands déplacements.

mots-clefs : flux optique. estimation de mouvement, quadrature, corrélation, phase, énergie.

ABSTRACT

We present a new method for motion estimation between two images that does not constraint the motion by the wavelength of the filters used. This approach uses signatures of phases computed for every pixel using a localized filter bank.

To find the displacement, we first correlate the signatures and choose the best match as an initial integer displacement. Since our filters are very thin and have a support limited to one wavelength, they are robust to motion discontinuities. In addition, we show that the correlation can be made invariant to change in brightness and change in orientation.

Once an initial displacement is obtained, a subsequent step refines the displacement to sub-pixel accuracy using the gradient of the signatures. Instead of using a least-square minimization for this refinement, we propose a new vote-based method. Our method is more robust to noise and can detect multiple motion, as observed in areas of motion discontinuity.

Our approach, although using direct search with no regularization, compares favourably with other methods with regularization. Results are shown for sequences with ground truth and sequences with large displacements.

keywords : optical flow, motion estimation, quadrature, correlation, phase, energy.

TABLE DES MATIÈRES

LISTE DES FIGURES

Chapitre 1

INTRODUCTION

L'estimation du mouvement est utilisée dans une multitude d'applications en imagerie : pour réduire la redondance temporelle en compression vidéo, stabiliser des séquences vidéo, générer des mosaïques, augmenter la résolution par super-résolution, pour du tracking, de la composition d'objets sur des scènes réelles, de la restauration, calibrage, de la reconstruction et pour beaucoup d'autres. Malgré son rôle clef en vision par ordinateur, il est difficile encore aujourd'hui de construire un système d'estimation de mouvement robuste qui se compare au système visuel humain. Le système visuel humain interprète très bien les occlusions, la transparence, les réflexion, changements d'illumination, de teinte, de forme, d'orientation, d'échelle, le bruit ou encore la mauvaise résolution temporelle – alors que chacun de ces phénomènes est un défi pour n'importe lequel des algorithme d'estimation de mouvement.

Plusieurs méthodes ont été proposées au cours des dernières années. Quatre familles de méthodes ont été identifiées par Barron *et al* : par gradient, par énergie, par phase et par corrélation. Plusieurs méthodes n'appartiennent pas strictement à une seule famille – c'est le cas de celle que nous présentons. Notre méthode effectue une analyse par phase, une recherche semblable à la corrélation et un raffinement au sous-pixel par gradient. De plus, comme la méthode par vote que nous proposons pour résoudre le sous-pixel a initialement été développée pour les méthodes par énergie, nous commencerons par un résumé de ces quatre familles de méthodes.

Par la suite, nous survolerons le développement du flux optique à travers les 25 dernières années incluant les méthodes les plus récentes. Cette revue permettra d'ob-

server que la recherche en estimation de mouvement s'est principalement préoccupée de trouver un modèle de régularisation et qu'il y a eu très peu d'évolution du côté du terme d'information (défini dans la prochaine section). La méthode présentée en §3 utilise seulement le terme d'information pour estimer le mouvement et obtient des résultats meilleurs que la plupart des méthodes avec régularisation. Enfin, en §4, nous présentons une méthode robuste développée afin d'estimer le mouvement par vote plutôt que par minimisation, ce qui lui permet d'être robuste aux aberrants et aux mouvements multiples.

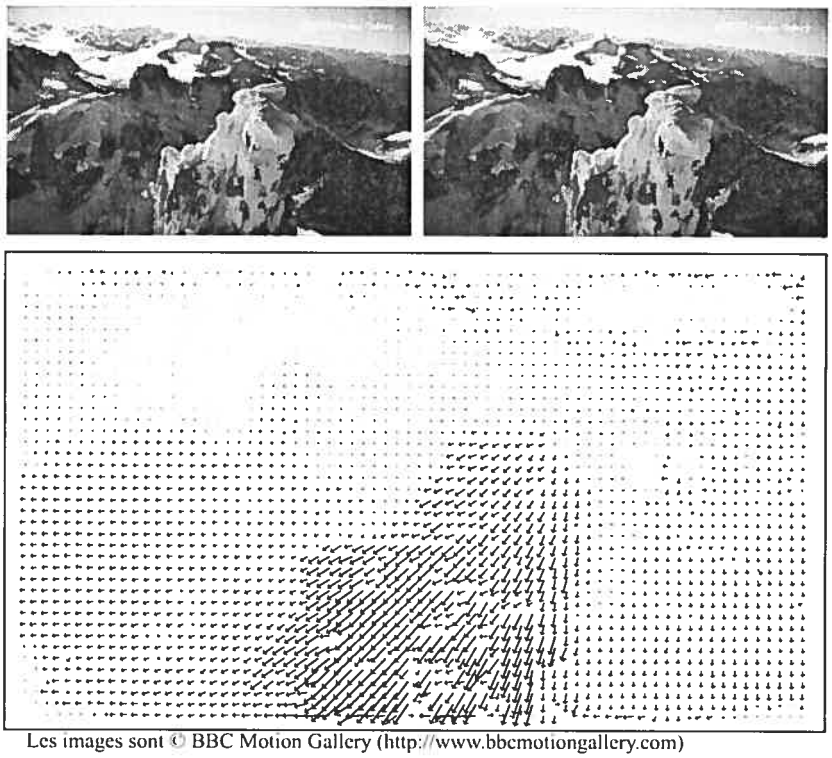
1.1 Familles de flux optique

Comme beaucoup d'autres problèmes en imagerie et en vision par ordinateur, l'estimation de mouvement dans une séquence d'images est un problème mal posé au sens de Hadamar [1]. C'est-à-dire qu'il manque d'information et qu'il existe une multitude de solutions possibles. Afin de résoudre ce genre de problèmes, il faut fixer des contraintes qui permettront au système de converger vers une solution acceptable. Trouver les contraintes idéales revient à résoudre le problème d'estimation de mouvement.

La grande variété de types de séquences (voir figures 1.1, 1.2 et 1.3) rend difficile le choix de contraintes pertinentes applicables dans tous les cas. De façon générale, Bertero, Poggio et Torre [2] on identifiés en 1988 trois types de contraintes qu'il faut satisfaire simultanément :

Conservation d'information : Il existe une certaine redondance dans l'information contenue entre deux images. Que ce soit par constance d'intensité, de gradient, de phase, etc., les deux images doivent être reliées l'une à l'autre temporellement. La solution devrait donc minimiser le changement de cette information.

Cohérence spatiale : les objets ont un certain support spatial, et les points d'un



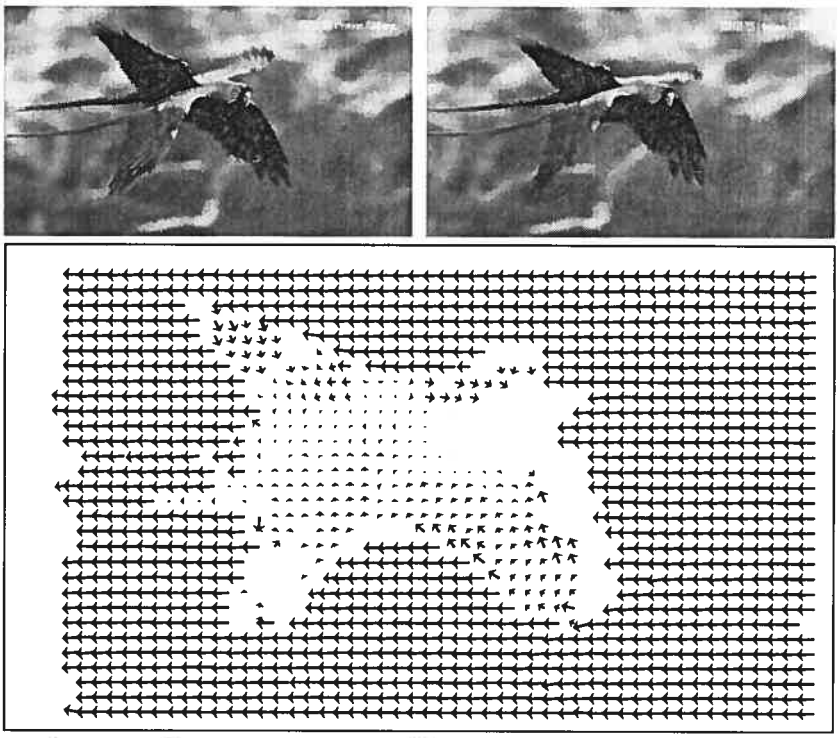
Les images sont © BBC Motion Gallery (http://www.bbcmotiongallery.com)

FIG. **1.1.** Deux images d'une séquences contenant des mouvements rigides. Seule la caméra se déplace, mais comme les surfaces sont à des profondeurs différentes, elles se déplacent à des vitesse différentes le long des lignes épipolaires. (**Bas**) Exemple de carte de mouvement pour ces deux images à partir de la méthode présentée en §3.

même objet se déplacent de façon semblable dans un voisinage. La solution devrait favoriser la cohésion du mouvement.

Cohérence temporelle : Le déplacement d'un objet dans le temps devrait être relativement continu (souvent simplifié à une accélération constante). La solution devrait donc maximiser la continuité du mouvement.

Les deux premières contraintes ont été introduites en 1981 lorsque Horn et Schunck [3] ainsi que Lucas et Kanade [4] ont présenté leur méthodes par gradient. Yachida [5] utilisait déjà la contrainte temporelle en 1981, mais ce n'est que 10 ans plus tard, en



Les images sont © BBC Motion Gallery (http://www.bbcmotiongallery.com)

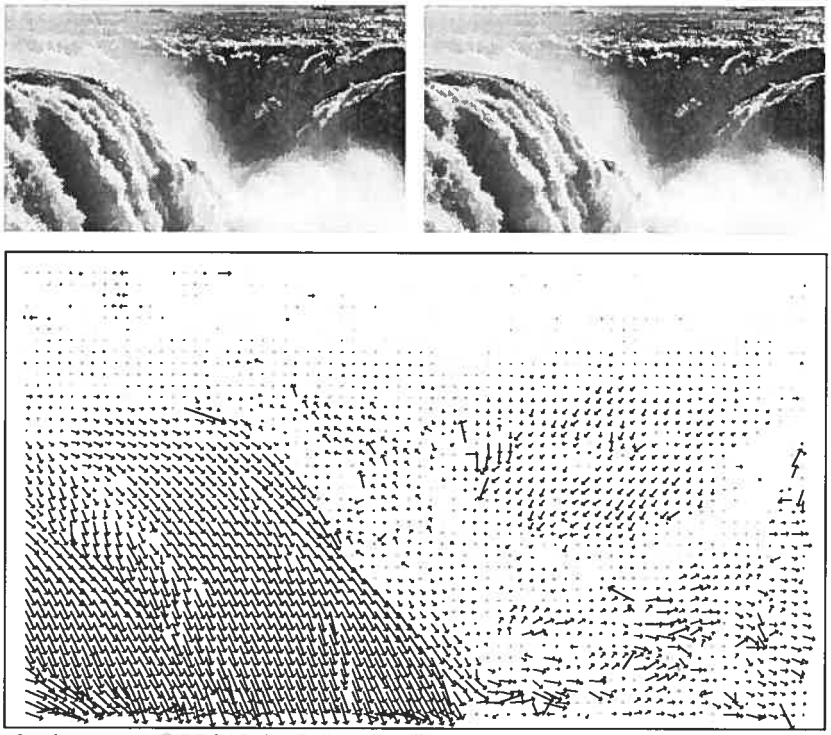
FIG. **1.2.** (**Haut**) Deux images d'une séquences contenant des mouvements rigides articulés. En plus de contenir de grands mouvements (le fond se déplace à plus de 60 pixels par image), la texture de l'aile change et une partie de la première image est cachée par la seconde (et vice et versa). (**Bas**) Exemple de carte de mouvement pour ces deux images à partir de la méthode présentée en §3.

1991 qu'elle a été plus formellement réintroduite par Black et Anandan [6].

Ces contraintes se représentent sous la forme d'une fonctionnelle $\mathcal{H}$ à minimiser :

$$\mathcal{H} \triangleq \alpha_1 \mathcal{H}_{\text{info}} + \alpha_2 \mathcal{H}_{\text{spatial}} + \alpha_3 \mathcal{H}_{\text{temporel}} \tag{1.1}$$

où $\mathcal{H}_{\text{info}}$, $\mathcal{H}_{\text{spatial}}$ et $\mathcal{H}_{\text{temporel}}$ sont les pénalités des contraintes d'information de cohérence spatiale et de cohérence temporelle et où α_1, α_2 et α_3 sont utilisés pour ajuster l'importance de chacune de ces contraintes.



Les images sont © BBC Motion Gallery (http://www.bbcmotiongallery.com)

FIG. **1.3.** (**Haut**) Deux images d'une séquences contenant des mouvements non rigide. En plus de contenir un mouvement difficile à modéliser, cette scène contient de la transparence ainsi que de la spécularité. (**Bas**) Exemple de carte de mouvement pour ces deux images à partir de la méthode présentée en §3.

Les contraintes sont évaluées comme une distance à la condition idéale $s = 0$:

$$\mathcal{H}_i = \Psi_i(s_i^2). \tag{1.2}$$

Traditionnellement, la fonction $\Psi(s^2)$ était prise comme simplement s^2, ce qui avait tendance à donner trop d'importance aux aberrations, au bruit et rendre trop «improbables» les discontinuités de mouvement. Black et Anandan [6] ont apporté une contribution majeure en 1991 en introduisant des distances redescendantes (figure 1.4) au domaine de l'estimation du mouvement.

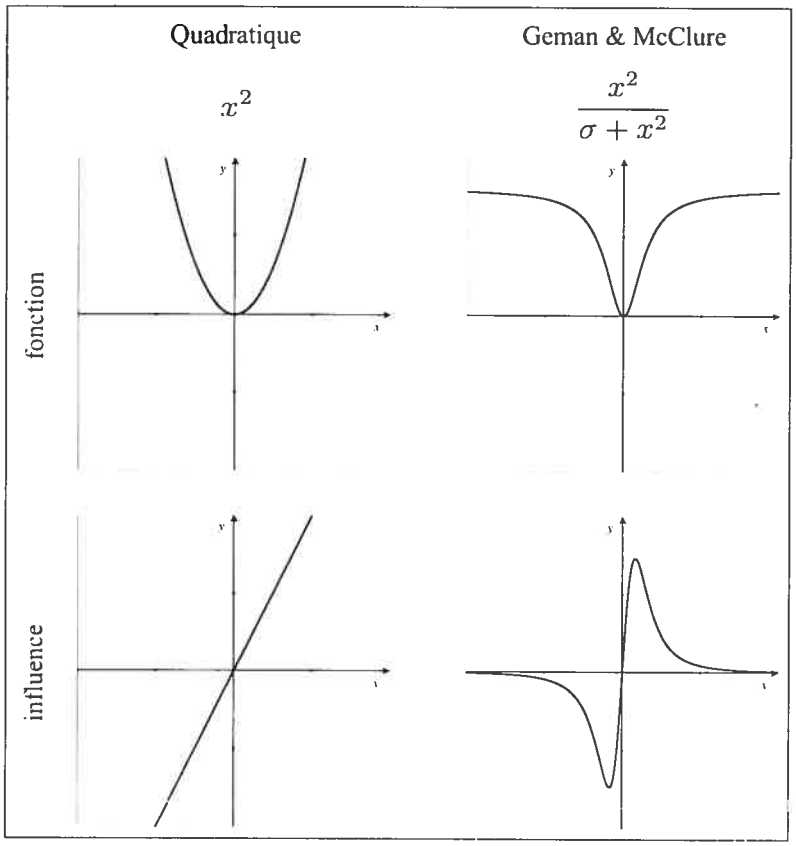


FIG. **1.4.** Une fonction Ψ quadratique et une fonction Ψ robuste ainsi que leurs dérivées respectives. D'autres fonctions robustes sont présentées dans la figure 2.1 à la page 47.

1.2 Régularisation

Le terme de régularisation temporelle reste le terme le moins utilisé et étudié comme l'a fait remarquer Weicker [7] en 2001. Ceci s'explique probablement par les efforts de calculs considérables qui y sont associés, mais aussi probablement parce que l'échantillonnage temporel n'assure pas toujours une fluidité de mouvement. Le modèle utilisé est généralement très simple : on suppose un mouvement constant ou une accélération constante.

Par contre, une pléthore de régularisations spatiales ont été proposées. Bergen,

Anandan, Hanna et Hingorani [8] identifient trois types de régularisation :

Pleinement paramétrique : Chaque pixel suit un mouvement paramétrique, selon un modèle global, par exemple. une scène planaire avec un mouvement de caméra.

Quasi-paramétrique : Chaque pixel est à la fois contraint par un mouvement paramétrique et une information locale. Par exemple, si le mouvement de caméra est connu, le mouvement paramétrique contraint les déplacements le long des lignes épipolaires, mais chaque pixel a une certaine liberté quant à sa vitesse.

Non-paramétrique : Aucun modèle n'est utilisé, mais généralement il existe tout de même une contrainte de lissage. Certains mouvements tels que les mouvements non rigides et articulés sont difficilement paramétrisables.

La paramétrisation peut se faire de façon globale (*e.g.* pour retrouver un mouvement de caméra ou prédominant), par régions ou encore localement. Parmi les paramétrisations les plus utilisées, on retrouve :

Paramétrisation rigide : constante, de sorte que les voisins soient contraints de se déplacer de la même façon.

Paramétrisation planaire : permet de représenter la projection du mouvement d'un plan 3D en 2D. Il s'agit d'un cas spécifique d'une paramétrisation affine.

Paramétrisation affine : permet de représenter une transformation affine 2D, c'est à dire la projection 2D d'une déformation linéaire 3D quelconque.

Enfin, très tôt il a semblé avantageux de coupler la régularisation spatiale avec le contenu de l'image. Déjà en 1983, Nagel [9] proposait d'utiliser le gradient de l'image dans le terme de régularisation. Cette idée a été reprise plusieurs fois et formalisée comme une méthode de diffusion anisotropique par Alvarez, Weickert et Sánchez en 2000 [10].

Plusieurs chercheurs ont également tenté de représenter le flux optique à l'aide d'un modèle probabiliste – bien adaptés aux problèmes mal posés. En 1990, Singh [11]

propose une méthode de corrélation de régions probabiliste. en 1991 Simoncelli, Adelson et Heeger [12] présentent un modèle semblable pour les méthodes par gradients, où la covariance d'une gaussienne est définie en fonction du gradient de l'image et en 2000, Roy et Govindu [13] présentent un modèle plus réaliste non gaussien qui tient compte de la variance à la fois spatiale mais aussi temporelle. Les approches probabilistes ouvrent la porte à des minimisations de type markovienne [14] ou encore l'utilisation de filtres de Kalman [12].

L'estimation de mouvement étant un problème mal posé, plusieurs chercheurs ont tenté de trouver un modèle qui puisse améliorer les solutions. Alors que beaucoup de recherche a déjà été faite dans le domaine des statistiques d'images [15, 16], Roth et Black [17] ont démontré récemment qu'il est possible d'utiliser de telles statistiques à des fins de régularisation.

Mais au coeur de l'estimation du mouvement, il y a le terme d'information. Quatre familles de termes d'informations ont été répertoriées par Barron, Fleet et Beauchemin [18]. Plusieurs méthodes appartiennent à plusieurs familles à la fois. Par exemple, les méthodes par phase et par gradient emploient une approche semblable mais utilisent des filtres différents : un filtre en quadrature (une exponentielle complexe modulée) dans le cas des phases, et une différence de gaussienne pour les gradients. Les méthodes par corrélation peuvent également être comparées aux méthodes par énergie : calculer la somme des différences au carré (SSD) dans le domaine spatial est semblable à trouver la distance entre des échantillons et un plan de mouvement dans le domaine spectral. Similairement, les méthodes par phase se comparent aux méthodes par énergie – comme nous l'expliquons en §1.5, il s'agit du changement phase dans le spectre 2D qui génère un plan dans le spectre 3D. Trouver le plan en 3D revient donc à trouver le déphasage en 2D. Bref, ces familles ne sont que des modèles différents représentant une information parfois très semblable.

1.3 *Approches par dérivées spatio-temporelles*

Les méthodes par dérivées spatio-temporelles (aussi appelées méthodes différentielles, variationelles ou encore par gradients) estiment le mouvement d'un signal à l'aide de ses dérivées. Le déplacement v entre un signal $f(x)$ au temps t et un signal $g(x) = f(x - v(x))$ peut être calculé *via* une expansion de Taylor :

$$g(x) = f(x - v(x)) = f(x) - v(x)\frac{f'(x)}{1!} + v(x)^2\frac{f''(x)}{2!} - v(x)^3\frac{f'''(x)}{3!} + ...$$

En supposant que le signal est purement linéaire (*i.e.* $f^{(n)}(x) = 0$ pour $n > 1$), cette série se simplifie à

$$g(x) = f(x) + v(x)f'(x)$$

et s'exprime généralement comme

$$f_t(x) + v(x)f_x(x) = 0 \quad \text{ou} \quad \nabla f \cdot \mathbf{v} = 0 \quad (\text{avec } v_t = 1)$$

Cette contrainte est appelée **contrainte de d'intensité constante** (*constant brightness contraint - CBC*), ou encore **contrainte de flux optique** (*optical flow constraint - OFC*), ou bien **supposition d'intensité constante** (*constant brightness assumption - CBA*) c'est à dire, que l'intensité d'un pixel ne devrait pas changer le long de son déplacement.

Résoudre cette équation en une dimension est simple (figure 1.5) mais la résoudre en deux dimensions implique une équation à deux inconnues :

$$f_t(x, y) + v_x(x, y)f_x(x, y) + v_y(x, y)f_y(x, y) = 0. \tag{1.3}$$

où f_x, f_y et f_t sont les gradients spatiaux-temporels. Plusieurs méthodes existent pour résoudre ce système d'équations. Une approche consiste à supposer que le mouvement est perpendiculaire au gradient, appelé **flux «normal»** (figure 1.6) et donc de poser v_x comme une fonction de v_y. Le déplacement peut alors être calculé avec

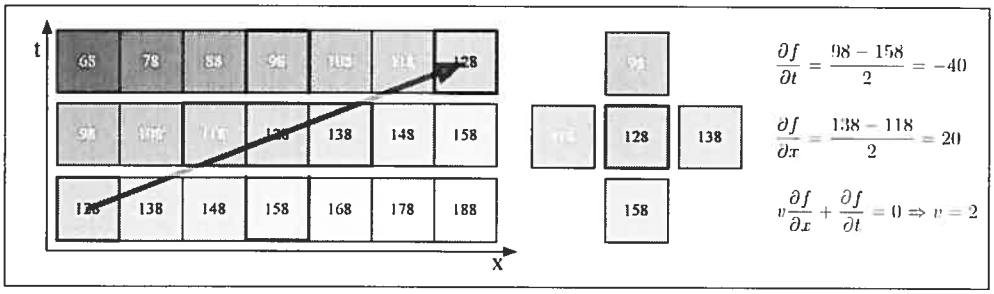


FIG. **1.5. On peut estimer le mouvement d'un signal 1D à partir de ses dérivées spatio-temporelles.** Un signal en translation de 2 pixels par image vers la droite. La dérivée en t et en x permet d'estimer le mouvement.

$$\mathbf{v} = f_t \frac{\langle f_x, f_y \rangle}{\| \langle f_x, f_y \rangle \|^2} \tag{1.4}$$

Le flux normal correspond à la perception du signal à travers une fenêtre très petite. Ce phénomène est appelé **problème d'ouverture** (*aperture problem*) et est aussi rencontré lorsque la texture ne permet pas une estimation du mouvement sans ambiguïté (figure 1.7). Le problème d'ouverture rend nécessaire la propagation d'information entre voisins jusqu'à ce que le mouvement puisse être estimé sans ambiguïté. Le système humain doit aussi faire face à ce problème, ce qui explique plusieurs illusions d'optiques, par exemple celle du barbier (figure 1.8). Dans le cas de cette illusion, les seuls points ne présentant pas d'ambiguïté sont les points d'intersection entre la texture et le cadre du fenêtrage. Si le fenêtrage est vertical, il y a plus de points se déplaçant verticalement qu'horizontalement et le mouvement global est perçu comme vertical. Si le fenêtrage est horizontal, il y a plus de points se déplaçant horizontalement et le mouvement global semble horizontal. Fermüller, Shilman et Aloimonos [19] discutent d'une autre illusion qu'ils expliquent par le problème d'ouverture : l'illusion d'Ouchi (figure 1.9). Lorsque la texture se déplace légèrement, l'orientation des rectangles amène l'oeil à résoudre le mouvement différemment dans les deux régions, ce qui donne un mouvement apparent et une segmentation de la région centrale.

Résoudre le problème d'ouverture peut se faire de deux façons. La première

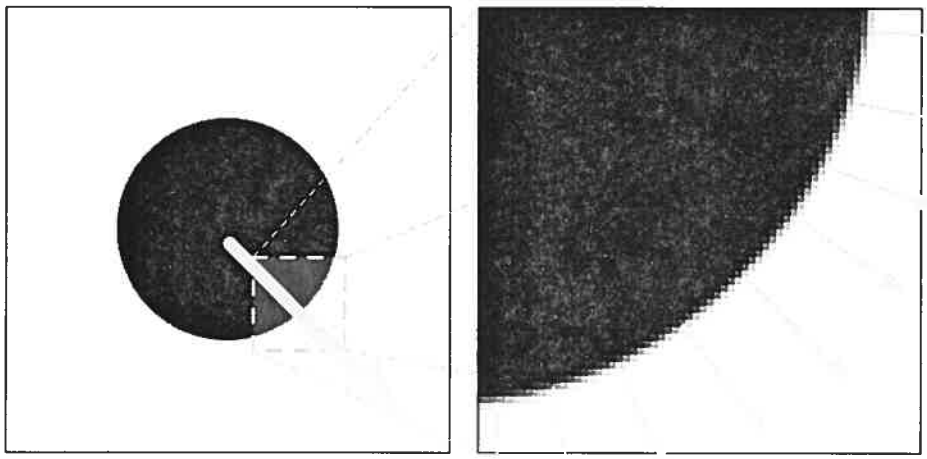


FIG. **1.6.** Le flux normal est le mouvement estimé le long de la normale du gradient de l'image.

consiste à choisir le flux normal et à laisser un terme de régularisation spatiale propager et corriger le mouvement réel. La deuxième consiste à choisir un fenêtrage et à utiliser plusieurs échantillons pour résoudre l'équation 1.3. Cette approche est généralement résolue par moindre carré. Tel que traité par Okutomi et Kanade [20], cette méthode fonctionne dans la mesure où la fenêtre est assez grande de sorte que le système d'équation ait une réponse unique, mais cette dernière ne doit pas contenir de mouvements multiples sans quoi le moindre carré donnera un mouvement moyen qui n'a probablement rien à voir avec aucun des mouvements réels. Bergen *et al* [21] ainsi que Jepson et Black [22] proposent des méthodes itératives afin de séparer les mouvements multiples. Des méthodes plus sophistiquées ont été proposées par la suite [23, 24, 25, 26]. La méthode proposée en §4 s'inscrit dans cette liste et se démarque par sa robustesse et son approche non itérative et non paramétrique.

Les méthodes spatio-temporelles fonctionnent sous deux conditions :

- $f(x) = f(x + v(x)) = g(x)$ implique que l'intensité du signal reste la même (un bruit peut être modélisé). Dans plusieurs situations, cette condition n'est pas

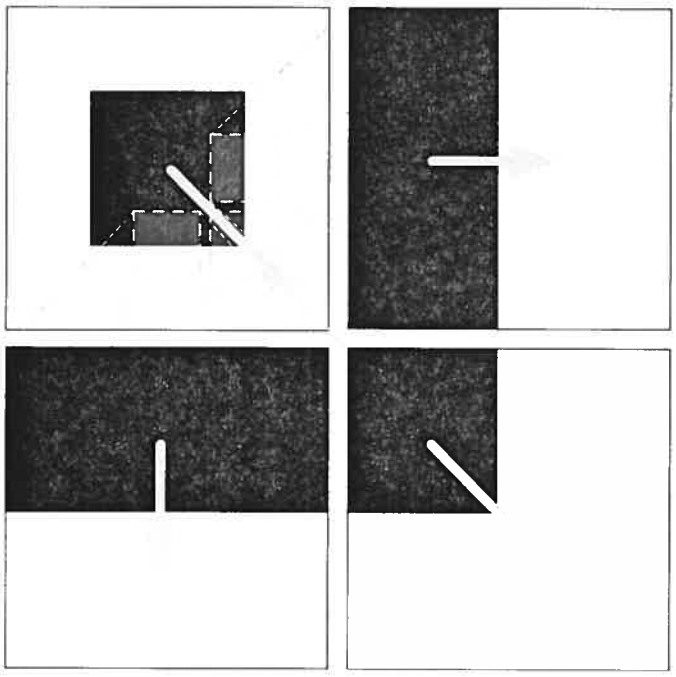


FIG. **1.7. Une texture ambiguë ne permet pas de déterminer le mouvement avec certitude.** Un carré noir se déplace en diagonal. Une petite fenêtre sur le côté droit perçoit un mouvement apparent (normal) horizontal. Une petite fenêtre sur le bas perçoit un mouvement apparent (normal) vertical. Seul les coins ont une texture assez bien définie pour résoudre le mouvement diagonal.

acceptable (e.g. changement d'exposition, zones d'ombres, réflexion, aliassage, ...)

- Le signal doit être linéaire. Entre deux pixels, pour des mouvements de moins de 1 pixel, le signal peut être interpolé linéairement. Cependant, lorsque les pixels se déplacent de plus de 1 pixel, le signal pourrait ne plus être linéaire, surtout en présence de hautes fréquences.

Des filtres passe-bas sont souvent utilisés pour linéariser le signal, mais ce filtrage se fait au détriment de la localisation. Jusqu'à récemment, ces méthodes ne pouvaient qu'être utilisées pour estimer des déplacements de moins de un pixel. Des approches hiérarchiques et par focus progressif (figure 1.10) permettent d'estimer précisément

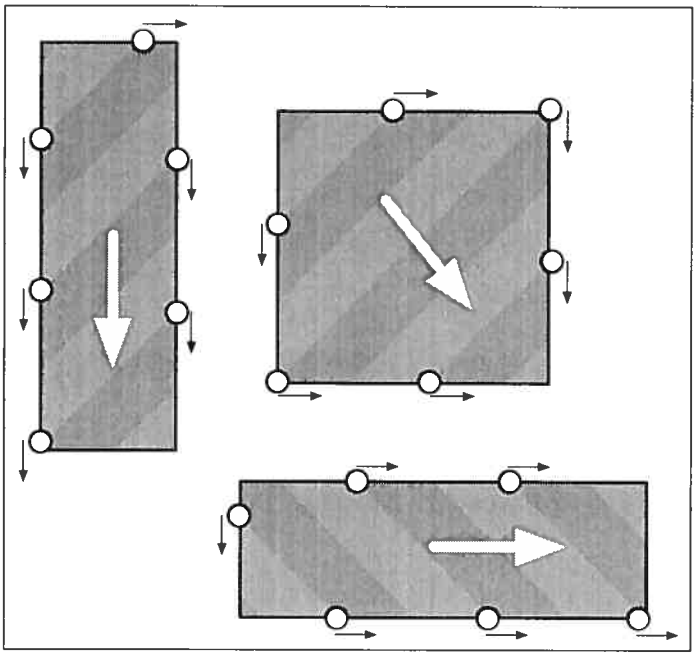


FIG. **1.8. L'illusion du barbier.** Une texture se déplaçant vers la droite mais qui est perçue à travers une fenêtre verticale, carrée ou horizontale n'aura pas le même mouvement apparent.

sur plusieurs pixels – chaque itération corrige pour la non linéarité du signal comme le ferait une méthode de Newton. Pour de meilleurs résultats. Brox, Bruhns, Papenberg et Weickert [27, 28] proposent de déformer par *warping* l'image à chaque niveau de la hiérarchie. L'idée de déformer l'image à partir de la carte de mouvement avait été proposée par Black afin d'identifier les mouvements multiples [29] et consiste à inverser la carte de flux optique, de créer une image déformée à partir de l'image au temps t et du mouvement estimé et de la comparer avec l'image au temps $t + 1$.

Pour plus de robustesse. quelques méthodes utilisent également les deuxièmes dérivées :

$$g'(x) = f'(x) + f''(x)v(x)$$

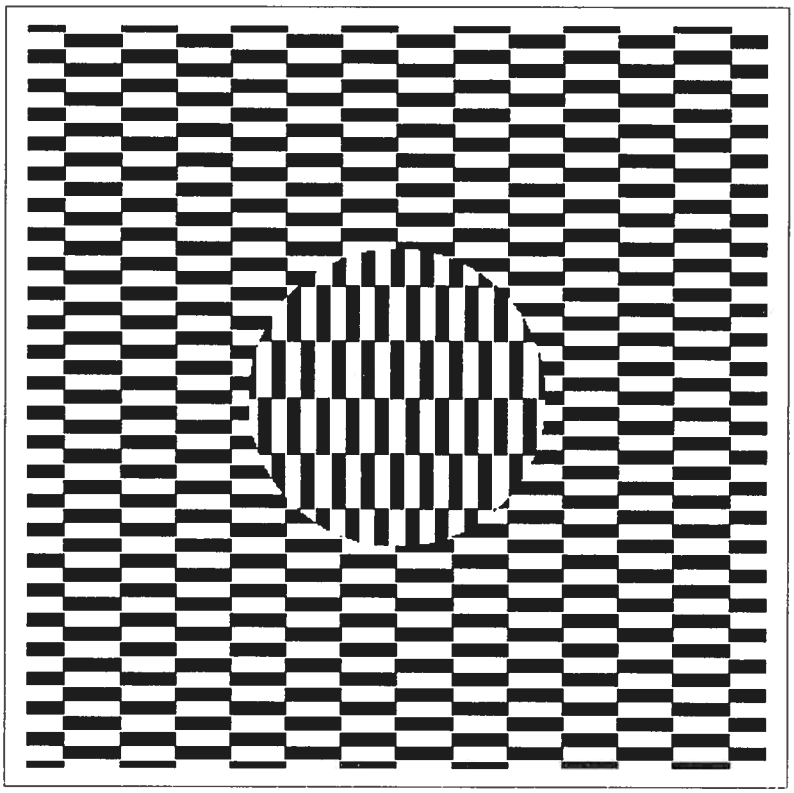


FIG. **1.9. L'illusion de Ouchi.** Inventée par l'artiste japonais Hajime Ouchi, l'illusion consiste de rectangles horizontaux contenant un cercle de rectangles verticaux. Un léger mouvement des yeux fait apparaître un mouvement apparent. L'image provient de [19].

et plutôt que de simplement considérer le gradient en x et y, l'expression devient

$$\begin{bmatrix} f_{x^2} & f_{yx} & f_{tx} \\ f_{xy} & f_{dy^2} & f_{ty} \end{bmatrix} v = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{1.5}$$

La contrainte d'intensité constante devient une contrainte de gradient constant ce qui rend cette approche robuste aux changements d'illumination, mais plus sensible aux changements d'orientation et d'échelle. De plus, à chaque dérivation, le bruit devient de plus en plus important. Ce genre d'approches est donc plus sensible au bruit.

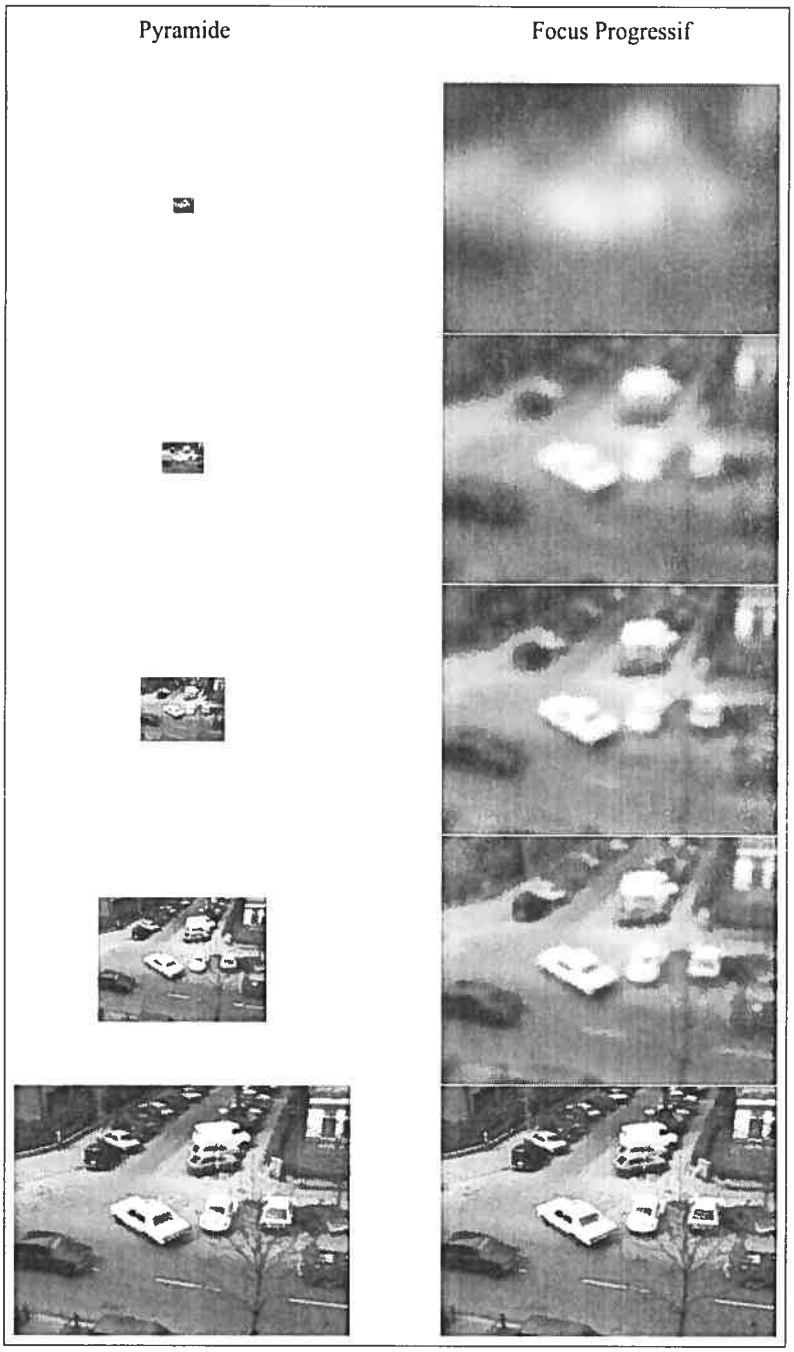


FIG. **1.10.** Une approche hiérarchique peut être implantée par pyramide (**à gauche**) ou focus progressif (**à droite**).

1.4 *Approches par phase*

Les méthodes par phase estiment le mouvement en observant le changement de phase d'un signal convolué avec un **filtre en quadrature** (typiquement $e^{-2\pi ix}$ ou une variante). Les filtres en quadrature sont une paire de filtres orthogonaux (le produit scalaire entre les deux est 0). L'angle formé par la réponse du signal à ces filtres indique la phase, alors que la norme donne l'énergie.

Si on représente le filtre en coordonnés polaires, on peut imaginer que le signal convolué $f(x)$ est enroulé autour d'un cercle dont la circonférence correspond à la longueur d'onde du filtre (figure 1.11). $f(x)$ correspond alors à la norme d'un vecteur orienté autours de ce cercle et la convolution consiste à additionner tous ces vecteurs. La phase est donc l'orientation du vecteur résultant (en rouge sur la figure) alors que l'énergie est sa norme.

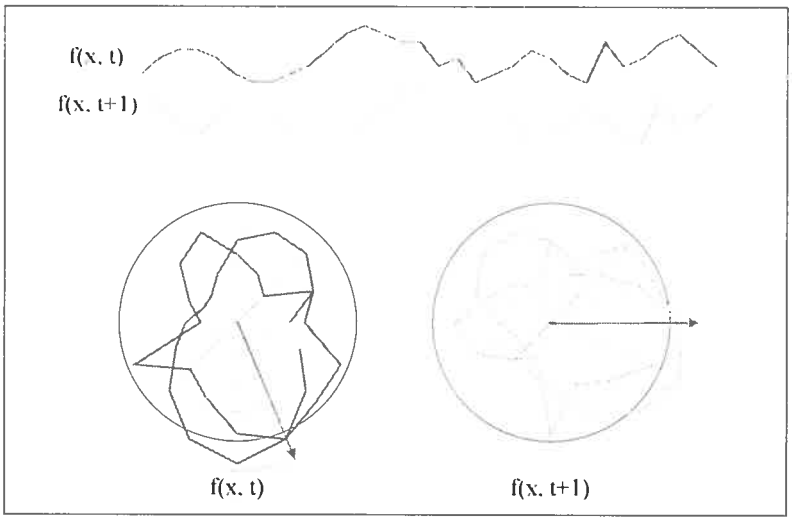


FIG. **1.11.** La convolution d'une fonction $f(x,t)$ avec un filtre en quadrature tel que $e^{-2\pi iwx}$ peut être vue comme l'enroulement de $f(x,t)$ autours d'un cercle de circonférence $1/w$. Une translation au temps $t+1$ introduit un changement de phase dans le vecteur résultant.

Pour la transformée de Fourier $\mathcal{F}$ d'un signal f, on peut démontrer que la trans-

lation d'un signal induit un changement de phase proportionnel à la fréquence et à la distance du déplacement :

$$\begin{aligned} \mathcal{F}(f(t+1),\omega) &= \textstyle\sum_x f(x-v,t)e^{-2\pi\imath\omega x} \\ &= \textstyle\sum_x f(x,t)e^{-2\pi\imath\omega(x+v)} \\ &= e^{-2\pi\imath\omega v}\textstyle\sum_x f(x,t)e^{-2\pi\imath\omega x} \\ &= e^{-2\pi\imath\omega v}\mathcal{F}(f(t),\omega). \end{aligned} \tag{1.6}$$

Pour une fréquence ω, un déplacement v introduit un déphasage

$$\Delta\phi_\omega = -2\pi\omega v \tag{1.7}$$

où $\Delta\phi_\omega$ peut être mesuré en multipliant la réponse du filtre au temps t avec le conjugué complexe (opérateur *) du filtre au temps $t+1$:

$$\Delta\phi_\omega = \arg\left(\mathcal{F}(f(t),\omega)\mathcal{F}^*(f(t+1),\omega)\right). \tag{1.8}$$

Combiner 1.8 et 1.7 permet de retrouver v. Idéalement, toutes les fréquences devraient suggérer le même déplacement. La **corrélation de phase**, introduite par De Castro et Morandi [30] est une façon simple de retrouver un mouvement dans un signal périodique. Le mouvement correspond alors à

$$\arg\max_x \left|\mathcal{F}^{-1}\left(\overline{\mathcal{F}}(f(t),\omega)\overline{\mathcal{F}^*}(f(t+1),\omega),x\right)\right|$$

ou $\overline{\mathcal{F}}$ est une transformée ou la norme de chacune des fréquences est normalisée à 1.

Plusieurs travaux ont étudié l'effet de transformation rigides sur le spectre d'une image [31, 32] et d'autres méthodes ont été développées afin de supporter également la rotation [30] et même le changement d'échelle [33].

Ces méthode fonctionnent si le signal est périodique ou si le support du filtre est assez large de sorte que la proportion de signal entrant et sortant au temps $t+1$ soit petite par rapport à la quantité de signal toujours présent. De plus, elles se limitent à des estimés de mouvements entiers. Comme les images ne sont généralement pas

périodiques, plus le mouvement est grand, moins la corrélation de phase est robuste. Figure 1.12 illustre bien ce problème : à mesure que l'image se déplace, le ratio de pixel qui sont présents dans l'image au temps t et au temps $t+1$ diminue et il devient de plus en plus difficile de discriminer le maximum des autres valeurs. On observe aussi que la méthode est très sensible au bruit. Ces observations sont vraies en général pour les méthodes par phase.

Dans le cas ou on s'intéresse au flux optique dense, une vitesse locale plutôt que globale doit être calculée et un fenêtrage doit donc être effectué afin de localiser la réponse. Lorsque la fenêtre devient petite, les effets de la non périodicité du signal deviennent plus importants. Pour palier à ce problème, le fenêtrage est souvent modulé par d'une gaussienne afin de donner moins d'importance aux bords. Les filtres deviennent alors des filtres de Gabor (figure 1.13) :

$$\mathcal{G}_{\omega,\sigma}(t) = \sum_x f(x,t)\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}e^{-2\pi i\omega x}.$$

Tout comme la transformé de Fourier, la réponse d'un signal convolué par des filtres de Gabor permet d'obtenir un spectre de fréquences. Contrairement à Fourier cependant, ce spectre est local et peut être calculé pour chaque point de l'image (autours duquel on centre la gaussienne).

L'introduction de la gaussienne permet de résoudre le problème de non-périodicité de f mais introduit un biais dans la réponse vers une phase de $\phi = 0$. En effet, pour un signal d'intensité constante $f(x) = 1$, un filtre de Gabor continu obtient une réponse de :

$$\int f(x)\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}e^{-2\pi i\omega x} = e^{-2\pi^2\sigma^2\omega^2}$$

Contrairement à la transformée de Fourier, qui aurait produit un Dirac à $\omega = 0$, un filtrage par Gabor aura des réponses pour toutes les fréquences. La gaussienne vient convoluer le spectre, distribuant ainsi la composante DC à travers ce dernier. Pour

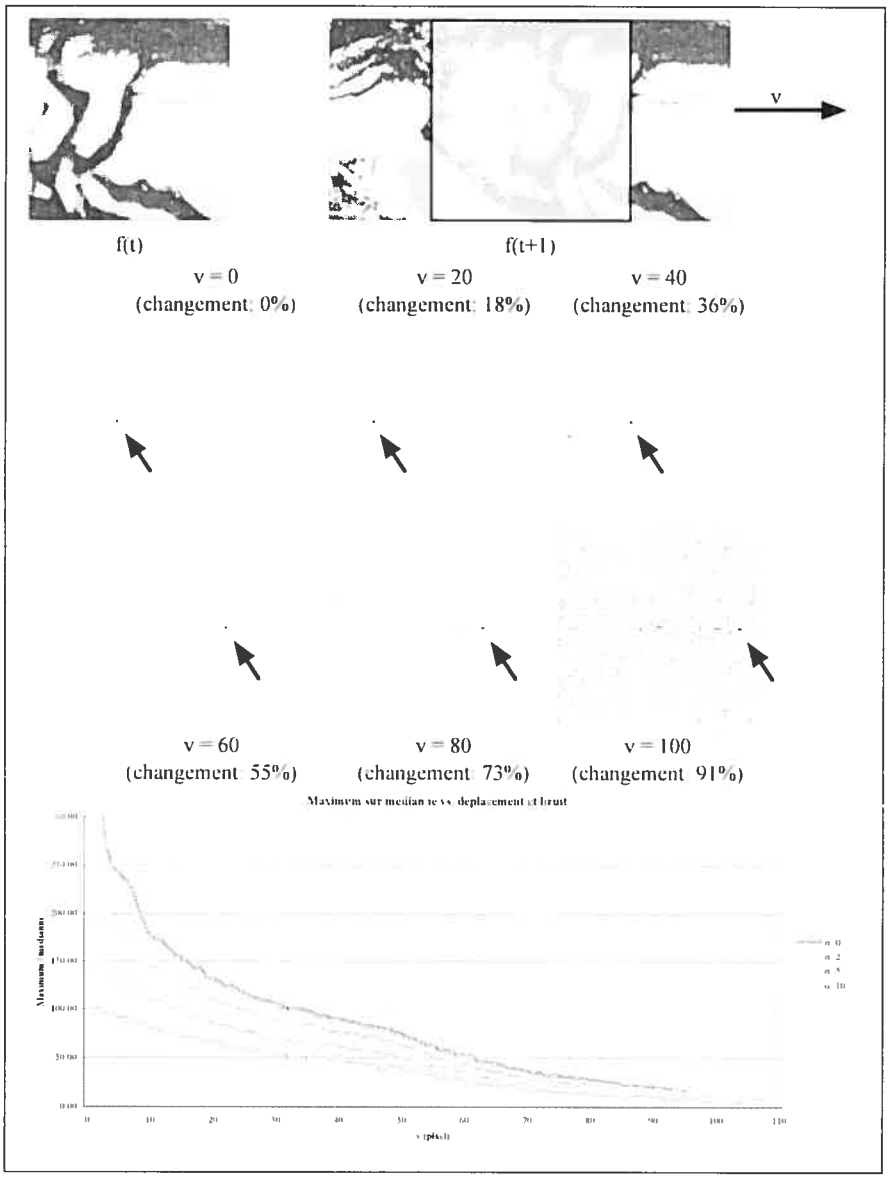


FIG. **1.12. Haut** : Corrélation de phase appliquée sur une image au temps t et au temps $t+1$ ayant subit une translation v. L'image fait 110x110 pixels. **Milieu** : Résultat de la corrélation de phases (les valeurs élevées sont en noir) pour diverses translations. La coordonnée du point maximum donne la translation estimée. **Bas** : Maximum sur la médiane pour diverses translations et bruit gaussien de diverses variances.

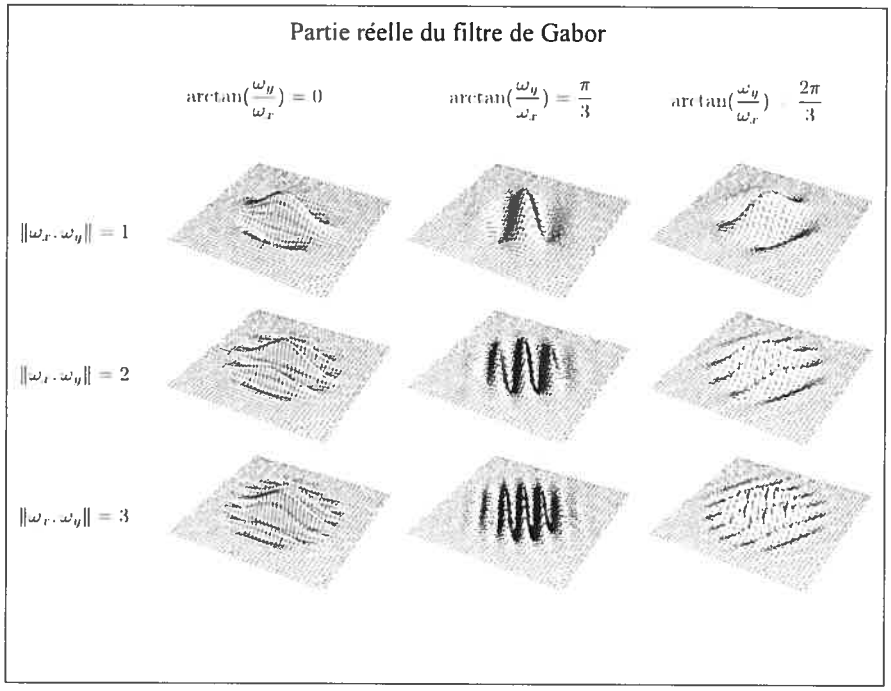


FIG. **1.13.** Partie réelle du filtre de gabor 2D pour diverses fréquences et orientations.

cette raison,

$$\mathcal{G}_{\omega,\sigma}(t+1) \neq e^{-2\pi\imath\omega v}\mathcal{G}_{\omega,\sigma}(t).$$

Cependant, on remarque que plus σ est grand, moins ce biais est important. On peut donc considérer

$$\mathcal{G}_{\omega,\sigma}(t+1) \approx e^{-2\pi\imath\omega v}\mathcal{G}_{\omega,\sigma}(t).$$

en se rappelant que le support σ ne devrait pas être trop petit et que la fréquence ω devrait être élevée. Dans ces conditions, la phase change de façon quasi linéaire avec le déplacement et on peut utiliser une approche semblable aux gradients discutés en §1.3. L'équation 1.3 devient

$$\phi_t(x,y) + v_x(x,y)\phi_x(x,y) + v_y(x,y)\phi_y(x,y) = 0 \tag{1.9}$$

et est appelée **contrainte de phase constante** (constant phase constraint). Les mêmes stratégies pour résoudre le mouvement s'appliquent donc, mais nous pouvons utiliser la réponse à plusieurs fréquences et orientations plutôt que d'utiliser le voisinage.

1.5 Approches par énergie

Les méthodes par énergie utilisent une propriété de la transformée de Fourier, où les lignes dans le domaine spatio-temporel deviennent des plans dans l'espace spectral.

Puisque Fourier est séparable, commençons par effectuer la transformée sur chacun des plans 2D (x, y). Si un plan au temps $t+1$ contient une translation de la même texture qu'au temps t, leur spectre aura la même énergie mais une phase différente (tel que démontré dans l'équation 1.6). En 2D, on obtient :

$$\mathcal{F}_{\omega_x\omega_y}(t+1) = e^{-2\pi\imath(\omega_x v_x + \omega_y v_y)}\mathcal{F}_{\omega_x\omega_y}(t)$$

S'il n'y avait pas eu de translation et que les plans avaient été identiques, la transformée Fourier 1D restante (en t) produirait un plan de Diracs sur la fréquence $\omega_t = 0$, ce que nous appelons un plan de mouvement en $\omega_t = 0$. Par contre, la différence dans la phase résulte en une translation de ces Diracs (l'inverse de ce qui a été démontré en 1.6) :

$$\sum_{\omega_y}\sum_{\omega_x} e^{-2\pi\imath(\omega_x v_x + \omega_y v_y)}\mathcal{F}_{\omega_x\omega_y}(t) = F(t - (\omega_x v_x + \omega_y v_y))$$

Autrement dit, le plan de Diracs se trouve maintenant à

$$v_x\omega_x + v_y\omega_y - \omega_t = 0. \tag{1.10}$$

La normale de ce plan, $(v_x, v_y, 1)$ permet de retrouver la translation 2D du signal (figure 1.14). Il est intéressant d'observer que l'équation 1.10 est quasiment identique à l'équation 1.3.

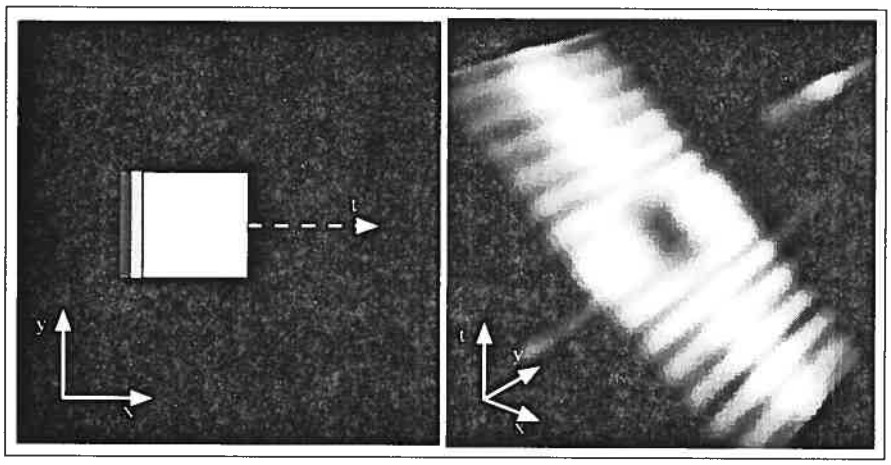


FIG. **1.14.** La translation d'une texture (**à gauche**) génère un plan dans l'espace fréquentiel $(\omega_x, \omega_y, \omega_t)$ (**à droite**) avec les paramètres du mouvement $\langle v_x, v_y, v_t \rangle$ comme normale (généralement, $v_t = 1$).

Tout comme les méthodes par phase, cette approche est sensible à la non périodicité d'un signal. Pour cette raison, un fenêtrage gaussien est généralement effectué (spatialement et temporellement).

Plusieurs méthodes existent pour retrouver le plan et sa normale. Heeger [34] propose d'utiliser des filtres distribués dans le spectre. La réponse de ces filtres permet de retrouver analytiquement la position du plan. D'autres méthodes de paramétrisation ont été proposées, certaines permettant aussi de retrouver une distribution de plans plutôt qu'un seul plan. Mann et Langer [24] récupèrent une paramétrisation de plans afin de déduire le mouvement de caméra (*egomotion*) dans des scènes contenant le parallaxe d'une multitude d'objets à diverses profondeurs. Ils réfèrent à ce type de scène comme de la «neige optique». La méthode en §4 permet de retrouver une distribution de ces plans de façon non paramétrique.

Le problème d'ouverture discuté en §1.3 est également présent dans les méthodes par énergie. L'explication donnée au début de cette section suppose qu'il existe une réponse non-nulle pour presque chaque fréquence des images 2D. Par contre, dans

le cas d'une texture ambiguë, par exemple avec une texture purement horizontale, la transformée 2D de ce signal donnera une ligne dans l'espace fréquentiel. Cette ligne restera une ligne de Diracs plutôt qu'un plan après la transformée de Fourier temporelle. Cette ligne ne permet pas de résoudre un plan unique.

1.6 Approches par corrélation

Les méthodes par corrélation, aussi dites par région, essaient de résoudre $f(x, y) = g(x + v_x, y + v_y)$ en choisissant la translation qui maximise la corrélation pondérée des signaux :

$$\mathcal{C}(v_x, v_y) = \sum W(v_i, v_j)\mathcal{D}\left(f(i, j), g(i + v_x, j + v_y)\right)$$

où W est une fonction de fenêtrage et $\mathcal{D}$ est une fonction qui compare f et g. Une corrélation au sens statistique peut être calculée, mais on utilise généralement une somme de différence au carré (*sum of squared differences* – *SSD*) ou encore une somme de différences absolues (*sum of absolute differences* – *SAD*). Puisque ces méthodes sont potentiellement coûteuses en effort de calcul, elle sont souvent couplées avec une approche hiérarchique. Plusieurs auteurs préfèrent également travailler sur une corrélation des laplaciens ou autres prétraitements passe-bande afin d'être résistant aux changements d'illumination et donner plus d'importance aux bordures.

À moins de tester le sous-pixel explicitement, cette approche fonctionne pour des valeurs entières de déplacements $\mathbf{v}$ et une étape additionnelle est nécessaire pour estimer le sous-pixel. Plusieurs méthodes existent, par exemple, par modélisation de la surface (linéaire ou quadratique) et maximisation de la SSD sur cette surface.

1.7 Méthodes multicontraintes

Des méthodes multicontraintes ont émergé vers la fin des années 1980, mais les résultats n'étaient pas d'assez bonne qualité pour qu'elles deviennent populaires [35].

Ces méthodes consistent à utiliser plusieurs contraintes – par exemple la contrainte d'intensité constante - afin d'obtenir un terme d'information plus robuste :

$$f_x^i v_x + f_y^i v_y + f_t^i = 0, \quad i = 1, \cdots, n.$$

Parmi les fonctions proposées : les dérivées directionnelles, le contraste, l'entropie, la moyenne, etc. [36]. En ce sens, les méthodes fréquentielles (phase et énergie) sont également des méthodes multicontraintes puisqu'elles résolvent simultanément pour plusieurs fréquences et orientations. Les SIFTs (*Scale Invariant Feature Transform*) [37] sont aussi en quelque sorte une méthode multicontrainte ; nous en discutons en §5.2.

Ces méthodes ont récemment refait surface [38, 27]. Brox *et al* améliorent leur résultats en combinant la constance d'intensité et la constance du gradient et la méthode présentée en §3 utilise aussi une approche multicontrainte.

Chapitre 2

SURVOL DES MÉTHODES RÉCENTES

Le problème étant fondamental à la vision, une multitude d'algorithmes de flux optique ont été présentés dans les dernières années. Nous proposons de survoler quelques uns des principaux articles publiés au cours des 25 dernières années. L'estimation du mouvement a beaucoup de ramifications, par exemple, la reconstruction de structure, segmentation, le tracking et les features, la stéréoscopie, ... Nous ne pouvons évidemment pas couvrir tout ce qui à été écrit en lien avec l'estimation de mouvement, voilà pourquoi nous nous concentrons sur le terme de conservation d'information pour du mouvement 2D et les terme de lissage.

1981 : Determining optical flow [3]

L'approche de Horn et Schunck est l'une des plus référée dans la littérature. Il s'agit d'une méthode par gradient tel que présenté en §1.3. La résolution de la contrainte d'intensité constante (equation 1.3) se fait globalement avec une minimisation itérative d'énergie sur une region D :

$$\int_D (\nabla I \cdot v + I_t)^2 + \lambda^2(\|\nabla u\|^2 + \|\nabla v\|^2)dx \tag{2.1}$$

où ∇I est le gradient spatial, I_t est le gradient temporel, ∇u et ∇v sont les gradients des vecteurs de déplacement horizontaux et verticaux et où λ pondère le lissage de la solution. Cette méthode permet de générer des flots denses : là ou l'information est ambiguë, le terme de lissage modulé par λ, appelé terme de régularisation, permet de désambigüer avec le voisinage. Plusieurs améliorations ont été proposées par la suite, notamment une régularisation temporelle et une régularisation qui tient compte du contraste afin de permettre des discontinuité de mouvement. Par exemple, Yachida

publit un article la même année intitulé «*Determining velocity map by 3-d iterative estimation*» [5] qui ajoute un terme de régularisation temporelle à la méthode de Horn et Schunck.

1981 : An iterative image registration technique with an application to stereo vision [4]

Lucas et Kanade [4] proposent également une méthode par gradient mais, cette fois, avec une résolution locale. Un modèle de mouvement constant est utilisé sur un voisinage Ω pour résoudre l'équation 1.3 *via* une minimisation de distances au carré avec pondération en fonction de la distance des voisins :

$$\sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x})[\nabla I(x,t)\cdot\mathbf{v} + I_t(\mathbf{x},t)]^2$$

Ici, W est une fonction de fenêtrage. Comme dans le cas de Horn et Schunck, cette méthode a été plusieurs fois reprise et modifiée [18]. Elle ne garantit pas un flot dense comme Horn et Schunck, mais résiste mieux au bruit [18, 39].

1983 : Displacement vectors derived from second-order intensity variations in image sequences [9]

Nagel propose un modèle semblable à Horn and Schunck (minimisation globale d'energie) qui utilise les dérivées secondes. Il développe un opérateur permettant de détecter les zones où la texture permet une estimation de mouvement sans ambiguïté (détecteur de coins) et utilise le flux de ces zones comme estimé initial du mouvement. De plus, il modifie le terme de régularisation en fonction de la norme du gradient et de son orientation afin de permettre une meilleur tolérance aux discontinuités de mouvement. La fonctionnelle à minimiser devient :

$$\int\int (\nabla I_T\mathbf{v} + I_t)^2 + \frac{\alpha^2}{\|\nabla I\|^2 + 2\delta}[(u_xI_y - u_yI_x)^2 + (v_xI_y - v_yI_x)^2 + \delta(u_x^2 + u_y^2 + v_x^2 + v_y^2)]dxdy.$$

Dans son article de 1987, Nagel met en relation son propre modèle en relation avec ceux de Horn et Schunck [3], Haralick et Lee [40], Tretiak et Pastor [41], de Hildreth [42].

1983 : The facet approach to optic flow [40]

Haralick et Lee ajoutent à la CBC (equation 1.3) une contrainte sur le gradient qui doit, lui aussi, rester constant :

$$\begin{bmatrix} f_x & f_y & f_t \\ f_{xx} & f_{xy} & f_{xt} \\ f_{yx} & f_{yy} & f_{yt} \\ f_{tx} & f_{ty} & f_{tt} \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

obtenant de cette façon un système surdéterminé.

1980/1985 : "Combining motion and contrast for segmentation" [43] et "Dynamic occlusion analysis in optical flow fields" [44]

Thompson propose de segmenter en utilisant le flux optique et le contraste de l'image, en faisant grandir des régions correspondants aux surfaces en mouvement. Dans son article de 1985, Thompson *et al.* proposent de segmenter les zones de mouvement en combinant le flux optique et le contrast passages par zéro par filtrage Marr-Hildreth.

1986 : Scaling theorems for zero crossings [45]

Yuille et Poggio démontrent que le bruit est amplifié dans les dérivées de plus haut niveaux. Ce problème décourage donc l'utilisation de dérivées de plus haut degrés pour les algorithmes par dérivées spatio-temporels. Dans le même ordre d'idée, en 1987 Kearney *et al.* démontrent dans leur article « *Optical flow estimation : an error*

analysis of gradient-based methods with local optimization » [46] l'importance de filtrer les images par un filtre passe-bas afin d'augmenter la stabilité des gradients.

1987 : Computations underlying the measurement of visual motion [42]

Hildreth propose d'utiliser les contours afin de calculer le mouvement. Les contours sont obtenus par un filtrage gaussien suivit du laplacien (*Laplacian of Gaussian - LOG*). Le mouvement est mesuré à partir des zones de passage par zéro (*zero crossing*). Initialement, le mouvement est pris comme étant perpendiculaire aux contours, mais un terme de régularisation assure la propagation de l'information et un flux lisse le long des contours :

$$\sum_i \|\mathbf{v}_i\|^2 + \beta(\mathbf{v}_i \cdot \mathbf{u}^n - \mathbf{v}_i^n)$$

où $\mathbf{u}^n$ est un vecteur unitaire dans la direction perpendiculaire au contour et $\mathbf{v}^n$ est le déplacement mesuré perpendiculairement au contour.

1987 : Scene segmentation from visial motion using global optimization [47]

Murray et Buxton proposent de segmenter le mouvement avec une approche bayesienne afin d'identifier les régions avec un déplacement propre. Comme ils utilisent également une contrainte temporelle, ils représentent le flux optique comme un champ Markovien temporel et résolvent par recuit simulé.

1988 : Optical flow using spatiotemporal filters [34]

Heeger présente une approche par énergie. Des plans d'énergie sont retrouvés dans le domaine spectral à l'aide de filtres de Gabors avec un support stratégique de sorte qu'une solution analytique puisse être calculée à partir de chacune de leur réponse.

1988 : On the Computation of Motion from Sequences of Images – A Review [35]

Aggarwal et Nandhakumar présentent une revue du progrès en estimation de mouvement entre 1979 et 1988. En plus de discuter des approches par de flux optique, ils présentent plusieurs approches par points d'intérêt ainsi que les méthodes de flux optique et de stéréoscopie orientée vers la reconstruction 3D (*structure from motion*).

1988 : Computational approach to motion perception [48]

Uras *et al.* présentent un méthode de gradient second degré, où on calcule la Hessienne pour un bloc 8×8 pour résoudre l'équation 1.5 en page 14. On prend les 8 meilleurs candidats du bloc ($\|M\nabla I\| \ll \|\nabla I_t\|$ où $M \equiv (\nabla v)^T$, et la solution est utilisée pour les 64 pixels.

1988 : Ill-posed problems in early vision [2]

Bertero, Poggio et Torre redéfinissent le problème de flux optique comme étant mal posé au sense de Hadamar [1]. Ils expliquent que la dérivation d'un signal discret est proscrit puisque que l'ajout d'une faible bruit de haute fréquence serait amplifié par une dérivation. Le flux optique est également mal-posé à cause du problème d'ouverture. Ils justifient donc l'utilisation du terme régularisateur utilisé par Horn et Schunck et Hilberth.

1989 : Regularization of discontinuous flow fields [49]

Shulmann et Hervé développent une version de Horn et Schunck offrant une meilleur tolérance aux discontinuités et changements d'illumination. Suite aux récents travaux de Blake et Zimmerman au MIT en 1987, à la thèse de Marroquin [50], des travaux de Mumford et Shah [51] et de Geman et Geman [52], ils introduisent un seuil T à la fonctionnelle à minimiser, permettant une gestion différente du lissage lorsque le premier et second dérivées suggère une discontinuité dans l'image. Le terme

de régularisation de Horn et Schunck (equation 2.1) devient :

$$\lambda \left(g_{T_1}(\|\nabla\|_2) + g_{T_2}(\|\nabla_2\|_2) \right)$$

où

$$\begin{aligned} g_{T_\epsilon}(x) &= x^2 && \text{pour x} \leq \text{T}, \\ &= T^2 + 2T|x - T| + \epsilon(x - T)^2 && \text{pour x} \geq \text{T}. \end{aligned}$$

Ici, ϵ assure que la fonctionnelle soit strictement convexe. En pratique, Shulman et Hervé mentionnent que ce terme peut être négligé (ou avoir une très petite valeur positive) puisque la fonctionnelle reste convexe sans ϵ.

Ils permettent également un legère relaxation de la constante d'intensité constante en y ajoutant un terme α. L'équation 1.3 devient

$$f_t + v_x f_x + v_y f_y + \alpha f = 0.$$

1989 : A computational framework and an algorithm for the measurement of visual motion [53]

Anadan propose une méthode hiérarchique par corrélation. Le laplacien est utilisée afin de donner une plus grande importance aux bords. Au premier niveau de la pyramide, le mouvement est comparé à un voisinage de 3×3 ($-1, 0, +1$ horizontalement et verticalement) pour une région W de 5×5 modulé par une gaussienne. Le meilleur déplacement est trouvé par une minimization de SSD. Un estimé du sous-pixel est ensuite obtenu en modélisant une surface quadratique des SSD autours du mouvement entier trouvé et en trouvant le minimum. Un lissage global *via* une méthode itérative est ajouté afin de minimiser $\nabla \mathbf{v}$.

1990 : An estimation-theoretic framework for image-flow computation [11]

Singh propose une méthode par corrélation des laplaciens et introduit une covariance à ses réponses locales. Le mouvement est une distribution de probabilité avec

une moyenne

$$\overline{\mathbf{v}} = \frac{\sum_{\mathbf{d}\in\Omega} \mathcal{R}(\mathbf{d})\mathbf{d}}{\sum_{\mathbf{d}\in\Omega} \mathcal{R}(\mathbf{d})}$$

où $\mathcal{R}$ est une probabilité de déplacement basé sur la SSD entre $\mathbf{x}$ et un déplacement $\mathbf{d}$:

$$\mathcal{R}(\mathbf{d}) = e^{-k\mathrm{SSD}(\mathbf{x},\mathbf{x}+\mathbf{d})}.$$

La covariance de la distribution est obtenue de façon similaire :

$$\Sigma = \frac{1}{\sum \mathcal{R}(\mathbf{d})} \begin{pmatrix} \sum \mathcal{R}(\mathbf{d})(d_x - \overline{v}_x)^2 & \sum \mathcal{R}(\mathbf{d})(d_x - \overline{v}_x)(d_y - \overline{v}_y) \\ \sum \mathcal{R}(\mathbf{d})(d_y - \overline{v}_y)(d_x - \overline{v}_x) & \sum \mathcal{R}(\mathbf{d})(d_y - \overline{v}_y)^2 \end{pmatrix}$$

Un processus itératif tente par la suite de minimiser

$$\int (\mathbf{v} - \overline{\mathbf{v}}_\mathbf{n})^T \Sigma_n^{-1} (\mathbf{v} - \overline{\mathbf{v}}_\mathbf{n}) + (\mathbf{v} - \overline{\mathbf{v}}_\mathbf{0})^T \Sigma_0^{-1} (\mathbf{v} - \overline{\mathbf{v}}_\mathbf{0}) d\mathbf{x}$$

où $\overline{\mathbf{v}}_\mathbf{n}$ et Σ_n sont les moyennes et covariances des voisins alors que $\overline{\mathbf{v}}_\mathbf{0}$ et Σ_0 est la moyenne et covariance à $\mathbf{x}$.

Conscient qu'une distribution gaussienne n'est qu'une approximation de la distribution de SSD, Singh propose également d'utiliser trois images plutôt que deux afin d'éviter les distributions multimodales. Les SSD sont modifiées comme suit :

$$SSD_0(\mathbf{x}, \mathbf{d}) = SSD_{0,1}(\mathbf{x}, \mathbf{d}) + SSD_{0,-1}(\mathbf{x}, -\mathbf{d}).$$

1990 : Computation of component image velocity from local phase information [54]

Fleet et Jepson proposent une méthode basée sur la phase plutôt que sur l'intensité. La phase est obtenue par convolution avec un filtre Gabor spatio-temporel orienté spatialement :

$$\mathcal{G}(\mathbf{x}, t; \mathbf{k}_0, \omega_0, \Sigma) = \mathcal{N}(\mathbf{x}, t; \Sigma) e^{-2\pi\imath(\mathbf{x},t)\cdot(\mathbf{k}_0,\omega_0)}$$

où $\mathbf{k}_0$, ω_0 sont l'orientation et la fréquence du filtre et où $\mathcal{N}$ est une gaussienne avec covariance Σ. Seulement la phase de la réponse complexe du filtre est utilisée, car la norme tolère mal les changements de contraste. Une méthode similaire à celle des gradients est par la suite utilisée afin de retrouver le déplacement qui préserve la phase (tel qu'à l'équation 1.9 à la page 20). La longueur $|\mathbf{v}|$ et l'orientation de ce déplacement se retrouvent par flot normal à la fréquence utilisée $(\mathbf{k}_0, \omega_0)$ (semblable à l'équation 1.4 à la page 10) :

$$||\mathbf{v}|| = \tilde{v} = \frac{\phi_t}{\|<\phi_x, \phi_y>\|}, \quad \arg \mathbf{v} = \arctan \frac{\tilde{n}_y}{\tilde{n}_x}, \quad \tilde{\mathbf{n}} = \frac{<\phi_x, \phi_y>}{\|<\phi_x, \phi_y>\|}$$

Comme la phase devrait varier de façon linéaire, on peut résoudre $\mathbf{v}$ par un système similaire aux méthodes par gradients. Fleet et Jepson suggèrent de miminiser un système d'équations en utilisant un voisinage de 5×5.

Fleet et Jepson introduisent également une mesure d'erreur, qui sera reprise et popularisée par l'article de Barron *et al.* [18]. Cette mesure dite, d'**erreur angulaire** (*angular error*) consiste à mettre le vecteur 2D $\langle x, y \rangle$ en 3D $\langle x, y, 1 \rangle$ et à comparer les angles avec un vecteur $\tilde{\mathbf{v}}$ de référence :

$$\Psi = \arccos \left[\frac{< \mathbf{v}, 1 > \cdot < \tilde{\mathbf{v}}, 1 >}{\sqrt{1 + \|\mathbf{v}\|^2} \sqrt{1 + \|\tilde{\mathbf{v}}\|^2}} \right] \tag{2.2}$$

1990 : Computing two motions from three frames [21]

Bergen *et al.* proposent un modèle simple pour retrouver deux mouvements simultanés. Ils identifient quelques cas impliquant des mouvements multiples :

- Deux objets séparés par une bordure
- Deux objets avec surface transparente, ombre ou réflection
- Deux objets entrecroisés
- Un mouvement dominant avec un objet à faible contraste ou petit avec mouvement indépendant (e.g balle partiellement suivie par une caméra)

Supposant qu'il y a deux mouvements p (prédominant) et q affectant deux zones P et Q possiblement en superposition, ils proposent une méthode itérative pour résoudre

ces deux mouvements à partir de trois images f_1, f_2 et f_3. Si $p_{1,2}$ et $p_{2,3}$ sont connus on peut obtenir deux images de différence $D_{1,2}$ et $D_{2,3}$ qui représentent l'erreur entre la prédiction f^p et l'image f réelle :

$$D_{1,2} \equiv f_2 - f_1^{p_{1,2}}$$
$$D_{2,3} \equiv f_3 - f_2^{p_{2,3}}$$

alors, on peut retrouver les régions P et Q ainsi que q à partir de ces deux images de différence. En pratique, p n'est pas connu, alors on peut l'estimer et utiliser une approche itérative :

1. Estimer p_0 avec une méthode traditionnelle
2. Générer $D_{1,2}$ et $D_{2,3}$ en utilisant p_n
3. Estimer q_{n+1} en appliquant une méthode de flux optique traditionnelle entre D_1 et D_2.
4. Générer $D_{1,2}$ et $D_{2,3}$ en utilisant q_{n+1}
5. Estimer p_{n+2} à partir de $D_{1,2}$ et $D_{2,3}$
6. Répéter à partir de l'étape 2

Bergen *et al.* rapportent une convergence rapide, de trois à cinq itérations. La méthode, fort simple et élégante, se limite à deux mouvements et un opérateur $\oplus$ qui doit être connu. L'idée que des mouvements multiples puissent être découpés en une superposition de régions est reprise par Wang *et al.* [55] qui propose un modèle par couche pour les mouvements multiples.

1991 : Probability distribution of optical flow [12]

En 1991, Simoncelli *et al.* [12] présentent un modèle probabiliste à l'estimation du mouvement. Le modèle présenté permet des imprécisions lors du calculs du mouvement, tenant compte du bruit de l'image et du gradient. Il est alors possible d'obtenir

une distribution de probabilité pour le mouvement avec covariance Λ. Pour une image contenant un bruit gaussien, à mesure que le contraste augmente, la précision de l'estimé du mouvement devrait également augmenter. Ce modèle permet non seulement d'adapter la forme du voisinage utilisé dans la méthode de Lucas et Kanade en fonction d'un seuil de probabilité désiré, mais ouvre également la voie à des méthodes de minimisation de type markoviennes.

Les résultats expérimentaux de Barron *et al.* ont cependant démontré que l'utilisation de valeurs propres du système linéaire à résoudre donnent une meilleure indication de la qualité de l'estimé.

1991 : Robust dynamic motion estimation over time [6]

Pour les méthodes par corrélation, Black et Anadan expliquent qu'une simple SSD a tendance à amplifier le bruit. Plutôt que d'utiliser x^2, ils proposent d'utiliser $\psi(x^2) = \frac{1}{1+x^2/\lambda^2}$ à la fois pour la comparaison de régions, mais également pour le lissage, lors de la comparaison avec les vitesses voisines. Ils ajoutent une notion temporelle à leur modèle en prédisant les vitesses au temps t à partir des vitesses au temps $t-1$ et en tenant compte d'une accélération $\Delta\mathbf{v}$ calculée par moyennage avec les images précédentes. Un lissage temporel est effectué avec la fonction de coût présentée. Enfin, ils proposent également d'utiliser une surface bicubique pour estimer le sous-pixel. Le tout est résolu avec un modèle probabiliste en utilisant une distribution de Gibbs et un recuit simulé continu [56] modifié afin de pouvoir réutiliser l'état à $t-1$.

1992 : A locally adaptive window for signal matching [20]

Okutomi et Kanade analysent le comportement des réponses par minimisation de carré lorsqu'il existe plusieurs mouvements dans la fenêtre de résolution. Plus la taille de la fenêtre est grande, plus il y a de chance d'y retrouver plusieurs mouvements et ainsi de trouver un minimum déplacé, mais si la fenêtre est petite, le ratio signal-bruit

diminue et la solution devient instable. Ils proposent donc une méthode ou la taille de la fenêtre est variable.

1992 : Hierarchical model-based motion estimation [8]

Bergen *et al.* proposent une approche hiérarchique pouvant s'appliquer à plusieurs modèles de mouvement. Les modèles présentés sont :

Modèle affine : Le mouvement peut être décrit comme une simple transformation affine 2D

$$\mathbf{v}(x,y) = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Modèle planaire : Le mouvement peut être modélisé par un plan. Il faut 8 paramètres pour décrire les mouvements possibles :

$$\mathbf{v(x)} = \frac{1}{Z(\mathbf{x})}\mathbf{A(x)t} + \mathbf{B(x)}\omega$$

ou $\mathbf{Z}$ est la profondeur, $\mathbf{t}$ et ω sont les matrices de translation et de rotation du plan, et $\mathbf{A}$ et $\mathbf{B}$ sont :

$$\mathbf{A(x)} = \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix}, \qquad \mathbf{B(x)} = \begin{bmatrix} \frac{xy}{f} & \frac{-(f+x^2)}{f} & y \\ \frac{(f+y^2)}{f} & -\frac{xy}{f} & -x \end{bmatrix}. \tag{2.3}$$

Si on prend $\mathbf{r} = (\frac{x}{f}, \frac{y}{f}, 1)$, on peut substituer

$$\frac{1}{\mathbf{Z(x)}} = \mathbf{r(x)}^T\mathbf{k}$$

où $\mathbf{k}$ correspond à la normal du plan normalisé tel que $k_1X + k_2Y + k_3Z = 1$. L'équation 2.3 devient

$$\mathbf{v(x)} = (\mathbf{A(x)t})(\mathbf{r(x)}^T\mathbf{k}) + \mathbf{B(x)}\omega$$

Modèle de corps rigide : Semblable au modèle planaire, ce modèle permet plusieurs plans dans la même image.

Modèle non contraint : Aucune paramétrisation, mais les discontinuités de flux sont pénalisées en supposant que $\nabla \mathbf{v}$ devrait être petit dans un petit voisinage.

1993 : Multimodal estimation of discontinuous optical flow using Markov random fields [57]

Heitz et Bouthemy expliquent que le problème d'ouverture rend impossible un flux optique purement local, mais qu'une minimisation globale résoud les ambiguïtés sans tenir compte des discontinuités de mouvement. Le mouvement devrait être traité comme continu par morceau (*piecewise continuous*). Ils proposent une méthode dans le même ordre d'idée que ce qui avant été proposé par Nagel [9] et Hildreth [42]. Ils utilisent le déplacement des contours pour contraindre le mouvement. La région du côté où le mouvement est le plus compatible avec le déplacement du contour est considérée comme cachante et est contrainte par son contour. La région cachée n'est cependant pas contrainte. Ils utilisent une formulation bayesienne qui contient cinq contraintes :

- une contrainte d'intensité constante (utilisée en zone lisse)
- une contrainte de contour et un lissage (utilisée en zone de contour)
- une contrainte de lissage (qui ne traverse pas les contours)
- une contrainte qui gère la détection de contours
- une contrainte qui gère la géométrie des contours

Le tout est présenté comme un problème markovien et résolut par «Mode conditionel itéré» (*Iterated Conditional Modes*) [58], une alternative déterministe au recuit simulé.

1993 : Mixture models for optical flow computation, in Partitioning Data Sets [22]

Lorsque plusieurs mouvements sont présents dans un voisinage (pour cause d'occlusion ou de transparance), les contraintes d'information convergent vers plus d'une réponse. Une minimisation de carré dans une telle situation donne une solution qui se situe souvent entre deux solutions possibles, mais qui n'en est pas une en soit. Jepson et Black proposent une approche Espérance-Maximisation (EM) [59] pour déterminer l'appartenance des contraintes à un mouvement et ainsi résoudre pour des mouvements multiples.

1994 : Performance of optical flow techniques [18]

L'article de Barron, Fleet et Beauchemin marque un point marquant dans le développement des méthodes d'estimation de mouvement. En plus de recenser les principales méthodes de flux optique de l'époque, l'article de Barron, Fleet et Beauchemin parut en 1994 [18] présente pour la première fois une comparaison de plusieurs algorithmes en utilisant une métrique bien définie. Les méthodes présentées sont :

- Horn et Schunck [3]
- Lucas et Kanade [4]
- Nagel [9]
- Uras, Girosi, Verri et Torre [48]
- Anandan [53]
- Singh [11]
- Heeger [34]
- Waxman, Wu et Bergholm [60]
- Fleet et Jepson [54]

La contribution de cet article est incontestable car elle permet une comparaison raisonnable entre les différentes approches. La métrique d'erreur angulaire (eq. 2.2) avait déjà été présentée par Fleet [54] mais a été popularisée par l'article de Barron

et al.

1994 : Segmentation of visual motion by minimizing convex non-quadratic functionals [61]

Schnörr présente une méthode qui essaie de généraliser Horn et Schunck et introduit une détection de contours dans le terme de régularisation spatiale afin de mieux traiter les discontinuités de mouvement. Il obtient une fonctionnelle à minimiser :

$$\int_\Omega |(A^T A + \alpha I)^{-1} A^T (A\mathbf{v} - b)|^2 + \lambda \|\nabla \mathbf{v}\|^2 dx$$

où A et b contiennent les contraintes

$$A = \begin{bmatrix} f_x^0 & f_y^0 \\ f_x^1 & f_y^0 \\ \cdots & \\ f_x^n & f_y^n \end{bmatrix}, \quad b = \begin{bmatrix} f_t^0 \\ f_t^1 \\ \cdots \\ f_t^n \end{bmatrix}.$$

Il propose aussi de remplacer $\nabla \mathbf{x}$ par

$$\begin{aligned} &\frac{1}{2}\lambda(\mathrm{div}(\mathbf{v}) + \mathrm{rot}(\mathbf{v}) + \mathrm{sh}(\mathbf{v})) \\ \mathrm{div}(\mathbf{v}) &= \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} \\ \mathrm{rot}(\mathbf{v}) &= \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \\ \mathrm{sh}(\mathbf{v}) &= \begin{bmatrix} \frac{\partial v_y}{\partial y} - \frac{\partial v_x}{\partial x} \\ \frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \end{bmatrix} \end{aligned}$$

1994 : Real-Time Optical Flow [62]

Dans l'idée de pouvoir estimer du flux optique en temps réel (pour des applications en robotique), Camus propose d'utiliser une méthode par corrélation (plus robuste au bruit et au changement de luminosité selon lui) et une approche qui permet de linéariser le temps de recherche, normalement quadratique.

Plutôt que de chercher dans une fenêtre $(2n+1) \times (2n+1)$ entre t et $t-1$ et de prendre la meilleure corrélation, il propose de fixer la recherche à une certaine distance (par exemple 1 pixel) et d'effectuer la recherche dans le temps. Cette stratégie permet de réutiliser les résultats des images précédentes, et donc, naturellement, le temps de recherche est linéaire plutôt que quadratique.

1996 : The robust estimation of multiple motions : Parametric and piecewise smooth flow fields [29]

Black et Anandan proposent une approche multi-échelle pouvant détecter des mouvements multiples. Tel que mentionné en 1991, ils expliquent qu'une fonction quadratique, telle qu'utilisée dans une différence au carré (SSD) pour une corrélation, n'est pas robuste au bruit et aux échantillons aberrant . Ils encouragent donc d'utiliser des fonctions robustes telles que présentées à la figure 2.1. Ces fonctions permettent aussi d'identifier les échantillons aberrants. Par exemple avec une Lorentzienne, un point est aberrant si $|x| > \sqrt{2}\sigma$, et pour la Gean-McClure, si $|x| > \frac{\sigma}{3}$ – dans ces deux cas, les échantillons tombent alors dans la partie non-convexe de la fonction.

Un première approximation de la solution est trouvée avec une fonction sans point aberrants : σ est choisi de façon à ce que tous échantillons tombent dans la partir convexe afin que la fonctionnelle soit également convexe. La solution est ensuite améliorée en itérant avec des valeurs de σ de plus en plus petites, jusqu'à-ce qu'il ne reste plus qu'un certain nombre d'échantillons non-aberrants.

Une approche multi-échelle est également employée afin de résoudre de grand déplacements, où l'image est déformée à chaque niveau avec le mouvement trouvé.

Avec une telle méthode et un modèle de flot affine, ils arrivent à retrouver plusieurs mouvements dans une scène, par exemple, un avant et arrière plan, ou encore deux images transparentes superposées.

1996 : Skin and Bones : Multi-layer, Locally Affine, Optical Flow and Regularization with Transparency [63]

Ju *et al.* présentent un modèle qui combine une régularisation locale et globale. Le flux est résolu localement avec un contrainte de régularisation affine. Une autre contrainte, cette fois globale, vient propager l'information et permet d'obtenir des mouvements aux endroits où le contraste de l'image ne le permet pas. Ils parlent du terme local comme des «os» et du terme global comme de la «peau» ; la peau et les os se contraignent mutuellement. En plus d'utiliser une des fonctions pénalisantes ψ robustes, ils traitent les mouvements multiples en permettant plusieurs couches et en résolvant par EM. Leur résultats restent encore aujourd'hui parmi les meilleurs.

1998 : Recovering motion fields : An analysis of eight optical flow algorithms [39]

Galvin *et al.* proposent de recomparer plusieurs algorithmes déjà comparés par Barron *et al.* [18] ainsi que deux nouveaux algorithmes. En tout, on y retrouve :

- Anandan [53]
- Horn et Schunck [3]
- Lucas et Kanade [4]
- Nagel [9]
- Singh [11]
- Uras, Girosi, Verri et Torre [48]
- Camus [62]
- Proesmans, Van Gool, Pauwels et Oosterlinck [64]

Les comparaisons sont effectuées avec des scènes synthéthiques plus complexes qu'avec Barron *et al.*. avec des mouvements connus. De plus, ils évaluent la robustesse des méthode au bruit.

1998 : A multigrid approach for hierarchical motion estimation [65]

Mémin et Pérez proposent une approche multirésolution avec grille adaptative dans laquelle le modèle de régularisation est construit itérativement. Une régularisation constante (2 paramètres) est d'abort trouvée, puis une régularisation affine simplifiée (4 paramètres) est estimée à partir du mouvement constant. puis une régularisation affine (6 paramètres). Pour chaque région, la meilleure régularisation est conservée. La grille est découpée en fonction de l'erreur et les modèles de paramétrisation sont ajustés aux frontières. Ce découpage permet une convergence plus rapide et une meilleure gestion des discontinuités.

La qualité du flux optique dense sur la séquence `Yosemite` se compare aux meilleures méthodes non denses étudiées par Barron *et al.*.

1999 : Estimating motion in image sequences [66]

Stiller et Konrad présentent une synthèse des modèles de flux optique. L'article est un excellent point de départ pour quiconque s'intéresse au flux optique et à un modèle actuel.

Ils abordent :

les modèles de mouvements translationnel, affine, projectif linéaire, quadratique, interpolé par échantillon et par polynôme.

le support global, par régions ou par pixel.

le multi-échelle avec sous-échantillonage ou filtrage.

le modèle d'intensité constante qui peut aussi être remplacé par un modèle de gradient constant.

de la régularisation par formulation bayesienne ou markovienne et des méthodes de résolutions.

L'article traite aussi des approches par blocs (utilisées en compression vidéo) et mentionne les approches par énergie.

1999 : Computing optical flow via variational techniques [67]

Aubert, Deriche et Kornprobst offrent une revue des modèles par gradients et de leur terme de régularisation. Ils proposent un modèle à la Horn et Schunk mais robuste aux occlusions et justifient leur modèle mathématiquement. Ils démontrent également que leur méthode de résolution converge vers une seule réponse.

2000 : Reliable estimation of dense optical flow fields with large displacements [10]

Alvarez, Weickert et Sánchez révisent le modèle de Nagel et Enkelmann [68] et observent que le modèle de régularisation qui prend en considération le gradient de l'image est en fait un modèle de diffusion anisotropique développé par Iijima [69] pour la reconnaissance de caractères et semblable à celui utilisé par Weikert [70] pour la restauration d'images. Ils proposent une méthode de flux optique révisée et couplée avec une approche de focus d'échelle (*Scale-Space Focussing*). Cette approche propage l'information à l'aide d'un filtrage gaussien dont la variance suit une fonction décroissante à chaque itération. Cette méthode est semblable à une approche multi-échelle, mais permet un ajustement plus graduel et une meilleure convergence.

2000 : Fast Computation of a Boundary Preserving Estimate of Optical Flow [71]

El-Feghali et Mitiche comparent trois termes de régularisation.

ϕ_I : tel que le poids γ_j du pixel j dans la régularisation dépend du gradient de l'image à j dans la direction i du mouvement :

$$\gamma_j = \frac{1}{k_j^1}\left(\frac{1}{1+|I_j - I_i|}\right)$$

où k_j est un terme de normalisation sur calculé sur les 8 pixels voisins.

ϕ_W : tel que le poids γ_j du pixel j dans la régularisation dépend du gradient des vitesses à j dans la direction i du mouvement :

$$\gamma_j = \frac{1}{k_j^2}\left(\frac{1}{1+|v_{x,j} - v_{x,i}|}\right)^{\beta}$$

où k_j^2 est un terme de normalisation sur calculé sur les 8 pixels voisins et $\beta > 1$ est utilisé pour une meilleur différenciation lorsque l'étendue des mouvements est petite. Un terme en v_x et en v_y est calculé.

ϕ_M : tel que le poids γ_j du pixel j dans la régularisation dépend de la distance à la médiane dans le voisinage.

Ils initialisent la première itération en estimant le mouvement sur les bordures (détectées comme dans [42]) et propagent l'information avec ϕ_I, ϕ_W ou ϕ_M. D'après leur résultats, ϕ_W offre les meilleurs résultats.

2001 : Variational optic flow computation with a spatio-temporal smoothness constraint [7]

Comme le suggérait Black [6], Schnörr [61] et Weickert [72], la méthode présentée utilise une fonction pénalisante robuste aux aberrations, permettant de conserver les discontinuités de mouvement. La pénalité utilisée est

$$\Psi(s^2) = \epsilon s^2 + (1-\epsilon)\beta^2\sqrt{1+\frac{s^2}{\beta^2}}$$

où $0 < \epsilon \ll 1$ pondère entre une quadratique et une courbe telle qu'utilisée par Charbonnier [73] (fig. 2.1).

Il montrent que l'ajout d'une régularisation temporelle (comme l'avait suggéré Yachida [5], Murray [47], Nagel [74] et Black [6]) améliore de beaucoup la qualité des résultats.

2001 : The statistics of optical flow [19]

Fermüller, Shilman et Aloimonos modélisent le biais introduit par le bruit dans l'estimation de mouvement. Ils montrent que ce biais est le même avec les méthodes par gradient, par énergie ou par corrélation : les vitesses tendent à être plus petites et les orientations à être orientées vers le flux normal prédominant dans la région.

2002/2004 : Lucas/Kanade Meets Horn/Schunck : Combining Local and Global Optic Flow Methods [28]

Bruhn *et al.* proposent une approche hybride entre Horn/Schunck et Lucas/Kanade. L'intention est d'être robuste au bruit comme Lucas/Kanade mais d'avoir également l'avantage du terme de régularisation pour remplir les zones ambigues et ainsi avoir un flot dense. Cette méthode a beaucoup en commun avec celle développée par Ju *et al.* [63] en 1996 qui combinait aussi une régularisation locale et globale pour obtenir d'excellents résultats.

La méthode reprend la fonctionnelle

$$\int_D (\nabla I \cdot v + I_t)^2 + \lambda^2(\|\nabla u\|^2 + \|\nabla v\|^2)dx$$

et explique le terme $(\nabla I \cdot v + I_t)^2$ comme un cas de zéro voisinage de la méthode de Lucas/Kanade. La fonctionnelle devient alors :

$$\int_D \left(\sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x})[\nabla I(x,t)\cdot\mathbf{v} + I_t(\mathbf{x},t)]^2\right) + \lambda^2(\|\nabla u\|^2 + \|\nabla v\|^2)dx$$

De plus, ils utilisent une approche de minimisation non quadratique qui permet de pénaliser les aberrants mois sévèrement. Les méthodes non-linéaires traitent mieux les discontinuités [75, 76].

$$\int_D \Psi_1\left(\sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x})[\nabla I(x,t)\cdot\mathbf{v} + I_t(\mathbf{x},t)]^2\right) + \Psi_2(\lambda^2(\|\nabla u\|^2 + \|\nabla v\|^2))dx$$

où $\Psi_1(s)$ et $\Psi_2(s)$ sont des pénalisateurs non quadratiques convexes en s. Ils utilisent un modèle de diffusion proposé par Charbonnier *et al.* [73] (voir figure 2.1)

$$\Psi_i(s^2) = 2\beta_i^2\sqrt{1+\frac{s^2}{\beta_i^2}}, i \in 1,2$$

Ils utilisent également une approche hiérarchique qui consiste à déformer l'image à mesure que le mouvement est trouvé et à recorriger par processus itératif.

2005 : On the spatial statistics of optical flow [17]

Roth et Black présentent un modèle «FoE» (Field of Experts) afin d'améliorer la qualité du flux optique. Ils reprennent la méthode locale-globale de Bruhn *et al.* [28] et modifient la fonctionnelle de façon à ce que la fonction de coût soit lorentzienne (comme dans [29]) et un lissage spatial par FoE. À partir d'une base de donnée de séquences, ils expliquent que les mouvements de caméra sont généralement biaisés : les mouvements horizontaux sont plus fréquents que verticaux, un déplacement de caméra vers l'avant est plus fréquent que vers l'arrière. Ils effectuent aussi une analyse en composante principale afin de trouver les distributions de mouvement les plus fréquentes dans une fenêtre donnée (*e.g.* 5×5 ou, pour leur expériences, 3×3) et observent également que la distribution du gradient des vitesses suit une distribution *Student t.* Ce dernier résultat semble justifier l'utilisation de fonctions de coût robustes et ces statistiques sont utilisées pour guider le lissage.

2.1 Remarques pertinentes

Cette revue permet plusieurs observations sur les tendances et l'état de l'art en estimation de mouvement et nous permet de situer notre contribution.

Tout d'abord, il est frappant d'observer qu'une somme considérable de travail s'est concentrée sur le terme de régularisation et sur la modélisation du mouvement. Le terme d'information constante est resté quasi-inchangé : les méthodes les plus récentes utilisent toujours la contrainte d'intensité (eq. 1.3) pour estimer le mouvement. Ceci est inquiétant puisqu'un grand nombre d'articles admettent que cette contrainte est rarement respectée. Quelques tentatives ont été faites en vue de modéliser le changement d'intensité non relié au mouvement (*e.g.* changement d'illumination ou des propriétés réflectives des objets), mais ces modèles ne sont pas populaires à cause de leur complexité.

Changer la contrainte d'intensité constante par une contrainte de gradient con-

stant n'est pas une solution populaire puisqu'elle est plus sensible au bruit et ne tolère pas la rotation et le changement d'échelle.

Weickert *et al.* montrent que d'autres termes d'information constante peuvent être utilisés pour plus de robustesse [77]. Dans cette optique, il semble que les travaux de Heeger [34] et de son terme par énergie ainsi que les travaux de [54] et de leur terme par phase soient restés presque ignorés de la recherche. Ces deux termes sont robustes au bruit et au changement d'intensité mais souffrent d'une mauvaise localisation spatiale.

Un autre problème non résolu et souvent invoqué est celui des mouvements multiples. Les approches supposent souvent qu'en réduisant le support du terme d'information (ou encore, en utilisant un plan de plus haute résolution dans les approches hiérarchiques), les mouvements multiples deviennent de plus en plus séparables. Cela n'est évidemment pas le cas dans les zones de discontinuité de mouvement, ni dans les cas de transparence, d'ombres, de réflection ou d'entrelaçage d'objets.

Notre contribution se situe dans ces deux domaines. Nous proposons un terme d'information constante robuste, inspiré des travaux de Fleet [54]. Notre première contribution se situe principalement dans la localisation des filtres utilisés pour calculer la phase qui permet de retrouver les discontinuité de mouvement avec grande précision (section §3). Notre deuxième contribution est une méthode permettant de résoudre des mouvements multiples sans paramétrisation. Contrairement aux autres méthodes présentées, la méthode n'est pas itérative et n'a pas besoin de connaître le nombre de mouvements présents. Le résultat est une carte de mouvement permettant d'identifier les mouvements présents ou encore de mesure le mouvement prédominant sans interférence des autres mouvements. Cette méthode s'applique aussi bien aux approches par gradient, par phase ou par énergie (section §4).

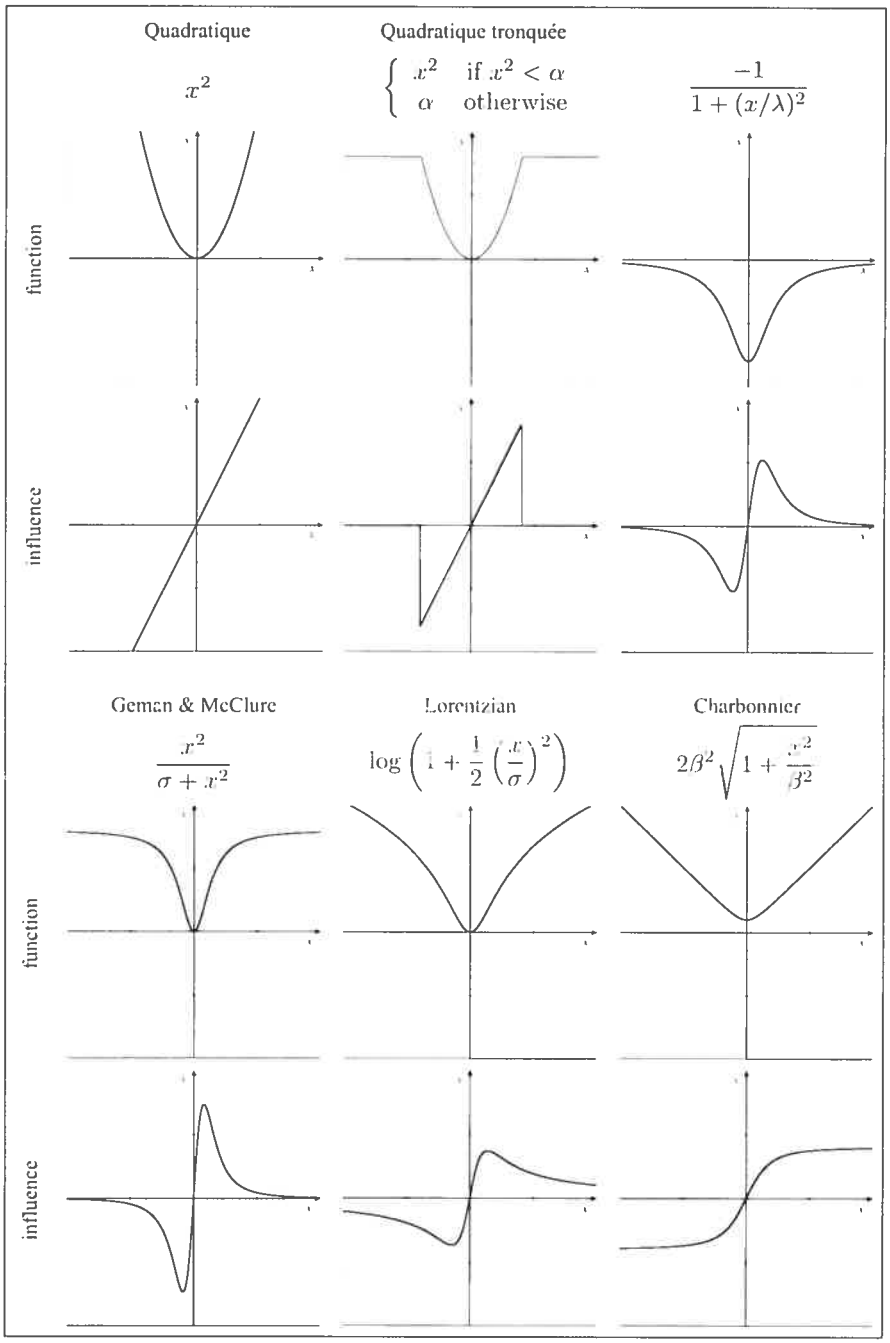


FIG. **2.1.** Quelques fonctions pénalisantes $\psi(s^2)$

Chapitre 3

ESTIMATION MOUVEMENT SANS RESTRICTION

Ce chapitre, décrivant notre méthode proposée d'estimation du mouvement, est rédigé en anglais et est présenté tel que soumis à la conférence *European Conference on Computer Vision (ECCV) 2006*. L'article s'intitule «Unconstrained Motion Estimation using Localized Quadrature Filters». Pour le mettre en contexte, Bergen a dit en 1992 :

> Because optical flow computation is an underconstrained problem, *all* motion estimation algorithms involve additional assumptions about the structure of the motion computed. In many case, however, this assumption is not expressed explicitly as such, rather it is presented as a regularization term in an objective function or described primarily as a computational issue [8].

L'algorithme présenté dans l'article qui suit ne comporte ni régularisation, ni modèle paramétrique, ni lissage. Ce que nous proposons, c'est un terme d'information robuste.

3.1 Introduction

Most spatio-temporal and phase based motion estimation methods compute instantaneous motion – that is, a motion vector that explains the observed change in brightness or phase of a pixel or neighbourhood. These methods work well for small motions. For large motions, the lack of temporal continuity and the presence of occlusions becomes a difficult challenge.

Hierarchical approaches have been proposed [8], but pyramids do not solve all

problems. They increase the proportion of pixels involved in occlusion and errors from higher levels of the pyramid propagate and are amplified in the lower levels. Also, while pyramids allow larger motions to be found, they do not help solving motion in cluttered scenes where the ordering constraint may be broken.

To address large occluded motion, we propose an oriented filter which has the minimal support possible : 1 pixel thick and 1 wavelength wide. Moreover, this filter is decentered to resists occlusion. The accumulation of responses for different orientations and frequencies of the filter makes up a pixel signature. This signature can then be compared between potential matching pixels to resolve large integral motions in a straightforward way.

The matching of two signatures can be made invariant to change in contrast, rotation, and very robust to occlusions by using only a fraction of the components for the comparison. This comes from an important property of the filters : as occlusions becomes more prevalent, the filter components do not degrade gracefully. Some components remain accurate while some others become rapidly wrong. In that context, keeping the best 50 % matching components, for example, ensures an exceptional resistance to occlusion.

In order to no only resist occlusion but also to detect it, we propose the compute a "forward" motion field for two images and a reverse field obtained by reordering the images. Imposing that the forward motion is exactly cancelled by the reverse motion is an effective way of detecting occlusion.

Signature matching provides only integral motions. The residual subpixel motion can be estimated directly using the spatio-temporal derivatives of the signature itself.

In the following section, we discuss instantaneous motion and why they cannot be used for large displacements. We proceed with an overview of phase methods and explain how they trade localization for stability. New quadrature filters are then presented, purposely designed for localization along with a method to match pixels with similar responses. Finally, we present some results, comparing the error with

other two-frame methods and showing the optical flow of some examples with large motion.

3.1.1 Instantaneous Velocities

Spatio-temporal derivative methods, also referred to as gradient based or differential, find velocities by estimating the translation of a signal using its derivative(s). Given $f(x,t)$[1], the spatial signal of a image at time t and $f(x,t+1) = f(x - v(x,t),t)$, the signal at time $t+1$, the translation v of the signal is approximated with :

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} = 0.$$

Similarly in phase based methods ([54]), the intensity f is replaced by the phase ϕ, which is assumed constant according to

$$\frac{d\phi}{\partial t} + v\frac{\partial \phi}{\partial x} = 0.$$

All these methods have requirements. For spatio-temporal ones, the image must be smooth enough that the first degree approximation is good. For phase methods, the image must be periodic enough so that the phase varies linearly. Another assumption is made about the two consecutive images : in the case of spatio-temporal methods, the intensity of the pixel must not change with time, and in the case of phase methods, the texture must not change with time. As shown in figure 3.1, it is not difficult to imagine cases where these methods predict velocities that do not satisfy their basic constraint of constant brightness. Since it is acceptable to assume linearity between consecutive pixels but not if they are farther apart, these method are generally restricted to small velocity sequences or coupled with a coarse-to-fine strategy.

Low-pass filtering can be used to improve the results. However, while lower frequencies are better approximations of a linear signal and will produce more stable

[1] we use a 1D signal throughout to simplify the notation

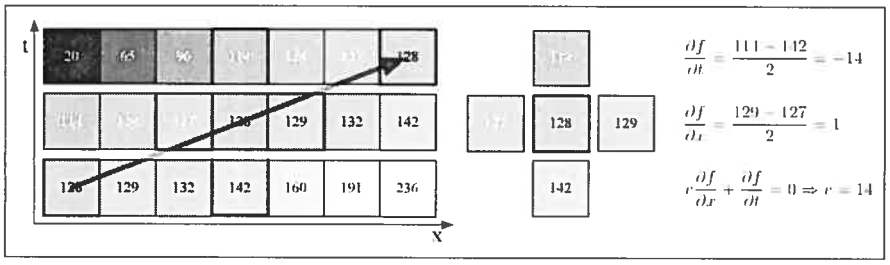


FIG. **3.1. Fast motion and the gradient method** Translation of fast moving (here, 3 pixels per frame) non-linear signal. The velocity is wrongly computed as 14 pixels.

results, they offer much less accurate localization than higher frequencies. For localization accuracy purposes, the proposed method uses a full search approach analogous to region matching, rather than instantaneous velocities.

3.1.2 Localization of Phase Methods

To understand our approach, it is important to recall some of the issues related to phase-based motion estimation. For a periodic signal with a constant velocity, one can find $v(x,t)$ by observing the phase change when convolving with a quadrature filter, for instance $e^{-2\pi\imath\omega x}$ (Fourier transform) :

$$\begin{aligned}\mathcal{F}(\omega, t+1) &= \sum_x f(x - v(x,t), t) e^{-2\pi\imath\omega x} \\ &= e^{-2\pi\imath\omega v(x,t)} \mathcal{F}(\omega, t)\end{aligned}$$

In polar coordinates, convolving $f(x,t)$ with $e^{-2\pi\imath\omega x}$ is equivalent to adding vectors of length $f(x,t)$ around a circle. The phase is the angle of the net sum of all the vectors (figure 3.2, left).

Convolving $f(x,t)$ with $e^{-2\pi\imath\omega x}$ corresponds to adding vectors of length $f(x,t)$ around a circle into a net vector (shown with an arrow), after f has been wrapped around the origin to do ω revolutions (2 in this example) . A translation of the signal is equivalent to a rotation around the circle. If the signal is not periodic (on the right),

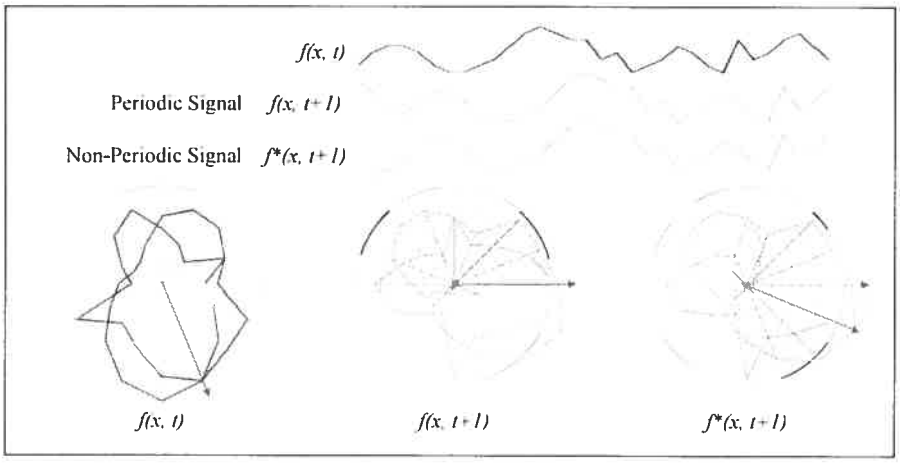


FIG. 3.2. **Bias induced by the non-periodicity of a signal.** On the left, the red arrow indicates the net vector from the convolution of $f(x,t)$. A shift in a periodic signal rotates the net vector (middle), but introduction of new data in a non-periodic case induces a bias (right).

new data appears as the signal is translated. If the proportion of new data is large, then the phase difference cannot be used anymore as a good estimate of the motion.

Although the signal of translating images is generally not periodic, if the support of the filter is large, the amount of new data introduced at the edges of each frame is usually small enough so as to have little effect over the samples present in both frames. When applying the method to recover local velocities however, this ratio increases and can become an important source of error. Figure 3.2 shows how the non-periodic data interferes with the method.

A gaussian filter can be used to circumvent the periodicity problem. The filter becomes a Gabor filter :

$$G(x,\omega,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} e^{-2\pi\imath\omega x}$$

where σ is generally set to be a function of ω, so that larger wavelengths have a wider support and higher frequencies are more localized in space.

Unfortunately, the introduction of the gaussian introduces a bias in the phase

shift observed from translating signals. This is problematic, especially since σ was introduced for localization and should be as small as possible. Another problem arises when using Gabor in 2D. Gabor filters tuned with the same σ horizontally and vertically (i.e. covariance of σI) provide poor localization perpendicular to the tuned orientation. As shown in figure 3.3-I, the motion in scenes involving occlusions will be blurred at motion discontinuities. A possible solution is to use different σ for both orientations (figure 3.3-II) : a larger one along the orientation of the detected motion, and a smaller one the perpendicular support.Unfortunately. when the filter gets very thin, it can no longer distinguish the direction of the motion. The phase of thicker filters remains relatively constant with perpendicular motion since the proportion of new data introduced by the perpendicular is small (figure 3.3-III). With thin filters, a perpendicular motion introduces a large proportion of new data and induces a change in the phase, indistinguishable form a motion oriented with the filter (figure 3.3-IV).

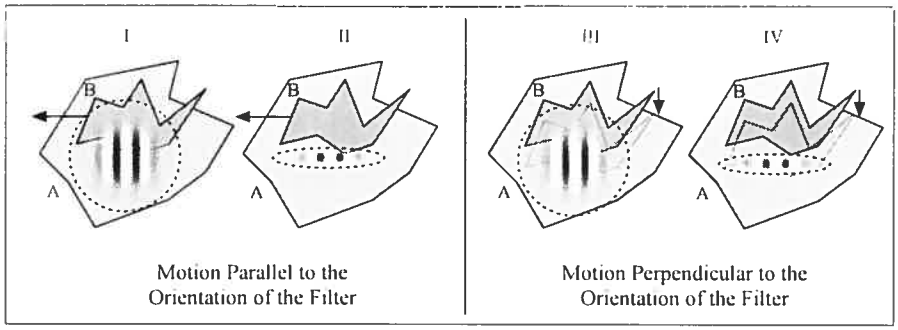


FIG. **3.3. Discontinuities in motion.** Shape B is moving and occludes the immobile shape A. (**I**) When the point of interest (the center of the gaussian) gets close to the occluder, the filter detects a phase change. (**II**) A smaller σ perpendicular to the orientation of the filter achieves better localization. (**III**) A vertical motion of B only slightly affects the phase. (**IV**) For thin filters, nearly all samples change, possibly inducing a large phase shift.

Clearly, there is a trade-off between stability and localization in the filter responses. We intend to favour localization.

3.2 Building Pixel Signatures

In order not to depend on temporal continuity, we intend to create a signature for each pixel by convolving thin filters in the spatial domain. The constant phase assumption assumption allows estimation a large motion field by matching these signatures. The disadvantage of a full search method is that it is limited to integral pixel motion. Thus, subpixel accuracy will have to be processed in subsequent step.

The filters have a support of a single wavelength, windowed using a rectangle function rather than a gaussian, making them bias free. The filter is one pixel thick, as if using a very small σ perpendicular to the orientation.

Low frequency filters, having a wider support, are the first to be affected by occlusions. In order to allow the detection to reach as close as possible to discontinuities, each filter is decentered by half a wavelength. This way, instead of having a lower match for a centered filter at a discontinuity, we have a good and a bad math for the two opposite decentered filters. Figure 3.6 shows how decentered filters outperform centered filters at motion edges.

Ideally, to estimate orientation reliably, we would test all possible orientations. In practice we select an angular aperture α and integrate the filters over this interval – yielding a radial filter. α must divide exactly the unit circle, for a total of $\frac{2\pi}{\alpha}$ filters equally distributed (figure 3.4). The filter is defined as

$$\mathcal{R}(r,\theta,i,\omega) = \begin{cases} \frac{1}{r}e^{-2\pi i\omega r} & \text{if } i-\frac{1}{2} < \frac{\theta}{\alpha} < i+\frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

where r goes from 0 to $\frac{1}{\omega}$. The normalization term $\frac{1}{r}$ is necessary to keep the sum of the filter equal to zero : $\int\int \mathcal{R} dr d\theta = 0$.

As illustrated in figure 3.5, tuned for the same frequency and orientation, our filters provide better localization than Gabor filters. This results in sharper edges at motion discontinuities (figure 3.6).

A set of such filters is chosen with different wavelengths and orientations. The

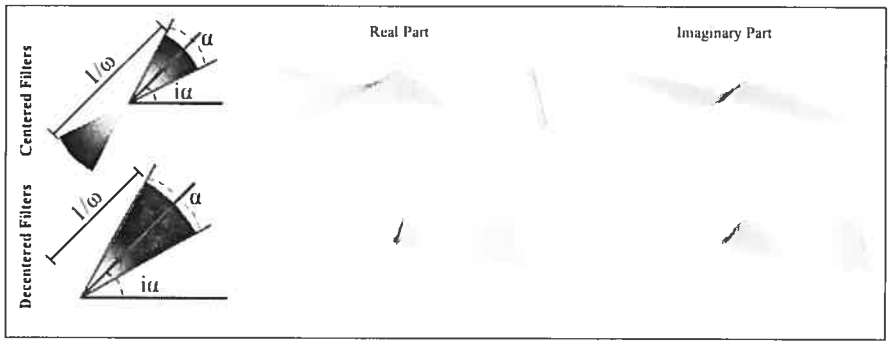


FIG. **3.4. The localized quadrature filter.** The filter used is a radial complex exponential normalized so that its sum is 0. It is oriented at $i\alpha$, has an aperture of α and a wavelength of $\frac{1}{\omega}$. We use decentered filters (**bottom**) for better response in cluttered scenes.

neighbourhood of each pixel of both frames is then convolved with each filter and the responses are kept as signature vector $\mathbf{s}$. Typically, signatures using 20 responses (5 wavelengths and 4 different orientations) work well in practice

3.2.1 Comparing Signatures

Once built, the signatures can be used to compare pixels with each another. This comparison can be made in several ways. In our case, we computed the quality of a match between pixels with signatures $\mathbf{s_0}$ and $\mathbf{s_1}$ as

$$\mathcal{Q}(\mathbf{s^0}, \mathbf{s^1}) = \frac{\sum_{ij} |s^0_{ij} - s^1_{ij}|}{||\mathbf{s^0}|| + ||\mathbf{s^1}||}, \tag{3.1}$$

where s^0_{ij} denotes the response of the filter $\mathbf{s}^0$ with orientation i and frequency ω_j. The displacement of a pixel is found by comparing its signature with those of all pixels in the other image.

This method is very similar to those using region matching, except that it is more discriminating. With a signature of 18 complex keys, 36 values are compared. Region matching with a 6x6 window has a similar complexity, its reliability is highly dependent on the quality of the textures.

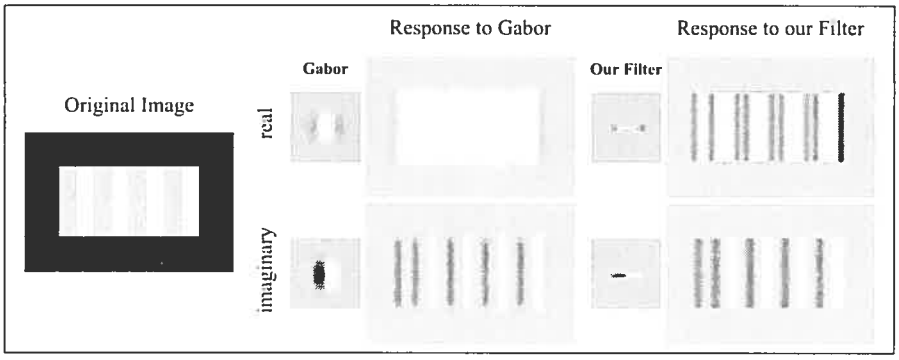


FIG. **3.5. Our filter vs. Gabor.** The original image (**left**) is filtered with Gabor ($\omega = 1/16$ and $\sigma = \omega/4$) (**middle**) and our filter ($\omega = 1/16$ and $\alpha = \frac{\pi}{8}$) (**right**). The filters are tuned horizontally. The Gabor filter has some spillover at the top and bottom while our filters preserve sharp edges. The real part of the Gabor image shows a bias.

As shown in figure 3.7 with the same number of samples to compare, our method outperforms simple region matching. In this example 36 samples were used (6x6 window vs 6 wavelengths and 3 orientations) to find the possible displacement of a pixel (at the left of the cube on the top image) between frame 1 and frame 18 of the Rubik sequence. The texture at the selected point contains little information and similar 6x6 regions can be found at several places in the other image. Our method resolves this ambiguity using lower frequencies. In this example, the 6 wavelengths are equally distributed from 8 to 58 pixels and the 3 orientations are 0, $\frac{\pi}{3}$ and $\frac{2\pi}{3}$.

3.2.2 Robustness to Local Change in Contrast

The signatures can also be compared in such a way that they are robust to local change in contrast (for instance, change in exposure or illumination). If we consider such a change as a simple multiplication of the original signal by a constant ($f'(x,t) = cf(x,t)$), it is easy to show that this constant appears in the norm of the filter's responses and has no effect on the phase.

Hence, normalizing the length of all responses to 1 would discard information

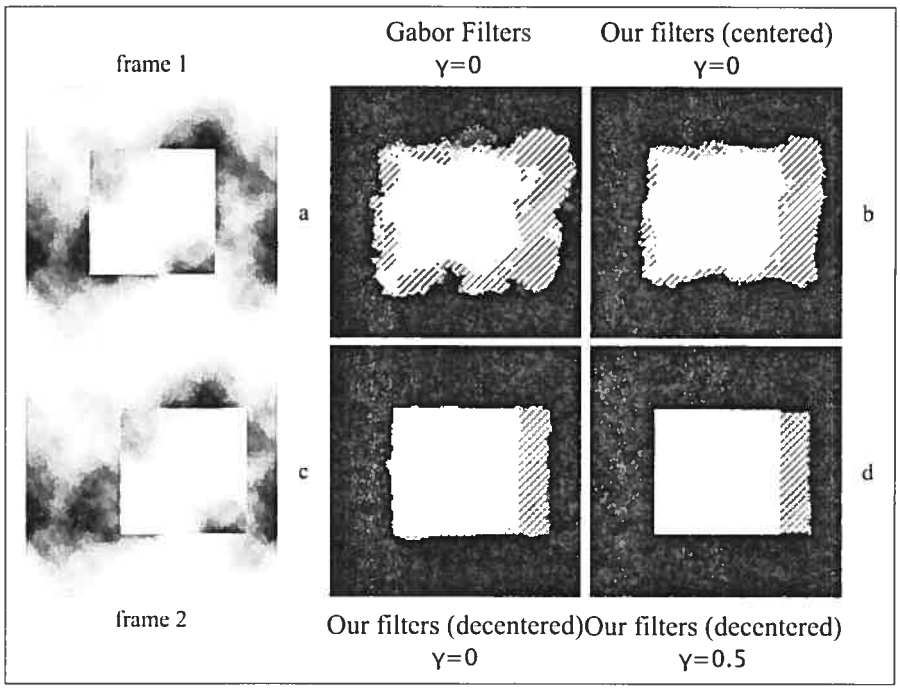


FIG. **3.6. Improving localization.** (**Left**) A translating square with fractal textures. (**a**) Results (brightness indicates norm of motion) from Gabor filters. (**b**) our centered filters, (**c**) our decentered filters and (**d**)our decentered filters with tolerance to occlusions ($\gamma = 0.5$, see §3.2.4). The diagonal pattern shows rejected motion after round-trip verification (§3.2.5).

about any multiplying factor. In practice, we chose to use a tolerance factor ζ instead of a straight normalization :

$$s'_{ij} = (1 - \zeta)s_{ij} + \zeta \frac{s_{ij}}{|s_{ij}|}$$

with ζ taken between 0 and 1. Figure 3.8 illustrates how ζ affects the matching.

3.2.3 Robustness to Local Change in Orientation

When the number of orientations used is small, the radial aperture α is large. Hence, if the object rotates a little, the response of the filter should not vary much. When better localization is required, α must be kept small and the tolerance to

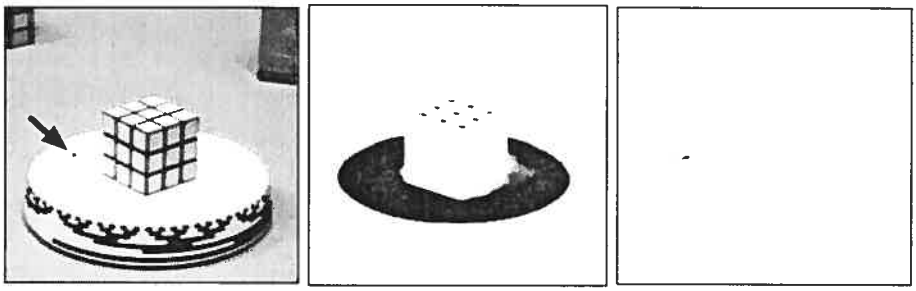


FIG. 3.7. **Our filters vs. correlation.** (**Left**) Image from the Rubik sequence with a selected pixel. (**Middle** and **right**) Potential matches, in black, shown for region matching (6×6 window, thus a correlation vector of size 36) and our method with 6 wavelengths equally distributed from 8 to 58 pixels and oriented at 0, $\frac{\pi}{3}$ and $\frac{2\pi}{3}$ (thus, 6 wavelengths $\times$ 3 orientations $\times$ 2 (real and complex response) $= 36$).

rotation is lost.

Since the filters are organized in polar coordinates, as the object rotates, the responses shift from one filter to next with the same wavelength but different orientation (figure 3.9). The shifted responses are periodic, so phase correlation can be used to recover this shift (figure 3.10). The angular translation is the index of the maximum response in

$$\mathcal{F}^{-1}(\overline{\mathcal{F}}(\mathbf{s}_j^0)\overline{\mathcal{F}}^*(\mathbf{s}_j^1))$$

where $\overline{\mathcal{F}}$ is the Fourier transform with each frequency normalized to a norm of 1, $\overline{\mathcal{F}}^*$ the complex conjugate of the normalized Fourier transform and s_i and s_j are the set of keys for the same wavelength but different angles in the same signature. Whether the angular translation should be solved for all wavelength simultaneously or for each wavelength individually remains to be investigated. In our implementation we resolved individually for each wavelength.

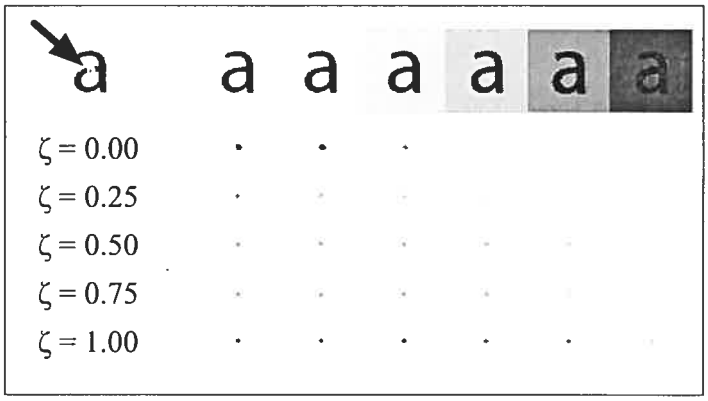


FIG. **3.8. Tolerance to change in contrast.** Effect of the factor ζ on the tolerance to change in contrast. The pixel searched for is at the center of the "a" pattern (top left) and the search space contains the same pattern with backgrounds of varying intensity. As ζ increases, the response of the similar pixels with darker backgrounds increases.

3.2.4 Robustness to Occlusions

Occlusion breaks the continuity of the optical flow and creates an edge that is hard to locate. This is where the localization of our filters has an advantage over standard Gabor filters. Since the filters have a radial width of only one wavelength with a sharp discontinuity at both ends, an occluder does not affect the estimated motion of a pixel when it is farther than half a wavelength from it since it is outside its support. In addition, the motion of an occluder above or below the filter has little or no influence on it since it is thin.

The only case where an occluder may interfere is when it is located at less than half a wavelength in the direction of the filter. This is usually not a problem when the wavelength is small, but can be inconvenient when we want to take advantage of the stability of lower frequencies. Other problems occur when an occluder is small and very close to the target pixel. In this case, smaller wavelengths would contain errors, but longer wavelengths remain relatively unaffected.

To increase robustness to all kinds of occluders, we compute the quality of a

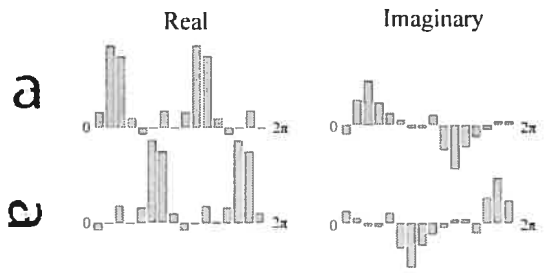


FIG. **3.9. Change in orientation and periodicity.** Rotation of a pattern will shift the responses of the filter for different angles but same wavelength. In this figure, the wavelength was 20 and 16 orientations were used ($\alpha = \frac{2\pi}{16}$). A rotation of the pattern by $\frac{\pi}{2}$ resulted in a shift of the responses by 4. Since the signal is periodic, the shift can easily be recovered with phase correlation.

match between two signatures by comparing only a subset of the keys. A constant γ is chosen between 0 and 1 to determine what fraction of the best matching keys are used for comparison. The selection of keys is done independently from one comparison to another. As demonstrated in figure 3.6, this method obtains valid matches much closer to discontinuities. Unfortunately, as the number of keys used decreases, the ambiguity in the response increases also.

3.2.5 Occlusion Detection

When computing motion from one frame to another, this method has an obvious limitation : pixels must be present in both images. If we consider the boundaries of the image as an occluding frame, this situation happens when a pixel which was visible is now occluded. In this situation, the method finds an erroneous match.

To detect these mismatches, after the displacements from t to $t+1$ have been computed, we compute the displacements from $t+1$ to t. If the forward and backward vectors agree with one another within a certain range (scaling or aliasing may produce many-to-one matches, therefore we cannot simply look for perfect match), then the forward displacement is flagged visible and valid. The invalid displacements assigned

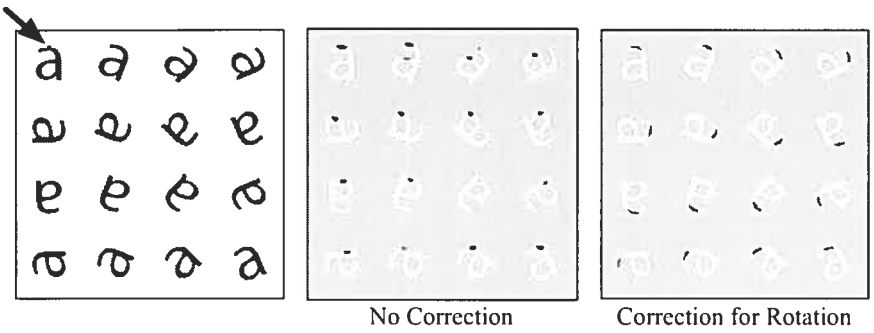


FIG. **3.10. Tolerance to change in orientation.** (**Left**) A pixel selected from an image containing "a" patterns at various orientation. (**Middle**) Potential matches without tolerance to rotation. (**Right**) Potential matches with tolerance to rotation. 5 wavelengths from 4 to 36 pixels and 16 orientations were used. Note that the original image is blended with the responses to show the location of the matches with respect to the original patterns.

to an occluded pixel is discarded and can be recomputed or approximated from nearby valid displacements. In our experiments, we chose to approximate.

3.3 Improving Accuracy : Subpixel Motion

The method presented so far estimates integral displacements only. Subpixel accuracy is recovered as a subsequent step. After an integral displacement is found, the remaining motion is assumed to be within 1 pixel. This subpixel motion is obtained through a standard gradient method over the signature. The gradient of the phase and norm of the complex keys are computed in horizontal, vertical and temporal directions. The subpixel displacement in x and y must satisfy

$$\begin{aligned} \arg(r)_x v_x + \arg(r)_y v_y &= -\arg(r)_t \\ |r|_x v_x + |r|_y v_y &= -|r|_t \end{aligned}$$

In theory, a single key is enough to solve the system (two equations, two unknowns), but since we have several keys, we try to find a solution that satisfies them all,

either using a least square minimization or, as in our implementation, using a voting scheme [78] .

3.4 Results

Our method was compared with various two-frame optical flow methods on three synthetic sequences with ground truth. Results for the sequences *Translating Tree*, *Diverging Tree* and *Yosemite* are shown in tables 1, 2 and 3. In these tree cases, we used 5 wavelengths and 6 orientations, with no tolerance to change in contrast, brightness or occlusions ($\zeta = 0$ and $\gamma = 0$). The running time varies depending if we choose to search the whole image (nearly an hour in the case of *Yosemite*) or if we restrict the search to a 16×16 window (a few minutes). Most of the time is spent on the signature comparison and the subpixel estimation (computing the signatures themselves takes only a few seconds). Moreover, since there is no dependency between the matches once the signatures have been computed, each displacement can be computed in parallel.

Better results are available in the literature, but we compare only to other two frame methods. Moreover, our method has no smoothing or parametric component, essentially performs a direct search. This methods could be improved with such a model.

The method was also run on three real sequences (figure 3.12, 3.13 and 3.14), a synthetic sequence involving rotation (figure 3.15) and the *Yosemite* sequence (figure 3.16). These sequences have very large displacements, given that we have used frames far apart in the original sequence. All but 3.15 feature a large amount of occlusions. These discontinuities are sharp and well localized. This is better illustrated with the magnitude of the full motion fields of figure 3.11(computed with 7 wavelength and 16 orientations and $\gamma = 0.5$).

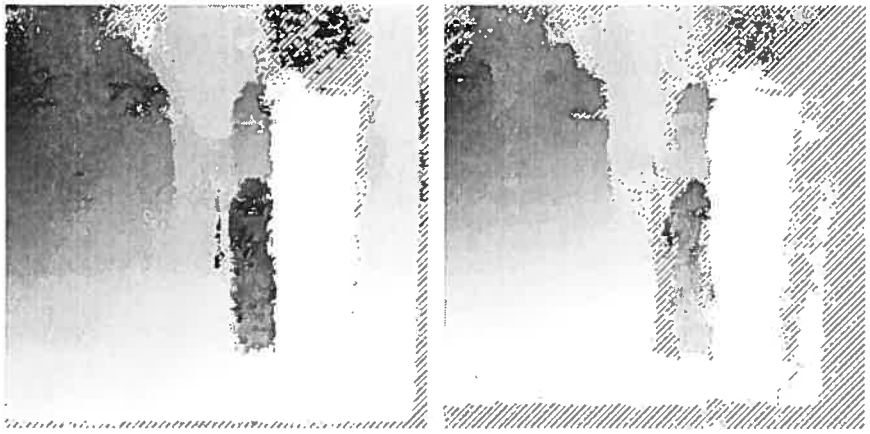


FIG. **3.11. Marbled-block sequence.** Magnitude of the motion field estimated for the marbled-block. (**left**) frame 1 and 10 and (**right**) frame 1 and 31. The diagonal pattern shows detected occlusions.

3.5 Conclusions

This paper presented a new approach to phase-based optical flow estimation in the context of large motion and a high level of occlusions. A set of filters was designed with minimal spatial support in order to achieve the best possible localization. The pixel signatures used in the matching can be made invariant to rotation and change in contrast. In addition, large integral displacements are refined by estimating subpixel motion from the gradient of the signatures and relying on the constant phase assumption. Results, even though computed on only two frames and without any parametric model are very good and feature excellent motion discontinuity localization.

TAB. 3.1. **Translating Tree**

Method	Average Error	Standard Deviation	Density (%)
Liu *et al* [79]	3.67^o	2.18^o	100
Horn and Schunck (original) [18]	38.72^o	27.67^o	100
Anandan [18]	4.54^o	3.10^o	100
Singh ($n = 2, \omega = 2$) [18]	1.25^o	3.29^o	100
Heeger (level 0) [18]	8.10^o	12.30^o	77.9
Heeger (level 1) [18]	4.53^o	2.14^o	57.8
Margarey (version 1, faster) [80]	2.31^o	—	100
Margarey (version 2, more accurate) [80]	1.32^o	—	100
Bernard [81]	0.78^o	—	99.30
Our method	$\mathbf{0.28^o}$	$\mathbf{0.19^o}$	**100**

TAB. 3.2. **Diverging Tree**

Method	Average Error	Standard Deviation	Density (%)
Liu *et al* [79]	1.67^o	0.88^o	100
Horn and Schunck (original) [18]	12.02^o	11.72^o	100
Anandan [18]	7.64^o	4.96^o	100
Singh ($n = 2, \omega = 2$) [18]	8.60^o	4.78^o	100
Heeger (level 0) [18]	4.95^o	3.09^o	73.8
Margarey (version 1, faster) [80]	3.92^o	—	100
Margarey (version 2, more accurate) [80]	3.12^o	—	100
Our method	$\mathbf{3.94^o}$	$\mathbf{2.83^o}$	**100**

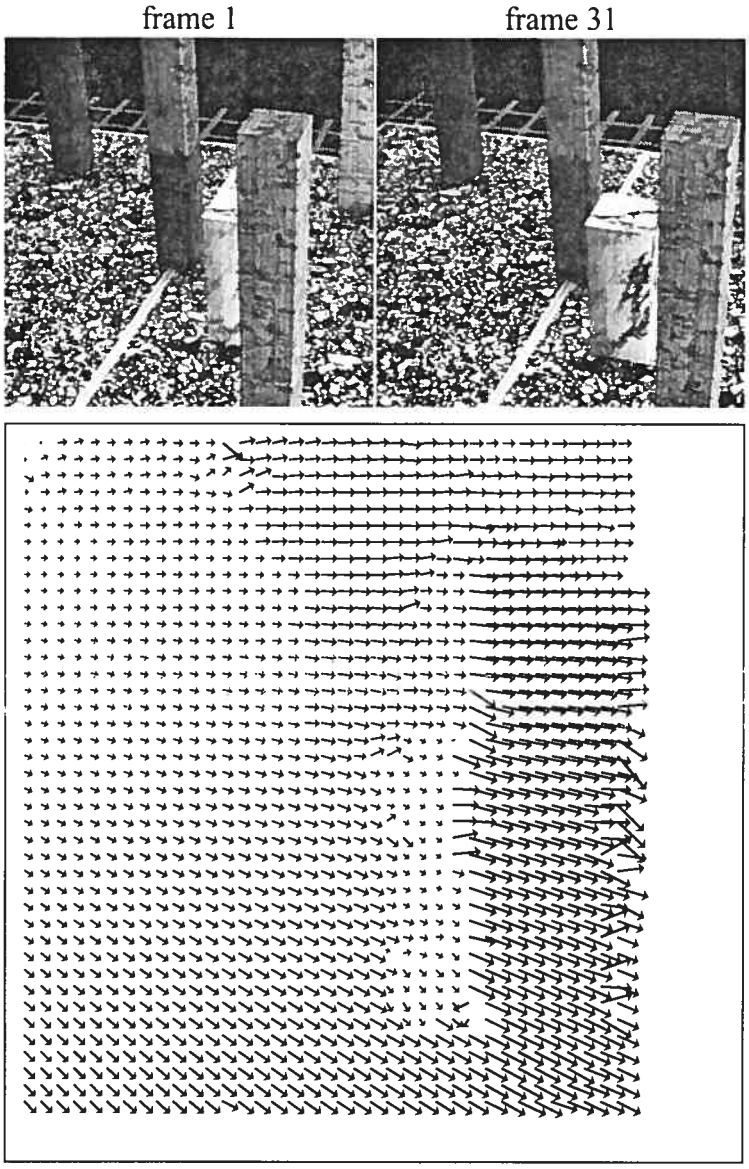


FIG. **3.12. Results on the sequences Marbled.** Real sequences taken several frames apart.

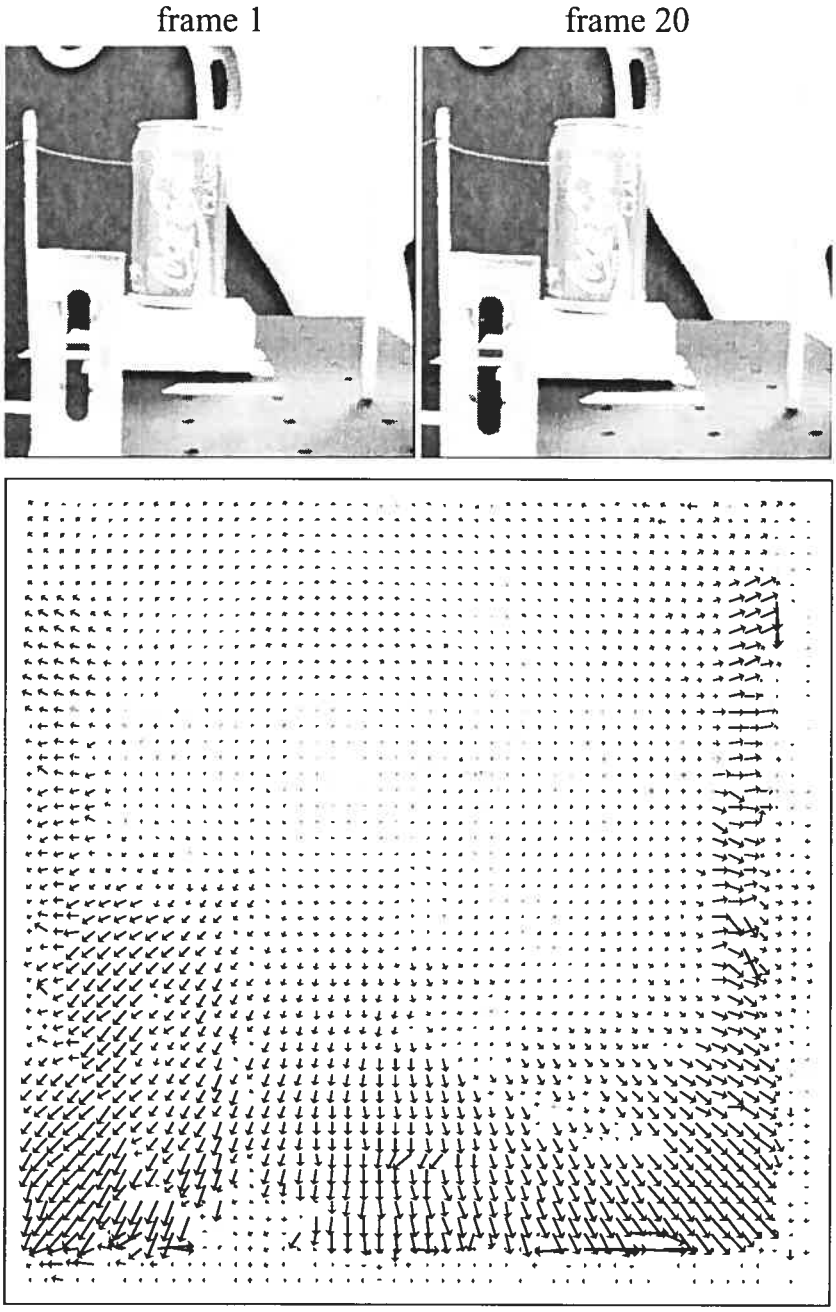


FIG. **3.13. Results on the sequences Nassa.** Real sequences taken several frames apart.

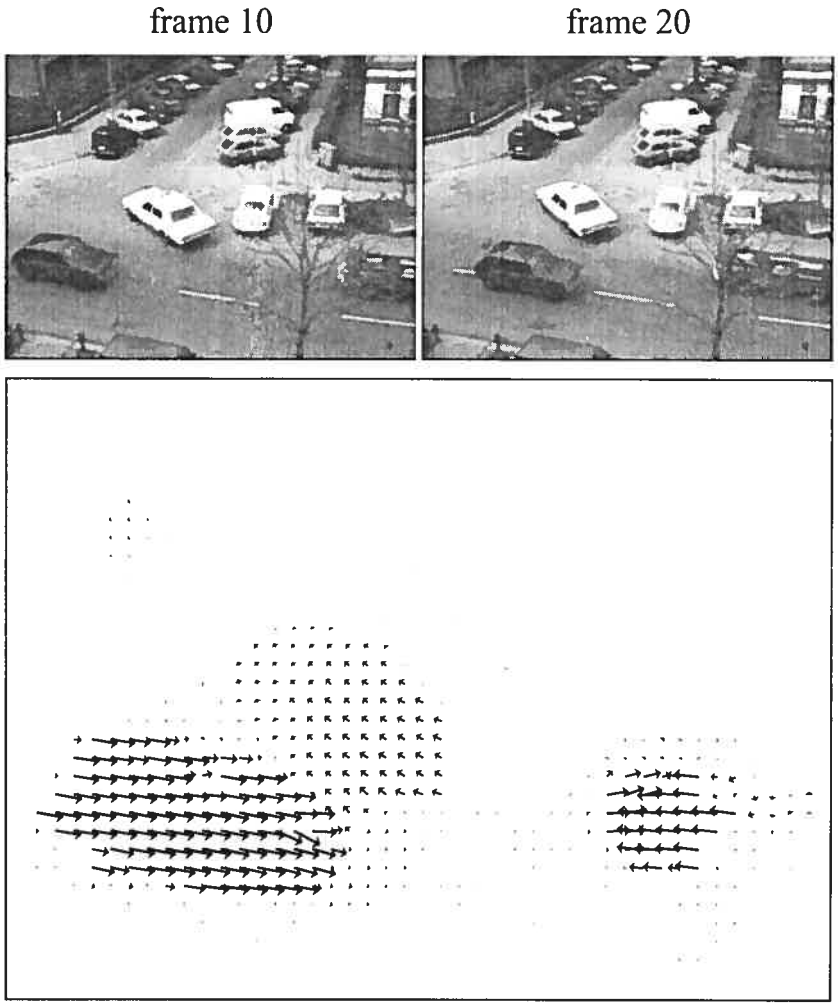
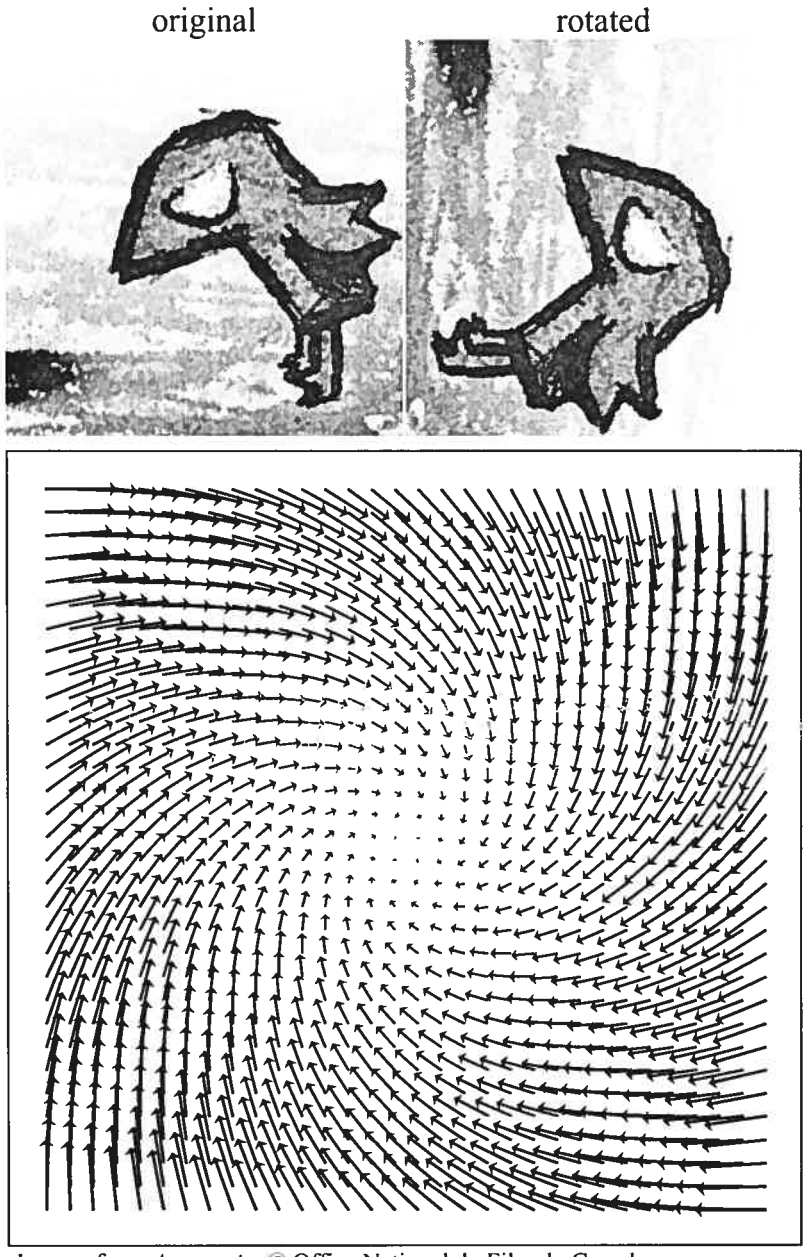


FIG. **3.14. Results on the sequences Taxi.** Real sequences taken several frames apart.



Images from *Antagonia*, © Office National du Film du Canada

FIG. **3.15. Results on the sequences Antagonia.** Synthetic sequences showing rotation.

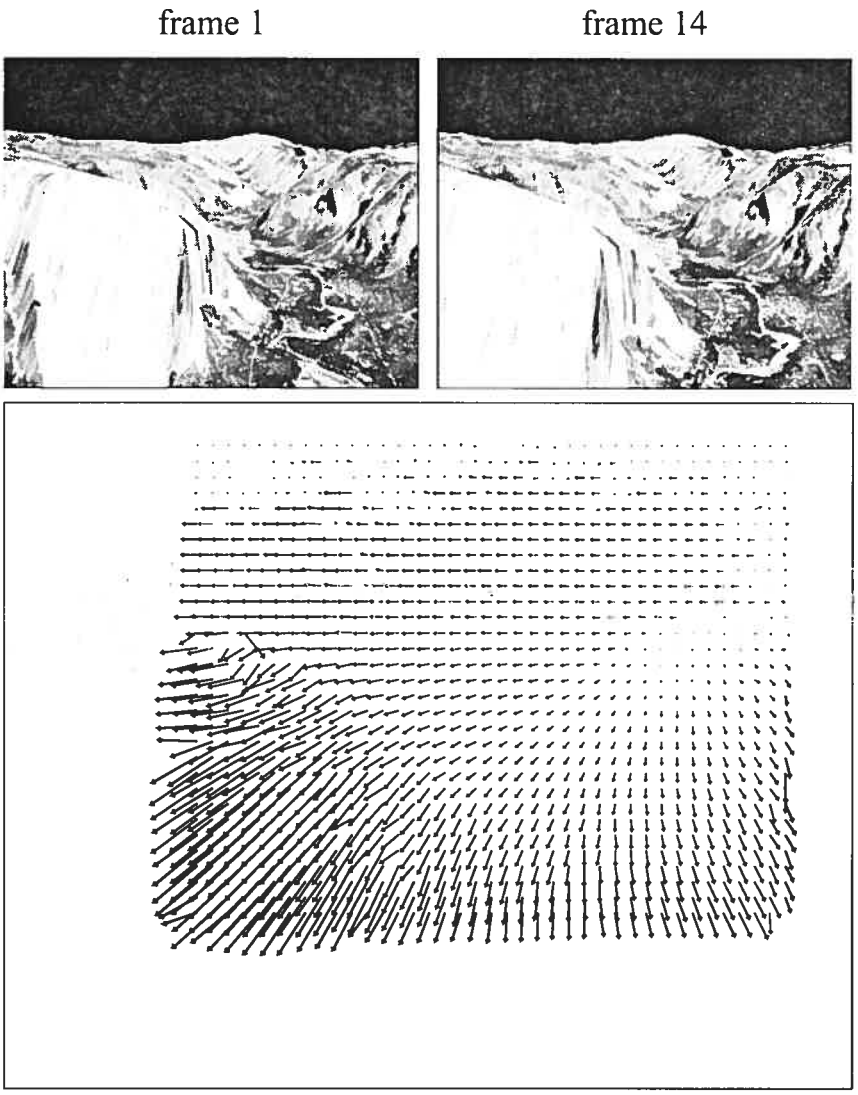


FIG. **3.16. Results on the sequences Yosemite.** Synthetic sequences.

TAB. **3.3. Yosemite (with clouds)**

Method	Average Error	Standard Deviation	Density (%)
Liu *et al* [79]	8.43^o	10.12^o	100
Horn and Schunck (original) [18]	32.43^o	30.28^o	100
Anandan [18]	15.84^o	13.46^o	100
Singh ($n = 2, \omega = 2$) [18]	13.16^o	12.07^o	100
Heeger (level 0) [18]	20.89^o	34.26^o	64.2
Margarey (version 1, faster) [80]	7.70^o	—	100
Margarey (version 2, more accurate) [80]	6.20^o	—	100
Bernard [81]	6.5^o	—	96.50
Our method	$\mathbf{5.19^o}$	$\mathbf{10.74^o}$	**100**

Chapitre 4

RÉSOLUTION DE PLANS DE MOUVEMENTS PAR VOTES

Ce chapitre, couvrant la méthode de résolution de mouvement par vote, est rédigé en anglais et est présenté tel que publié dans les proceedings pour la conférence *IAPR Conference on Machine Vision Applications (MVA) 2005*. L'article s'intitule "Solving Motion Planes by Projection and Ring Integration".

La méthode avait initialement été développée pour résoudre des plans de mouvement pour du flux optique par énergie (ou le plan apparaît après une transformée Fourier 3D tel qu'expliqué en §1.5 à la page 21). La méthode s'est révélée utile dans un contexte où l'on veut résoudre $[x, y, z]^t$ dans

$$A \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0}.$$

Dans bien des cas, on peut résoudre par moindre carré (par exemple, *via* une pseudo-inverse ou une décomposition en valeurs singulières). Dans d'autres cas, A est multimodale et si on la résout par moindre carré, on obtient une solution moyenne qui n'appartient à aucun des modes. Il est possible d'identifier les échantillons aberrants dans A et, par processus itératif, de converger vers une meilleure réponse, mais cette solution ne nous paraissait pas pratique. Voilà l'utilité de cette méthode, et pourquoi elle a été développée.

4.1 Introduction

When computing optical flow involving occlusions or transparency, local motion is often ambiguous and needs to be resolved in a larger window or globally as an energy minimization problem. The method presented in this paper was developed to preserve information in the form of distributions of local motions and provides a way to estimate ambiguity. This information can then be used to resolve the system globally. The method resolves planes that pass through the origin in a 3D sampled space. Such planes occur naturally in energy and spatio-temporal derivative methods.

4.1.1 Energy motion planes

Energy based motion estimation approaches rely on the principle that a linear motion of textures will draw oriented lines in time. These lines, in turn, form planes that intersect at the origin in the sequence's spectrum. Energy based motion estimation is effective for egomotion [24] but can also be used for optical flow where Gabor-like filters are used to locally estimate frequencies [34].

Parametrizing energy motion planes is not trivial. In the frequency domain, low frequencies are close to the origin, giving little information about the orientation of the plane while high frequencies give accurate information but are sensitive to noise. In addition, when motion is large, spectral overlapping occurs and the signal appear to "wrap around" (fig.4.1).

The quality of the motion distribution recovered depends mainly on the support of the filters used (usually determined by the window size in the spatial domain) and the response to the spatial textures, the resolution in time and wether the assumption that motion is constant over time is true or not. Taking more samples in time provides higher accuracy (especially for large motion) but may break the assumption of constant motion.

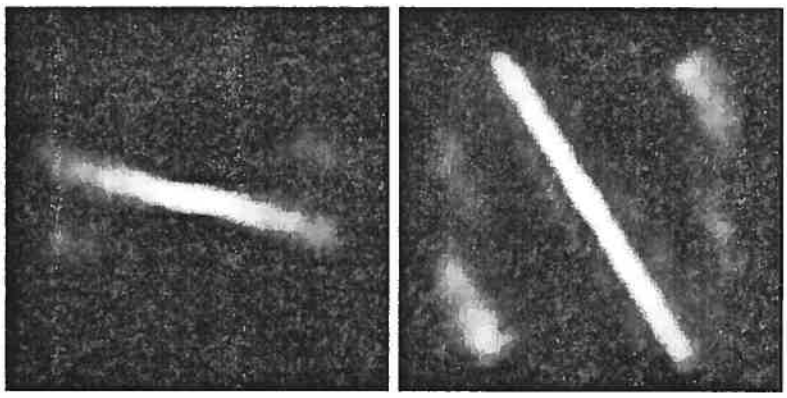


FIG. **4.1. Warping artifacts in the frequency domain :** when motion is not exactly one pixel per frame (**left :** less than one pixel, **right :** more than one pixel) warping artifacts begin to appear.

4.1.2 Spatio-temporal derivative planes

Spatio-temporal derivatives plotted in 3D will also lie on the same plane if they represent the same motion. This can easily be seen from the constant brightness constraint which describes a plane where the normal is the motion in the spatial domain :

$$\langle v_x, v_y, v_t \rangle \cdot \nabla I = 0$$

Each sample draws a line in the derivative space. Two samples or more are necessary to solve the plane. Assuming that motion is constant in a small neighborhood, samples may be taken inside a window. They could also be taken at different scales, thus, the spatio-temporal derivative filters would be "tuned" to different frequencies in the image texture.

4.1.3 Existing work

Traditional implementations make assumptions about the number or type of motions present in the sampling window. Heeger [34] assumes a single motion plane that

can be solved analytically using gabor filters. Chen *et al.* [23] makes no assumption as far as the number of motions is concerned, but the method is sensitive to noise and under-sampling artifacts. Mann and Langer [24] support various motion speeds but the orientation has to be the same for all. Pingault [25] makes an *a priori* estimate for the number of motions and uses 3D gaussians and expectation maximization (EM) to model the motion planes. Extra planes are then discarded *a posteriori* using thresholding. Our method is most similar to Yu *et al.* [26] which takes responses of conic filters to map precisely the planes in a spherical (θ, ϕ) space. The main advantage of our method is that we do not need to find the number of motions by clustering the non-zero values near the θ axis, and counting the clusters and model the spherical signal using EM to recover the orientation of the plane corresponding to each cluster. Instead, we propose to integrate the energy along rings and generate a motion distribution where a simple voting scheme can then be used to identify the dominant motions. This removes the need for the iterative EM and, because we do not make the assumptions that the motion distribution is gaussian, we obtain a distribution that can be more complex and allows a motion that is not purely translational.

4.2 Pre-Filtering

In energy based methods, if the filter used does not naturally have a limited support, the image should be filtered to prevent discontinuities on the edges. For example, in a windowed Fourier transform we multiply the signal in our window by a sine : $I'(x) = I(x) \sin\left(\frac{2\pi x}{size_x}\right) + 1$. While several authors use a gaussian filter, we find that it tends to blur the spectrum (see fig.4.2). If this step is neglected, the sharp borders will interfere with the motion analysis : they will be considered as non-moving discontinuities with full range spectrum and will add an artificial plane at $t = 0$.

The next step consists in attenuating the low frequencies using a high-pass filter. This filter will get rid of the DC component as well as frequencies that provide little

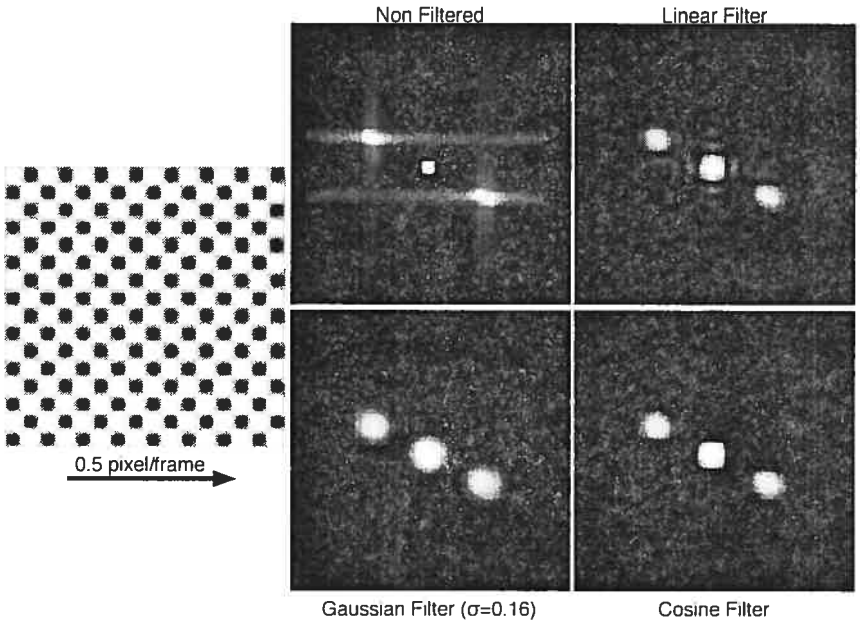


FIG. **4.2. Effect of various filters on energy :** neglecting spatio-temporal filtering or using the wrong image filter generates artifacts in frequency space.

information about the motion and tend to interfere with the rest of the process. A typical standard deviation for this filter is 20% of the nyquist frequency.

4.3 Projection

The projection remaps the energy of the sequence onto the surface of a sphere. This step is similar to [26] except that instead of using conic filters, it simply projects by casting rays along the normals of the sphere thus integrating the spectrum F onto the surface $\mathcal{S}$:

$$N_{\theta,\phi} = \langle \sin\theta\cos\phi, \sin\phi, \cos\theta\cos\phi \rangle$$

$$\mathcal{S}(N) = \int F(rN)dr$$

Where F is the 3D Fourier or derivative samples. Only half the sphere needs to be processed, and rays may be cast halfway inside since the energy is even-symmetric. All motion planes pass through the origin, therefore we expect each plane to appear a line on the surface of the sphere (fig.4.3).

The projection step is not necessary for the derivative approach. The spatio-temporal derivatives already describe a line in 3D that goes through the origin. Therefore, instead of casting rays, we can project the derivative lines on the surface of the sphere directly.

4.4 Integration

The motion distribution is found by integrating rings around the sphere. We define an axis $\langle u, v, 1 - \sqrt{u^2 + v^2} \rangle$ on the surface of the sphere and find a point p_0 perpendicular to this axis :

$$\begin{aligned} p_0 = \langle \quad & (\sqrt{u^2 + v^2} - 1)\cos(\arctan \tfrac{v}{u}), \\ & (\sqrt{u^2 + v^2} - 1)\sin(\arctan \tfrac{v}{u}), \\ & \sqrt{u^2 + v^2} \rangle \end{aligned}$$

We use quaternions to rotate p_0 around the axis integrate $\mathcal{S}$ along the ring :

$$\mathcal{P}(u, v) = \int_0^{2\pi} \mathcal{S}(Rot_\theta \cdot p_0) d\theta$$

Again, in practice, because the signal is even-symmetric, only one hemisphere needs to be computed and rings can be limited to 180^o instead of 360^o.

The result is a gauss map $\mathcal{P}(u, v)$ that gives the response of each motion plane oriented with a normal $\langle u, v, \sqrt{1 - u^2 - v^2} \rangle$. To enhance this map, its minimum value is substracted from all other responses. This minimum value corresponds to white noise in the original signal and low frequencies that contributed to several or all orientations (thus, giving no relevant information about motion). The gauss map can

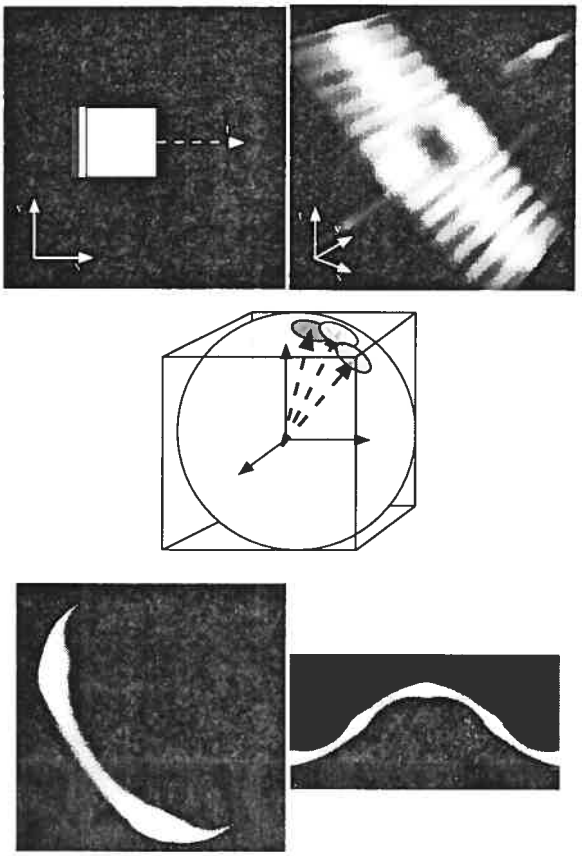


FIG. **4.3. Projection of the energy on the surface of a sphere :** the **top row** shows the sequence of a square that moves to the right at a speed of $\langle 1.0 \rangle$ along with the *log*-energy of its 3D Fourier transform. The black hole in the center is a result of the high-pass filter. Rays are then cast from the center of the cube and projected onto the surface of a sphere. **Bottom left :** projection on an actual sphere. **Bottom right :** unwrapped (θ, ϕ) texture map.

be represented as a planar map, as shown in fig.4.4.

After normalization, the motion distribution can be used as a probability distribution. The range of values $(\max(\mathcal{P}) - \min(\mathcal{P}))$ also provides an indication of the ambiguity of the motion. Further analysis can be performed from the motion distribution. The window size could be readjusted from the motion ambiguity (unless the spatial signal is a superposition of different signals, a more localized support should

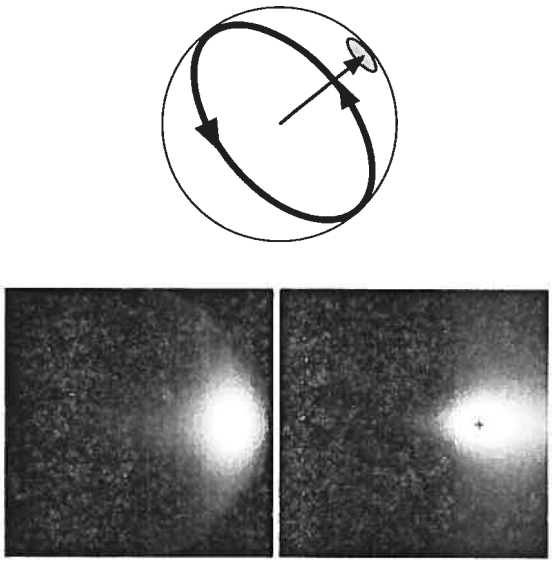


FIG. **4.4. Integration of rings around the sphere :** we define a normal $\langle x, y, z \rangle$ with coordinates $\langle x, y, \sqrt{1-x^2-y^2} \rangle$ and integrate the energy on the ring that is perpendicular. **Middle :** the gauss map where each point is the responses of a motion plane. **Right :** the equivalent planar map where we find the actual coordinates of the motion : the marked maximum corresponds to the vector $\langle 1.017, 0 \rangle$

result in less ambiguous motion). Energy minimization could be used to resolve motion globally from local patches, or additional parametrization could be applied for specific motion models. For instance, one can analyze the translation stretch caused by parallax of a sequence (see fig. 4.7).

4.5 Spectral Aliasing in Large and Small Motion

With energy based motion analysis, spectral aliasing usually occurs when the motion speed is larger than one pixel per frame. Since the filter is periodic, the motion that is seen at a given frequency of the filter is actually a modulus of that frequency. In general, we assume that the motion is smaller than half the wavelength,

but this is not always true. Aliasing can be temporal or spatial, but as pointed out by Mann and Langer [24] , in natural sequences, spatial aliasing is less important because of optical blur.

Mann and Langer describes a way to rectify the plane when there is single motion (or a bow-tie). This rectification cannot be used in our method since rectifying for one motion plane might interfere with another valid plane.

Instead, we choose to wrap the rays that we use when projecting the energy on the surface of our sphere. This way, the rays will follow the planes as they wrap and should hold a greater amount of energy when they reach the surface of the sphere. In fig.4.6. we show how this method allows us to detect motion of 8 pixels using a sampling of $31 \times 31 \times 9$ even if the temporal aliasing is quite important at that speed.

Even without the warping of the rays, experimental results indicate that spectral overlap has little effect on our method. Since the wrapped planes are not aligned with the origin, they become diffused on the surface of the sphere during the projection step.

4.6 Conclusions

We presented a simple yet robust method to find multiple motion planes for energy-based motion estimation. The method makes no assumption about the number or the type of motion in the sampled window and is robust to the spatial and temporal aliasing. We showed that the maximum response in the motion distribution map corresponds to the plane of the dominant motion, thereby making the method suitable for simple motion estimation. In addition, it is possible to perform further analysis on the local maxima to learn and take into account other motions present in the sampling window. Depending on the resolution of the motion density map, an implementation of this method takes a fraction of a second to compute and one could easily imagine further optimizations where the rings are evaluated from coarse to fine over a region

of interest.

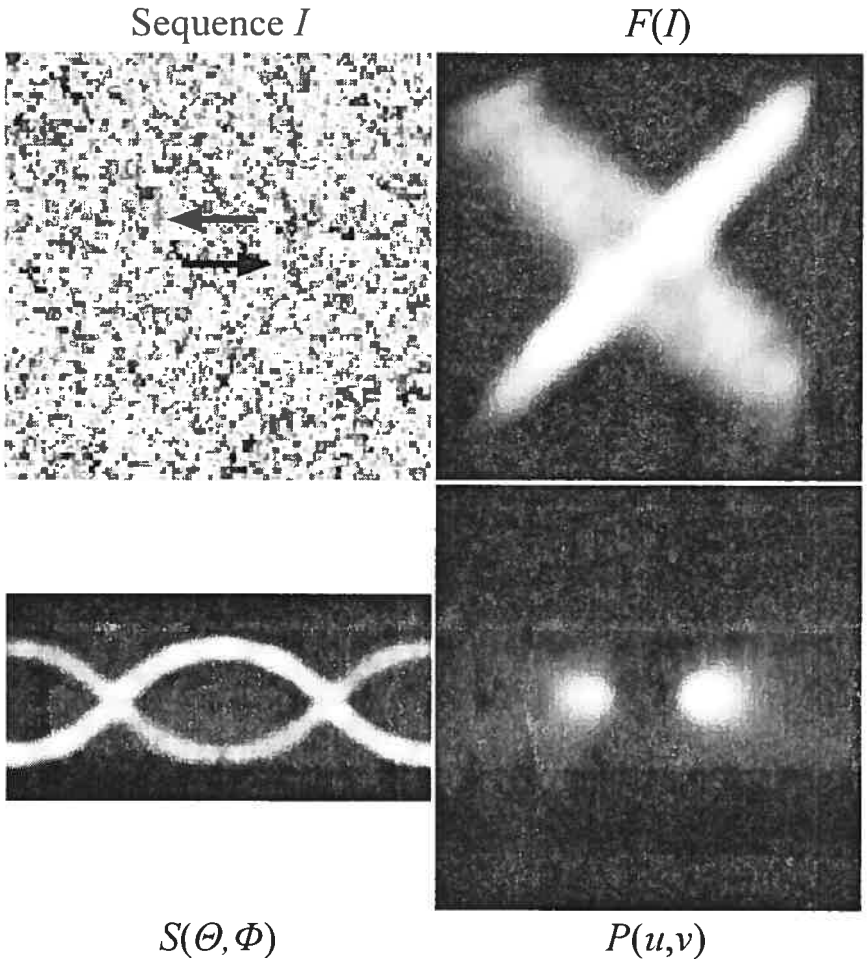


FIG. **4.5. Results on superposition of motion.** All results used a $31 \times 31 \times 9$ window and an energy approach. The resolution for $\mathcal{S}(\theta, \phi)$ and of the density map $\mathcal{P}(u, v)$ were 64×32 and 64×64 respectively. Here, two images of noise with motion $\langle 1, 0 \rangle$ and $\langle -1, 0 \rangle$ added together . The two motions are recovered (one maximum at $\langle 1.004, 0 \rangle$ and another one at $\langle -0.992, 0 \rangle$). (**Top left**) The sequence. (**Top right**) 3D Fourier transform ; two planes appear. (**Bottom left**) Projection on the sphere. (**Top right**) Ring integration : two maxima appear.

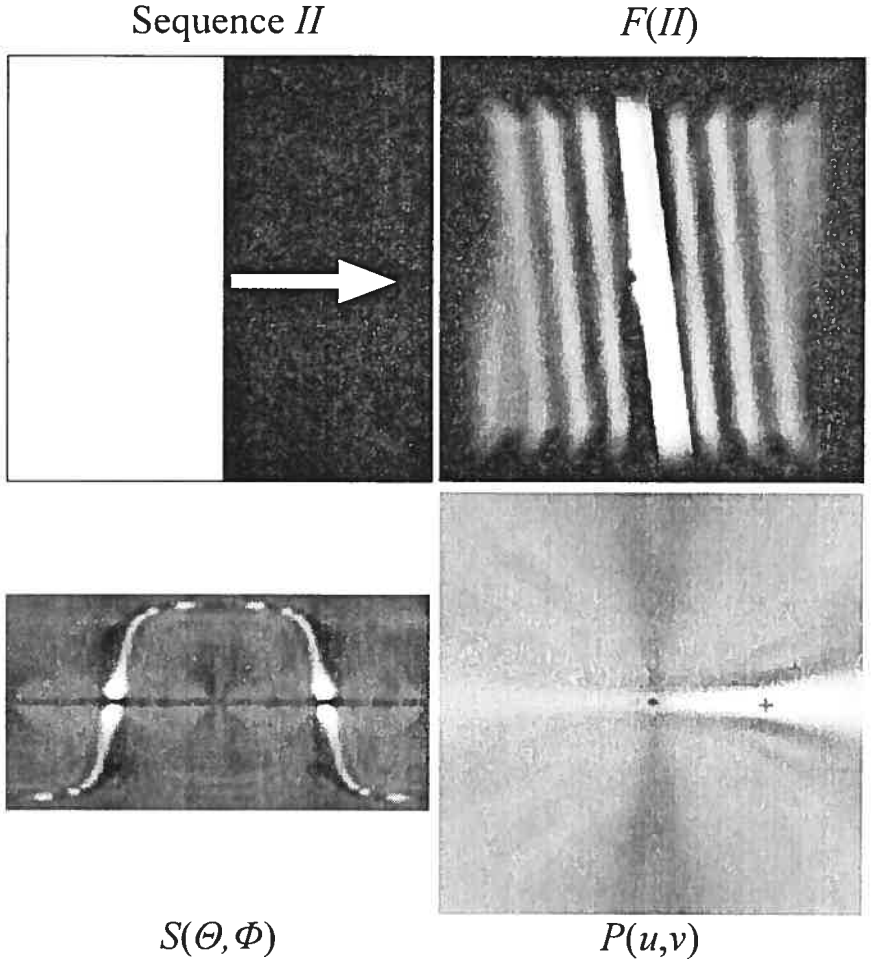


FIG. **4.6. Results on large motion.** All results used a $31 \times 31 \times 9$ window and an energy approach. The resolution for $\mathcal{S}(\theta, \phi)$ and of the density map $\mathcal{P}(u, v)$ were 64×32 and 64×64 respectively. Here, a black rectangle on a white background moving at $\langle 8, 0 \rangle$ pixels per frame. The large motion creates severe aliasing along the temporal axis. Yet, motion is found at $\langle 7.46, 0 \rangle$. For such a sequence, a better approach would be to first subsample the image - reducing the speed and thus the aliasing. but we wanted to show how aliasing affects our method.

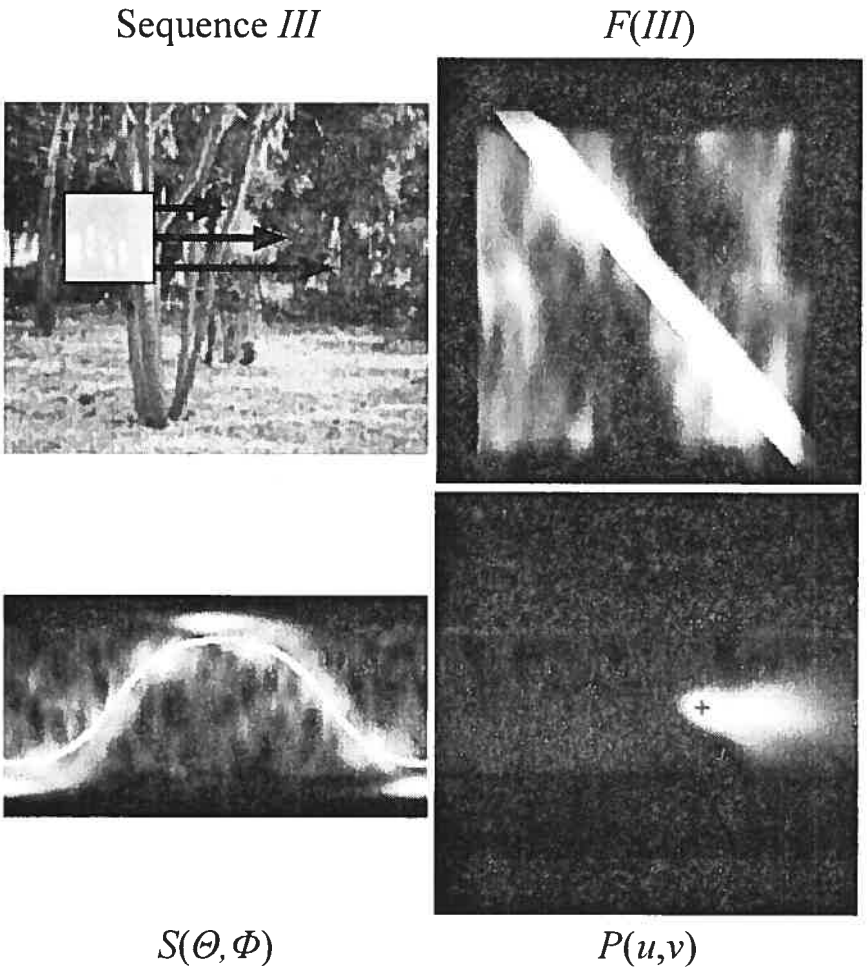


FIG. **4.7. Results on parallax.** All results used a $31 \times 31 \times 9$ window and an energy approach. The resolution for $S(\theta, \phi)$ and of the density map $\mathcal{P}(u, v)$ were 64×32 and 64×64 respectively. **a :** Two images of noise with motion $\langle 1, 0\rangle$ and $\langle -1, 0\rangle$ added together . The two motions are recovered (one maximum at $\langle 1.004, 0\rangle$ and another one at $\langle -0.992, 0\rangle$). Here, the window chosen contains multiple motions going to the right (the camera translates to the left and trees of different depth move at different speeds). In the rightmost image, the parallax appears as stretch along the x axis.

Chapitre 5

AUTRES APPLICATIONS

Nous avons présenté notre méthode dans un contexte d'estimation de mouvement, mais elle pourrait très bien être utilisée dans d'autres contextes. Nous en présentons deux autres brièvement : la reconstruction par stéréoscopie et la mise en correspondance.

5.1 Reconstruction stéréo par filtres en quadrature localisés

La reconstruction par stéréo utilise l'effet de parallaxe suite à un déplacement de caméra pour déterminer la profondeur des pixels pour des fins de reconstruction. Le problème de stéréo à beaucoup de similitudes avec problème de flux optique : comme il s'agit d'un problème mal posée, on utilise un terme d'information comme fonction de coût et une fonction de régularisation spatiale. Cependant, puisqu'on connaît le mouvement de caméra et qu'il s'agit du seul mouvement présent, le point p_c se déplace au point p'_c le long des lignes épipolaires dans l'espace caméra.

$$p_c^t E p'_c = 0$$

où E est la matrice essentielle (translation et rotation). Si les paramètres intrinsèques de la caméra ne changent pas, les points de l'image sont alors reliés par :

$$p_i^t F p'_i = 0$$

où F est la matrice fondamentale. Cette nouvelle contrainte permet d'exprimer le problème en une seule dimension plutôt que deux. Aussi, contrairement au flux optique ou on permet des disparités arbitraires, en stéréo on fixe généralement le nombre

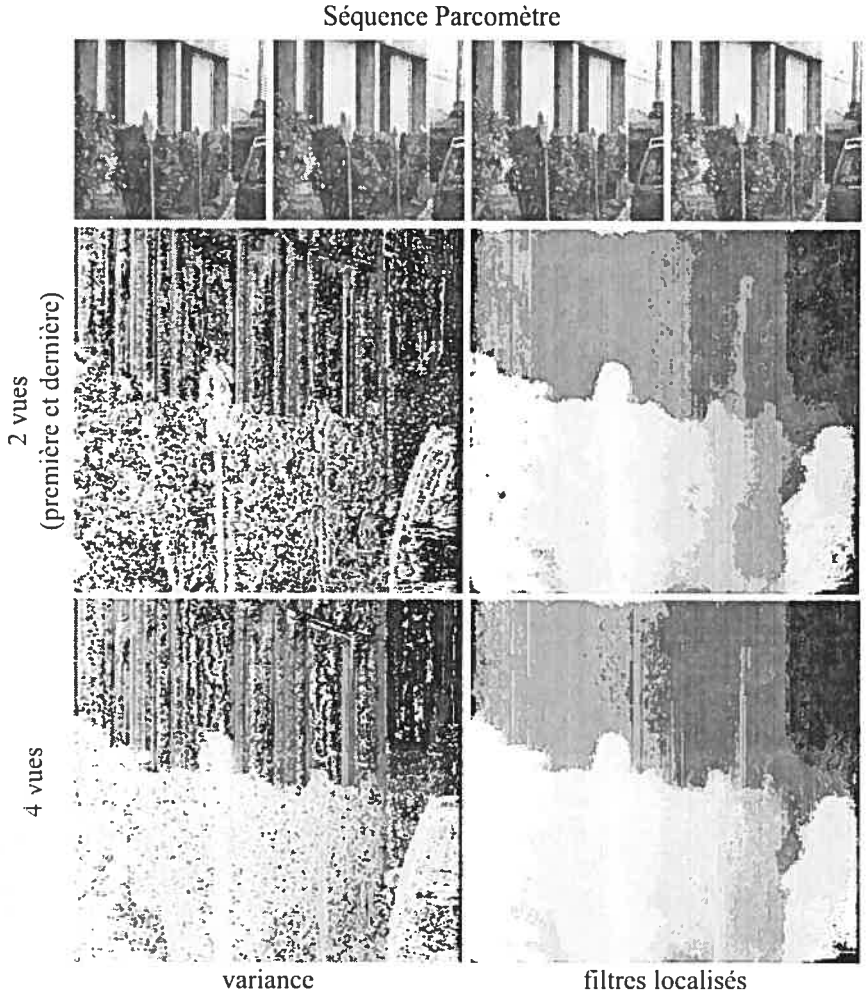


FIG. **5.1. Reconstruction de la scène Parcomètre.** (**Haut**) Reconstruction par recherche directe (aucun lissage) à partir de deux images avec une fonction de coût de variance et notre fonction de coût. (**Bas**) Reconstruction à partir des quatre images. La reconstruction s'est faite avec 13 disparités ; pour notre fonction de coût. 40 filtres ont été utilisés, avec $\alpha = 2\pi/8$ et $\omega = \{4, 9, 14, 19, 24\}$.

de disparités afin de pouvoir optimiser le résultats avec des graphe ou encore par programmation dynamique.

En stéréo, la fonction de coût est tout aussi importante que le modèle de régularisation et nous proposons d'utiliser la même fonction de coût présentée en §3.

Une étude plus approfondie révélerait le véritable potentiel de notre méthode par rapport aux autres fonctions de coûts existantes, mais les quelques tests que nous avons effectués semblent encourageants.

Les figures 5.1, 5.2, 5.3, 5.4 et 5.5 comparent notre fonction de coût à une recherche par minimisation de la variance. Dans le cas de notre fonction de coût, la disparité retenue était celle pour laquelle la somme des corrélation de signature entre toutes les caméras était maximum ($\mathcal{Q}$, tel que vu dans l'équation 3.1 à la page 55). Dans le cas de la fonction de coût par variance, la disparité retenue était celle pour laquelle la variance de l'intensité des pixels était la plus petite à travers toutes les caméras. Aucune forme de lissage n'a été utilisée. Les résultats obtenus montrent que notre fonction de coût est nettement supérieure à une simple fonction de variance. Les erreurs sont présentées dans la table 5.1.

5.1.1 Gestion des occlusions par reconstruction géométrique

Puisque la géométrie est reconstruite en stéréo, il serait possible de combiner l'information de la géométrie pour ignorer la réponses de certains filtres dans la signature. En effet, si on observe un changement de profondeur à l'intérieur du support d'un des filtres, on sait que celui-ci est en zone d'occusion. Par exemple, dans la figure

TAB. **5.1. Erreurs de reconstruction** Pour une comparaison avec les autres méthodes et plus de détails sur les métriques, le lecteur peut se référer à [82] où encore [83]

	sans occlusion	image entière	discontinuités
Tsukuba	4.74	6.67	19.6
Venus	9.91	11.4	37.5
Teddy	14.3	23.1	30.7
Cones	8.42	18.7	20.7

5.6. une occlusion couvre une partie du support du filtre. En tenant compte de la géométrie de la scène, on peut déterminer quelle section du filtre utiliser lors de la corrélation. Cette méthode tire avantage du fait que nos filtres ne soient pas centrés et aient un support radial limité. Une approche semblable pourrait aussi être utilisée en estimation du mouvement, mais plutôt que détecter les discontinuité de profondeur, il faudrait détecter les discontinuité de mouvement. Aucun résultat n'a encore été produit et cette idée est laissée pour des travaux futurs.

5.2 Identification de point d'intérêt

Les signatures utilisées par notre méthode ressemblent au vecteur de traits caractéristiques utilisés par les SIFTs (*Scale Invariant Feature Transform*) développés par Lowe [37]. Il n'est pas dans notre intention de couvrir les détails concernant l'extraction et l'identification des points d'intérêts, mais il semble indiqué de comparer la performance de notre méthode avec celle des SIFTs. Pour plus d'information sur les SIFTs, le lecteur devrait se référer à [37].

Les vecteurs de traits caractéristiques obtenus par SIFTs peuvent être comparés les uns aux autres pour obtenir des correspondances de points d'intérêts. Nous avons comparé de façon qualitative ces correspondances avec celles obtenues par nos filtres. Les résultats en figure 5.7, 5.8 montrent quelques différences entre les deux méthodes. Pour des séquences de flux optique relativement difficile (la deux paires d'images sont prises à plusieurs images d'intervalle et comportent beaucoup d'occlusion), les deux méthodes ont plutôt bien performé. Il n'y a presque pas de différence pour la première paire d'images. Par contre, notre méthode semble légèrement mieux performer sur la deuxième paire, en partie grâce à la détection d'occlusion.

Il faut tenir compte du fait que nous sous utilisons les SIFTs dans un contexte de flux optique. Les SIFTs permettent de retrouver des caractéristiques sur deux images d'échelle différente, ce que notre méthode ne permet pas. Par contre, contrai-

rement aux SIFTs, notre méthode est conçue pour retrouver des correspondances denses, plutôt que seulement aux points d'intérêts et la recherche, dans le cas de notre méthode, se fait sur toute l'image plutôt que sur des points d'intérêts ce qui nous désavantage d'une certaine façon.

Cette expérience n'a pour but que de s'assurer que notre méthode fonctionne au moins aussi bien que les SIFTs dans un contexte de flux optique pour retrouver des points d'intérêts. En ce sens, les résultats semblent convaincants.

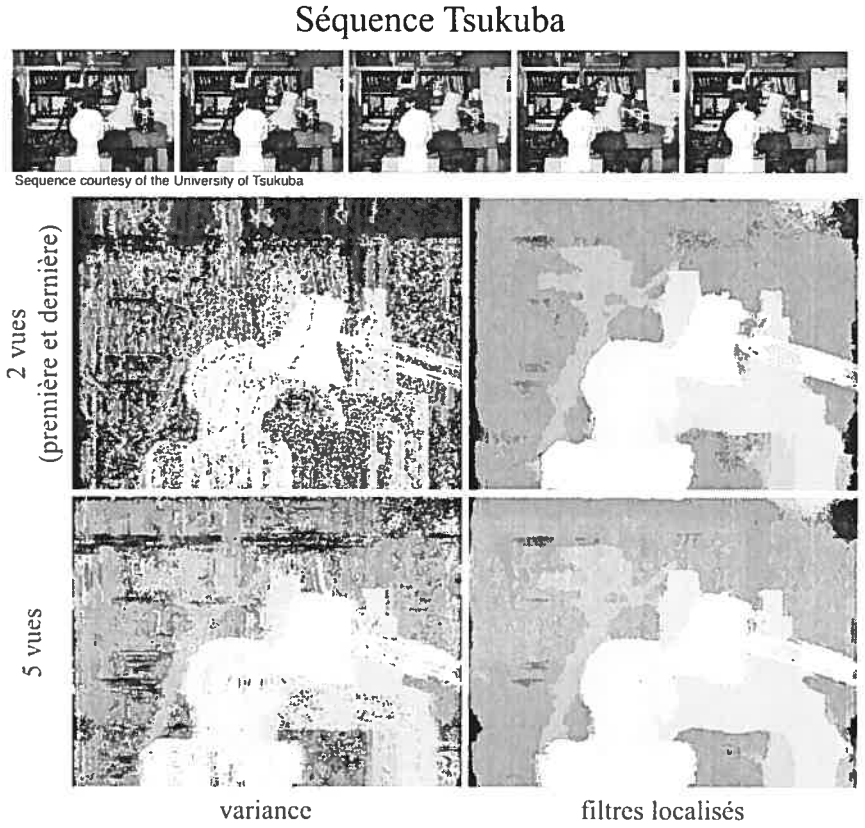


FIG. **5.2. Reconstruction de la scène Tsukuba.** (**Haut**) Reconstruction par recherche directe (aucun lissage) à partir de deux images avec une fonction de coût de variance et notre fonction de coût. (**Bas**) Reconstruction à partir des cinq images. La reconstruction s'est faite avec 16 disparités ; pour notre fonction de coût, 40 filtres ont été utilisés, avec $\alpha = 2\pi/8$ et $\omega = \{4, 9, 14, 19, 24\}$.

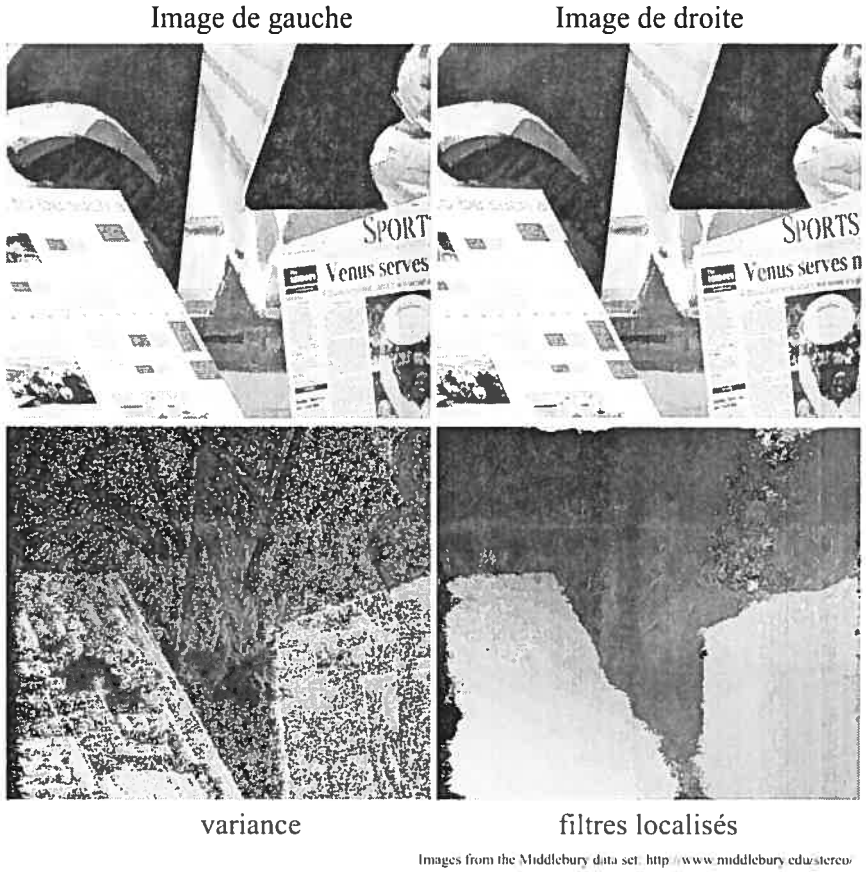


FIG. **5.3. Reconstruction de la scène Venus.** (**Haut**) Pair d'images stéréo «Venus». (**Bas**) Reconstruction par recherche directe (aucun lissage) à partir de deux images avec une fonction de coût de variance et notre fonction de coût. La reconstruction s'est faite avec 20 disparités ; pour notre fonction de coût. 40 filtres ont été utilisés, avec $\alpha = 2\pi/8$ et $\omega = \{4, 9, 14, 19, 24\}$.

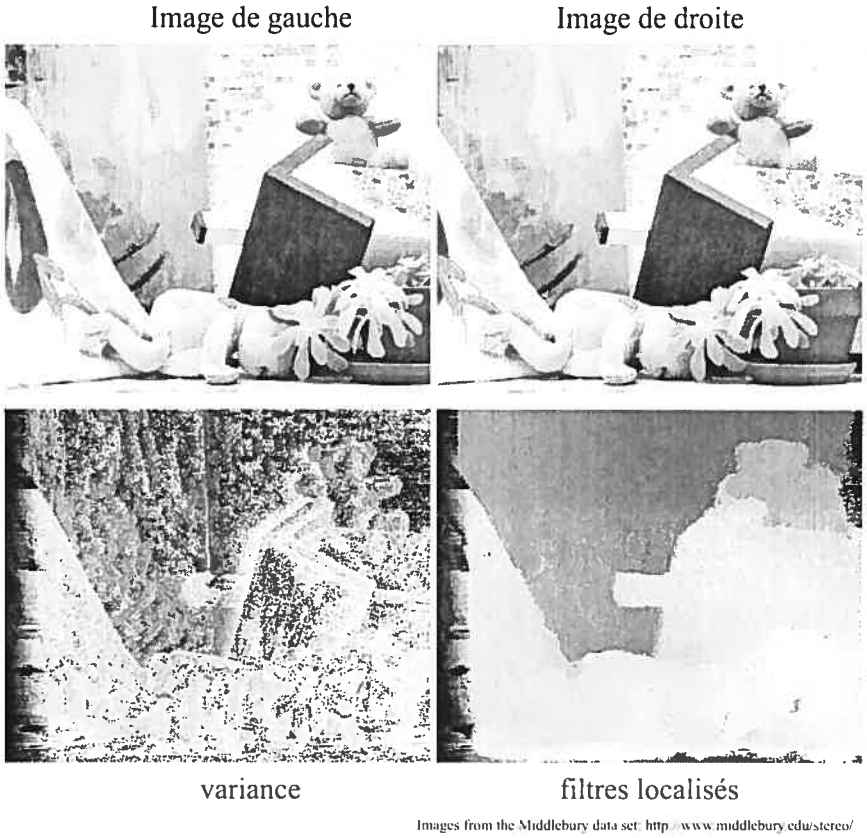


FIG. **5.4. Reconstruction de la scène Teddy. Haut** Pair d'images stéréo «Teddy». **Bas** Reconstruction par recherche directe (aucun lissage) à partir de deux images avec une fonction de coût de variance et notre fonction de coût. La reconstruction s'est faite avec 60 disparités ; pour notre fonction de coût. 40 filtres ont été utilisés, avec $\alpha = 2\pi/8$ et $\omega = \{4, 9, 14, 19, 24\}$.

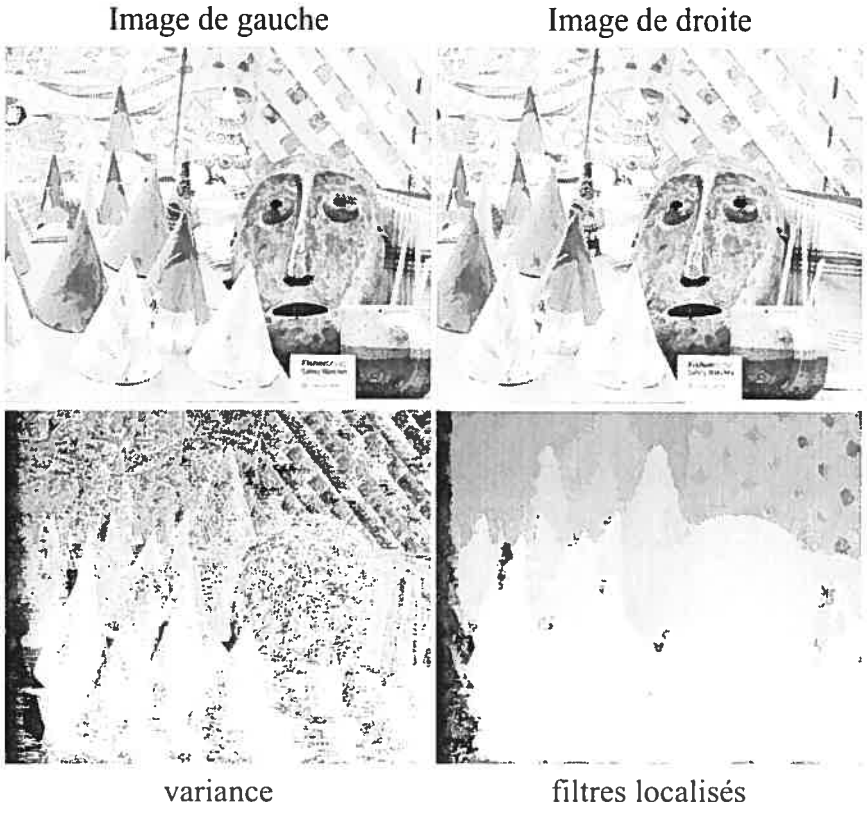


FIG. **5.5. Reconstruction de la scène Cone. Haut** Pair d'images stéréo «Cone». **Bas** Reconstruction par recherche directe (aucun lissage) à partir de deux images avec une fonction de coût de variance et notre fonction de coût. La reconstruction s'est faite avec 60 disparités ; pour notre fonction de coût, 60 filtres ont été utilisés, avec $\alpha = 2\pi/8$ et $\omega = \{4, 9, 14, 19, 24\}$.

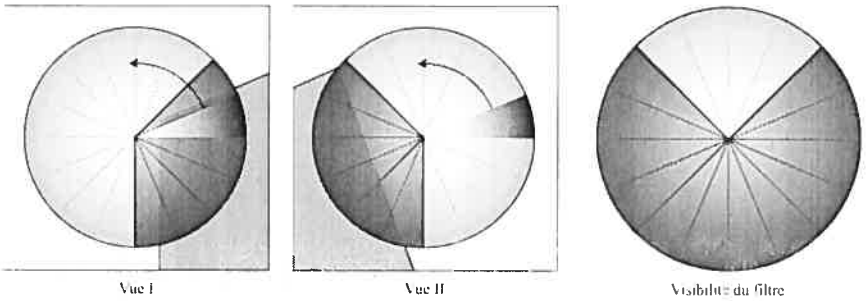


FIG. **5.6. Gestion des occlusions en fonction de la géométrie. Gauche et milieu** La signature de deux pixels est comparée sur deux vues de la même scène. On détecte sur chaque vue une discontinuité de profondeur sur le support de certains filtres. **Droite** Certains filtres de notre signature sont ignorés et la corrélation des signatures se fait sur la portion restante.

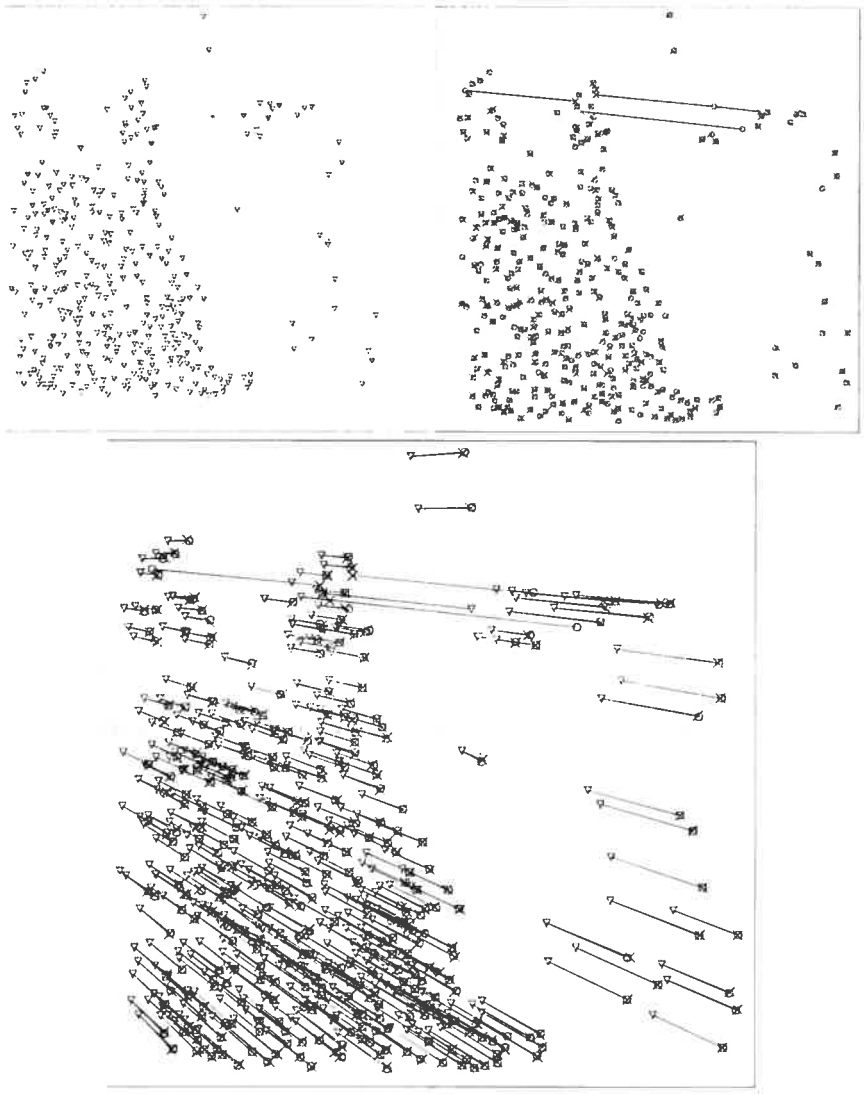


FIG. **5.7.** (**Haut à gauche**) Points d'intérêts pour lesquels une correspondance à été recherchée. (**Haut à droite**) Différence entre les correspondances trouvées par SIFT (cercles) et notre méthode (croix). Certaines croix n'apparaissent pas lorsque le point de d'intérêt a été détecté comme étant en occlusion avec notre méthode.

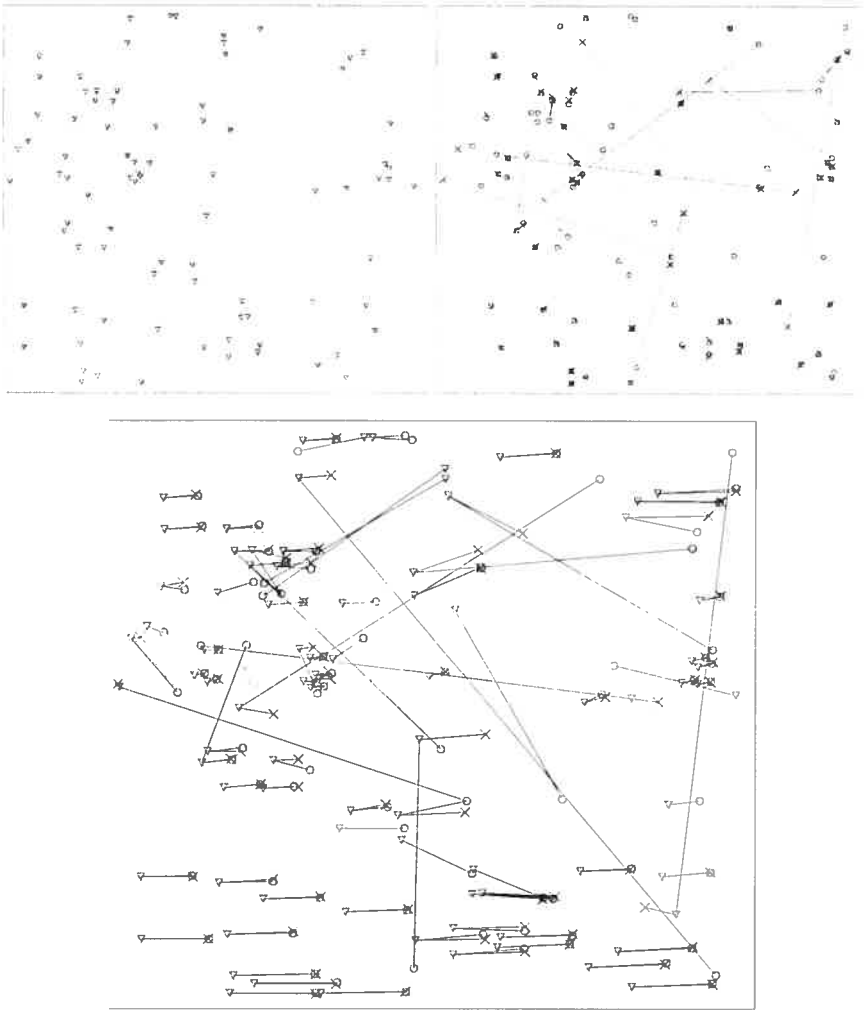


FIG. **5.8.** (**Haut à gauche**) Points d'intérêts pour lesquels une correspondance à été recherchée. (**Haut à droite**) Différence entre les correspondances trouvées par SIFT (cercles) et notre méthode (croix). Certaines croix n'apparaissent pas lorsque le point de d'intérêt a été détecté comme étant en occlusion avec notre méthode.

Chapitre 6

CONCLUSION

Nous avons présenté en §3 une nouvelle méthode qui permet de résoudre le flux optique sans terme de régularisation. Nous y sommes parvenu en nous inspirant des méthodes de flux optique par phase, et en modifiant le support des filtre de sorte que ce dernier soit réduit au minimum. De plus, le décentrage de nos filtre permet une excellente localisation et une gestion des occlusions qui était impossible avec des filtres symétriques. Enfin puisque la méthode utilise des filtres de longueur d'onde variable, elle intègre directement les avantage des techniques multirésolutions.

Pour obtenir une méthode réellement robuste, il faudrait ajouter un terme de régularisation. Il reste beaucoup de travail à faire de ce côté, puisqu'il n'est pas évident que notre terme d'information réagira au lissage de la même façon qu'un terme traditionnel de gradient.

Nous avons également proposé en §4 une nouvelle méthode pour résoudre les plans de mouvements. La méthode présentée utilise une approche par vote plutôt que de trouver une solution analytique, ce qui lui permet d'être robuste au bruit et de détecter les mouvements multiples. Il est fort probable que cette méthode puisse également s'appliquer à d'autres problème où on doit résoudre un système linéaire dont les échantillons sont bruités ou proviennent de plusieurs classes.

Notre méthode s'applique à l'estimation du mouvement, mais comme nous l'avons montré en §5. la stéréo et la correspondance de points pourraient aussi en bénéficier.

RÉFÉRENCES

[1] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, 1902.

[2] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 869–889, 1988.

[3] B. K. P. Horn and B. G. Schunck, "Determining optical flow.," *Artificial. Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[4] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of International Joint Conference on Artificial Intelligence*, (Vancouver, Canada), pp. 674–679, 1981.

[5] M. Yachida, "Determining velocity map by 3-d iterative estimation.," in *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 24–28, 1981.

[6] M. Black and P. Anandan, "Robust dynamic motion estimation over time," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (Maui, Hawaii, USA), pp. 296–302, June 1991.

[7] J. Weickert and C. Schnörr, "Variational optic flow computation with a spatio-temporal smoothness constraint," *Journal of Mathematical Imaging and Vision*, vol. 14, no. 3, pp. 245–255, 2001.

[8] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proceedings of European Conference on Computer Vision*, (London, UK), pp. 237–252, Springer-Verlag, 1992.

[9] N. H. H., "Displacement vectors derived from second-order intensity variations in image sequences," *Computer Graphics Image Processing*, vol. 21, pp. 85–117, 1983.

[10] L. Alvarez, J. Weickert, and J. Sánchez, "Reliable estimation of dense optical flow fields with large displacements," *International Journal of Computer Vision*, vol. 39, no. 1, pp. 41–56, 2000.

[11] A. Singh. "An estimation-theoretic framework for image-flow computation," in *International Conference on Computer Vision*, pp. 168–177, 1990.

[12] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger, "Probability distributions of optical flow," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (Maui, Hawaii, USA), pp. 310–315, IEEE Computer Society, 1991.

[13] S. Roy and V. Govindu, "Mrf solutions for probabilistic optical flow formulations," in *International Conference on Pattern Recognition*, pp. 7053–7059, 2000.

[14] J. Konrad and E. Dubois, "Estimation of image motion fields : Bayesian formulation and stochastic solution," in *Proceedings of Int. Conf. Accoust. Speech Signal Processing*, pp. 1072–1075, 1988

[15] B. Y. Betsch, W. Einhäuser, K. P. Körding, and P. König, "The world from a cat's perspective : Statistics of natural videos," *Biological Cybernetics*, vol. 90, pp. 41–50, 2004.

[16] J. Huang and D. Mumford, "Statistics of natural images and models," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 23–25, June 1999.

[17] S. Roth and M. J. Black, "On the spatial statistics of optical flow," in *International Conference on Computer Vision*, vol. I, pp. 42–49, October 2005.

[18] J. Barron. D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[19] C. Fermüller, D. Shulman, . and Y. Aloimonos, "The statistics of optical flow," *Computer Vision and Image Understanding*, vol. 82, no. 1, pp. 1–32, 2001.

[20] M. Okutomi and T. Kanade, "A locally adaptive window for signal matching," *International Journal of Computer Vision*, vol. 7, no. 2, pp. 143–162, 1992.

[21] J. Bergen, P. Burt, R. Hingorani, and S. Peleg, "Computing two motions from three frames," in *International Conference on Computer Vision*, pp. 27–32, 1990.

[22] J. A. and M. Black, "Mixture models for optical flow computation, in partitioning data sets," *DIMACS Workshop*, pp. 271–286, April 1993.

[23] W.-G. Chen, G. Giannakis, and N. Nandhakumar, "A harmonic retrieval framework for discontinuous motion estimation," *IEEE Transactions on Image Processing*, vol. 7, no. 9, pp. 1242–1257, 1998.

[24] R. Mann and M. Langer, "Estimating camera motion through a 3d cluttered scene," in *First Canadian Conference on Computer and Robot Vision (CVR 2004)*, (London, Ontario, Canada), pp. 472–479, 2004.

[25] M. Pingault, *Estimations fréquentielle et temporelle du mouvement en transparence additive dans les séquences d'images*. PhD thesis, Laboratoire des Images et des Signaux, Grenoble, France, 2003.

[26] W. Yu, G. Sommer, and K. Daniilidis, "Three dimensional orientation signatures with conic kernel filtering for multiple motion analysis," *Image and Vision Computing*, vol. 21, no. 5, pp. 447–458, 2003.

[27] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping.," in *Proceedings of European Conference on Computer Vision*, vol. 4, (Heidelberg), pp. 25–36, Springer-Verlag Berlin, May 2004.

[28] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck : Combining local and global optic flow methods." *International Journal of Computer Vision*, vol. 61, no. 3, pp. 211–231, 2005.

[29] M. J. Black and P. Anandan, "The robust estimation of multiple motions :

parametric and piecewise-smooth flow fields." *Comput. Vis. Image Underst.*, vol. 63, no. 1, pp. 75–104, 1996.

[30] E. D. Castro and C. Morandi, "Registration of translated and rotated images using finite fourier transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 700–703, 1987.

[31] S. Alliney and C. Morandi, "Digital image registration using projections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 222–233, March 1986.

[32] H. Stone, B. Tao, and M. McGuire, "Analysis of image registration noise due to rotationally dependent aliasing," *Journal of Visual Communication and Image Representation*, vol. 14, pp. 114–135, June 2003.

[33] M. McGuire and H. S. Stone, "Techniques for multiresolution image registration in the presence of occlusions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 1476–1479, May 2000.

[34] D. Heeger, "Optical flow from spatiotemporal filters," *International Journal of Computer Vision*, vol. 1, pp. 279–302, January 1988.

[35] J. K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images – a review," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 917–935, 1988.

[36] A. Mitiche, Y. F. Wang, and J. K. Aggarwal, "Experiments in computing optical flow with the gradient-based, multiconstraint method," *Pattern Recognition*, vol. 20, no. 2, pp. 173–179, 1987.

[37] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, November 2004.

[38] D. B. A. and N. P., "Optical flow computation using extended constraints," *IEEE Transactions on Image Processing*, vol. 5, no. 5, 1996.

[39] B. Galvin, B. McCane, K. Novins, D. Mason, and S. Mills, "Recovering motion fields : An evaluation of eight optical flow algorithms.," in *Electronic Proceedings of the British Machine Vision Conference*, 1998.

[40] R. Haralick and J. Lee, "The facet approach to optic flow," in *Image Understanding Workshop*, pp. 84–93, 1983.

[41] O. Tretiak and L. Pastor, "Velocity estimation from image sequences with second order differential operators.," in *International Conference on Pattern Recognition*, (Montreal, Canada), pp. 16–19, 1984.

[42] E. C. Hildreth, *Computations underlying the measurement of visual motion*. Norwood, NJ, USA : Ablex Publishing Corp., 1987.

[43] W. B. Thompson, "Combining motion and contrast for segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 543–549, 1980.

[44] W. Thompson, K. Mutch, and V. Berzins, "Dynamic occlusion analysis in optical flow fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 374–383, July 1985.

[45] A. Yuille and T. Poggio, "Scaling theorems for zero crossings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 15–25, 1986.

[46] J. K. Kearney, W. B. Thompson, and D. L. Boley, "Optical flow estimation : an error analysis of gradient-based methods with local optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 2, pp. 229–244, 1987.

[47] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 2, pp. 220–228, 1987.

[48] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," *Biological Cybernetics*, vol. 60, pp. 79–87, 1988.

[49] D. Shulman and J. Herve, "Regularization of discontinuous flow fields," in *IEEE Workshop on Visual Motion*, (Irvine, CA), pp. 81–90, IEEE Computer Society Press, 1989.

[50] J. Marroquin, *Probabilistic Solution of Inverse Problems*. PhD thesis, MIT AI-TR, 1985.

[51] D. Mumford and J. Shah, "Boundary detection by minimizing functionals, I," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 22–26, 1985.

[52] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[53] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," in *International Journal of Computer Vision*, vol. 2, pp. 283–310, 1989.

[54] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 77–104, 1990.

[55] J. Y. A. Wang and E. H. Adelson, "Layered representation for motion analysis," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (New York), pp. 361–366, 1993.

[56] D. Vanderbilt and D. G. Louie, "A monte carlo simululated annealing approach to optimization over continuous variables," *Journal of Computer Physics*, vol. 56, pp. 259–271, 1984.

[57] F. Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow

using markov random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 12, pp. 1217–1232, 1993.

[58] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society*, vol. B, no. 48, pp. 282–284, 1986.

[59] G. J. McLachlan and K. E. Basford, *Mixture models : Inference and applications to clustering*. Statistics : Textbooks and Monographs. New York : Dekker, 1988, 1988.

[60] A. M. Waxman, J. Wu, and F. Bergholm, "Convected activation profiles and receptive fields for real time measurement of short range visual motion," in *CVPR*, pp. 717–723, 1988.

[61] C. Schnörr, "Segmentation of visual motion by minimizing convex non-quadratic functionals," in *International Conference on Pattern Recognition*, vol. A, pp. 661–663, 1994.

[62] T. Camus, *Real-Time Optical Flow*. PhD thesis, Brown University, 1994.

[63] A. D. J. Shanon X. Ju, Michael J. Black, "Skin and bones : Multi-layer, locally affine, optical flow and regularization with transparency," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (Washington, DC, USA), p. 307, IEEE Computer Society, 1996.

[64] M. Proesmans, L. van Gool, E. Pauwels, and A. Oosterlinck, "Determination of optical flow and its discontinuities using non-linear diffusion," in *Proceedings of European Conference on Computer Vision*, vol. 2, (Secaucus, NJ, USA), pp. 295–304, Springer-Verlag New York, Inc., 1994.

[65] E. Mémin and P. Pérez, "A multigrid approach for hierarchical motion estimation," in *International Conference on Computer Vision*, (Washington, DC, USA), pp. 933–938, IEEE Computer Society, 1998.

[66] C. Stiller and J. Konrad, "Estimating motion in image sequences," *IEEE Signal Processing Magazine*, vol. 16, pp. 70–91, 1999.

[67] G. Aubert, R. Deriche, and P. Kornprobst, "Computing optical flow via variational techniques," *SIAM Journal on Applied Mathematics*, vol. 60, no. 1, pp. 156–182, 1999.

[68] H. H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 5, pp. 565–593, 1986.

[69] T. Iijimam, "Basic theory on normalization of pattern (in case of typical one-dimentional pattern)," *Bulletin of the Electrotechnical Laboratory*, vol. 26, pp. 368–388, 1962.

[70] J. Weickert, "Foundations and applications of nonlinear anisotropic diffusion filtering," *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 76, no. 1, pp. 283–286, 1996.

[71] R. El-Feghali and A. Mitiche, "Fast computation of a boundary preserving estimate of optical flow.," in *Electronic Proceedings of the British Machine Vision Conference*, 2000.

[72] J. Weickert, "On discontinuity-preserving optic flow," in *Proceedings of the Computer Vision and Mobile Robotics Workshop*, pp. 115–122, September 1998.

[73] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half quadratic regularization algorithms for computed imaging," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, (Austin, TX), pp. 168–172, IEEE Computer Society Press, 1994.

[74] H. H. Nagel, "Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow," in *Proceedings of European Conference on Computer Vision*, (New York, NY, USA), pp. 139–148, Springer-Verlag New York, Inc., 1990.

[75] P. J. Huber, *Robust Statistics*. New York : John Wiley and Sons, 1981.

[76] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics : The Approach Based on Influence Functions.* New York, NY : John Wiley and Sons, 1986.

[77] J. Weickert, A. Bruhn, T. Brox, and N. Papenberg, "A survey on variational methods for small displacements," tech. rep., Department of Mathematics, Saarland University, Saarbrücken, Germany, September 2005.

[78] G. Petit and S. Roy, "Solving motion planes by projection and ring integration," in *IAPR Conference on Machine Vision Applications*, (Tsukuba, Japan), pp. 1–4, May 2005.

[79] H. Liu, R. Chellappa, and A. Rosenfeld, "Fast two-frame multi-scale dense optical flow estimation using discrete wavelet filters," *Journal of Optical Society of America*, vol. 20, no. 8, pp. 1505–1515, 2003.

[80] J. Magarey and N. Kingsbury, "Motion estimation using a complex-valued wavelet transform," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1069–1084, April 1998.

[81] C. P. Bernard, "Discrete wavelet analysis : A new framework for fast optic flow computation," in *Proceedings of European Conference on Computer Vision*, (London, UK), pp. 354–368, Springer-Verlag, 1998.

[82] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, April-June 2002.

[83] Middlebury, "Middlebury college : Stereo vision research page," 2005.