

DETECTION DE MOUVEMENT PAR FILTRAGES SPATIO-TEMPORELS

Olivier BREMOND^(1,2)
Hugues WALDBURGER⁽¹⁾

- (1) T.R.T., Centre technique du PLESSIS-ROBINSON, 5, avenue Réaumur, 92350 Le PLESSIS-ROBINSON
(2) Ecole Supérieure d'Electricité, 2 rue Edouard Belin, 57078 METZ cedex 3

RÉSUMÉ

L'estimation locale de l'orientation 3D dominante où se concentre l'énergie spatio-temporelle permet de quantifier le mouvement en direction et en vitesse. Trois approches fondées sur des opérations locales de filtrage sont proposées sous forme d'une étude comparative. La première évalue l'orientation par le calcul du gradient spatio-temporel. La seconde consiste à combiner linéairement les réponses de filtres directionnels en quadrature. La dernière enfin, cherche dans l'espace de Fourier la direction optimale minimisant un critère quadratique.

SUMMARY

The local estimation of the main 3D orientation where spatio-temporal energy is concentrated allows to quantify motion both in direction and velocity. Three approaches founded on local filtering are compared in this paper. The first one derives the orientation from the spatio-temporal gradient. The second one consists in a linear combination of quadrature directional filters outputs. The last one is based on the computation of the optimal direction in the Fourier domain by minimization of a quadratic criterion.

1 Introduction

Le mouvement des lignes et contours est caractérisé par une variation spatio-temporelle de l'intensité dans une orientation 3D de l'espace (x,y,t) [1]. Dans cet article, nous présentons des méthodes d'estimation de cette orientation par filtrage local fondés sur une généralisation des approches maintenant classiques de détection des structures statiques bidimensionnelles [2][3]. Le résultat de chaque traitement se présentera sous la forme d'un champ de vecteurs attribués 3D exprimés en coordonnées sphériques:

$$\bar{V}(x,y,t) = (r, \phi, \theta)$$

- (x,y,t) est la position locale.
- (ϕ, θ) représente l'estimation de l'orientation définie sur le demi-espace D :

$$-\pi \leq \phi \leq \pi \quad ; \quad 0 \leq \theta \leq \pi/2$$

- r le module, exprime le degré de confiance accordée à l'estimation, lié principalement à l'énergie orientée extraite.

2 Modélisation

Les structures mobiles sont modélisées par des images linéairement symétriques ou ILS dont il faut estimer l'orientation afin de quantifier le mouvement.

Définition:

Une séquence d'image est une ILS si les surfaces de même niveau de gris (isogris) sont des plans parallèles de direction normale commune \bar{k}_0 , ce qui s'exprime par:

$$f(\bar{x}) = g(\bar{x} \cdot \bar{k}_0)$$

g étant une fonction scalaire. \bar{k}_0 est appelée orientation 3D de l'ILS.



Théorème:

La transformée de Fourier d'une ILS d'orientation \bar{k}_0 a pour support une droite passant par l'origine O et de direction \bar{k}_0 .

Dans le cas du mouvement, si f est une ILS d'orientation $\bar{k}_0 = (a,b,c) = (\phi_0, \theta_0)$ avec $c \geq 0$ alors:

$$f(x,y,t) = g(ax+by+ct)$$

et l'équation du mouvement:

$$f(x+dx,y+dy,t+dt) = f(x,y,t)$$

$$\text{est équivalente à: } a \cdot dx + b \cdot dy + c \cdot dt = 0.$$

En notant \bar{v} le vecteur vitesse et $\bar{k}_r = (a,b)$ l'orientation spatiale, nous obtenons l'équation de propagation:

$$\bar{k}_r \cdot \bar{v} + c = 0$$

ce qui nous permet d'exprimer la composante normale de la vitesse, colinéaire à \bar{k}_r , par:

$$\bar{v}_n = -c \cdot \bar{k}_r / |\bar{k}_r|^2 \quad (1)$$

où $\text{direction}(\bar{v}_n) = \phi_0 \pm \pi$ et

$$|\bar{v}_n| = \text{tg}(\theta_0)$$

3 Détection de mouvement par calcul du gradient spatio-temporel (GST)

Cette approche modélise localement l'image par une ILS grâce à un développement de Taylor du 1er ordre.

$$f(\bar{x}) = f(\bar{0}) + \bar{\nabla} f \cdot \bar{x} + \varepsilon(\bar{x}) \quad (2)$$

avec $\varepsilon(\bar{x}) \rightarrow 0$ quand $|\bar{x}| \rightarrow 0$

Dans ce cas, l'attribut V d'orientation n'est rien d'autre que le gradient spatio-temporel. Dans le cas discret, le gradient se calcule par convolution de l'image avec le gradient de la fonction d'interpolation de Shannon:

$$\bar{\nabla} f = f(x,y,t) * \bar{\nabla} \mu \quad (3), \text{ avec:}$$

$$\mu(x,y,t) = \text{sinc}(\pi x) \cdot \text{sinc}(\pi y) \cdot \text{sinc}(\pi t)$$

Malheureusement, en pratique, le filtre $\bar{\nabla} \mu$ n'est pas acceptable car il amplifie fortement le bruit et sa limitation en taille entraîne des erreurs assez importantes sur $\bar{\nabla} f$. Pour palier ces inconvénients, la fonction μ est remplacée par une gaussienne:

$$g(x,y,t) = \exp[-kr^2 \cdot (x^2 + y^2) - kt^2 \cdot t^2]$$

ce qui revient à calculer le gradient de l'image moyennée par cette gaussienne. Le choix des paramètres k_r et k_t détermine le gabarit fréquentiel mais impose également la taille des filtres 3D.

4 Détection de mouvement par combinaison de filtres directionnels (CFD)

Partant d'une analyse dans l'espace de Fourier, l'idée est de combiner linéairement les réponses de plusieurs filtres directionnels complexes pour estimer une orientation dominante, support de l'énergie du mouvement.

Nous savons qu'il faut un jeu d'au moins cinq filtres régulièrement distribués dans le demi-espace de Fourier pour estimer l'orientation d'une ILS. Cela implique que les orientations principales \bar{n}_k des filtres doivent pointer vers les six sommets d'un demi-icosaèdre ou vers les dix sommets d'un demi-dodécaèdre [4]. Les K (K=6 ou K=10) filtres qui sont en quadrature afin d'obtenir des amplitudes de sorties indépendantes de la phase sont de la forme:

$$H_k(\bar{u}) = A \cdot (\bar{\omega} \cdot \bar{n}_k)^2 \cdot [1 + j \cdot \text{sign}(\bar{\omega} \cdot \bar{n}_k)] \quad (4)$$

$$\text{avec } A = \exp\{-4 \cdot \ln 2 \cdot \ln^2(\rho/\rho_0)/B^2\}$$

$$\bar{u} = \rho \cdot \bar{\omega} \quad ; \quad \rho_0 = \text{fréquence centrale} \quad ;$$

$$B = \text{largeur de bande à 6 dB}$$

Soit une ILS d'orientation $\bar{k}_0 (\phi_0, \theta_0)$. Alors le module de la sortie du kième filtre est donné par:

$$q_k = Q_0 \cdot (\bar{\omega}_0 \cdot \bar{n}_k)^2 \quad (5)$$

$$\bar{\omega}_0 = (\cos \theta_0 \cdot \cos \phi_0; \cos \theta_0 \cdot \sin \phi_0; \sin \theta_0)$$

Comme l'orientation est définie modulo π uniquement sur le demi-espace D, nous sommes confrontés à un problème de discontinuité. Une transformation T préservant les propriétés métriques sur D et offrant une représentation continue en dimension 5 a été proposée [5]; l'attribut d'orientation \bar{k}_0 est donné par son représentant dans l'espace 5D en effectuant la combinaison linéaire suivante:

$$\bar{Z} = T(\bar{V}) = q_k \cdot T(\bar{n}_k) = \alpha \cdot Q_0 \cdot T(\bar{\omega}_0) \quad (6)$$

$$\{ \alpha = 4/5 \quad \text{si } K=6 \quad ; \quad \alpha = 4/3 \quad \text{si } K=10 \}$$

Par la transformation inverse, l'attribut \bar{V} est calculé explicitement puisque T est bijective de D sur $R \times R \times R$.

Remarque:

Si le signal est isotrope dans l'espace fréquentiel, les sorties q_k sont égales pour tous les filtres; le module de \bar{Z} compte tenu du choix des filtres et des propriétés de T devient nul. Ce qui exprime l'incertitude totale à estimer une orientation dominante.



5 Détection de mouvement par minimisation de variance (MV)

Cette méthode consiste à rechercher dans l'espace de Fourier, l'orientation \bar{k}_0 optimale minimisant une variance [6]:

$$\sigma^2(\bar{k}_0) = \min\{\sigma^2(\bar{k}); \bar{k} \in D\}$$

$$\text{où } \sigma^2(\bar{k}) = \int d^2(\bar{u}, \bar{k}) \cdot |F(\bar{u})|^2 d\bar{u} \quad (7)$$

Ici $d(\bar{u}, \bar{k})$ est la distance euclidienne entre un point \bar{u} et une direction candidate \bar{k} . La variance σ^2 peut s'exprimer également à l'aide d'une matrice J symétrique définie positive.

$$\sigma^2(\bar{k}) = \bar{k} J \bar{k} \quad (8)$$

Ainsi le problème se résume à une diagonalisation de matrice (3x3) où l'orientation est la direction du sous-espace correspondant à la plus petite valeur propre $\lambda_0 \geq 0$. Le principe est analogue au problème de recherche de l'axe principal d'inertie d'un solide en mécanique où l'énergie serait considérée comme une distribution de masse.

Les six coefficients de J s'obtiennent par filtrage spatio-temporel grâce à l'égalité de Parseval:

$$\int u_i \cdot u_j \cdot |F(\bar{u})|^2 d\bar{u} = a \cdot \int (\partial f / \partial x_i) \cdot (\partial f / \partial x_j) \cdot d\bar{x}$$

$$i, j = 1, 2, 3 \quad \text{et} \quad a = 1/4 \cdot \pi \cdot 2 \quad (9)$$

Le calcul se fait en deux étapes successives de convolution:

- Les produits des dérivées partielles sont calculés par des opérateurs gradient.
- Puis un moyennage des résultats précédents sur un voisinage exprime le fait que nous étudions le comportement local.

Le module de l'orientation estimée est lié à la dispersion des deux autres valeurs propres λ_1 et λ_2 relativement à λ_0 .

$$r = |\bar{V}| = \delta_1 \cdot \delta_2 \cdot \text{Trace}[J] \quad (10)$$

$$\delta_i = (\lambda_i - \lambda_0) / (\lambda_i + \lambda_0) \quad i=1, 2$$

6 Résultats expérimentaux

Afin de comparer les performances des trois approches, nous avons procédé à des simulations sur station de travail GOP-300 [7]. Une même séquence test contenant cinq images est utilisée avec des rapports signal à bruit différents. Ce signal test théorique contient toutes les orientations (ϕ, θ) possibles, tout en conservant une fréquence spatio-temporelle constante. L'image attribut $\bar{V}(x, y, 0)$ de référence présente une répartition linéaire angulaire en ϕ et radiale en θ . Ainsi pour chaque

estimateur, nous pouvons mesurer statistiquement la précision et la sensibilité au bruit par le calcul de leurs biais (b_ϕ, b_θ) et déviations ($\sigma_\phi, \sigma_\theta$).

Les résultats sont regroupés dans un tableau tenant compte en outre de deux paramètres supplémentaires: le temps de calcul T_c et une mesure de résolution de détection τ . Nous remarquons que l'utilisation de filtres non-cubiques (9x9x5 par exemple) produit un léger biais en θ . L'approche par minimisation de variance (MV) fournit la meilleure précision tandis que la méthode de combinaison de filtres directionnels (CFD) testée avec $K=6$ et $B=2$, semble donner le meilleur compromis résolution de détection / résistance au bruit. Enfin, à cause de sa simplicité de mise en oeuvre et ses performances relativement bonnes, l'approche gradient spatio-temporel (GST) avec le filtre \sqrt{g} présente un intérêt pratique évident.

7 Conclusion

Dans cet article, nous avons montré que la généralisation aux lignes et contours mobiles des principes de détection par filtrage 2D des structures statiques était tout à fait réalisable. Cette étude comparative nous permet de choisir une des trois méthodes selon la nature des images et la disponibilité en temps de calcul. Elle nous ouvre également une voie pour des traitements plus complexe sur les séquences d'images comme la restauration, le codage et l'analyse de scène 3D (x,y,t) pour lesquels le champ d'orientation 3D s'avère être une information contextuelle à la fois simple et pertinente.

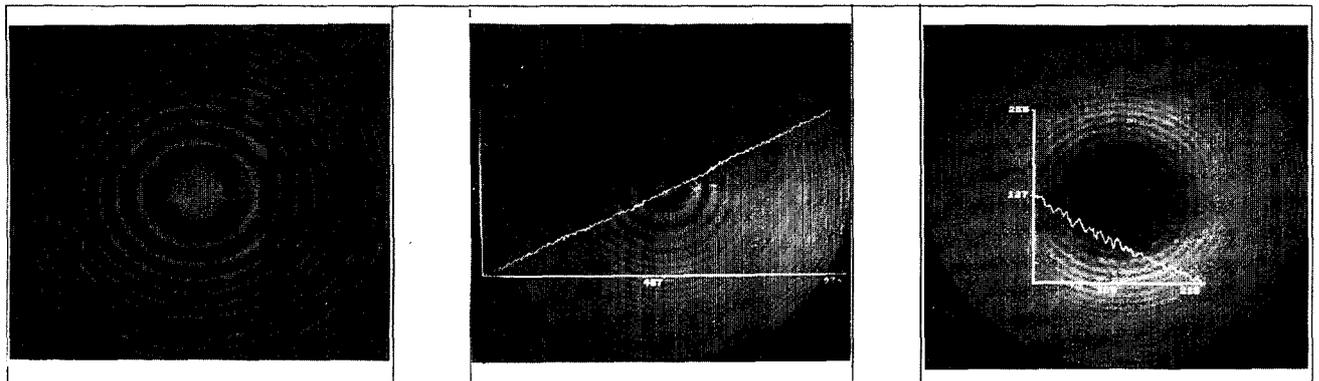
8 Bibliographie

- [1] ADELSON H. & BERGER R. : "Spatio-temporal energy model", optical society of America, vol2, n2, February 1985.
- [2] GRANLUND G.H. & KNUTSSON H.: "Fourier domain design of line and edge detectors", Proceedings of the 5th International Conference on Pattern Recognition, Miami, 1980.
- [3] KNUTSSON H. : "Filtering and reconstruction in image processing", Dissertation N88, Linköpings studies in Science and Technology, 1982.
- [4] BERGER M. : "Géométrie", Tome 3 & 5, CEDIC/Fernand Nathan, 1978.
- [5] KNUTSSON H. : "Producing a continuous and distance preserving 5D vector representation of 3D orientation", IEEE Computer Society Workshop on CAPAIDM, pp. 175-182 ; November 1985.
- [6] BIGUN J. & GRANLUND G.H. : "Optimal orientation detection of linear symmetry", IEEE International Conference on Computer Vision, 3, PP.433-438, 1987.
- [7] GRANLUND G.H.: "In search of a general picture processing operator", Computer Graphics an Image Processing, 8, PP. 155-173, 1978.



		GST			CFD				MV	
		1 mn 20 s		35 s	2 mn 10 s		2 mn 45 s		1 mn 30 s	1 mn 15 s
filtres		$\bar{\nu}_\mu$	$\bar{\nu}_g$	$\bar{\nu}_g$	$\rho_0 = 2^{-7/2}$	$\rho_0 = 2^{-5/2}$	$\rho_0 = 2^{-7/2}$	$\rho_0 = 2^{-5/2}$	$\bar{\nu}_g$ (5x5x3)	$\bar{\nu}_g$ (5x5x3)
SNR		(9x9x5)	(9x9x5)	(5x5x5)	(5x5x5)	(5x5x5)	(9x9x5)	(9x9x5)	g (5x5x3)	g (5x5x1)
χ (%)	0 dB	20	39	48	51	38	64	50	32	36
		-2.8	0.0	-0.1	0.0	-0.1	0.0	-0.1	-0.2	0.0
		-0.1	0.8	0.0	0.1	0.6	-4.0	-1.1	0.2	-0.2
		29.9	8.0	12.4	12.5	18.4	7.5	14.0	5.7	5.4
		21.9	7.1	8.0	6.8	11.5	5.1	9.1	4.6	4.2
b_φ (deg)	10 dB	21	42	56	61	58	76	68	41	45
		-0.3	0.0	0.0	0.0	0.0	0.0	0.0	-0.2	0.0
		-3.1	-1.4	-0.2	0.0	0.0	-4.1	-1.5	-0.2	-0.2
		28.4	3.9	5.9	7.7	9.6	4.4	7.3	2.6	2.5
		19.7	3.6	3.6	3.1	4.3	3.1	3.6	1.8	1.7
b_θ (deg)	100 dB	29	43	60	66	72	79	81	45	48
		-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.2	0.0
		-8.9	-1.5	-0.2	0.0	0.1	-4.1	-1.5	-0.3	-0.2
		20.4	1.8	2.6	6.5	6.3	3.1	4.0	1.5	1.2
		13.2	2.4	1.6	2.2	2.0	2.7	2.0	0.9	0.9
σ_φ (deg)	dB	40	44	62	68	76	80	86	47	49
		0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.2	0.0
		-13.2	-1.5	-0.2	0.0	-0.1	-4.1	-1.5	-0.3	-0.2
		11.6	1.1	0.8	6.3	5.5	3.0	2.4	1.2	0.8
		8.6	2.2	1.1	2.0	1.6	2.7	1.7	0.7	0.8

Tableau des résultats



Traitement CFD (9x9x5) $\rho_0 = 2^{-7/2}$

Image centrale de la sequence test : SNR = 10 dB

Image (r, φ) avec profil angulaire de φ .
r = luminance ; φ = couleur

-180 -90 0 90 180
rouge orange vert bleu rouge

Image (r, θ) avec profil radial de θ .
r = luminance ; θ = couleur

0 45 90
vert bleu rouge