

Calcul du flot optique par filtrages de Gabor combinés

Optical Flow Based on Combined Gabor Filters

par Denis PELLERIN, Adrian SPINÉI, Anne GUÉRIN-DUGUÉ

Laboratoire de Traitement d'Images et Reconnaissance de Formes
I.N.P.G. - 46, avenue Félix Viallet F-38031 Grenoble cedex

résumé et mots clés

Nous présentons une méthode fréquentielle de calcul du flot optique à base de filtres spatio-temporels de Gabor, inspirée de l'architecture introduite par Heeger en 1987. Cette méthode, dite de filtrages combinés, associe des filtres de sélectivités différentes. Elle permet de réduire les erreurs d'estimation dans le cas d'une caméra fixe, en s'appuyant sur la détermination préalable des zones en mouvement. Nous montrons comment réaliser l'intégration automatique des résultats d'estimation de trois niveaux d'une pyramide spatiale pour étendre la plage de vitesse admissible. Nous proposons enfin une implantation numérique efficace exploitant les symétries temporelles et spatiales des filtres de Gabor 3D.

Estimation du mouvement, Flot optique, Méthode fréquentielle, Filtre de Gabor.

abstract and key words

We shall present a frequential method to calculate optical flow based on spatio-temporal Gabor filters, inspired from the architecture of Heeger introduced in 1987. This method, named combined filtering, associates filters of different selectivities. A previous determination of the moving regions permits, in the case of a fixed camera, to reduce the estimation errors. We show how to automatically integrate the estimation results for three levels of a spatial pyramid in order to extend the domain of the allowed velocities. We propose also an efficient numerical implementation exploiting the temporal and spatial symmetries of the 3D Gabor filters.

Motion estimation, Optical flow, Frequential method, Gabor filter.

1. introduction

Les problèmes abordés par l'analyse du mouvement dans les séquences d'images sont la détection et l'estimation du mouvement, la segmentation en régions de mouvement homogène (au sens d'un modèle de mouvement donné), l'interprétation tridimensionnelle du mouvement. Notre intérêt s'est plus particulièrement porté sur les méthodes fréquentielles d'estimation du mouvement, relativement récentes et très prometteuses.

Dans cet article, nous proposons une nouvelle méthode de calcul du flot optique à base de filtres spatio-temporels de Gabor, inspirée de l'architecture introduite par Heeger [1]. Nous présentons le principe des méthodes fréquentielles et l'approche par banc de filtres de Gabor 3D (paragraphe 2). Nous décrivons notre méthode, dite de filtrages combinés, de calcul du flot optique qui repose sur la combinaison de filtres de sélectivités différentes et la détermination préalable des zones en mouvement (paragraphe 3). Nous montrons aussi comment réaliser l'intégration automatique

des résultats d'estimation de plusieurs niveaux de pyramide spatiale pour étendre la plage de vitesse autorisée. Nous proposons enfin une implantation numérique efficace exploitant les symétries temporelles et spatiales des filtres de Gabor 3D (paragraphe 4).

2. une méthode fréquentielle à base de filtres de Gabor

2.1. principe des méthodes fréquentielles

Les méthodes fréquentielles sont fondées sur une caractérisation du mouvement dans le domaine des fréquences [2]. Elles ont pour origine des recherches concernant la vision des mammifères. Ces

recherches avaient mis en évidence la présence de cellules simples dans l'aire corticale V1 [3] et de cellules complexes dans l'aire MT [4], qui se comportent comme des filtres passe-bande spatio-temporels. Les méthodes fréquentielles présentent de nombreux atouts tels que la simplicité de l'interprétation physique ou la cohérence avec d'autres traitements (compression, restauration d'images) utilisant les mêmes modèles.

Soit $g(x, y, t)$ une image en mouvement et $G(f_x, f_y, f_t)$ sa transformée de Fourier, où x et y représentent les coordonnées spatiales, t le temps, f_x et f_y les fréquences spatiales et f_t la fréquence temporelle. Soit $i(x, y)$ la même image mais fixe, et $I(f_x, f_y)$ sa transformée de Fourier. Lorsque toute l'image est en translation à vitesse uniforme $v(v_x, v_y)$, on peut écrire la relation :

$$g(x, y, t) = i(x + v_x t, y + v_y t)$$

d'où l'on déduit par transformée de Fourier,

$$G(f_x, f_y, f_t) = I(f_x, f_y) \cdot \delta(f_t - v_x f_x - v_y f_y)$$

où δ représente la distribution de Dirac. Cette dernière expression montre que le spectre d'énergie pour une image 2D en translation a des valeurs nulles partout sauf dans un plan passant par l'origine dont l'équation est :

$$f_t = v_x f_x + v_y f_y$$

Le principe des méthodes fréquentielles consiste à localiser le plan d'énergie non nulle ou plan de vitesse au moyen de filtres spatio-temporels orientés, pour obtenir ensuite directement les composantes du vecteur vitesse (v_x, v_y) . Plusieurs transformations ont été utilisées pour construire de tels filtres comme les transformées de Fourier court terme [5], de Wigner-Ville [6], de Gabor [2, 7, 8, 9] et les tenseurs [10].

2.2. approche par banc de filtres de gabor 3D

Les filtres spatio-temporels ne sont pas sélectifs à une vitesse particulière mais plutôt réglés sur des fréquences spatio-temporelles : ils ne permettent donc pas à eux seul l'estimation de vitesse. Seul Heeger [1, 9] a proposé une méthode de calcul du flot optique en combinant les réponses fournies par de tels filtres. L'approche consiste à échantillonner l'espace des fréquences spatio-temporelles au moyen de filtres de Gabor 3D (étape « filtrages ») puis, par une technique de minimisation, de déterminer le plan de vitesse correspondant le mieux aux réponses fournies par l'ensemble des filtres (étape « minimisation »). Cette méthode suppose que l'image est suffisamment texturée pour que le spectre associé soit distribué sur une large gamme de fréquences spatio-temporelles. Un filtre de Gabor 3D à phase cosinus est représenté comme le produit d'une fonction spatio-temporelle Gaussienne et d'une fonction cosinus :

$$g(x, y, t) = \frac{1}{(\sqrt{2\pi})^3 \sigma_x \sigma_y \sigma_t} \exp\left(-\left\{\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} + \frac{t^2}{2\sigma_t^2}\right\}\right) \cos(2\pi(f_{x0}x + f_{y0}y + f_{t0}t))$$

où (f_{x0}, f_{y0}, f_{t0}) est la fréquence centrale du filtre et $(\sigma_x, \sigma_y, \sigma_t)$ l'écart-type associé à la fonction spatio-temporelle Gaussienne. La réponse fréquentielle d'un filtre de Gabor à phase cosinus est donnée par :

$$G(f_x, f_y, f_t) = \frac{1}{2} \exp\{-2\pi^2[\sigma_x^2(f_x - f_{x0})^2 + \sigma_y^2(f_y - f_{y0})^2 + \sigma_t^2(f_t - f_{t0})^2]\} + \frac{1}{2} \exp\{-2\pi^2[\sigma_x^2(f_x + f_{x0})^2 + \sigma_y^2(f_y + f_{y0})^2 + \sigma_t^2(f_t + f_{t0})^2]\}$$

La somme des carrés des réponses de deux filtres de Gabor en quadrature (l'un à phase cosinus, l'autre à phase sinus mais d'orientation et de largeur de bande identique) donne une mesure d'énergie indépendante de la phase du signal. Le filtre obtenu est appelé filtre d'énergie de Gabor.

L'architecture de filtrage consiste à placer un ensemble de 36 filtres d'énergie de Gabor orientés, répartis en trois gammes de vitesse. Pour chaque gamme de vitesse, les 12 filtres sont positionnés sur un cylindre à différentes fréquences spatiales et temporelles (Figures 1 et 2). Dans le domaine des fréquences spatiales, la fréquence centrale (ou la position) d'un filtre pourra s'exprimer en coordonnées cartésiennes (f_{x0}, f_{y0}) ou plus naturellement en coordonnées polaires (ρ, φ) avec $\rho = \sqrt{f_x^2 + f_y^2}$ et $\varphi = \text{Arctg}(f_y/f_x)$.

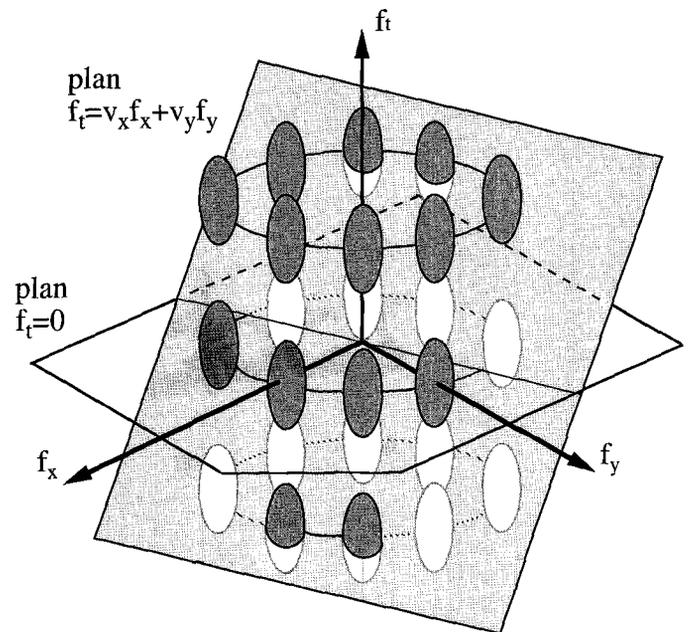


Figure 1. - Spectres de 12 filtres spatio-temporels orientés de Gabor pour une seule gamme de vitesse. Chaque filtre a la forme d'une paire d'ellipsoïdes (Gaussienne 3D) centrées à (f_{x0}, f_{y0}, f_{t0}) et $(-f_{x0}, -f_{y0}, -f_{t0})$.

L'estimation de vitesse est formulée comme une minimisation de l'écart quadratique entre les réponses mesurées des filtres pour une image quelconque et les réponses prédites pour une image de

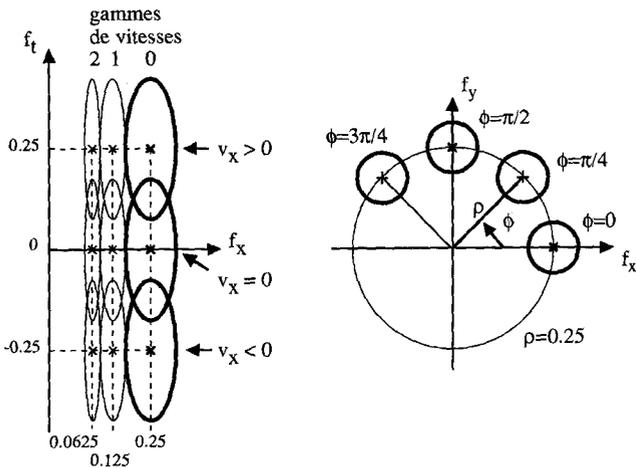


Figure 2. – Positions des filtres : a) 3 filtres dans le plan (f_x, f_t) : un pour chaque groupe de vitesses ($v_x > 0$, $v_x = 0$ et $v_x < 0$), ceci est répété pour les 3 gammes de vitesses 0, 1 et 2, b) 4 filtres à des orientations différentes dans le plan (f_x, f_y) (une seule gamme de vitesse est représentée).

bruit blanc en translation. Dans ce dernier cas, le plan de vitesse correspond à un spectre d'énergie constant. La réponse prédite $R(v_x, v_y)$ fournie par un filtre d'énergie de Gabor positionné aux fréquences (f_{x0}, f_{y0}, f_{t0}) et d'écart-type $(\sigma_x, \sigma_y, \sigma_t)$ pour une image de bruit blanc en translation, est exprimée comme une fonction de la vitesse [9] :

$$R(v_x, v_y) = \exp \left[\frac{-4\pi^2 \sigma_x^2 \sigma_y^2 \sigma_t^2 (v_x f_{x0} + v_y f_{y0} + f_{t0})^2}{(v_x \sigma_x \sigma_t)^2 + (v_y \sigma_y \sigma_t)^2 + (\sigma_x \sigma_y)^2} \right]$$

Pour éliminer les problèmes liés aux images dont le contraste est différent suivant les directions, la réponse fournie par chaque filtre est normalisée par la somme des réponses des filtres ayant la même orientation spatiale (même colonne dans la figure 2a). Soit M_i (avec $1 \leq i \leq 12$) l'ensemble des filtres ayant la même orientation spatiale et soient \bar{m}_i et \bar{R}_i la somme des réponses mesurées m_j et prédites R_j , délivrées par les filtres j de l'ensemble M_i d'où :

$$\bar{m}_i = \sum_{j \in M_i} m_j \quad \text{et} \quad \bar{R}_i = \sum_{j \in M_i} R_j$$

La vitesse s'obtient en cherchant le minimum de l'expression :

$$f(v_x, v_y) = \sum_{i=1}^{12} \left[m_i - \bar{m}_i \frac{R_i(v_x, v_y)}{\bar{R}_i(v_x, v_y)} \right]^2$$

Pour cette recherche de minimum, Heeger suggère une méthode parallèle reposant sur la quantification du domaine des vitesses en module et en orientation.

De manière à limiter la taille des noyaux de convolution des filtres de Gabor pour les fréquences spatiales les plus faibles, une décomposition pyramidale à trois niveaux est utilisée. Les niveaux 0, 1 et 2 couvrent respectivement les gammes de vitesse 0 – 1.25, 1.25 – 2.5 et 2.5 – 5 pixels par image (figure 2a).

Les écart-types des filtres sont $\sigma_s = \sigma_x = \sigma_y = 4$ en spatial et $\sigma_t = 1$ en temporel. Ces paramètres permettent de minimiser l'incertitude sur le mouvement estimé [11, 12]. La bande passante des fréquences temporelles est alors plus large que celle des fréquences spatiales, en accord avec ce qui est observé pour les cellules du cortex sensibles au mouvement [3].

L'ensemble de la méthode repose sur la connaissance de l'énergie locale liée au mouvement : cela impose donc un filtrage par convolution spatiale plutôt qu'une implantation dans le domaine fréquentiel utilisant des transformées de Fourier rapides. Les filtres de Gabor spatio-temporels utilisés sont des ellipsoïdes dans le domaine fréquentiel dont les axes sont parallèles aux axes f_x, f_y, f_t , ce qui autorise une implantation par somme de convolutions séparables (Annexe A).

L'approche par banc de filtres de Gabor 3D, avec les paramètres initiaux, fournit une très bonne estimation des vitesses sur les objets en mouvement mais l'utilisation de filtres avec un grand écart-type spatial ($\sigma_s = 4$) entraîne un effet de lissage spatial excessif et donc une mauvaise localisation des objets en mouvement. De plus, la taille élevée du noyau de convolution (23 coefficients) conduit à un coût de calcul important.

3. méthode de filtrages combinés

Nous proposons une méthode, dite de filtrages combinés, qui permet de réduire les erreurs d'estimation pour le cas d'une caméra fixe en s'appuyant sur la détermination préalable des zones en mouvement. Cette méthode consiste à utiliser conjointement deux mécanismes distincts pour la *détection des zones en mouvement* et l'*estimation de vitesse*. Nous avons choisi d'agir seulement sur la sélectivité des filtres spatiaux, c'est-à-dire l'écart-type σ_s pour exécuter chacune de ces deux tâches, et de conserver globalement l'architecture initiale (nombre et position des filtres) qui réalise un très bon compromis entre la précision de l'estimation et la complexité algorithmique.

3.1. principe et illustration

Notre méthode de filtrage combiné (figure 3) consiste à effectuer successivement deux estimations de vitesses :

- la première, réalisée avec des filtres peu sélectifs (écart-type faible), permet dans le cas d'une caméra fixe, la détection des zones en mouvement. Un pixel est détecté en mouvement lorsque le vecteur vitesse associé admet un module supérieur au plus petit niveau de quantification soit 1.25/8 pixels par image.
- la seconde, réalisée avec des filtres très sélectifs (écart-type important), permet d'estimer correctement le flot optique mais

Calcul du flot optique par filtrage de Gabor

seulement sur les objets en mouvement. Le flot optique final est alors constitué du second flot calculé affiché seulement sur les zones détectées en mouvement.

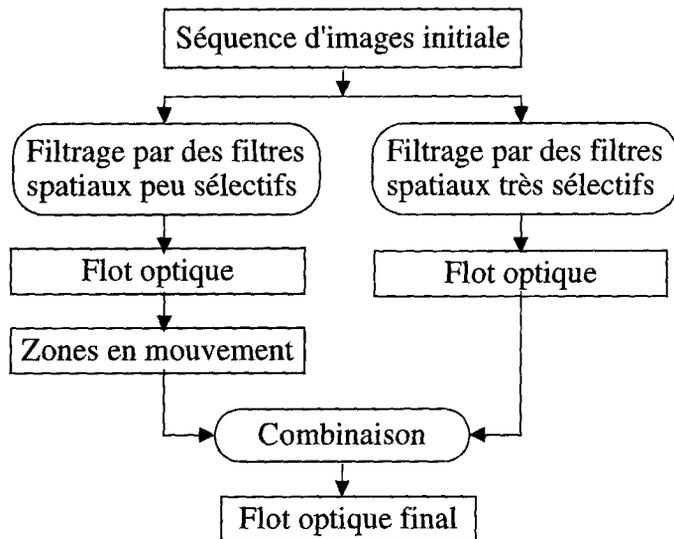


Figure 3. – Principe de la méthode de filtrages combinés.

Nous présentons les résultats obtenus pour la séquence d'images réelle intitulée « Taxi de Hambourg » (Figure 4). Dans cette scène de rue, filmée avec une caméra fixe, quatre objets sont en mouvement : un piéton en haut, un taxi au centre, une voiture à gauche et une camionnette à droite. Les vitesses approximatives de ces objets sont respectivement 0,3, 1, 2 et 2 pixels/image. Pour les deux objets (piéton et taxi) dont les vitesses se situent dans la gamme 0 – 1,25 pixels/image (niveau 0 de la pyramide), le flot optique est bien estimé. Au contraire, pour les deux objets (voiture et camionnette) dont les vitesses sortent de cette gamme, le flot optique est mal estimé. Les résultats détaillés des estimations du flot optique au niveau 0 de la pyramide pour le taxi et la voiture sont fournis figures 8a, 8b et tableau 2. Le champ de vecteurs est lisse pour le taxi et les composantes de vitesse bien estimées, sauf dans les zones de discontinuité de mouvement (moyennes légèrement sous-estimées pour les composantes v_x et v_y). Inversement, les orientations et les modules des vecteurs vitesse pour la voiture sont extrêmement variés (moyennes très faibles pour les composantes v_x et v_y). Afin d'autoriser une large plage de vitesses, nous décrivons au paragraphe 3.3 une méthode d'intégration automatique des résultats des trois niveaux de pyramide.

3.2. choix des écarts-types

Le choix des écart-types spatiaux est un problème complexe, car il dépend en toute rigueur des dimensions des objets en mouvement dans les séquences d'images. Pour nous guider, nous avons utilisé

des séquences d'images artificielles test qui satisfont au mieux les hypothèses liées à la méthode : scènes fortement texturées (spectre uniforme) et mouvements de type translation. Par exemple, la séquence test présentée dans la figure 5a, se compose d'un carré en translation à droite et vers le bas sur un fond immobile. Le carré et le fond sont constitués de textures de type bruit blanc (spectre quasi uniforme). Afin d'analyser ces séquences test, nous avons défini deux critères d'erreur :

– l'erreur sur le flot optique (EF) qui s'exprime comme l'écart quadratique moyen (normalisé) entre les vecteurs vitesse réelle et vitesse estimée en chaque pixel du carré en mouvement (C) :

$$EF(\%) = \frac{\sum_{i \in C} |v_{ri} - v_{ei}|^2}{\sum_{i \in C} |v_{ri}|^2} \times 100$$

– l'erreur sur les zones de changements temporels (EZ) qui s'exprime comme l'écart (normalisé) entre le nombre total de pixels détectés en mouvement dans l'image (D) et le nombre exact de pixels du carré en mouvement (C) :

$$EZ(\%) = \frac{\sum_{i \in D} p_i - \sum_{j \in C} p_j}{\sum_{j \in C} p_j} \times 100$$

Des essais sur plusieurs séquences d'images artificielles et naturelles nous ont permis de sélectionner les écart-types $\sigma_s = 4$ pour l'estimation de vitesse (comme la méthode initiale) et $\sigma_s = 2.4$ pour la détermination des zones en mouvement. Pour ce second cas, nous montrerons dans le paragraphe 4.1 que la valeur d'écart-type est étroitement liée à la taille du noyau de convolution.

Les résultats obtenus sur une séquence test pour $\sigma_s = 4$ et $\sigma_s = 2.4$ sont présentés dans la figure 5 et le tableau 1. Nous vérifions que :

– pour des filtres spatiaux d'écart-type important ($\sigma_s = 4$), le flot optique est correctement estimé (EF faible). Le phénomène de lissage spatial conduit malheureusement à un flot optique qui déborde largement les objets en mouvement.

– à l'inverse, pour les filtres spatiaux d'écart-type plus faible ($\sigma_s = 2.4$), le flot obtenu comporte de nombreux vecteurs vitesse pour lesquels le module est mal estimé. Ces filtres permettent toutefois une bonne détection des zones de changements temporels (EZ faible).

σ_s	$EF(\%)$	$EZ(\%)$
4	3.98	17.57
2.4	5.05	10.28

Tableau 1 : Erreurs sur le flot optique (EF) et sur les zones de changements temporels (EZ) en fonction de l'écart-type pour la séquence test.

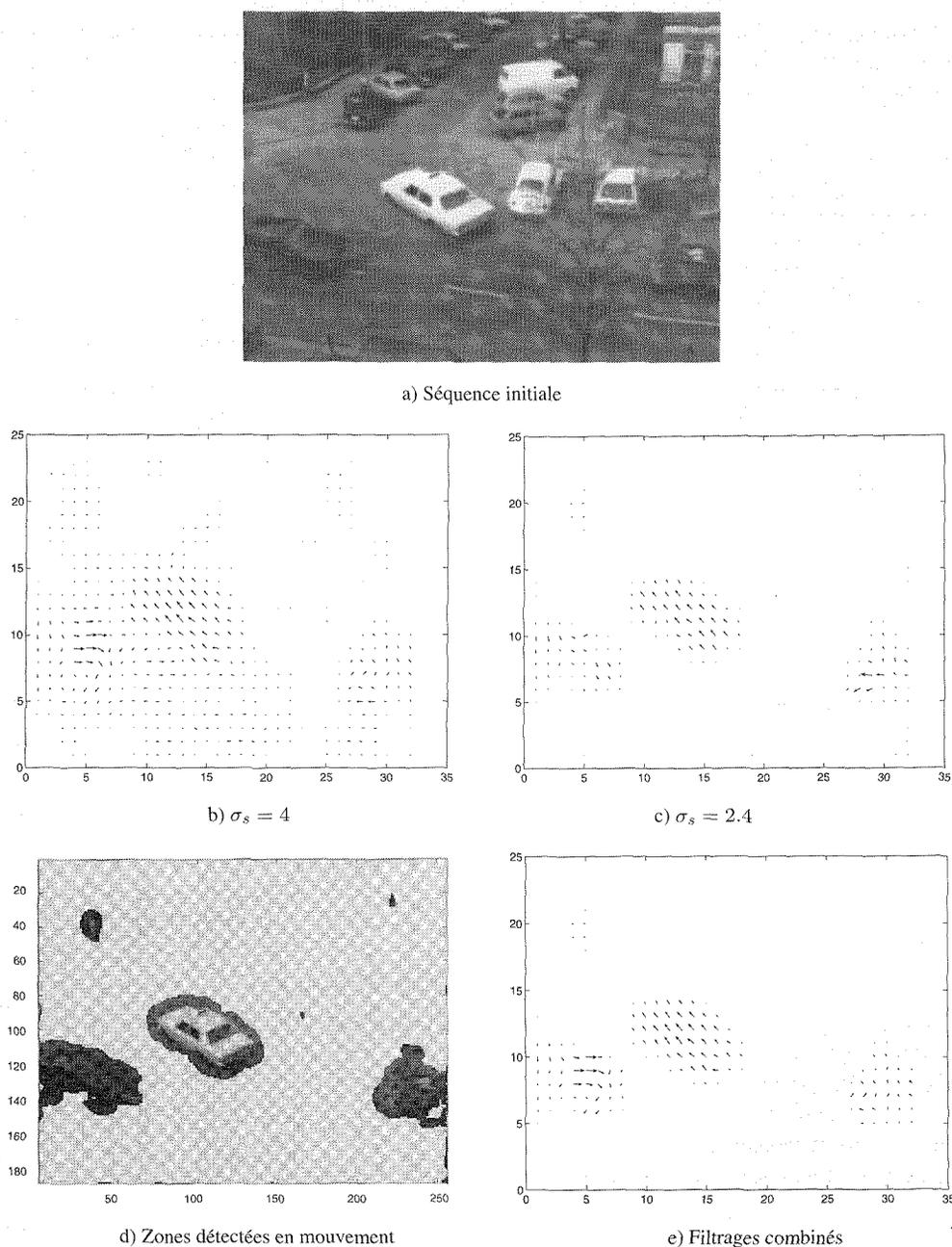


Figure 4. – a) Séquence d'images «Taxi de Hambourg», Flots optiques résultants pour des filtres : b) très sélectifs ($\sigma_s = 4$) et c) peu sélectifs ($\sigma_s = 2.4$), d) Zones détectées en mouvement obtenues à partir de c), e) Flot optique par la méthode des filtres combinés (les flots optiques sont calculés au niveau 0 de la pyramide et moyennés pour l'affichage sur des carrés de taille 8×8 pixels).

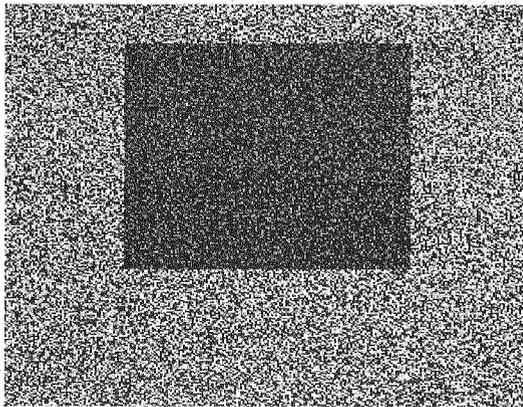
3.3. intégration des trois niveaux de pyramide

La décomposition pyramidale d'images, ou représentation multi-résolution, est une technique classique en traitement d'images. Elle permet à la fois d'améliorer les résultats des traitements et de diminuer le volume des calculs [13]. La construction d'une

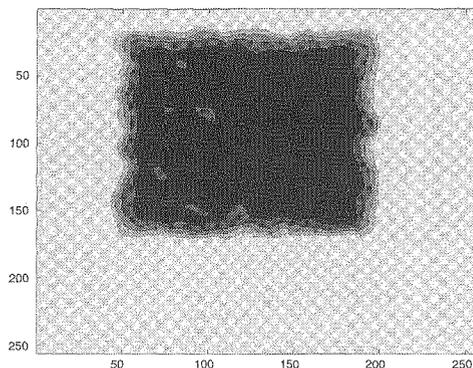
pyramide d'images passe-bas résulte d'une suite de filtres et de re-échantillonnages produisant un ensemble d'images de taille $N \times N, N/2 \times N/2, N/4 \times N/4...$ où N est le nombre de lignes et de colonnes de l'image initiale.

L'approche multi-résolution est aujourd'hui massivement utilisée, dans l'analyse du mouvement, à des fins d'estimation des grands déplacements [14]. En effet, un déplacement important dans l'image initiale (haute résolution) apparaît comme un

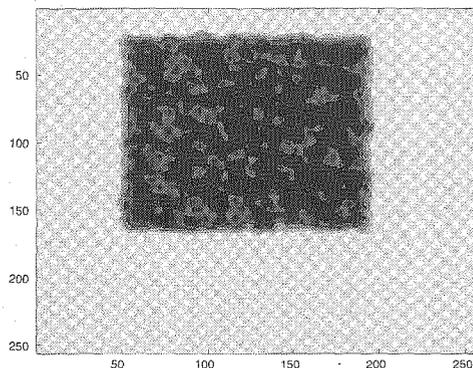
Calcul du flot optique par filtrage de Gabor



a) Séquence test



b) $\sigma_s = 4$



c) $\sigma_s = 2.4$

Figure 5. – a) Séquence test d'un carré tecturé en translation, Répartition des modules des vecteurs vitesse représentés sous forme de niveaux de gris pour b) $\sigma = 4$, c) $\sigma_S = 2.4$, la couleur noire correspond à une estimation exacte du module (les orientations ne comportent aucune erreur).

petit déplacement dans une image à plus faible résolution. Cela permet de respecter la règle suivante : pour être estimé correctement, le déplacement maximum des objets dans une image doit être inférieur à la demi-période de la plus haute fréquence spatiale présente dans l'image [15].

Dans notre cas, l'estimation de vitesse pour chaque objet en mouvement ne peut fournir de résultats satisfaisants que si elle est effectuée dans la gamme de vitesse adéquate (niveau de pyramide

0, 1 ou 2 respectivement pour la gamme de vitesse 0 – 1.25, 1.25 – 2.5 ou 2.5 – 5 pixels/image). Le choix du niveau de la pyramide doit être réalisé sans connaissance a priori des vitesses des objets dans la scène. Heeger n'a pas proposé de solution pour intégrer les résultats de l'estimation obtenus avec plusieurs niveaux de pyramide.

Nous observons que lorsque la vitesse d'un objet en mouvement est correctement estimée à un niveau donné de pyramide (gamme de vitesse convenable), le flot optique calculé dans cette même zone est nul aux niveaux supérieurs de la pyramide. Inversement, le flot optique calculé pour ce même objet aux niveaux inférieurs n'est pas nul, mais le phénomène de repliement temporel entraîne des erreurs extrêmement importantes. Par exemple, une grande vitesse de sens positif peut être estimée au niveau 0 comme une petite vitesse de sens négatif (Figure 2a).

La méthode, à la fois simple et efficace, que nous présentons, comporte trois étapes (figure 6) :

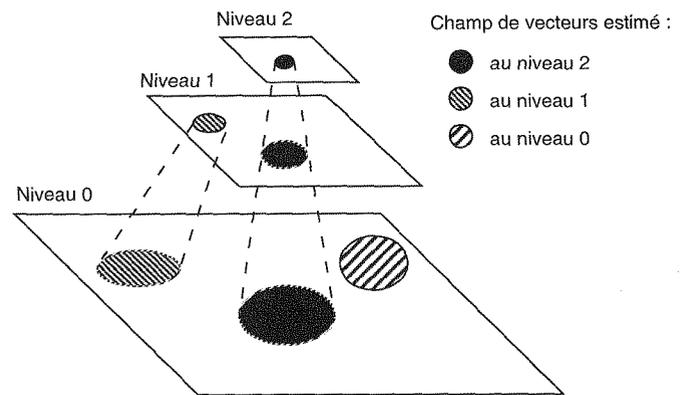
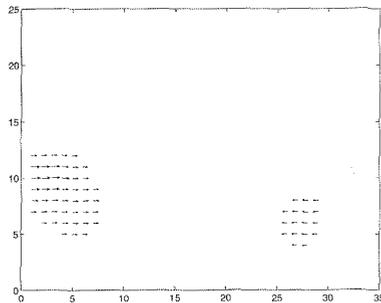


Figure 6. – Propagation des champs de vitesse au sein de la pyramide.

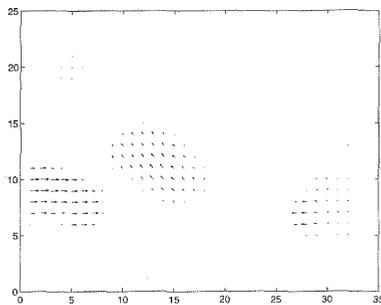
- Estimation séparée sur chaque niveau de pyramide (0, 1 et 2) par filtrages combinés,
- Propagation des vecteurs vitesse du niveau 2 vers le niveau 1. Les vecteurs vitesse non nuls qui apparaissent au niveau 2 vont remplacer les vecteurs vitesse correspondants au niveau 1. La perte de résolution fait que chaque vecteur vitesse au niveau 2 se transforme au niveau 1 en 4 vecteurs identiques disposés aux quatre coins d'un carré.
- Propagation des vecteurs vitesse du niveau 1 vers le niveau 0 (y compris les nouveaux vecteurs) de la même manière.

Ainsi, pour la séquence Taxi de Hambourg, le flot optique est bien estimé au niveau 1 de la pyramide pour les deux objets (voiture et camionnette) dont les vitesses se situent dans la gamme 1.25 – 2.5 pixels/image (figure 7a). L'intégration des niveaux de la pyramide conduit à une bonne estimation de vitesse pour l'ensemble des objets en mouvement (figure 7b). La figure 8c et le tableau 2 présentent plus précisément les résultats obtenus pour la voiture. Le champ de vecteurs est lisse et les composantes de vitesse bien

estimés sauf au voisinage du contour (moyennes légèrement sous-estimées pour les composantes v_x et v_y du fait des discontinuités de mouvement mais aussi des effets de bords liés à la pyramide).

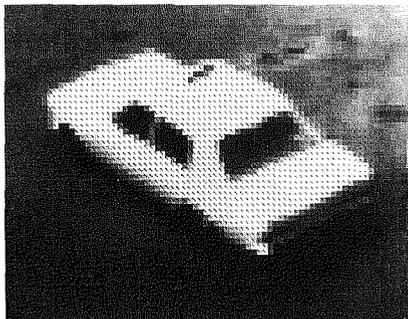


a) Niveau 1

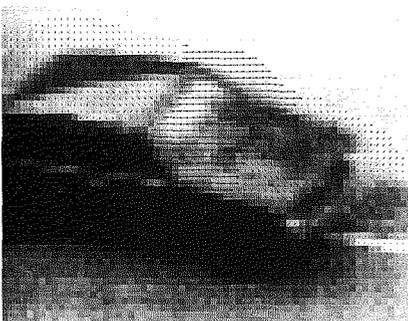


b) Intégration des niveaux 0 et 1

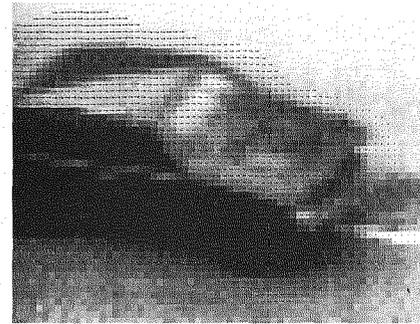
Figure 7. – Flots optiques par la méthode des filtrages combinés (séquence Taxi de Hambourg).



a) Niveau 0



b) Niveau 0



c) Intégration des niveaux 0 et 1

Figure 8. – Flots optiques par la méthode des filtrages combinés dans la séquence Taxi de Hambourg pour a) le taxi, b) et c) la voiture.

Type de véhicule	Taxi		Voiture		Voiture	
	Niveau 0		Niveau 0		Niveaux 0 et 1	
Composantes	v_x	v_y	v_x	v_y	v_x	v_y
Moyennes	0.4	0.4	0	0.29	1.1	0.2

Tableau 2. – Moyennes sur les composantes v_x et v_y des vecteurs vitesse (filtrages combinés) pour le taxi et la voiture.

4. implantation numérique des filtres de Gabor

Nous proposons une implantation numérique plus efficace des filtres, reposant sur le choix de la taille du noyau de convolution, et sur l'utilisation de symétries.

4.1. taille des noyaux de convolution

L'implantation pratique des filtres de Gabor dans le domaine spatial nécessite de tronquer l'enveloppe Gaussienne, d'où un phénomène bien connu d'ondulations parasites ou effet de Gibbs, dans le domaine fréquentiel. Pour faciliter le choix de la taille des noyaux de convolution, nous utilisons un critère d'erreur quadratique moyenne (EQM) permettant de comparer les réponses du filtre idéal et des filtres implantés pratiquement :

$$EQM(\%) = \frac{\int_{[-0.5,0.5]} |G_{\text{tronqué}}(f_x) - G_{\text{idéal}}(f_x)|^2}{\int_{[-0.5,0.5]} |G_{\text{idéal}}(f_x)|^2} \cdot 100$$

Calcul du flot optique par filtrage de Gabor

Heeger a utilisé, sans le justifier, une taille de noyau de convolution égale à 7 pour la dimension temporelle et à 23 pour chacune des dimensions spatiales. Nous vérifions aisément à la figure 9 que l'*EQM* tend à diminuer lorsque la taille du noyau, exprimée en nombre de colonnes ou de lignes, croît. Les tracés ont été effectués seulement pour l'enveloppe gaussienne des fonctions de Gabor. En réalité, la taille minimale de noyau de convolution à considérer est de 7, ce qui correspond à 4 échantillons par période pour une fréquence centrale $f = 0.25$.

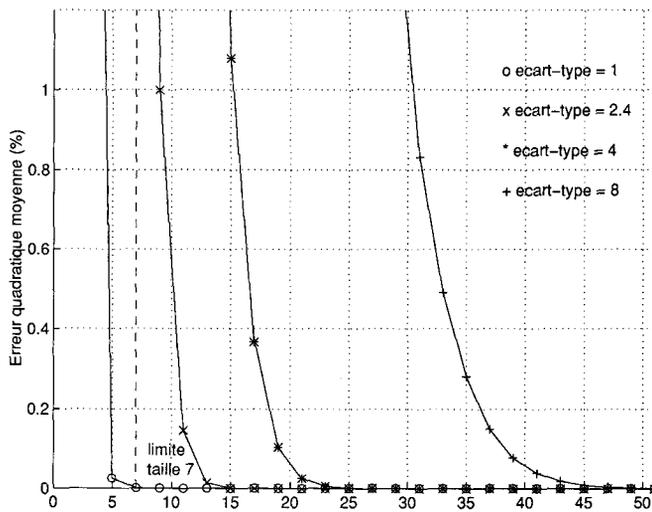


Figure 9. – Erreur quadratique moyenne en fonction de la taille du noyau de convolution pour les écart-types $\sigma = 1, 2.4, 4$ et 8 .

a) Dimension temporelle des filtres

Pour l'écart-type $\sigma = \sigma_t = 1$, nous n'observons pas d'amélioration significative de l'*EQM* à partir de la taille minimale de 7 (Figure 9). Ce filtre étant peu sélectif, la troncature n'introduit pas d'ondulations parasites importantes. Aussi, nous avons choisi un noyau de convolution en temporel de taille minimale, soit 7.

b) Dimensions spatiales des filtres :

• Estimation de vitesse

Pour l'écart-type $\sigma = \sigma_s = 4$, la décroissance régulière du critère d'*EQM* ne permet pas réellement le choix de la taille du noyau de convolution (Figure 9). Par contre, le critère d'erreur sur le flot optique (*EF*) pour la séquence test croît brutalement lorsque la taille du noyau de convolution devient strictement inférieure à 17 (tableau 3). Nous avons donc sélectionné la taille 17 et non pas 23, qui offre un bon compromis entre l'erreur d'estimation du flot optique et le coût de calcul. Pour cette taille, l'amplitude des ondulations parasites dues à la troncature est très réduite (figure 10) et l'*EQM* égale à 0.37%. Un exemple de noyau de convolution spatial de taille 17 est donné dans l'annexe B.

Taille	<i>EQM</i> (%)	<i>EF</i> (%)
23	0.006	4.6
21	0.026	4.6
19	0.1	4.9
17	0.37	4.9
15	1.08	7.7
13	3.2	7.4
11	8.79	26.4
9	21.85	24.7
7	50.98	48.4

Tableau 3. – Erreur quadratique moyenne en fonction de la taille du noyau de convolution pour des filtres d'écart-type $\sigma = 4$, erreur sur le flot optique pour la séquence test.

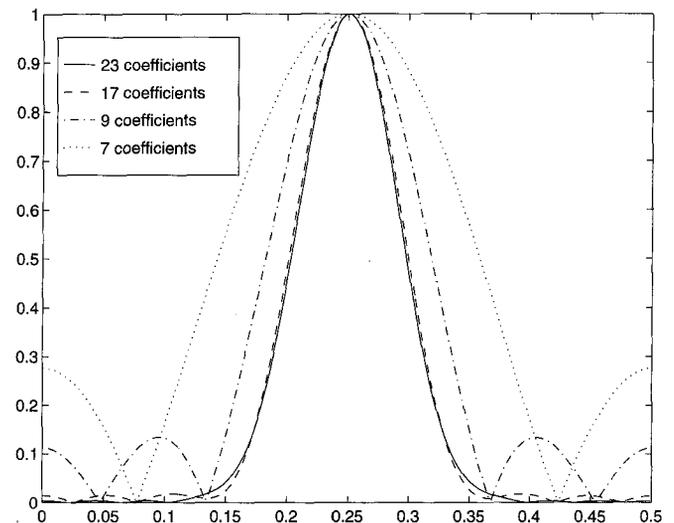


Figure 10. – Spectres (1D) des filtres de Gabor à phase cosinus tronqués de taille 7, 9, 17, 23 pour l'écart-type $\sigma = 4$.

• Détermination des zones en mouvement

Il s'agit dans ce cas de tester simplement la présence ou l'absence de mouvement. Pour limiter l'effet de lissage spatial, nous nous sommes intéressés aux noyaux de convolution de petites tailles (taille minimale 7). Pour ces noyaux, l'effet de troncature peut entraîner l'élargissement de la bande passante du filtre et donc la diminution de l'écart-type équivalent : taille et écart-type sont donc couplés. Le meilleur compromis pour des zones en mouvement à la fois peu morcelées et pas trop étendues, sur diverses séquences d'images artificielles et naturelles, a été obtenu

pour le couple : taille 9, écart-type $\sigma_s = 2.4$ (Figure 5d), et dans une moindre mesure le couple : taille 7, écart-type $\sigma_s = 1.8$ (Figure 11). Les valeurs d'*EQM*, non critiques pour la détermination des zones en mouvement, valent respectivement 0.99% et 0.67%.

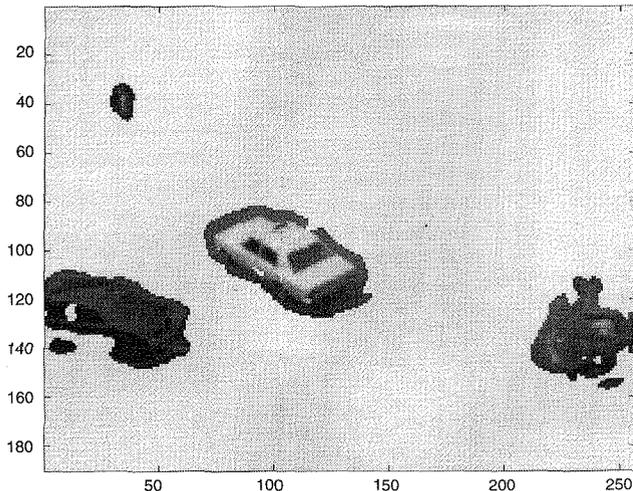


Figure 11. – Exemple de zones détectées en mouvement dans la séquence Taxi de Hambourg pour un filtre de taille 7 et d'écart-type $\sigma_s = 1.8$.

Afin de limiter la taille des noyaux de convolution pour l'estimation des vitesses élevées, nous avons utilisé l'algorithme de décomposition pyramidale d'images de Chéhikian [16] qui permet, par une séquence de filtrages re-échantillonnage, la génération d'une pyramide d'octave avec un minimum de bruit de repliement de spectre. Les images de la séquence sont donc préalablement décomposées en une pyramide à trois niveaux. Le filtrage s'effectue ensuite avec le même noyau de convolution pour chacun des niveaux. Ainsi, dans le cas de l'estimation de vitesse, le filtrage réalisé est équivalent à celui de trois filtres dont les fréquences centrales (en 1D) sont respectivement $f = 1/4, 1/8, 1/16$ et les écart-types $\sigma = 4, 8, 16$. L'implantation de tels filtres sans décomposition pyramidale nécessiterait des noyaux de convolution de taille extrêmement importante. La figure 9 montre, à titre d'exemple, que pour $\sigma = 8$, la taille du noyau de convolution doit être 33 pour que l'*EQM*, égal à 0.49%, soit proche de l'*EQM* pour $\sigma = 4$.

4.2. élimination de la composante continue

La troncature du filtre de Gabor à phase cosinus conduit pour le noyau de convolution à une somme de coefficients non nulle et donc l'apparition d'une composante continue parasite d'autant plus importante que la taille du filtre est réduite et l'écart-type élevé (par exemple $\sigma = 4$) (figure 10). La solution proposée consiste simplement à soustraire à chaque coefficient du filtre la valeur moyenne du noyau de convolution de manière à obtenir

une somme des coefficients nulle. Les filtres étant implantés par somme de convolutions séparables 1D, plusieurs cas sont à distinguer suivant l'orientation φ du filtre dans le domaine des fréquences spatiales (Figure 2b). Pour l'orientation $\varphi = 0$ (respectivement $\varphi = \pi/2$), l'effet de troncature apparaît suivant l'axe des fréquences f_x (respectivement f_y). La composante continue est alors éliminée en annulant seulement la somme des coefficients du noyau de convolution pour la dimension x (respectivement y). Par contre, pour les orientations $\varphi = \pi/4$ et $\varphi = 3\pi/4$ il est inutile de modifier les noyaux de convolution puisque l'effet de la troncature ne s'observe que suivant les directions parallèles aux axes f_x et f_y , et donc que la composante continue induite à l'origine est extrêmement faible.

4.3. symétries temporelles et spatiales

Une paire de filtres de Gabor 3D (phases cosinus et sinus) implantée par convolution séparable nécessite le calcul de 8 termes qui sont les 8 combinaisons possibles des convolutions 1D à phases cosinus et sinus en t , en x et en y (annexe A). Lorsque l'ordre des opérations est t , puis x et enfin y , le nombre de convolutions requis pour une paire de filtre est de 2 en t , $2 \times 2 = 4$ en x et $4 \times 2 = 8$ en y . Bien entendu les résultats partiels de filtrage doivent être conservés en mémoire. Le résultat final pour les filtres de Gabor 3D cosinus ou sinus s'obtient par sommation algébrique des quatre termes adéquats.

L'architecture de Heeger comporte 12 paires de filtres de Gabor pour la gamme de vitesses 0 (figure 2). Le nombre de convolutions requis pour le calcul complet de ces 12 paires de filtres est donc de $2 \times 3 = 6$ en t , $6 \times 8 = 48$ en x et $48 \times 8 = 384$ en y (tableau 4). En fait, le nombre d'opérations nécessaires au filtrage peut être diminué en utilisant les *symétries temporelles et spatiales* présentes entre filtres de Gabor. Ces simplifications reposent sur des relations trigonométriques élémentaires. Le nombre de convolutions se réduit alors à 3 en t , $3 \times 5 = 15$ en x et $15 \times 5 = 75$ en y (tableau 4).

4.4. synthèse : volume de calculs

Nous évaluons tout d'abord le nombre d'opérations nécessaires pour chaque convolution. Soit N le nombre de pixels dans une image carrée, M' la taille d'un noyau de convolution temporel et M la taille de chacun des noyaux spatiaux. Une convolution temporelle (respectivement spatiale 1D) comporte environ $2NM'$ (respectivement $2NM$) opérations (additions et multiplications). Sachant que $M' = 7$ et $M \geq 7$, le nombre total d'opérations pour 12 paires de filtres peut donc s'écrire d'après le paragraphe 4.3 :

$$n = 6N(30M + M') \# 180M$$

en utilisant les symétries, au lieu de

$$n' = 6N(144M + 2M') \# 864M$$

Calcul du flot optique par filtrage de Gabor

Séquence d'images								
t	$g_c(\times 3 \rightarrow \underline{\times 2})$				$g_s(\times 3 \rightarrow \underline{\times 1})$			
x	$g_c(\times 4 \rightarrow \underline{\times 3})$		$g_s(\times 4 \rightarrow \underline{\times 2})$		$g_c(\times 4 \rightarrow \underline{\times 3})$		$g_s(\times 4 \rightarrow \underline{\times 2})$	
y	g_c	g_s	g_c	g_s	g_c	g_s	g_c	g_s
	$(\times 4 \rightarrow \underline{\times 3})$	$(\times 4 \rightarrow \underline{\times 2})$	$(\times 4 \rightarrow \underline{\times 3})$	$(\times 4 \rightarrow \underline{\times 2})$	$(\times 4 \rightarrow \underline{\times 3})$	$(\times 4 \rightarrow \underline{\times 2})$	$(\times 4 \rightarrow \underline{\times 3})$	$(\times 4 \rightarrow \underline{\times 2})$

Tableau 4. – Nombre de convolutions nécessaires pour l'implantation de 12 paires de filtres de Gabor 3D : sans ($\times P$) et avec ($\underline{\times R}$) les symétries (g_c et g_s sont les filtres de Gabor 1D à phases respectivement cosinus et sinus).

dans le cas contraire. Le volume de calculs devient proportionnel à la taille M du noyau de convolution spatial 1D. En définitive, le facteur de gain théorique pour l'étape « filtrages » est de 2 par l'utilisation des symétries temporelles et de 4.8 pour l'ensemble des symétries. Le surcoût introduit par la méthode de filtrages combinés est presque entièrement compensé par la diminution de la taille des noyaux de convolution spatiaux (tailles 17 et 9 au lieu de 23).

Nous avons implanté la méthode avec des symétries temporelles partielles et estimé le gain apporté par l'ensemble des symétries (tableau 5). En effet, l'utilisation des symétries temporelles partielles seules conduit à un bon compromis entre le volume de calculs, la complexité de l'algorithme et la taille de la mémoire.

Taille des filtres	23	17	Filtrages combinés (9 et 17)
Sans symétrie	49	40	67
Symétries partielles	37	30	52
Toutes symétries (estimations)	20	18	33

Tableau 5. – Durée totale de calcul (étapes « filtrages » et « minimisation ») de chaque flot optique (en secondes sur des images 128×128 pixels avec une station de travail Sparc 10).

5. conclusion

La méthode de filtrages combinés que nous avons décrite, issue de l'approche par banc de filtres de Heeger, réduit les erreurs d'estimation de vitesse dans le cas d'une caméra fixe en s'appuyant sur la détermination préalable des zones en mouvement. L'intégration automatique des résultats d'estimation de trois niveaux d'une pyramide spatiale permet d'étendre la plage de vitesse admissible. Enfin, une implantation numérique efficace, exploitant les symétries temporelles et spatiales des filtres de Gabor 3D, accroît

la vitesse de calcul. Les approches de l'estimation du mouvement par filtrages se prêtent particulièrement bien à une implantation sur machine parallèle. Cela ouvre de nombreuses perspectives pour le développement d'applications réelles. Une voie pour accélérer encore la méthode serait l'utilisation d'un modèle de rétine artificielle de détection des zones en mouvement.

Remerciements

Nous remercions notre collègue Franck Luthon pour de fructueuses discussions.

6. annexes

Annexe A

Le principe de l'implantation d'un filtre de Gabor 3D par convolution séparable repose sur la décomposition dans l'expression $g(x, y, t)$ du terme exponentiel (avec la relation $\exp(-(m^2 + n^2)) = \exp(-m^2) \exp(-n^2)$ et du terme cosinus (ou sinus suivant la phase) en appliquant les relations trigonométriques usuelles d'où :

$$\begin{aligned} \cos(\omega_{t0}t + \omega_{x0}x + \omega_{y0}y) = & -\sin(\omega_{t0}t) \cos(\omega_{x0}x) \sin(\omega_{y0}y) \\ & -\sin(\omega_{t0}t) \sin(\omega_{x0}x) \cos(\omega_{y0}y) - \cos(\omega_{t0}t) \sin(\omega_{x0}x) \\ & \sin(\omega_{y0}y) + \cos(\omega_{t0}t) \cos(\omega_{x0}x) \cos(\omega_{y0}y) \end{aligned}$$

$$\begin{aligned} \sin(\omega_{t0}t + \omega_{x0}x + \omega_{y0}y) = & \sin(\omega_{t0}t) \cos(\omega_{x0}x) \cos(\omega_{y0}y) \\ & -\sin(\omega_{t0}t) \sin(\omega_{x0}x) \sin(\omega_{y0}y) + \cos(\omega_{t0}t) \sin(\omega_{x0}x) \\ & \cos(\omega_{y0}y) + \cos(\omega_{t0}t) \cos(\omega_{x0}x) \sin(\omega_{y0}y) \end{aligned}$$

En appelant respectivement g_c et g_s les filtres de Gabor 1D à phases respectivement cosinus et sinus, la décomposition globale s'écrit donc :

$$\begin{aligned} g_c(t, x, y) = & -g_s(t) \cdot g_c(x) \cdot g_s(y) - g_s(t) \cdot g_s(x) \cdot g_c(y) \\ & -g_c(t) \cdot g_s(x) \cdot g_s(y) + g_c(t) \cdot g_c(x) \cdot g_c(y) \\ g_s(t, x, y) = & -g_s(t) \cdot g_c(x) \cdot g_c(y) - g_s(t) \cdot g_s(x) \cdot g_s(y) \\ & -g_c(t) \cdot g_s(x) \cdot g_c(y) + g_c(t) \cdot g_c(x) \cdot g_s(y) \end{aligned}$$

Annexe B

Exemple du noyau spatial en x de taille 17 à phase cosinus ($f_{x0} = 0.25, f_{y0} = 0$)

0.1313 -0.0041 -0.3287 -0.0041 0.6024 -0.0041
 -0.8866 -0.0041 0.9959 -0.0041 -0.8866 -0.0041
 0.6024 -0.0041 -0.3287 -0.0041 0.1313

BIBLIOGRAPHIE

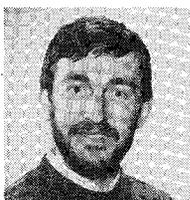
[1] D.J. Heeger, « Model for the extraction of image flow », *Journal of the Optical Society of America A*, vol. 4, n°8, 1987, pp. 1455-1471.
 [2] E.H. Adelson and J.R. Bergen, « Spatiotemporal energy models for the perception of motion », *Journal of the Optical Society of America A*, vol. 2, n°2, 1985, pp. 284-299.
 [3] J.McLean-Palmer, J. Jones and L. Palmer, « New degrees of freedom in the structure of simple receptive fields », *Invest. Ophthalmol. Vis. Sci. Suppl.* 26, 1985, 265 p.
 [4] W.T. Newsome, M. S. Gizzi and J. A. Movshon, « Spatial and temporal properties of neurons in macaque MT », *Invest. Ophthalmol. Vis. Sci. Suppl.* 24, 1983, 106 p.
 [5] J. Weng, « A theory of image matching », *Proc. 3rd International Conference on Computer Vision*, Osaka, Japon, 1990, pp. 200-209.
 [6] L. Jacobson and H. Wechler, « Derivation of optical flow using spatiotemporal frequency approach », *CVGIP*, 38, 1987, pp. 29-65.

[7] A.B. Watson and A.J. Ahumada, « Model of human visual-motion sensing », *Journal of the Optical Society of America A*, vol. 2., 1985, pp. 284-299.
 [8] D.J. Fleet and A.D. Jepson, « Hierarchical construction of orientation and velocity selective filters », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, n°3, 1989, pp. 315-325.
 [9] D.J. Heeger, « Optical Flow using Spatiotemporal Filters », *Int. Journal of Computer Vision*, 1988, pp. 279-302.
 [10] H. Barman, L. Haglung, H. Knutsson and G. Granlund, « Estimation of velocity, acceleration and disparity in time sequences », *Proc. Workshop on Visual Motion*, Princeton, NJ, 1991, pp. 44-51.
 [11] R.S. Jasinchi, « Energy filters, motion uncertainty, and motion sensitive cells in the visual cortex : a mathematical analysis », *Biol. Cybern.*, 65, 1991, pp. 515-523.
 [12] R.S. Jasinchi, « Intrinsic constraints in space-time filtering : a new approach to representing uncertainty in low level vision », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, n°3, 1992, pp. 353-366.
 [13] J.M. Jolion and A. Rosenfeld, « A pyramid framework for early vision », Kluwer Academic Publishers, 1994.
 [14] F. Heitz, P. Bouthemy, « Multimodal estimation of discontinuous optical flow using Markov random fields », *Trans. on PAMI*, vol. 15, n°12, 1993, pp. 1217-1232.
 [15] S.F. Wu and J. Kittler, « A differential method for simultaneous estimation of rotation, change of scale and translation », *Signal Processing : Image Communication* 2, Elsevier, 1990, pp. 69-80.
 [16] A. Chéhikian, « Algorithmes optimaux pour la génération de pyramides d'images passe-bas et laplaciennes », *Traitement du Signal*, vol. 9, n°4, 1992, p. 297-307.

Manuscrit reçu le 1er Février 1995.

LES AUTEURS

Denis PELLERIN



Denis Pellerin, Docteur de l'Institut National des Sciences Appliquées (INSA) de Lyon, est Maître de Conférences à l'Université Joseph Fourier de Grenoble (UJFG) depuis 1989. Ses recherches concernent la théorie et les applications des réseaux de neurones en traitement du signal et des images. Il s'intéresse notamment aux modèles biologiques de la vision et à l'analyse du mouvement.

Anne GUÉRIN-DUGUÉ



Anne Guérin-Dugué est Docteur de l'Institut National Polytechnique de Grenoble et Maître de Conférences à l'ENSERG depuis 1987. Ses travaux de recherche au Laboratoire de Traitement d'Images et de Reconnaissance des Formes concernent les réseaux neuronomimétiques sous les aspects architectures parallèles, analyse de données, auto-organisation, classification et fusion de données, ainsi que l'analyse de textures.

Adrian SPINÉI



Adrian Spinéi, ingénieur de l'Université Politehnica de Bucarest (Roumanie), prépare actuellement une thèse au laboratoire de traitement d'Images et Reconnaissance des formes sur l'étude des méthodes fréquentielles pour l'estimation du mouvement.