



## ESTIMATION DU MOUVEMENT PAR APPROCHE MIXTE : DROITES DE CONTRAINTE / MISE EN CORRESPONDANCE

Maciej ORKISZ \* et Denis PELLERIN \*\*

\* Laboratoire de Traitement du Signal et Ultrasons, INSA de Lyon, 20 av. Albert Einstein, 69100 Villeurbanne

\*\* Laboratoire de Traitement d'Images et Reconnaissance de Formes, INPG, 46 av. Félix Viallet, 38031 Grenoble cedex

### RÉSUMÉ

Le travail présenté concerne l'estimation du mouvement entre deux images consécutives extraites d'une séquence. Il s'agit de l'estimation initiale pour chaque pixel, à un niveau de résolution donné et avant toute régularisation. L'originalité de la méthode proposée réside dans l'utilisation combinée de deux techniques : de technique différentielle, rapide mais ayant un champ d'applications restreint, et la technique de mise en correspondance par blocs, plus lente, utilisée pour des points où elle est très efficace, situés hors du domaine de validité des méthodes différentielles. La méthode différentielle proposée consiste à chercher des intersections de "droites de contrainte du mouvement". Deux critères sont utilisés pour extraire les zones où cette méthode n'est pas applicable : passage par zéro de la dérivée temporelle ou valeur absolue du laplacien élevée.

### I. OBJECTIFS

Notre étude est consacrée à l'estimation du mouvement. Disposant de deux images extraites d'une séquence, nous souhaitons obtenir un champ dense de vecteurs de vitesse apparente. La qualité du résultat est jugée en termes de cohérence spatiale (possibilité de segmenter ce champ de vecteurs en régions homogènes du point de vue du mouvement) et temporelle (stabilité d'une telle segmentation dans le temps). Outre le critère de qualité, nous voulons satisfaire le critère de rapidité des calculs.

### II. HYPOTHESES

Conscients des limites de leur validité, nous sommes tout de même obligés de reprendre les hypothèses sur lesquelles s'appuient les méthodes existantes :

1° hypothèse d'invariance de l'intensité lumineuse [SCH'89], selon laquelle l'intensité lumineuse de chaque point de la scène reste quasiment constante au cours du mouvement (en tout cas entre deux images voisines) ;

2° hypothèse de déplacements petits par rapport à la taille des objets [WU'90], selon laquelle les déplacements doivent être plus petits que la demi période de la plus haute fréquence spatiale présente dans l'image <sup>1</sup> et tout mouvement doit être localement assimilable à une translation ;

3° hypothèse du caractère localement lisse du champ de vecteurs

<sup>1</sup> la validité des méthodes différentielles est cependant limitée aux déplacements inférieurs au quart de cette période.

### ABSTRACT

In our work we are dealing with the estimation of motion between two consecutive images extracted from a sequence. The problem addressed is a reliable computation of the initial estimates for every pixel, at a given resolution level and before any regularization. The original aspect of the proposed method is a joint use of two different techniques: the differential one, quick but having a limited application field, and the block matching one, slower but applied for points for which this technique is very efficient, outside the application field of the differential technique. We are proposing a differential method which consists in computation of intersections of "motion constraint lines". Two criteria are used in order to extract the regions where this method cannot be applied: zero crossings of the temporal derivative or large value of the laplacian.

vitesse [AIS'89], selon laquelle les points voisins situés sur un même objet sont animés d'un mouvement semblable ;

4° hypothèse du caractère non-rectiligne des éléments de la scène [SCH'89], selon laquelle les variations locales de l'orientation des motifs présents dans l'image permettent de résoudre le "problème d'ouverture".

### III. CONTEXTE BIBLIOGRAPHIQUE

Les techniques existantes utilisent soit l'approche différentielle, soit la mise en correspondance, soit encore l'approche fréquentielle. Les méthodes fréquentielles [HEE'88], [JAC'87], bien que prometteuses, demandent des calculs très lourds. C'est pourquoi nous avons rejeté cette approche. Chacune des deux autres techniques a ses avantages et ses inconvénients. Rappelons d'abord que l'hypothèse 2° s'exprime par l'équation suivante :

$$DFD(x, y, t) = I(x+dx, y+dy, t+dt) - I(x, y, t) = 0$$

où  $I(x,y,t)$  est l'intensité lumineuse du point  $(x,y)$  de l'image prise à l'instant  $t$  et  $(dx,dy)$  représente le déplacement du point de la scène correspondant pendant l'intervalle de temps  $dt$ .

La mise en correspondance consiste à minimiser directement la DFD (ang. *Displaced Frame Difference*). L'image considérée est découpée en blocs. Pour chaque bloc, on cherche le fragment de l'image voisine qui lui ressemble le plus [ANA'89], [ZHE'92]. Les estimations obtenues sont généralement de bonne qualité (fig.1a), mais les blocs doivent être suffisamment grands pour contenir des motifs discriminants, autrement dit pour que le maximum de



ressemblance soit unique et suffisamment "pointu". De même, la taille de la fenêtre dans laquelle on cherche la correspondance est d'autant plus grande que le déplacement maximum autorisé est grand. Or, le nombre de calculs croît avec le carré de chacune de ces tailles. Par conséquent, la bonne qualité des résultats s'obtient au prix de calculs longs. De plus, la taille des blocs détermine la résolution spatiale (fig.1b).

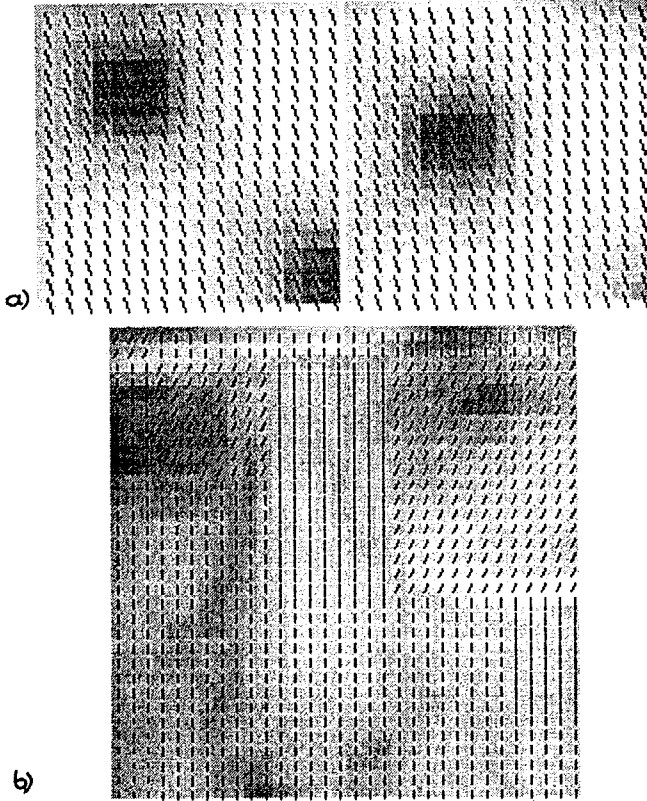


Figure 1. Vecteurs déplacement obtenus par mise en correspondance de blocs 8 sur 8 pixels :

a) exemple extrait d'une séquence synthétique représentant une texture sinusoidale de la forme  $I(x,y) = I_0 + I_1 \sin \omega_x x \sin \omega_y y$  dont le mouvement (1 pixel à droite, 3 pixels vers le bas) a été estimé sans erreur, b) exemple extrait d'une séquence naturelle TWY (cf fig.3) où la limitation de la résolution par "l'effet de bloc" est nettement perceptible.

L'approche différentielle, quant à elle, consiste à exploiter une relation entre les déplacements d'une part et les dérivées spatiales et temporelles de l'intensité lumineuse d'autre part. Cette relation, appelée **équation de contrainte du mouvement**, s'obtient par développement de la DFD en série limitée de Taylor. Généralement, ce développement est limité aux termes du premier ordre, en donnant :

$$\nabla I \cdot \vec{v} + I_t = 0$$

où  $\nabla I = (I_x, I_y)$  et  $I_t$  représentent respectivement le gradient spatial et la dérivée temporelle de la fonction intensité et  $\vec{v} = (u, v)$  est le vecteur vitesse <sup>2</sup> [SCH'89], [CHU'89]. L'équation de contrainte permet de déterminer très rapidement et pour chaque pixel la projection du vecteur vitesse sur la direction du gradient

spatial, donc une seule composante du vecteur vitesse. Cette composante est appelée "normale", car elle est perpendiculaire à l'orientation locale du contour. La même relation obtenue pour les points voisins est alors utilisée pour lever l'indétermination de la deuxième composante, généralement par un processus itératif de régularisation, qui est bien plus long que l'estimation de la composante normale. En pratique cependant, l'équation de contrainte mène à de nombreuses estimations erronées (fig.2), dues à l'amplification du bruit par la dérivation d'une part et à l'existence de zones où des termes d'ordre plus élevé ne peuvent pas être négligés d'autre part. Par conséquent, la régularisation sert souvent à minimiser l'influence de ces erreurs plutôt qu'à propager l'information utile.

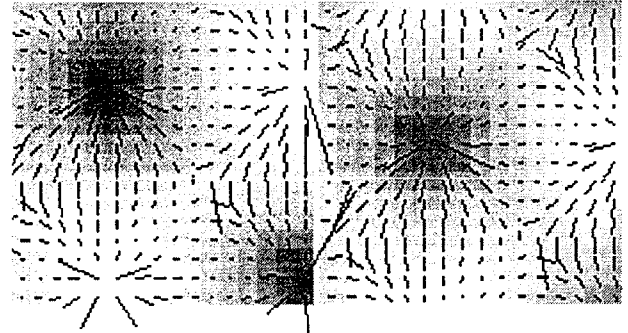


Figure 2. Composante normale des vecteurs vitesse : bien qu'il s'agisse d'une séquence synthétique où le seul bruit est dû à la quantification à 256 niveaux de gris, certains vecteurs sont aberrants (notamment au voisinage des extrema locaux de l'intensité). Pour juger la qualité des autres vecteurs, il faut préciser que tous les points de l'image effectuent un même mouvement de translation (cf fig.1a), par conséquent les vecteurs ayant une même orientation devraient avoir des longueurs identiques, ce qui n'est pas toujours le cas.

Avant de régulariser, il faut donc s'assurer que les estimations initiales sont fiables. Il nous semble plus judicieux de renoncer à l'estimation initiale pour certains pixels et de déduire leur mouvement du mouvement des points voisins, plutôt que de propager des valeurs aberrantes. En outre, dans certaines zones de l'image, il est possible de retrouver la composante manquante du vecteur vitesse avant même la régularisation. L'intérêt d'avoir de telles estimations, auxquelles on peut attribuer un coefficient de confiance élevé, est d'accélérer la convergence de la régularisation.

Citons notamment la méthode de SCHUNCK [SCH'89], qui utilise la notion de droite de contrainte, reflet graphique de l'équation de contrainte dans le plan des composantes orthogonales du vecteur vitesse. L'orientation de cette droite étant dépendante de l'orientation du gradient spatial pour le point considéré, si l'on trouve sur un même objet deux points où les orientations du gradient sont suffisamment différentes, l'intersection de leurs droites de contrainte détermine les deux composantes du vecteur vitesse de cet objet. La méthode de Schunck consiste à déterminer l'ensemble des intersections de la droite de contrainte du point considéré avec les droites de contrainte des points appartenant à un voisinage donné, et à chercher le centre de la partie la plus "dense" de cet ensemble.

<sup>2</sup> précisons que les termes vecteur vitesse et vecteur déplacement sont utilisés alternativement et ont le même sens du moment où l'intervalle de temps  $dt$  qui s'écoule entre deux images consécutives, est pris comme unité de temps.

#### IV. SOLUTION PROPOSEE

Compte tenu des remarques précédentes, nous avons fait les choix suivants :

- ne pas estimer le mouvement dans les zones homogènes où aucune méthode ne peut donner un résultat fiable (à l'échelle de résolution considérée), mais signaler ce fait par une étiquette ;
- calculer un maximum de vecteurs vitesse à l'aide de l'équation de contrainte du premier ordre, en déterminant auparavant les zones de validité de cette équation :
  - si l'orientation du gradient spatial varie suffisamment, calculer les deux composantes du vecteur vitesse par intersection de droites de contrainte ;
  - sinon <sup>3</sup>, estimer seulement la composante normale, mais signaler ce fait par une étiquette ;
- hors des zones de validité de l'équation de contrainte, utiliser la mise en correspondance.

La première étape de notre algorithme consiste donc à éliminer les zones homogènes en leur attribuant l'étiquette appropriée. Le critère utilisé est le module du gradient qui doit être supérieur à un seuil fixé.

Ensuite, afin de déterminer le domaine de validité de l'équation de contrainte du premier ordre, il faut en principe calculer, pour la fonction intensité, les dérivées partielles d'ordres plus élevés et retenir les zones où ces dérivées ont des valeurs faibles. Nous nous limitons aux dérivées du second ordre, car la sensibilité au bruit augmente à chaque dérivation. Pour évaluer les dérivées spatiales, nous calculons le laplacien, ceci par convolution de l'image avec un masque de taille  $3 \times 3$ . Le résultat est ensuite seuillé.

En revanche, ne disposant que de deux images consécutives, il n'est pas envisageable de calculer la dérivée partielle seconde selon l'axe du temps. De plus, la résolution temporelle dans les séquences d'images n'est généralement pas suffisamment fine. Cependant, nous avons constaté expérimentalement que les variations les plus rapides de la dérivée temporelle première ont lieu au voisinage du passage de cette dérivée par zéro, autrement dit là où l'orientation locale du gradient spatial est orthogonale à la direction du mouvement et par conséquent la composante normale est nulle. Une solution de remplacement consiste donc à seuiller la valeur absolue de la dérivée temporelle.

La mise en correspondance est alors effectuée pour chaque bloc contenant des pixels pour lesquels l'équation de contrainte n'est pas valable. Le vecteur vitesse obtenu pour le bloc est attribué seulement à ces pixels-là, car ils peuvent correspondre à une frontière entre deux objets, et les points situés de part et d'autre de la frontière n'effectuent pas forcément le même mouvement.

Pour tous les autres points, la dérivée temporelle et le gradient spatial calculés précédemment, permettent de déterminer les droites de contrainte correspondantes. Ces paramètres étant chargés d'erreurs dues à l'amplification du bruit par la dérivation, nous procédons d'abord à un lissage : la droite initialement calculée pour un point est remplacée par la "moyenne" des droites des points voisins les plus proches (voisinage  $3 \times 3$ ). Cette démarche est justifiée par le fait que les vitesses et les

<sup>3</sup> dans les zones où l'hypothèse 4° n'est pas respectée.

orientations du gradient pour des voisins très proches sont similaires. Bien entendu, seuls les points où la droite de contrainte est définie (donc jugée fiable) sont pris en compte.

Inversement, la différence d'orientation des gradients, nécessaire à un calcul fiable de l'intersection des droites de contrainte, ne peut être trouvée que pour des points assez éloignés. Nous avons choisi d'explorer le pourtour d'un voisinage de taille fixée, en évitant cependant que les points pris en compte ne se situent au delà d'une frontière, et de retenir le point dont la droite de contrainte forme avec la droite du point central l'angle le plus proche de l'angle droit. L'intersection des droites de contrainte correspond à la solution du système d'équations de contrainte obtenues pour le point central et le voisin retenu :

$$\begin{aligned} u I_{x1} + v I_{y1} &= -I_{t1} \\ u I_{x2} + v I_{y2} &= -I_{t2} \end{aligned} \Rightarrow \begin{aligned} u &= \frac{D_u}{D} \\ v &= \frac{D_v}{D} \end{aligned} \quad \text{avec} \quad \begin{aligned} D &= I_{x1}I_{y2} - I_{y1}I_{x2} \\ D_u &= I_{y1}I_{t2} - I_{t1}I_{y2} \\ D_v &= I_{t1}I_{x2} - I_{x1}I_{t2} \end{aligned}$$

Le déterminant D est nul lorsque les droites sont parallèles. Normé par le produit des modules du gradient  $d = |D| / (\sqrt{I_{x1}^2 + I_{y1}^2} \sqrt{I_{x2}^2 + I_{y2}^2})$ , il atteint au maximum la valeur 1, ceci lorsque les deux droites sont orthogonales. Il fournit donc une bonne mesure de l'angle formé par les deux droites.

#### V. RESULTATS

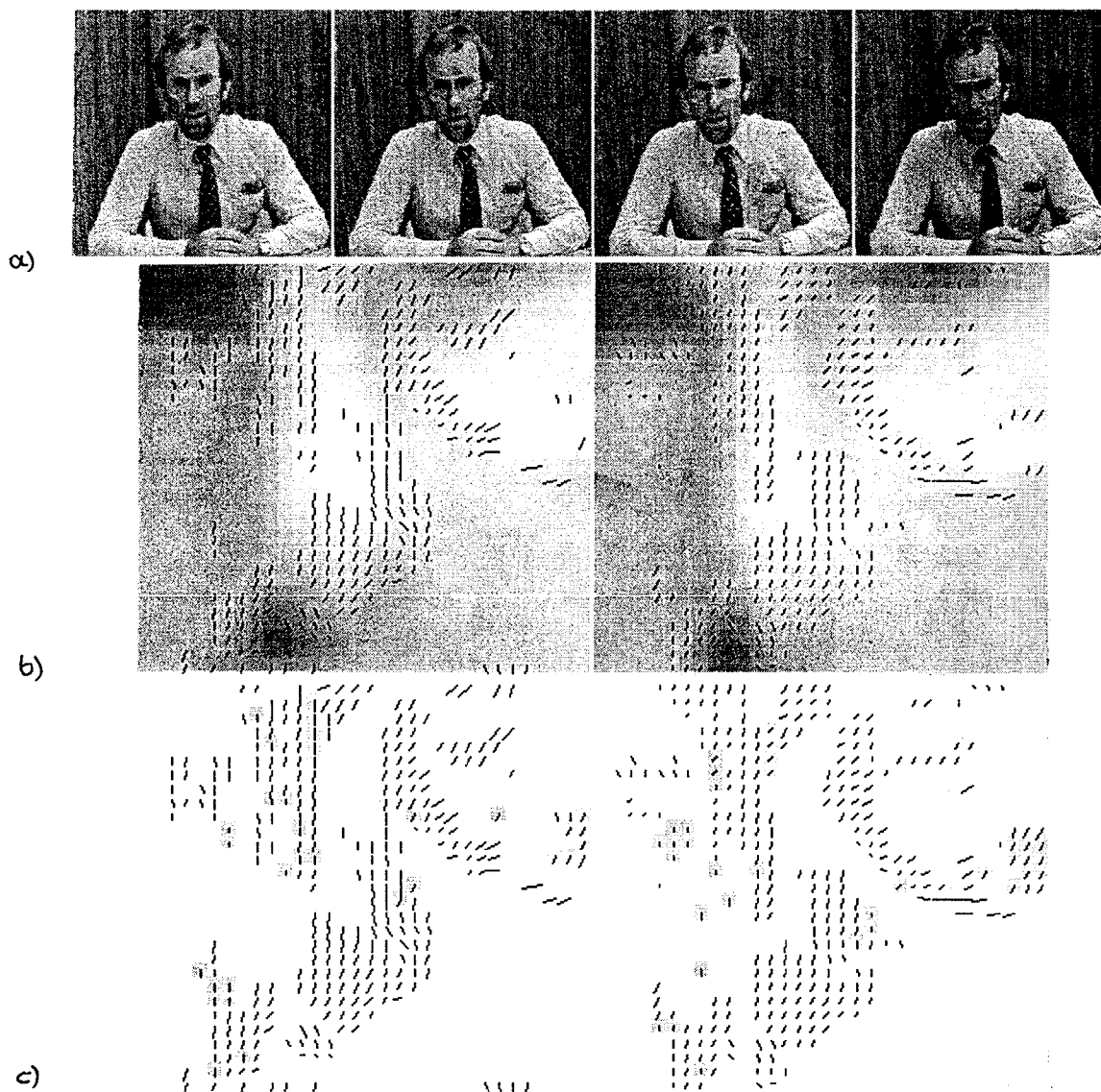
Comparons d'abord notre méthode différentielle avec la méthode de Schunck dont nous nous sommes inspirés. Dans la méthode de Schunck, le calcul d'une intersection "moyenne" permet de filtrer l'influence du bruit lorsque tous les points du voisinage considéré appartiennent à la zone de validité de l'équation de contrainte. Au lieu d'analyser un ensemble d'intersections, nous en calculons une seule qui, du fait d'un angle important entre les droites choisies, est la moins sensible au bruit. Par conséquent, notre méthode est plus rapide sans pour autant être ici moins fiable. Lorsque certains points se situent hors de la zone de validité de l'équation de contrainte, les résultats de la méthode de Schunck peuvent être erronés, voire complètement faux, surtout dans le cas où la droite de contrainte du point central est fautive. Nous nous efforçons de prendre en compte exclusivement les droites fiables. Pour les points où l'équation de contrainte ne s'applique pas, le déplacement est estimé par mise en correspondance, méthode plus lente mais très efficace.

La méthode proposée permet d'obtenir un champ de vecteurs vitesse relativement dense et de bonne cohérence temporelle : les mêmes configurations de vecteurs se retrouvent au long des trajectoires des objets (fig.3). Subsistent cependant quelques estimations aberrantes obtenues par la méthode différentielle. Ceci est probablement dû à l'imperfection de la sélection des zones de validité de cette méthode. Nous envisageons entre autres l'augmentation de la taille de l'opérateur laplacien utilisé et la prise en compte des dérivées mixtes.

#### VI. CONCLUSIONS

Nous avons proposé un algorithme d'estimation du mouvement, dont l'originalité réside dans l'utilisation combinée de deux techniques : la technique différentielle, rapide mais ayant un champ d'application restreint, et la technique de mise en





correspondance, plus lente, utilisée pour les points où elle est très efficace, situés hors du domaine de validité des méthodes différentielles. Le champ de vecteurs vitesse obtenus est déjà assez cohérent avant toute régularisation, mais nous espérons augmenter encore sa fiabilité par un meilleur choix du critère de validité de la technique différentielle. Dans un deuxième temps, nous envisageons une régularisation du champ de vecteurs obtenu, afin de propager les mesures vers les zones où le mouvement n'a pas pu être estimé de manière fiable (homogènes ou sujettes au problème d'ouverture). Enfin, disposant d'un champ dense de vecteurs à un niveau de résolution donné, il sera envisageable de propager l'information vers un niveau de résolution différent et de s'adapter ainsi à différentes fréquences spatiales et différentes vitesses présentes dans l'image.

#### REFERENCES

- [AIS'89] AISBETT J. "Optical flow with an intensity-weighted smoothing", *IEEE Trans. Pattern Anal. Mach. Intell.*, 11/5, 1989, 512-522.  
 [ANA'89] ANANDAN P. "A Computational Framework and an Algorithm for the Measurement of Visual Motion", *Int. J. Comp. Vision*, Vol. 2, 1989, pp 283-310.  
 [CHU'88] CHU C.H., DELP E.J. "Estimating displacement vectors from an image sequence",

*Figure 3. Exemple de résultats obtenus pour une séquence d'images naturelles : a) images 1 à 4 de la séquence standard TWY représentant une vidéoconférence, lissées par un filtre gaussien de taille  $7 \times 7$ , b) zoom sur le visage avec superposition des vecteurs vitesse, obtenus par l'intersection de droites de contrainte, c) image des étiquettes avec superposition des vecteurs vitesse, zoom sur le même fragment, les niveaux de gris représentent les étiquettes : gris foncé = mise en correspondance, gris clair = zones de validité de l'équation de contrainte, le fond blanc correspond aux régions homogènes. Pour la clarté, les vecteurs dont seule la composante normale a pu être déterminée, ne sont pas représentés.*

*J. Opt. Soc. Am. A*, 6/6, 1989, 871-878.

- [HEE'88] HEEGER D.J. "Optical Flow Using Spatio-temporal Filters", *Int. J. Comp. Vision*, 1988, 279-302.  
 [JAC'87] JACOBSON L., WECHSLER H. "Derivation of Optical Flow Using Spatiotemporal-Frequency Approach", *Comp. Vision, Graphics and Image Proc.*, 38, 1987, 29-65.  
 [WU'90] WU S.F., KITTLER J. "A differential method for simultaneous estimation of rotation, change of scale and translation", *Signal Processing : Image Comm.* 2, 1990, 69-80.  
 [ZHE'92] ZHENG W.J., BLOSTEIN S.D. "Adaptive Regularization for Motion Field Estimation", *Signal Processing VI: Theories and Appl.*, Elsevier, 1992, 1327-1330.