

Estimation rapide de vitesse à base de triades de filtres de Gabor

A. Spinéi, D. Pellerin, J. Hérault

Laboratoire de Traitement d'Images et de Reconnaissance des Formes
LTIRF, INPG, 46 Av. Félix-Viallet, 38031 Grenoble Cedex 1 / spinei@tirf.inpg.fr

RÉSUMÉ

Les méthodes fréquentielles d'estimation de vitesse dans les séquences d'images fournissent des flux optiques de qualité, mais sont réputées gourmandes en calculs. Nous présentons une nouvelle méthode fréquentielle qui repose sur quatre estimations de vitesse unidimensionnelle. Chaque estimation 1D est réalisée en combinant directement les réponses énergétiques d'une triade de filtres de Gabor spatio-temporels. Ce modèle possède deux propriétés essentielles : il représente un maillon important dans la chaîne de traitement du système visuel, et constitue un algorithme rapide et efficace.

1 Introduction

Les méthodes fréquentielles (ou méthodes énergétiques) d'estimation de vitesse dans les séquences d'images sont fondées sur une caractérisation du mouvement dans le domaine des fréquences [1]. Elles ont pour origine des recherches concernant la vision des mammifères, et la mise en évidence de la présence de cellules simples dans l'aire corticale V1, qui se comportent comme des filtres passe-bande spatio-temporels. Ces méthodes fournissent des flux optiques de qualité, mais sont réputées gourmandes en calculs. Nous présentons une méthode fréquentielle rapide qui repose sur quatre estimations de vitesse unidimensionnelle. Chaque estimation 1D est réalisée en combinant directement les réponses énergétiques d'une triade de filtres de Gabor spatio-temporels.

2 Principe des méthodes fréquentielles

Pour une image $i(x,y)$ en translation à vitesse uniforme $v(v_x, v_y)$, le spectre associé dans le domaine des fréquences spatio-temporelles occupe un plan passant par l'origine [1] d'équation :

$$v_x f_x + v_y f_y + f_t = 0$$

où f_x , f_y représentent les fréquences spatiales et f_t la fréquence temporelle. Le principe des méthodes fréquentielles consiste à localiser le plan de spectre non nul ou plan de vitesse au moyen d'un banc de filtres spatio-temporels

ABSTRACT

The frequential methods for motion estimation in image sequences provide good optical flows, but are reputed to require intensive calculations. We present a new frequential method based on four monodimensional motion estimation. Each 1D estimation is realized by directly combining the energetic responses of a triad of spatio-temporal Gabor filters. This model has two fundamental properties : it represents an important link in the processing chain of the visual system, and it constitutes a fast and efficient algorithm.

orientés, pour obtenir les composantes du vecteur vitesse (v_x , v_y).

L'estimation de vitesse est formulée généralement comme une minimisation de l'écart quadratique entre les réponses mesurées des filtres pour une séquence d'images quelconque et les réponses prédites pour une séquence d'images de bruit blanc en translation [4]. Le coût de calculs de cette procédure est très élevé et dépend étroitement de la précision requise pour l'estimation.

3 Description du modèle

Les filtres spatio-temporels ne sont pas sélectifs à une vitesse particulière mais plutôt réglés sur des fréquences spatio-temporelles. Ils ne peuvent donc pas être considérés séparément. Nous allons montrer comment les combiner pour estimer successivement la vitesse en 1D puis en 2D.

Un filtre de Gabor 3D à phase cosinus est représenté comme le produit d'une fonction spatio-temporelle Gaussienne et d'une fonction cosinus :

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

$$\times \cos(2\pi(f_{x0}x + f_{y0}y + f_{t0}t))$$

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 Gaussienne. 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.

A partir de l'expression générale de la réponse énergétique d'un filtre de Gabor 3D [4], nous pouvons exprimer la réponse d'un filtre de Gabor en fonction des composantes de vitesse v_d et v_{dp} respectivement dans la direction de l'onde de modulation spatiale et dans la direction perpendiculaire :

$$R_{f_{t0}} = \frac{1}{8\pi\sigma_s\sigma_t\sqrt{K}} \exp\left[\frac{-4\pi^2\sigma_s^2(v_d f_{d0} + f_{t0})^2}{K}\right]$$

en posant

$$\begin{aligned} \sigma_s &= \sigma_x = \sigma_y, \\ f_{d0} &= \sqrt{f_{x0}^2 + f_{y0}^2}, \\ \text{et } K &= v_d^2 + v_{dp}^2 + (\sigma_s/\sigma_t)^2 \end{aligned}$$

La vitesse v_d se calcule alors directement à partir des réponses énergétiques de trois filtres de Gabor (triade) de mêmes écart-types σ_s et σ_t , et placés à la même fréquence spatiale f_{d0} , mais à trois fréquences temporelles différentes $f_{t0} = +0,25; 0; \text{ et } -0,25$ (Fig. 1a) :

$$v_d = \frac{L_{-0,25} - L_{+0,25}}{6L_0}$$

avec comme variable intermédiaire de calcul

$$L_{f_{t0}} = \ln R_{f_{t0}} - \frac{1}{3}(\ln R_{-0,25} + \ln R_0 + \ln R_{0,25})$$

En théorie, deux composantes de vitesse suffisent pour décrire tout mouvement dans une image. En pratique, le plan d'énergie non nulle n'est pas uniforme, car le contraste dans les images est généralement différent suivant les directions. Cela peut conduire à d'importantes erreurs d'estimation. Nous avons donc choisi d'estimer la vitesse 1D dans quatre directions ($d = 0, \pi/4, \pi/2, 3\pi/4$) en utilisant quatre triades de filtres placées suivant les orientations spatiales $\phi = 0, \pi/4, \pi/2,$ et $3\pi/4$ (Fig. 1b). Le vecteur vitesse s'obtient par moyenne vectorielle des estimations 1D (Fig. 2).

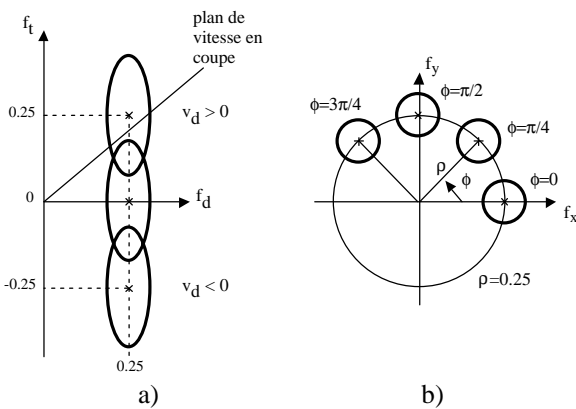


Figure 1 : a) Une triade de filtres de Gabor 3D dans le plan (f_d, f_t) , b) Quatre triades de filtres de Gabor 3D dans le plan (f_x, f_y) .

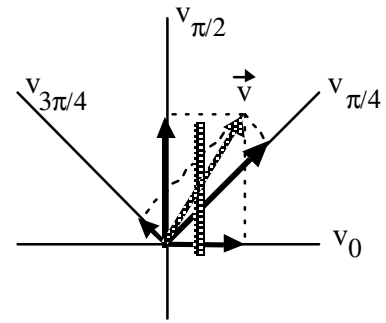


Figure 2 : Calcul du vecteur vitesse par moyenne vectorielle de quatre estimations de vitesse 1D.

Lorsque l'on traite des séquences d'images réelles, il apparaît le problème de la non satisfaction de la condition de spectre blanc. Atick [2] a en effet montré que le spectre des images naturelles étaient de type $1/f$, et qu'une compensation au moins partielle pouvait être obtenue par un préfiltrage similaire à celui réalisé par la rétine. La méthode comporte donc trois étapes fondamentales (Figure 3) qui pourraient correspondre aux traitements effectués dans le système visuel par la rétine, l'aire V1 et l'aire MT :

- Préfiltrage pour "blanchir" le spectre, et surtout limiter l'influence des basses fréquences. Cette étape est réalisée au moyen d'un filtre passe-bande spatial de type DOG ("Difference Of Gaussians").
- Extraction des énergies liées au mouvement au moyen de filtres spatio-temporels orientés.
- Estimation de vitesse 1D en combinant les énergies fournies par un ensemble de trois filtres, puis 2D en associant quatre triades.

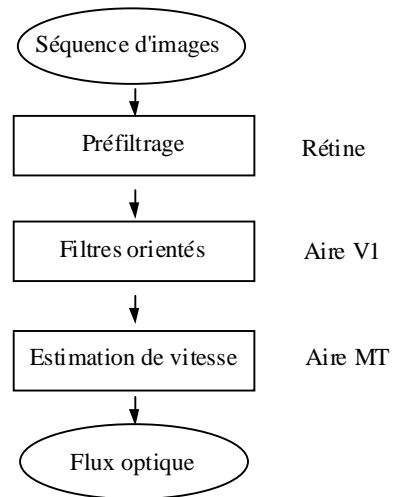


Figure 3 : Synoptique de la méthode

4. Implantation et résultats

La mise en oeuvre de la méthode s'appuie sur plusieurs travaux antérieurs. Les filtres de Gabor spatiaux sont implantés de manière récursive [6], et une décomposition pyramidale d'images à trois niveaux permet d'élargir la gamme de vitesses estimées [5].

Afin de pouvoir comparer nos résultats avec ceux obtenus par Barron et al. [3], nous avons choisi trois séquences d'images de référence avec caméra mobile ("Translating Tree", "Diverging Tree", "Yosemite") dont le flux optique est connu (Figures 4 et 5). Nous avons utilisé la même mesure de confiance qui permet d'éliminer en fonction d'un seuil, les vecteurs les moins fiables. De manière à comparer aussi les temps de calculs, nous avons implanté la méthode de minimisation par multirésolution (deux grilles de vitesse emboîtées) proposée par Barron.

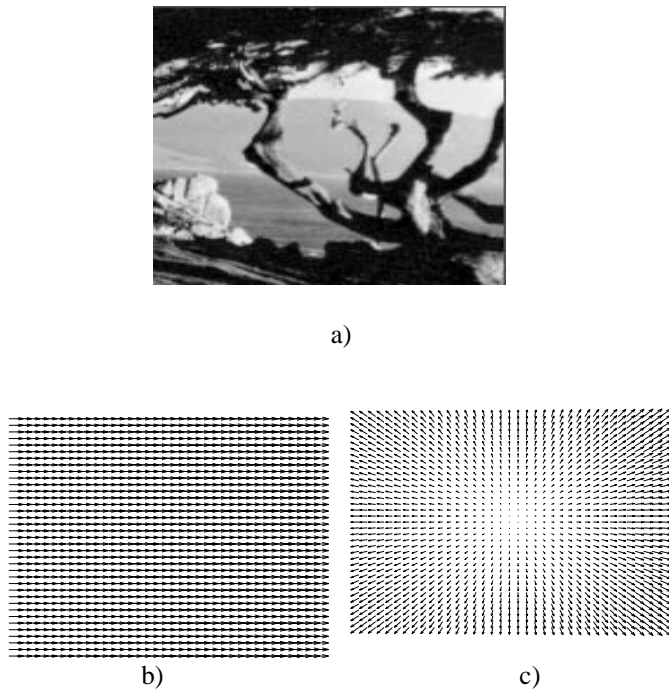


Figure 4 : a) Image initiale des séquences "Translating Tree" et "Diverging Tree", b) et c) flux optiques réels.

Les erreurs d'estimation obtenues par les deux approches sont assez voisines (Tableau 1), mais la durée d'exécution est beaucoup plus courte dans notre cas. Par exemple, les temps d'exécution pour un calcul de flux optique sur des images de taille 128×128 pixels (station de travail UltraSparc 1) sont de 8,4 s pour notre méthode et de 184 s environ pour la méthode de minimisation par multirésolution. Les flux optiques réel et estimé pour la séquence d'images "Yosemite" sont présentés figure 5.

Afin de réduire encore le volume de calculs, nous pouvons ne considérer que le niveau le plus grossier de la pyramide (niveau 2). Le temps de calcul, dans les mêmes conditions que précédemment, tombe alors à environ 0,4 s. Pour les séquences testées, les résultats obtenus restent de bonne qualité (Tableau 1).

5 Conclusion

Nous avons décrit un modèle original pour l'estimation de vitesse, à base de triades de filtres de Gabor implantés en récursif. Ce modèle possède deux propriétés essentielles. Il représente tout d'abord, dans la chaîne de traitement du système visuel, un maillon important qui peut être enrichi à partir des données biologiques les plus récentes. Il constitue ensuite un algorithme rapide et efficace, qui ouvre la voie vers une implantation en temps réel sur un processeur de traitement du signal d'usage général.

Références

- [1] E.H. Adelson E.H, J.R. Bergen, "Spatiotemporal energy models for the perception of motion", Journal of the Optical Society of America A., vol. 2, n° 2, 284-299, 1985.
- [2] J. Atick, A. Redlich, "What does the retina know about natural scenes?", Neural Comput., 4, 196-210, 1992.
- [3] J.L. Barron, D.J. Fleet and S.S. Beauchemin, "Performances of optical flow techniques", Proc. Conf. on Computer Vision and Pattern Recognition, Champaign, June, 236-242, 1992.
- [4] D.J. Heeger, "Model for the extraction of image flow", Journal of the Optical Society of America A., vol. 4, n° 8, 1455-1471, 1987.
- [5] D. Pellerin, A. Spinéi, A. Guérin-Dugué, "Calcul du flot optique par filtrages de Gabor combinés", Revue Traitement du Signal, vol. 13, n°1, 13-23, 1996.
- [6] D. Pellerin, A. Spinéi, A. Guérin-Dugué, "Estimation du mouvement par filtres spatiotemporels récursifs", 15ème colloque GRETSI, Juan-les-pins, 841-844, septembre 1995.

Nom de la séquence		Barron et al.	3 niveaux de pyramide	niveau 2 de pyramide
Translating Tree	Moy. des erreurs	4.79°	4.06°	2.42°
	Dispersion	2.39°	2.8°	5.79°
Diverging Tree	Moy. des erreurs	4.95°	4.6°	8.3°
	Dispersion	3.09°	3.5°	6.8°
Yosemite	Moy. des erreurs	11.97°	14.9°	14.95°
	Dispersion	19.0°	17.3°	12.6°

Tableau 1 : Comparaison des erreurs d'estimation de vitesse pour trois séquences d'images

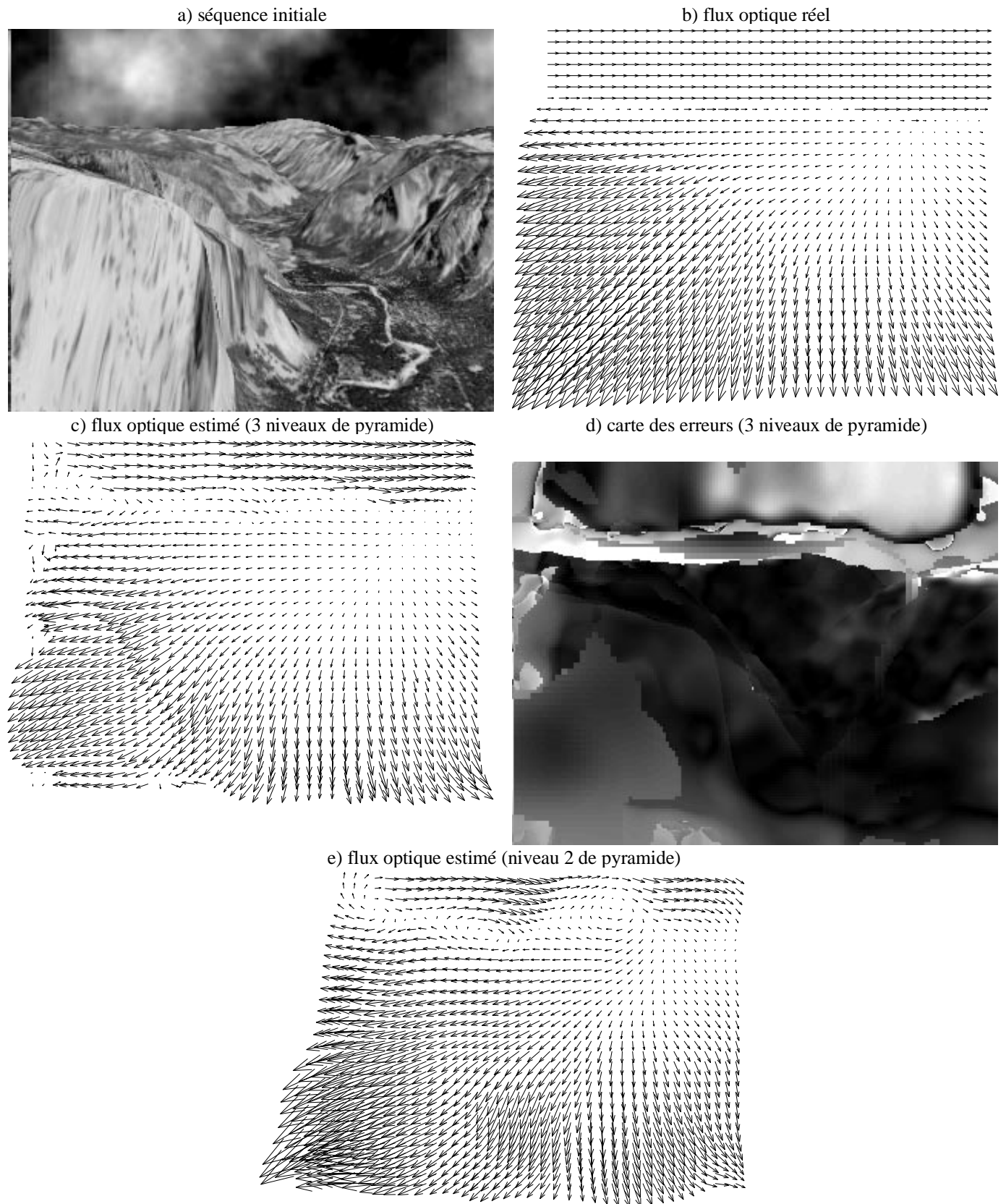


Figure 5 : a) Séquence d'images "Yosemite", Flux optiques b) réel, c) estimé (3 niveaux de pyramide), d) Carte des erreurs, e) Flux optique estimé (niveau 2 de pyramide).