

Mise en correspondance par invariants locaux

Matching Using Local Invariant

par Cordelia SCHMID et Roger MOHR

GRAVIR - INRIA*
655 avenue de l'Europe
38330 Montbonnot, France
2 004, route des Lucioles, BP 93
06 902 Sophia Antipolis Cedex, France
tél : +33 76 61 52 30 fax : +33 76 61 52 52
e-mail : Cordelia.Schmid@imag.fr

résumé et mots clés

Le but de cet article est de proposer une alternative aux méthodes de mise en correspondance. Cette dernière repose sur le calcul d'invariants de la fonction de luminosité. Ceci permet d'effectuer une mise en correspondance dans le cas d'une transformation importante entre deux images. Pour ce faire, nous utilisons des mesures différentielles invariantes en rotation et une approche multi-échelle. Pour autoriser la mise en correspondance entre des images éclairées de façon différente, les invariants utilisés sont également robustes à de tel changement. Une mise en œuvre simple de la méthode fournit des résultats de bonne qualité autorisant des applications variées.

Mise en correspondance, Invariants différentiels, multi-échelle, Point d'intérêt

abstract and key words

This paper presents an alternative to standard matching methods. Its main novel feature is to rely on greylevel invariants that are computed locally. This allows matching even in presence of large image change in rotation or scaling. These invariants are the set of differential invariants in rotation with an additional multiscale approach for taking into account scale changes. Such invariants were also expanded towards illumination changes allowing therefore large illumination variations. The overall method is quite easy to be understood and implemented and provides surprisingly good results.

Matching, Differential invariant, Multiscale, Interest point.

1. le problème et ses approches

Nous nous intéressons ici au problème de la mise en correspondance entre des points d'une image. C'est un problème majeur en vision par ordinateur; c'est par exemple la clé de l'appariement en stéréo-vision. La mise en correspondance image-modèle est aussi le moyen le plus utilisé pour identifier des formes complexes. Pour voir la difficulté du problème que nous cherchons à résoudre, on pourra se reporter aux figures 12 à 17.

* Ce travail est effectué au sein du projet MOVI appartenant au CNRS, à l'INPG, l'INRIA et l'UJF.

1.1. alternatives existantes

Passer en revue un thème aussi largement traité dans la vision par ordinateur est pratiquement impossible. Aussi nous allons ici simplement rappeler, et discuter, les principales façon d'opérer la mise correspondance entre images pour mieux situer notre contribution.

La méthode de mise en correspondance la plus ancienne est la corrélation du signal (voir par exemple [Fau 92] pour une comparaison des méthodes, voir aussi [Zha 89] ou [Zab 94] pour une corrélation binaire rapide). Pour réussir, le signal doit avoir été pris dans des conditions voisines pour que les fenêtres de corrélation se superposent correctement par simple translation. Si une rotation sensible de l'image est intervenue, alors cette méthode échoue. On notera cependant des études qui permettent, lorsque

la géométrie épipolaire est connue, de compenser des transformations géométriques et ainsi obtenir des mises en correspondance au niveau sub-pixel ([Ack 84], [Rem 94], [Lot 94]).

Pour restreindre la mise en correspondance sur les éléments les plus informatifs du signal, bon nombre d'auteurs se sont limités aux lignes de contraste. On peut alors obtenir une structure de contours (essentiellement la structure du graphe [Hor 90] extrait de ces contours) qui capture la structure globale de la scène. C'est l'utilisation de cette ressemblance globale qui va pouvoir être utilisée pour la mise en correspondance en utilisant des techniques d'optimisation comme la relaxation [Lon 86], ou alors des techniques combinatoires comme la recherche de cliques maximales [Hor 89]. On notera que dans toutes ces approches, la combinatoire reste si forte qu'il faut la contraindre par l'usage de la contrainte épipolaire afin de rester dans des temps d'exécution raisonnables.

Sur les contours on peut aussi vouloir utiliser des informations locales. En théorie, à un point d'un contour courbe peut être associé un invariant qui permet de discriminer ce point [Wei 91]. De fait ce genre de méthode n'est pas praticable. En revanche, si une partie suffisante du contour est visible, des méthodes semi-locales [vD 94], [Rot 92] permettent de calculer des invariants permettant de classer le morceau de contours observé. On peut ainsi faire une mise en correspondance qui est au moins grossière, même si cette application directe n'a pas été faite à notre connaissance.

Ce type de méthode n'est applicable que si la segmentation des images en contours est correctement faite. Sur des scènes aux structures simples, ces méthodes ont prouvé leur applicabilité, mais elles restent néanmoins difficiles à maîtriser.

1.2. buts poursuivis et approche proposée

Nous souhaitons avoir un outil de mise en correspondance suffisamment général pour pouvoir considérer des images prises dans des conditions très différentes. Ceci est par exemple nécessaire pour l'autocalibration caméra-pince d'un robot : ce type d'étalonnage implique différentes rotations avec des axes non parallèles conduisant à des images très différentes d'une même scène [Hor 95]. Il nous faut donc identifier des formes et rester peu sensible à un déplacement quelconque dans l'image (rotation et translation image). La caméra pouvant s'approcher, nous souhaitons aussi être capable d'absorber un changement d'échelle raisonnable, disons un facteur 2 ; il est clair que des facteurs d'échelle trop importants modifient la structure d'une forme. Si le facteur d'échelle devient trop petit, une courbe concave se transforme en courbe convexe puis en un simple point. Il faut donc le limiter.

Il nous faut enfin être capable de gérer des occultations, d'avoir des mouvements différents pour les formes en présence et opérer sur des scènes complexes (voir figure 14 et 15).

L'absence de structure globale nous oblige à rechercher une solution du côté des méthodes locales ; les grandes variétés de déformation excluent l'usage de la corrélation ; l'absence de contour fort ne permet pas de s'appuyer sur cette information.

Il est illusoire d'espérer un processus de mise en correspondance parfait avec un cahier de charges aussi contraignant. Les expériences menées avec des méthodes robustes [Sch 90] prouvent qu'avec entre 1/3 et 2/3 d'appariements initiaux corrects, on peut le plus souvent opérer de façon satisfaisante (voir par exemple une application à la détermination de la géométrie épipolaire : [Zha 94]). Nous nous sommes donc fixé comme contrainte d'obtenir au moins 70 % d'appariements corrects.

La nature géométrique des contraintes que nous nous sommes imposées nous a conduit à nous intéresser à l'étude du signal au voisinage de points particuliers, les points d'intérêts : ils fournissent des indications qui sont locales et en ces points on peut espérer avoir un signal suffisamment riche pour permettre une bonne identification. De fait, une telle démarche existe déjà dans la littérature scientifique.

Souvent la corrélation est appliquée uniquement au points d'intérêt, [Zha 94]. Pour améliorer une telle corrélation [Der 94] propose l'utilisation des caractéristiques suivantes : la direction du gradient, la courbure et la disparité qui est supposée en dessous d'un certain seuil. Malgré l'utilisation de telles caractéristiques locales, la méthode ne résiste pas aux rotations. Pour résoudre ce problème Hu [Hu 94] propose d'effectuer la corrélation en plusieurs directions.

Il est également possible de caractériser localement un signal en le représentant dans l'espace fréquentiel. Le filtre de Gabor est souvent utilisé pour des problèmes de stéréo-correspondance ou pour réaliser la vergence d'une tête stéréoscopique active [Wes 92, Fle 91, San 88]. Ce filtre permet d'obtenir une information locale de la phase du signal et sert ainsi à estimer la disparité entre deux images. Toutefois cette mesure est locale en fréquence et directionnelle en espace. Aussi, une telle approche résiste uniquement à des rotations et des changements d'échelle faibles. Pour résoudre ce problème, Wu [Wu 94] a proposé d'utiliser des filtres de Gabor dans plusieurs directions et à plusieurs échelles.

Les approches [Hu94] et [Wu 94] utilisent des mesures directionnelles pour déterminer des mesures invariantes en rotation. Ces solutions nécessitent la discrétisation de l'espace suivant plusieurs directions et reposent sur une hypothèse de linéarité entre deux directions adjacentes. Cette hypothèse est difficilement vérifiable. Ceci explique pourquoi ces trois méthodes ont des résultats médiocres. Afin de résoudre ce problème, nous proposons de caractériser les points d'intérêt par des mesures locales invariantes en rotation. Grossièrement, ceci est possible en effectuant des calculs symétriques en rotation. Nous n'avons donc pas besoin de discrétiser l'espace et nos mesures ne reposent pas sur l'évaluation d'un paramètre auxiliaire instable. En outre, notre méthode est robuste à un changement d'échelles par l'utilisation d'une approche multi-échelle.

1.3. plan de l'article

Cet article est organisé comme suit : En section 2 nous décrivons comment caractériser les points d'intérêt d'une manière locale. Ensuite une telle caractérisation est utilisée pour la mise en correspondance. La section 3 explique comment rendre la caractérisation locale invariante aux rotations images, aux changements d'échelle et aux changements de luminosité. En section 4 nous passons alors aux validations expérimentales avec les différents types de modifications que nous autorisons et avec différents types d'images.

2. principe de la méthode

La méthode consiste en trois étapes : l'extraction de points d'intérêt, la caractérisation de ces points d'intérêt et leur mise en correspondance. Dans cette section nous décrivons ces trois étapes. Ensuite en section 3 nous explicitons comment rendre une telle caractérisation invariante aux différents transformations images.

La figure 1 illustre notre approche : chacun des points d'intérêt est décrit localement à partir d'une caractérisation locale du signal. L'extraction de tels points d'intérêt est décrite en section 2.1, le principe d'une caractérisation locale du signal en section 2.2 et la procédure d'appariement proprement dite en section 3.

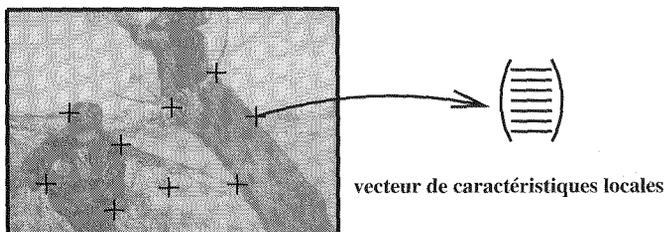


Figure 1. - Représentation d'une image.

2.1. extraction de points d'intérêt

L'extraction des points d'intérêt est la première étape du processus de mise en correspondance.

Pour ce faire il existe de nombreux algorithmes dans la littérature. De tels algorithmes sont soit basés sur une extraction des contours préalable soit utilisent directement le signal de niveaux de gris. L'avantage principal des détecteurs basés sur le signal de niveaux de gris est que leur performance n'est pas dépendante de la réussite ou de l'échec de l'extraction des contours.

Dans un contexte de mise en correspondance, il est important que le détecteur soit répétable, c'est-à-dire invariant aux transformations images. Une imprécision dans la localisation des points

d'intérêt mène à des caractérisations trop différentes qui sont difficilement utilisables dans un contexte d'appariement. Dans [Sch 96] nous avons comparé différents détecteurs en présence d'une rotation image, d'un changement d'échelle, d'un changement de luminosité et du bruit image. Ce travail a montré que la stabilité de méthodes existantes varie de façon importante et qu'on obtient les meilleurs résultats pour le détecteur de Harris [Har 88].

Ce détecteur considère un point d'intérêt comme un point lequel les courbures principales de la fonction d'autocorrélation est maximale, ce qui revient à considérer les valeurs propres de la matrice

$$\begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

où I désigne la fonction de luminance de l'image. Une implémentation stabilisée a été utilisée dans ce travail en utilisant des dérivées Gaussiennes. Une implémentation récursive des ces filtres assure une détection rapide.

La figure 2 montre les points d'intérêt détectés sur la même scène pour une rotation. Elle montre la répétabilité du détecteur obtenu. Il faut cependant noter que tous les points ne sont pas répétés, mais un grand nombre. Ceci est suffisant dans le cadre d'un processus statistiquement robuste.

2.2. caractérisation locale du signal

De nombreuses caractérisations locales sont possibles. Dans le cas d'images de niveaux de gris, des exemples en sont les dérivées, les filtres de Gabor - et plus généralement les ondelettes - et les moments. Dans ce travail nous avons utilisé des dérivées.

Une fonction peut être approximée localement par ses dérivées. Sachant calculer les dérivées d'une fonction en un point jusqu'à un ordre N , la série de Taylor décrit cette fonction localement jusqu'à cet ordre :

$$f(x_0 + x, y_0 + y) = f(x_0, y_0) + x f_x(x_0, y_0) + y f_y(x_0, y_0) \dots + O(x^N, y^N)$$

où $f_x = \frac{\partial f}{\partial x}$ et $f_y = \frac{\partial f}{\partial y}$

De ce fait il est possible de décrire une image en un point en stockant dans un vecteur l'ensemble des dérivées en ce point. Un tel vecteur a été utilisé par Koenderink [Koe 87] qui l'a nommé « jet local ». Koenderink calcule en outre les dérivées d'une manière stable en utilisant un filtre passe-bas. Nous allons d'abord décrire le calcul stable de ces dérivées et ensuite préciser le « jet local ».

2.2.1. calcul de dérivées

Le calcul des dérivées est mal conditionné dans le sens où il manque de robustesse vis-à-vis du bruit dans les données d'entrée. Même un bruit faible peut perturber ce calcul de façon importante. Pour illustrer ce manque de robustesse, considérons les fonctions $f(x)$ et $\hat{f}(x) = f(x) + \epsilon \sin(\omega x)$. Elles sont similaires, si ϵ est petit. Toutefois $f'(x)$ peut différer beaucoup de $\hat{f}'(x)$ pour un

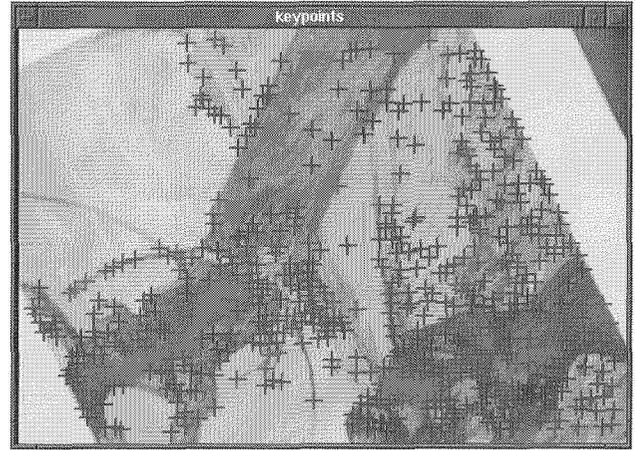
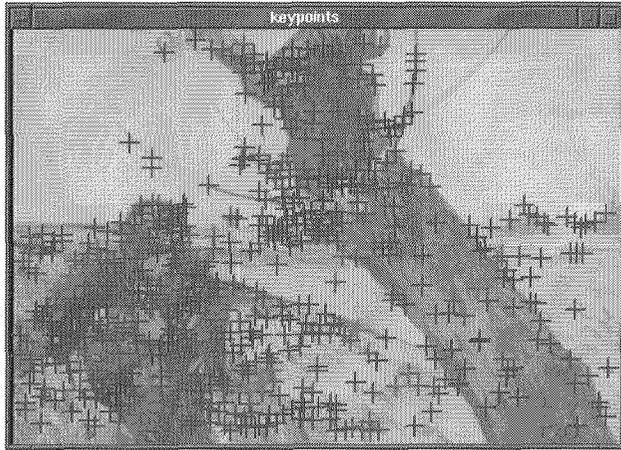


Figure 2. – Points d'intérêt détectés pour deux rotations différentes. Angle de rotation entre les deux images : 155 degrés. Taux de répétabilité obtenu : 92%.

ω grand puisque $\hat{f}'(x) = f'(x) + \omega \epsilon \cos(\omega x)$. En conséquence un bruit de relativement haute fréquence dans une image peut modifier considérablement la première dérivée et donc a fortiori les dérivées d'ordre supérieur.

Préalablement à tout calcul de dérivation, il est donc nécessaire d'effectuer un lissage. Puisque la différenciation commute avec la convolution : $\partial_i(g * f) = g * \partial_i f = \partial_i g * f$, un tel lissage peut être obtenu soit en lissant l'image soit en lissant l'opérateur de dérivation. Un moyen simple pour stabiliser les calculs de dérivation est donc d'utiliser les dérivées d'une fonction de lissage. Le choix le plus communément fait est alors de considérer la fonction Gaussienne [Wit 83, Tor 86, Rom 94a, Flo 93, Lin 94]. La formule de la Gaussienne $G(\vec{x}, \sigma)$ et de ses dérivées $G_{i_1 \dots i_n}(\vec{x}, \sigma)$ pour un \vec{x} de dimension 2 est :

$$G(\vec{x}, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\vec{x}^2}{2\sigma^2}\right) \quad (1)$$

$$G_{i_1 \dots i_n}(\vec{x}, \sigma) = \frac{\partial^n}{\partial_{i_1} \dots \partial_{i_n}} G(\vec{x}, \sigma) \quad (2)$$

avec $\vec{x} = (x_1, x_2)$, $1 \leq i_k \leq 2$

Le σ de la fonction Gaussienne détermine la quantité de lissage effectué. Ce σ peut également être interprété comme un facteur d'échelle. Par la suite nous appellerons σ la taille de la Gaussienne.

2.2.2. jet local

Ayant calculé les dérivées d'une manière robuste, ceci permet de définir le « jet local » :

Définition 2.1 Jet local

Soit I une image et σ un facteur d'échelle. Le « jet local » d'ordre N en un point \vec{x} , noté $J^N[I](\vec{x}, \sigma)$, est défini par

$$J^N[I](\vec{x}, \sigma) = \{L_{i_1 \dots i_n}(\vec{x}, \sigma) \mid (\vec{x}, \sigma) \in I \times \mathbb{R}^+, n = 0, \dots, N\}$$

Où $L_{i_1 \dots i_n}(\vec{x}, \sigma)$ est la convolution de I avec la différentielle $G_{i_1 \dots i_n}(\vec{x}, \sigma)$ de la fonction gaussienne :

$$L_{i_1 \dots i_n}(\vec{x}, \sigma) = G_{i_1 \dots i_n}(\vec{x}, \sigma) * I(\vec{x})$$

$G(\vec{x}, \sigma)$ et $G_{i_1 \dots i_n}(\vec{x}, \sigma)$ sont définis par les équations 1 et 2. L'opérateur $*$ représente l'opérateur de convolution.

Un « jet local » est calculé au voisinage d'un point et décrit la géométrie locale de ce voisinage ainsi que les caractéristiques géométriques qu'il contient. Pour un point donné le jet local est fonction d'un paramètre : la taille σ de la gaussienne. Ce paramètre permet de caractériser une fonction à plusieurs niveaux d'échelle ou d'adapter l'échelle à l'image considérée.

Le « jet local » est basé sur la dérivation du signal. Toutefois, comme on ne connaît pas la fonction du signal de manière analytique, les dérivées doivent être estimées de façon numérique. Elles sont calculées par convolution avec une gaussienne et ses dérivées. On peut donc interpréter le jet local comme la donnée du développement limité de la fonction de niveau de gris au point considéré.

3. caractérisation locale

Dans la suite nous détaillons comment rendre la caractérisation locale basée sur les dérivées invariante et robuste aux différentes transformations images, notamment une rotation image, un changement d'échelle et un changement de luminosité. Ensuite, nous discutons la possibilité d'extension à d'autres transformations.

3.1. invariance à une rotation image

Nous présentons d'abord une approche pour obtenir l'invariance à une rotation image. Cette approche est basée sur des mesures

différentielles invariantes au groupe des déplacements, c'est-à-dire des combinaisons de dérivées invariantes à une rotation image. Il est également possible d'obtenir une caractérisation invariante au groupe des déplacements en utilisant des filtres ajustables. Dans ce cas les dérivées sont ajustées dans la direction du gradient et donc invariantes à une rotation image, voir [Fre 91] pour plus de détails. Nous n'avons pas retenu une telle solution, car elle est fortement influencée par le calcul de l'angle.

3.1.1. invariants différentiels

À partir du « jet local » défini à la section 2.1, Koenderink [Koe 87] ainsi que Romeny *et al.* [Sal 92, Flo 93, Rom 94b, Rom 94a] proposent de calculer des invariants différentiels, invariants pour le groupe des déplacements $SO(2)$. Pour ce faire ils ont repris des résultats mathématiques, formulés entre autre par Hilbert [Hil 93]. Pour pouvoir calculer ces invariants à un ordre élevé, il est nécessaire d'implémenter le calcul des dérivées de manière stable. L'ensemble d'invariants que nous utilisons est regroupé dans un vecteur noté \vec{V} . Pour calculer ce vecteur, la dérivation est effectuée jusqu'au troisième ordre. La première partie de ce vecteur est constituée d'un ensemble complet et irréductible d'invariants différentiels jusqu'au deuxième ordre. La formulation de cette première partie du vecteur est donnée en notation d'Einstein (voir l'équation 3) et en coordonnées cartésiennes (voir l'équation 4). Les L_i sont les éléments du « jet local » défini par la définition 2.1. L représente la fonction de luminance convoluée avec une gaussienne. Il est donc possible de calculer les L_i et donc les invariants pour plusieurs tailles de σ de la gaussienne. En outre, un tel calcul des dérivées garantit leur stabilité.

$$\vec{V}[0..4] = \begin{bmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ji} \end{bmatrix} \quad (3)$$

En notation cartésienne, les indices x et y représentent respectivement la dérivation par rapport aux variables x et y , par exemple $L_{xy} = \frac{\partial^2}{\partial x \partial y} L$. En notation d'Einstein, un indice i signifie la sommation des dérivations par rapport à l'ensemble des variables :

$$L_i = \sum_i L_i = L_x + L_y$$

où les indices x et y représentent respectivement la dérivation par rapport aux variables x et y .

En notation cartésienne nous obtenons donc :

$$\vec{V}[0..4] = \begin{bmatrix} L \\ L_x L_x + L_y L_y \\ L_{xx} L_x L_x + 2L_{xy} L_x L_y + L_{yy} L_y L_y \\ L_{xx} + L_{yy} \\ L_{xx} L_{xx} + 2L_{xy} L_{xy} + L_{yy} L_{yy} \end{bmatrix} \quad (4)$$

On peut constater que la deuxième composante de ce vecteur est la magnitude du gradient et la quatrième le Laplacien.

La deuxième partie du vecteur est constituée d'un ensemble complet d'invariants du troisième ordre en notation d'Einstein (cf. 5).

$$\vec{V}[5..8] = \begin{bmatrix} \varepsilon_{ij} (L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{ijj} L_j L_k L_k - L_{ijjk} L_i L_j L_k \\ -\varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix} \quad (5)$$

où ε_{ij} représente le tenseur canonique anti-symétrique : $\varepsilon_{12} = -\varepsilon_{21} = 1$ et $\varepsilon_{11} = \varepsilon_{22} = 0$.

3.1.2. normalisation en taille de l'image

Pour obtenir effectivement des dérivées invariantes en rotation, il faut de plus constater que la taille des pixels doit être carrée. Sinon le calcul des dérivées est influencé par le rapport longueur/largeur d'un pixel. Il faut donc normaliser l'image. Ceci est fait par interpolation linéaire sur les colonnes de l'image en utilisant un facteur de réduction égal au facteur « α_v/α_u ». Ce facteur de normalisation « α_v/α_u » représente le ratio entre la largeur et la longueur d'un pixel. Différentes expérimentations ont montré que ce facteur est stable et peu dépendant du calibrage.

3.2. robustesse à l'échelle

Un changement d'échelle peut être dû soit à un changement de la distance entre la caméra et l'objet soit à un changement de la longueur focale de l'objectif (dans le cas d'un zoom). Nous noterons dans la suite un changement d'échelle par α de manière à le distinguer de la taille σ de la gaussienne utilisée pour effectuer les calculs de dérivation.

Bien qu'il existe des invariants théoriques à un changement d'échelle, de tels invariants ne sont pas valables si les dérivées sont calculées à partir du jet local. Pour obtenir une caractérisation robuste à un changement d'échelle, il est donc nécessaire d'utiliser une approche multi-échelle.

3.2.1. invariants à un changement d'échelle

Étant donnée une fonction f , un changement d'échelle α peut être décrit par un changement de variable : $f(x) = g(u)$ où $g(u) = g(u(x)) = g(\alpha x)$. De cette relation découlent les relations suivantes entre f et g :

$$f^{(n)}(x) = \alpha^n g^{(n)}(u) \quad (6)$$

où $f^{(n)}(x)$ représente la dérivée n -ième de f .

L'équation 6 montre que les dérivées premières de f et de g sont égales à un facteur multiplicatif α près. De façon équivalente,

les dérivés secondes (respectivement n -ièmes) sont égales à un facteur multiplicatif α^2 (respectivement α^n) près. Le rapport de deux dérivées de même ordre est donc un invariant théorique à un facteur d'échelle. Cet invariant s'exprime par l'équation suivante :

$$\frac{[f^{(n)}(x)]^{\frac{k}{n}}}{f^{(k)}(x)}$$

Cependant, les résultats en prenant des images à différentes échelles ont montré que de tels invariants sont peu stables à un changement d'échelle supérieur à 20%. Les résultats présentés à la section 4.3 ont montré que les invariants à la rotation étaient eux aussi robustes à un changement d'échelle de 20%. Cette robustesse est cohérente avec les observations faites par Fleet dans [Fle] dans le contexte des filtres de Gabor.

En fait, les invariants à l'échelle n'apportent pas d'amélioration parce que leur calcul numérique est nécessairement effectué sur un support. Ce support doit être adapté au changement d'échelle pour calculer effectivement l'invariant. En effet, dans le cas où les fonctions f et g sont représentées de façon discrète, l'équation 6 précédente se réécrit de la manière suivante :

$$\int_{-\infty}^{+\infty} I_1(\vec{x}) G_{i_1 \dots i_n}(\vec{x}, \sigma) d\vec{x} = \alpha^n \int_{-\infty}^{+\infty} I_2(\vec{u}) G_{i_1 \dots i_n}(\vec{u}, \sigma \alpha) d\vec{u} \quad (7)$$

où les $G_{i_1 \dots i_n}$ représentent les dérivées de la fonction gaussienne définie par l'équation 2.

Cette équation montre l'importance du support sur lequel sont effectués les calculs. Ceci est à l'origine des méthodes multi-échelle telle que celle que nous allons présenter à la section suivante.

3.2.2. approche multi-échelle

L'équation 7 a montré que la taille de la gaussienne, c'est-à-dire le support de calcul doit être ajusté pour calculer des invariants à l'échelle. Nous avons donc développé une approche multi-échelle. Une telle approche a été proposée initialement par [Wit 83] et Koenderink [Koe 84]. Plus tard, Lindeberg [Lin 94] a étendu et résumé leur approche. Il est cependant impossible de calculer les invariants à toutes les échelles. Il est donc nécessaire de discrétiser l'espace d'échelle en utilisant une approche multi-échelle. De nombreux auteurs ont proposé une discrétisation par demi-octave (voir par exemple Wu dans [Wu 95]). Avec un tel pas de discrétisation la caractérisation obtenue s'est révélée imprécise et instable. Puisque notre caractérisation est robuste à un changement d'échelle jusqu'à 20% (cf. section 4.3), nous avons choisi un pas de discrétisation qui garantit qu'entre deux échelles consécutives, le changement est inférieur à 20%. De manière à être résistant à un changement d'échelle jusqu'à un facteur 2, les différentes échelles retenues ont pour valeur : 0.48, 0.58, 0.69, 0.83, 1, 1.2, 1.44, 1.73, 2.07.

L'intégration des invariants différentiels présentés à la section 3.1 dans un cadre multi-échelle permet ainsi d'obtenir une caractérisation robuste au groupe des similitudes.

3.3. robustesse à la luminosité

Il est également intéressant d'obtenir une caractérisation qui soit robuste à un changement de luminosité. Il existe plusieurs possibilités pour modéliser un changement de luminosité, et il est possible de calculer des invariants pour chaque modèle correspondant. Ceci est illustré par la suite pour trois modèles : une simple translation, une transformation affine et une transformation monotone de la luminosité du signal.

3.3.1. translation des niveaux de gris

Une translation de la luminance se modélise par :

$$\tilde{I}(x, y) = I(x, y) + b$$

Il est facile de voir que par simple dérivation, le facteur b s'élimine et par conséquent les invariants différentiels, à part la moyenne des intensités lumineuses sont invariants à un tel changement. Le vecteur \vec{V} sans la composante $\vec{V}[0]$ est dans la suite référencé par \vec{V}_T .

3.3.2. transformation affine des niveaux de gris

Une transformation affine de la luminance se modélise par :

$$\tilde{I}(x, y) = aI(x, y) + b$$

Une telle transformation modifie les dérivées du signal de la manière suivante : $\tilde{I}^{(n)}(x, y) = aI^{(n)}(x, y)$. N'importe quel quotient de deux dérivées est donc invariant à une transformation affine de la luminance. Dans le cas des invariants différentiels, il y a différentes manières permettant d'obtenir des invariants à une transformation affine. Une possibilité est d'utiliser la magnitude du gradient. On obtient donc le vecteur \vec{V}_A :

$$\vec{V}_A[0..2] = \begin{bmatrix} \frac{L_i L_{ij} L_j}{(L_i L_i)^{3/2}} \\ \frac{L_{ii}}{(L_i L_i)^{1/2}} \\ \frac{L_{ij} L_{ji}}{L_i L_i} \end{bmatrix}$$

$$\vec{V}_A[3..6] = \frac{1}{(L_i L_i)^2} \begin{bmatrix} \varepsilon_{ij} (L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{ii} L_j L_k L_k - L_{ij} L_i L_j L_k \\ -\varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix}$$

3.3.3. transformation monotone

Un changement de luminosité peut également être modélisé par une fonction monotone et donc inversible. L'inversibilité de la fonction (ou sa stricte monotonie) est nécessaire pour éviter une perte d'information par rapport à celle contenue dans l'image de niveaux de gris. Florack dans [Flo 94] montre que le « jet local » permet également de calculer des invariants par rapport à n'importe quelle transformation inversible de la luminosité. Il fait remarquer que les isophotes ne sont pas modifiés sous l'action d'une transformation inversible de luminosité. Jusqu'au deuxième ordre il existe deux tels invariants, notamment la courbure des isophotes κ et la courbure des lignes de plus grande pente μ :

$$\kappa = \frac{\epsilon_{ij}\epsilon_{kl}L_iL_jL_kL_l}{(L_mL_n)^{3/2}} = \frac{2L_xL_yL_{xy} - L_x^2L_{yy} - L_y^2L_{xx} - L_y^2L_{xx}}{L_x^2 + L_y^2}$$

$$\mu = \frac{\epsilon_{ij}L_jL_kL_{ik}}{(L_mL_n)^{3/2}} = \frac{L_{xy}(L_yL_y - L_xL_x) + L_xL_y(L_{xx} - L_{yy})}{L_x^2 + L_y^2}$$

3.4. extension à d'autres transformations

Il est possible de calculer des invariants différentiels pour d'autres types de transformation image, par exemple pour le groupe des transformations affines.

Dans ce cas, une manière de faire est d'utiliser une transformation affine pour transformer les équations quadratiques du second ordre de dérivation de la forme $\vec{x}^T L_{ij} \vec{x} = const$ en cercle et d'effectuer les calculs dans le repère normalisé obtenu. Ceci est équivalent à calculer des invariants à partir de l'inverse de la matrice L_{ij} . On obtient alors l'ensemble suivant d'invariants jusqu'au troisième ordre :

$$\vec{V}[0..5] = \left[\begin{array}{c} L \\ L_iL_j(L^{-1})^{ij} \\ L_{ijl}L_lL_mL_n(L^{-1})^{il}(L^{-1})^{jm}(L^{-1})^{kn} \\ L_{ijk}L_l(L^{-1})^{ij}(L^{-1})^{kl} \\ L_{ijk}L_{lmn}(L^{-1})^{ij}(L^{-1})^{kl}(L^{-1})^{mn} \\ L_{ijk}L_{lmn}(L^{-1})^{il}(L^{-1})^{jm}(L^{-1})^{kn} \end{array} \right]$$

Toutefois, comme dans le cas d'un changement d'échelle, pour calculer de tels invariants, il est nécessaire de tenir compte du support de calcul. La difficulté majeure est alors de définir une méthodologie permettant de couvrir l'ensemble des paramètres admissibles pour ces transformations (comme on a pu le faire en réalisant une approche multi-échelle dans le cas d'un changement d'échelle). Étendre directement l'approche multi-échelle serait possible, mais extrêmement coûteux en ressources de calcul : il suffirait de calculer les invariants pour différents supports, chacun correspondant à un jeu de paramètres donnés, sachant qu'il y a trois paramètres à prendre en compte. Après évaluation, nous n'avons pas retenu ces invariants dans notre approche à cause de la difficulté de leur mise en œuvre.

4. expérimentations

Cette section présente les résultats obtenus. Afin de mieux cerner l'influence des différentes transformations image (rotation, changement d'échelle et changement de luminosité), nous présentons d'abord les résultats obtenus en ne considérant qu'un type de transformation. La section 4.2 illustre les résultats dans le cas d'une rotation, la section 4.3 dans le cas d'un changement d'échelle et la section 4.4 dans le cas d'un changement de luminosité. La section 4.5 montre les résultats obtenus lorsque ces changements sont combinés et les résultats pour des scènes réelles.

Une question souvent posée est l'utilité des dérivées troisième. La section 4.6) montre que les différentielles d'ordre troisième apportent une information non négligeable pour le calcul de la mise en correspondance.

La qualité des résultats est mesurée par le pourcentage d'appariements corrects par rapport au nombre total d'appariements. Ceci a été fait sur plus de 100 images. D'autre part le nombre total d'appariements est souvent supérieur à 100. Il était donc impossible d'évaluer manuellement les résultats. La section 4.1 présente la méthode automatique d'évaluation des résultats que nous avons utilisée.

Il est à noter que le processus d'appariement rejette en moyenne 50% des points détectés. Ce taux de rejet est surtout dû à l'imprécision du détecteur de points d'intérêt utilisé qui varie suivant les images considérées et la transformation entre deux images.

4.1. évaluation des résultats

Dans le cas d'une scène plane, n'importe quel changement entre deux images de cette scène peut être modélisé par une homographie [Sem 52]. Cette homographie définit une relation univoque point à point entre les deux images. À partir de cette homographie, on sait dire si un appariement est correct ou faux : les pourcentages affichés dans la suite correspondent aux taux d'appariements corrects par rapport aux appariements trouvés.

Dans le cas de scènes planes, nous avons calculé l'homographie existant entre deux images à partir des appariements trouvés. Le calcul de cette homographie utilise la méthode des moindres carrés médians et est donc robuste à la présence d'erreurs grossières si celles-ci représentent moins de 50% des données.

Dans le cas des transformations entre scènes non-planes, une évaluation automatique est plus difficile. Dans ce cas, nous utilisons la géométrie épipolaire existant entre deux images. Cette géométrie est également calculée par une méthode de moindres carrés médians. Un appariement est alors estimé exact si ses deux

points se trouvent sur des épipolaires conjuguées. Cette évaluation n'est pas parfaite puisque non univoque : deux points d'un appariement faux peuvent être sur des épipolaires conjuguées. C'est cependant un bon estimateur de la qualité du résultat.

4.2. invariance à une rotation

Pour tester l'invariance à une rotation image, nous avons effectué nos tests sur deux scènes planes différentes. La première est dénommée « Sanja ». La deuxième est le tableau « le semeur » de « Van Gogh ».

Pour chacune de ces deux scènes, nous avons pris une séquence d'images entre lesquelles nous avons fait une rotation de la caméra. La figure 3 montre les résultats obtenus pour une paire d'images de la séquence « Sanja » pour laquelle l'angle de rotation est de 133.0 degrés. Les points clairs indiquent les bons appariements, les points sombres les mauvais appariements. Le taux d'appariements corrects est de 88.52 %. La figure 4 montre les résultats pour la séquence « Van Gogh ».

La figure 5 montre le pourcentage d'appariements corrects en fonction de l'angle en degrés. Ce pourcentage a été calculé pour deux détecteurs différents de points d'intérêt : Heitger et Harris-Precis (voir [Sch 96] pour plus d'informations sur ce détecteur). Ce pourcentage a également été évalué à partir des points précis. Ces points correspondent aux points d'intérêt détectés dans une première image et transformés par homographie dans une seconde. En effet ces points sont sans erreur de détection.

La taille σ de la gaussienne utilisée est de 3 pour cette expérimentation. Les courbes diffèrent en fonction du détecteur utilisé. Pour les points précis, le taux de mise en correspondance est approximativement de 100%. Les résultats obtenus avec le détecteur Harris-Precis sont presque aussi bons. Le détecteur de Heitger donne des résultats nettement moins bons. En effet les résultats varient en fonction de l'angle de rotation. On obtient de moins bons résultats pour un angle de 90 degrés. Ceci est également vrai si l'on évalue la répétabilité de ce détecteur.

Cette expérience montre deux choses. Premièrement, elle montre que la caractérisation proposée permet de distinguer les différents points et qu'elle est invariante à la rotation. Deuxièmement, on peut observer que les résultats obtenus dépendent de la répétabilité du détecteur utilisé.

4.3. robustesse à l'échelle

Cette section montre la robustesse de notre algorithme d'appariement par rapport à un changement d'échelle. Pour ce faire, une séquence d'images a été prise pour chaque scène en faisant varier la longueur focale de notre objectif (un zoom). Les paragraphes suivants montrent la robustesse de la caractérisation lorsque l'approche multi-échelle est utilisée, puis la stabilité

des invariants différentiels lors d'un changement d'échelle. Ces expériences montrent la validité de l'approche multi-échelle et permettent de déterminer l'espacement nécessaire entre deux échelles.

4.3.1. approche multi-échelle

La figure 6 montre les résultats obtenus pour une paire d'images de la séquence « Sanja ». Le changement d'échelle entre ces deux images est d'un facteur 0.68. Les croix blanches indiquent les appariements corrects et les noires les faux appariements. Le pourcentage d'appariements corrects est de 90.91%. La figure 7 montre les résultats sur la séquence « Van Gogh » pour un facteur d'échelle de 0.68. Le pourcentage d'appariements corrects est alors de 70.31%.

La figure 8 trace le pourcentage d'appariements corrects en fonction du facteur d'échelle. Ce pourcentage a été calculé pour deux détecteurs différents de points d'intérêt : Heitger et Harris-Precis (voir [Sch 96] pour plus d'informations sur ce détecteur). Ce pourcentage a également été évalué à partir des points précis. Ces points correspondent aux points d'intérêt détectés dans une première image et transformés par homographie dans une seconde. En effet ces points sont sans erreur de détection. La taille moyenne σ de la gaussienne utilisée est de 5 pour cette expérimentation, ce qui correspond à des tailles utilisées commençant à 2.5.

Cette expérience montre qu'on obtient des résultats robustes à un changement d'échelle. On peut également observer que les résultats obtenus dépendent de la répétabilité du détecteur utilisé.

4.3.2. robustesse des invariants à un changement d'échelle

La figure 9 montre l'utilisation d'invariants sans approche multi-échelle. On peut voir que les résultats se dégradent rapidement au-dessus d'un facteur d'échelle de 1.2.

4.4. robustesse à la luminosité

Cette section présente la robustesse de notre algorithme d'appariement à un changement de luminosité. Pour ce faire des séquences d'images ont été prises en changeant l'ouverture de l'objectif. Pour mesurer ce changement nous avons introduit la notion de « niveau de gris relatif ». Cette mesure est le rapport de la moyenne des niveaux de gris d'une image de la séquence par rapport à une image de référence. L'image de référence est une image au milieu de la séquence, c'est à dire une image qui est ni très sombre ni très claire. La figure 10 montre deux images de la séquence, une très sombre avec un niveau de gris relatif de 0.6 et une claire avec un niveau de gris relatif de 1.7.

La figure 11 montre les résultats obtenus pour un changement de luminosité. Le détecteur utilisé est Harris-Precis. Cette figure

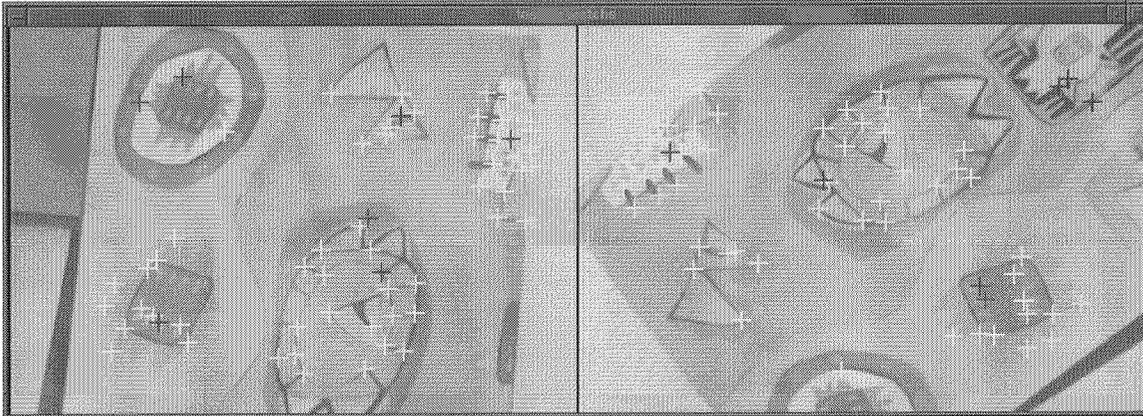


Figure 3. – 88.52 % d'appariements corrects dans le cas d'une rotation de 133.0 degrés.

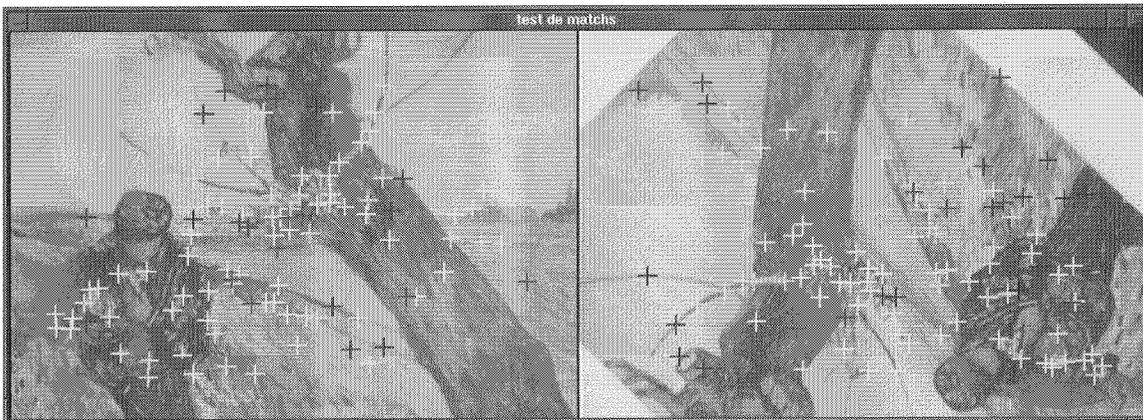


Figure 4. – 74.26 % d'appariements corrects dans le cas d'une rotation de 134.8 degrés.

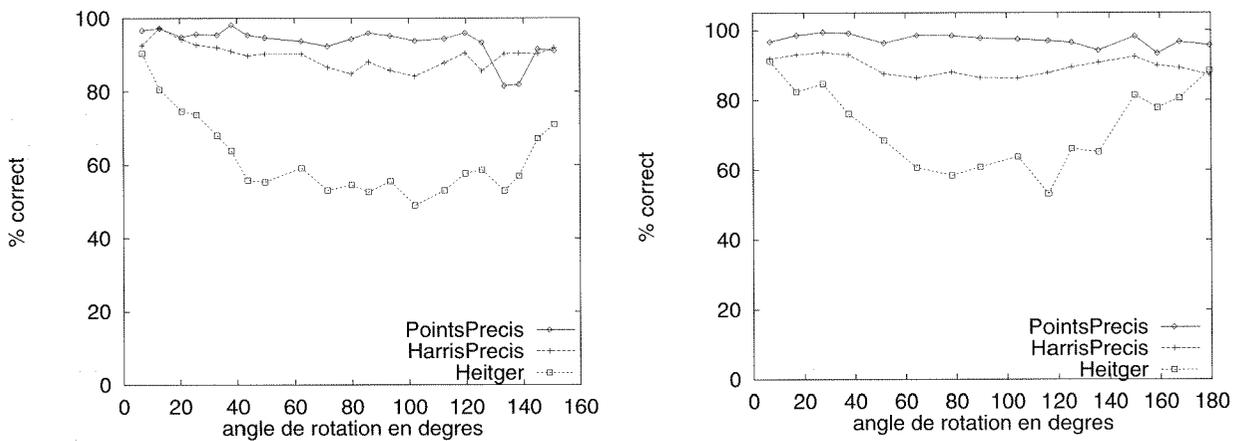


Figure 5. – Pourcentage d'appariements corrects en fonction de l'angle de rotation. À gauche pour la séquence Sanja, à droite pour la séquence Van Gogh. Les trois courbes correspondent aux différents détecteurs de point d'intérêt utilisés : Heitger, HarrisPrecis et les points précis. La taille σ de la gaussienne utilisée est de 3.

montre que les résultats obtenus en utilisant différents vecteurs, notamment \vec{V} , \vec{V}_T et \vec{V}_A . On peut observer que le vecteur \vec{V} est peu robuste à des changements de luminosité. Le vecteur \vec{V}_T est légèrement plus robuste à de tels changements. Il repose sur une modélisation par translation des niveaux de gris. On peut observer

qu'une telle modélisation n'est pas suffisante. Par contre si l'on modélise le changement de luminosité par une transformation affine, on obtient de très bons résultats, voir la courbe pour le vecteur \vec{V}_A . En effet, le pourcentage d'appariements corrects est supérieur à 80% pour un niveau de gris relatif allant de 0.5 à 1.7.

Mise en correspondance par invariants locaux

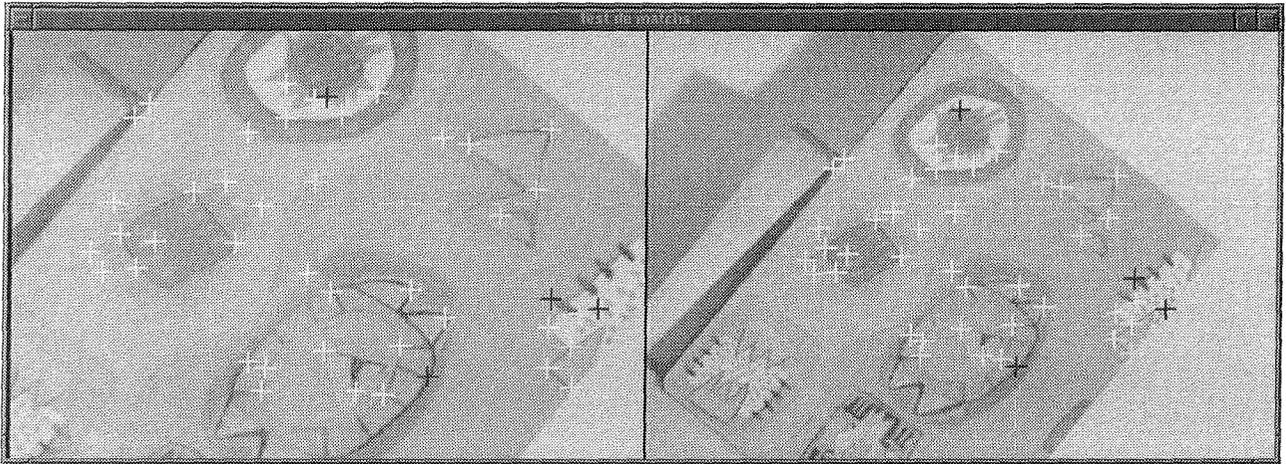


Figure 6. – 90.91 % d'appariements corrects dans le cas d'un facteur d'échelle de 0.68.

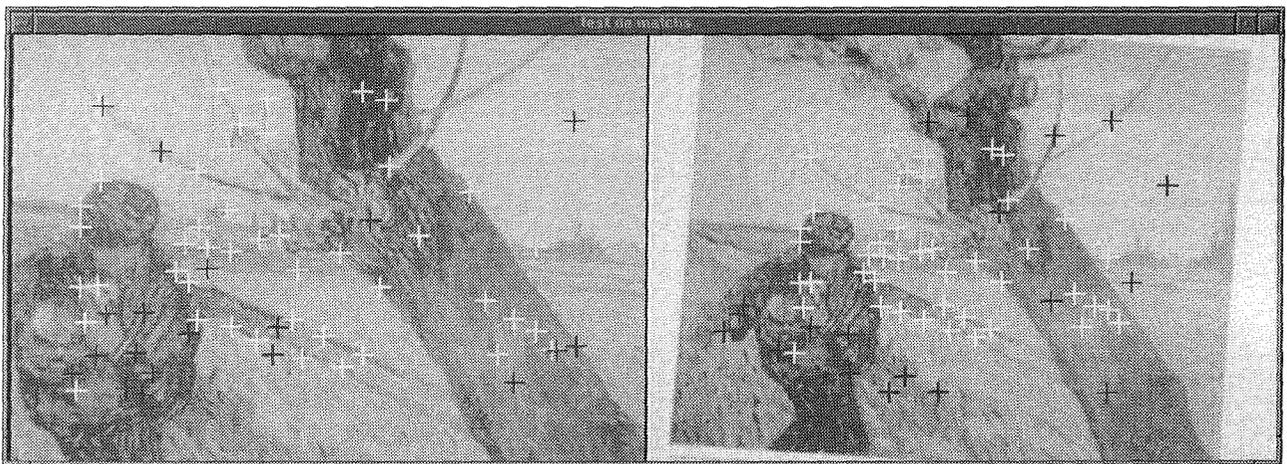


Figure 7. – 70.31 % d'appariements corrects dans le cas d'un facteur d'échelle de 0.68.

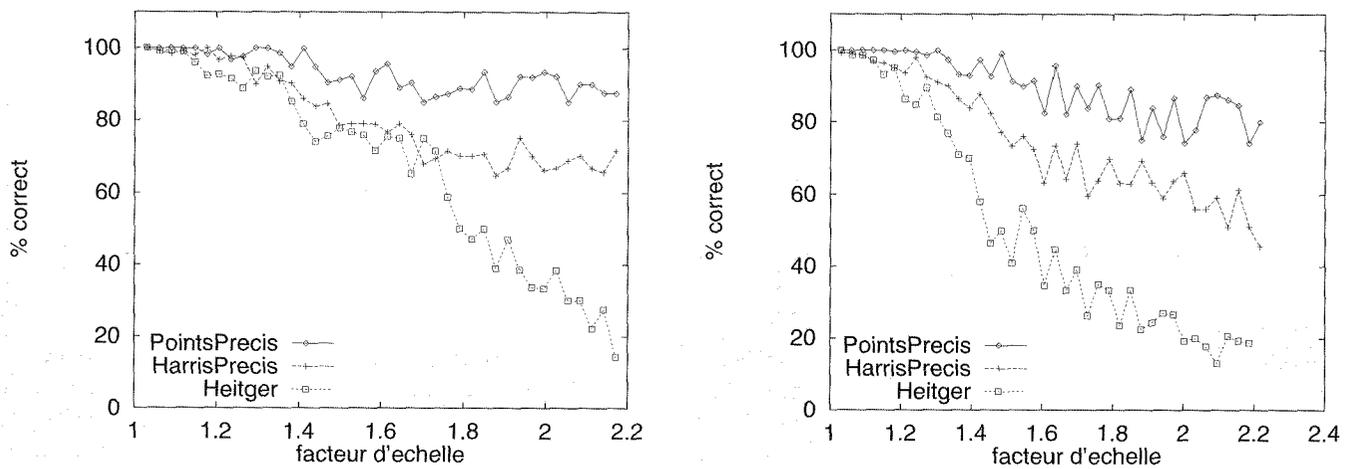


Figure 8. – Pourcentage d'appariements corrects en fonction du facteur d'échelle. À gauche pour la séquence Sanja, à droite pour la séquence Van Gogh. Les trois courbes correspondent aux différents détecteurs de point d'intérêt utilisés : Heitger, HarrisPrecis et les points précis. La taille moyenne σ de la gaussienne utilisée est de 5.

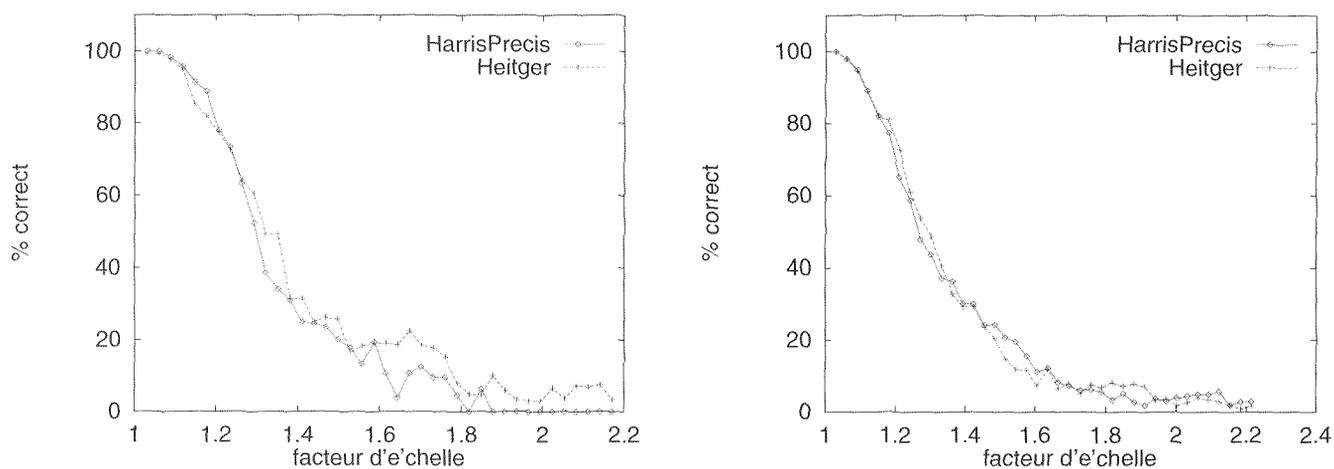


Figure 9. – Pourcentage d'appariements corrects en fonction du facteur d'échelle pour la séquence Van Gogh. Les deux courbes correspondent aux différents détecteurs de point d'intérêt utilisés : Heitger et HarrisPrecis. La taille σ de la gaussienne utilisée est de 5.



Figure 10. – À gauche l'image de la séquence "VanGogh_lum" de changement de luminosité avec un niveau de gris relatif de 0.6, à droite l'image avec un niveau de gris relatif de 1.7.

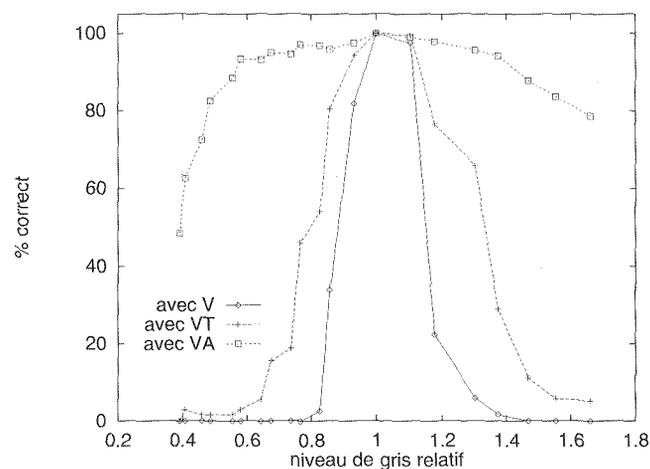
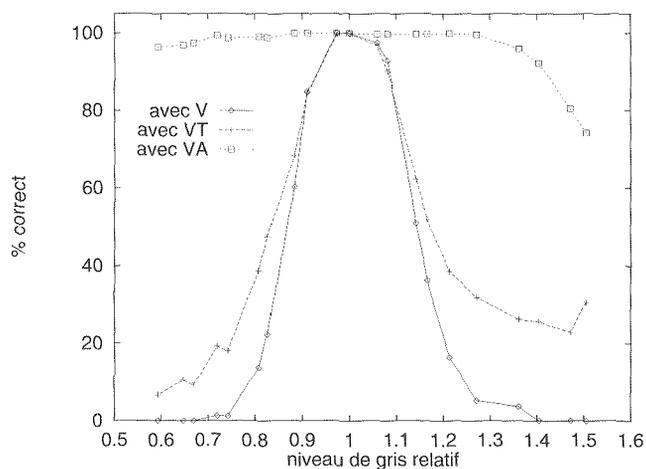


Figure 11. – Pourcentage d'appariements corrects en fonction du changement de luminosité en utilisant différents invariants. À gauche pour la séquence Sanja, à droite pour la séquence Van Gogh.

4.5. exemples plus complexes

Jusqu'à présent nous avons examiné la robustesse de la caractérisation proposée séparément à chaque transformation image et uniquement pour des scènes planes. Cette section présente les résultats obtenus pour la combinaison d'une rotation et d'un changement d'échelle ainsi que les résultats obtenus pour des scènes tridimensionnelles.

4.5.1. combinaison d'une rotation et d'un changement d'échelle

Ce paragraphe présente la robustesse de la caractérisation proposée à la combinaison simultanée d'une rotation et d'un changement d'échelle. La figure 12 montre les résultats obtenus avec la séquence « Sanja » dans le cas d'une rotation de 108.8 degrés et d'un changement d'échelle d'un facteur de 1.47. Comme précédemment, les croix blanches indiquent les appariements corrects et les noires les appariements faux. Le taux d'appariements corrects est de 82.35%. La figure 13 montre les résultats obtenus avec la séquence « Van Gogh » dans le cas d'une rotation de 99.8 degrés et d'un changement d'échelle d'un facteur 1.47. Le pourcentage d'appariements corrects est de 68.42%. Ces résultats prouvent que la méthode proposée permet de gérer les similitudes image.

4.5.2. scènes tridimensionnelles

Ce paragraphe présente les résultats obtenus pour des scènes tridimensionnelles. L'évaluation des résultats a été réalisée en utilisant la méthode décrite à la section 4.1. Le critère reste la satisfaction de la géométrie épipolaire, critère qui n'est pas parfait et qui explique que quelques erreurs peuvent subsister. La figure 14 montre les résultats obtenus sur une scène d'extérieur complexe. La transformation entre les deux images est relativement faible, cependant la présence de motifs répétitifs rend la mise en correspondance particulièrement difficile. Le pourcentage d'appariements corrects est de 84.34%, chiffre qui contient quelques erreurs cependant il n'en reste pas moins qu'il s'agit d'un bon résultat vu la difficulté de la scène. On peut d'ailleurs remarquer que certains des faux appariements sont dus aux motifs répétitifs. De fait les motifs répétitifs sont relativement bien positionnés car la fenêtre de calcul des invariants est relativement large, de l'ordre de 9 pixels de rayon, prenant ainsi en compte des aspects non répétitifs.

La figure 15 montre les résultats obtenus sur une autre scène. La transformation entre les deux images est constituée d'une rotation scène, d'une rotation image et d'un changement d'échelle. Le pourcentage d'appariements corrects est de 80%. Là encore, les faux appariements sont essentiellement dus aux motifs répétitifs contenus dans la scène, un coin clair ressemble en effet fort à un point coin clair d'angle similaire. La figure 16 montre le champ de déplacement calculé à partir des appariements corrects de la figure 15. Une méthode classique basée sur la corrélation échoue complètement sur une telle scène.

La figure 17 montre les résultats obtenus pour une scène d'intérieur. La transformation entre les deux images consiste en une rotation scène, une rotation image et un changement d'échelle. Le taux d'appariements corrects est de 84.9%. La figure 18 montre le champ de déplacement calculé à partir des appariements évalués comme corrects. On peut voir qu'un appariement a été faussement évalué. Ceci est dû à la méthode d'évaluation qui repose sur la contrainte épipolaire (cf. section 4.1).

4.6. influence des différentes composantes du vecteur

Cette section montre l'influence des différentes composantes du vecteur de caractéristiques. Pour ce faire, deux expériences ont été réalisées. La première expérience évalue le taux d'appariement lorsque l'on utilise une seule composante du vecteur de caractéristiques. La deuxième montre le taux d'appariement obtenu en utilisant un sous-ensemble de composantes. Ces expériences permettent de montrer la discriminance de chacune des composantes, notamment des invariants de troisième ordre.

La figure 19 montre le taux d'appariement en fonction de la composante utilisée pour les séquences Sanja et Van Gogh. Les valeurs de cette figure correspondent aux moyennes obtenues sur des images de séquences comportant un bruit d'image. Les taux d'appariement sont supérieurs pour la séquence Van Gogh ce qui s'explique par la texture de cette scène qui facilite le calcul d'appariement. En outre, pour la séquence Van Gogh les composantes correspondant aux premières et deuxièmes dérivées permettent un meilleur taux d'appariement. Ceci est dû à la présence de texture. Dans ce cas les dérivées de troisième ordre sont moins stables que les dérivées d'ordre inférieur. En ce qui concerne la séquence Sanja, toutes les composantes permettent d'obtenir approximativement le même taux d'appariement.

Le tableau 1 montre le taux d'appariement pour un sous-ensemble de composantes. La première colonne donne le sous-ensemble utilisé et les deux autres colonnes donnent les résultats pour les séquences Van Gogh et Sanja. On peut observer que certains sous-ensembles permettent d'obtenir de très bons résultats. D'autre part, les invariants de troisième ordre apportent le même taux d'appariement que l'ensemble des invariants de premier et de deuxième ordre.

Tableau 1. – Taux d'appariement en fonction du sous-ensemble de composantes utilisées.

invariants	Sanja	Van Gogh
0 ... 8	99.78	99.81
1 ... 8	99.55	99.67
0 ... 4	99.69	99.42
1 ... 4	97.78	98.91
2 ... 4	92.07	98.36
5 ... 8	97.30	99.21
0, 1	82.38	92.99

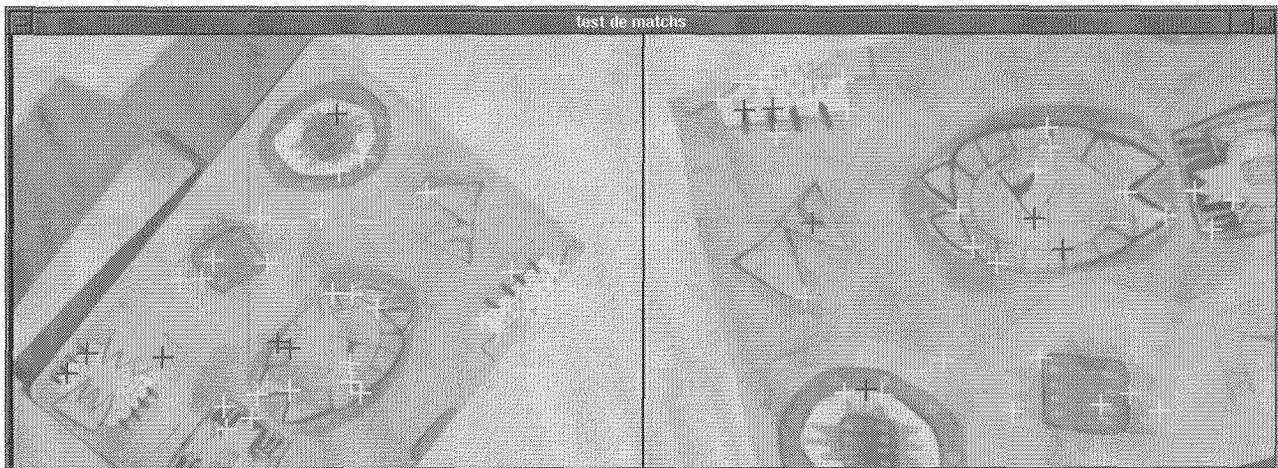


Figure 12. – 82.35 % d'appariements corrects dans le cas d'une rotation de 108.8 degrés et un facteur d'échelle de 1.47.

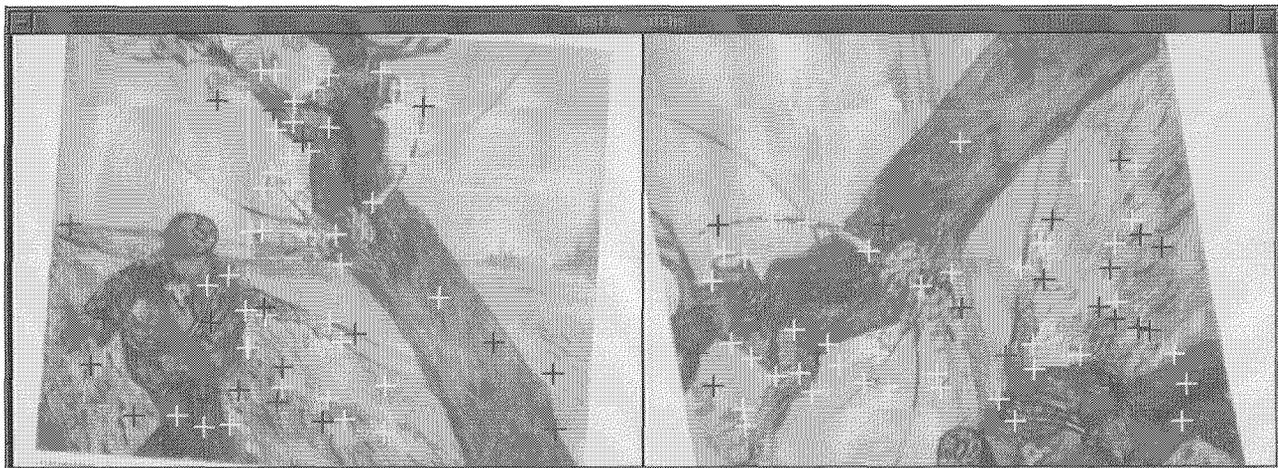


Figure 13. – 68.42 % d'appariements corrects dans le cas d'une rotation de 99.8 degrés et un facteur d'échelle de 1.47. À gauche pour la séquence Sanja, à droite pour la séquence Van Gogh.

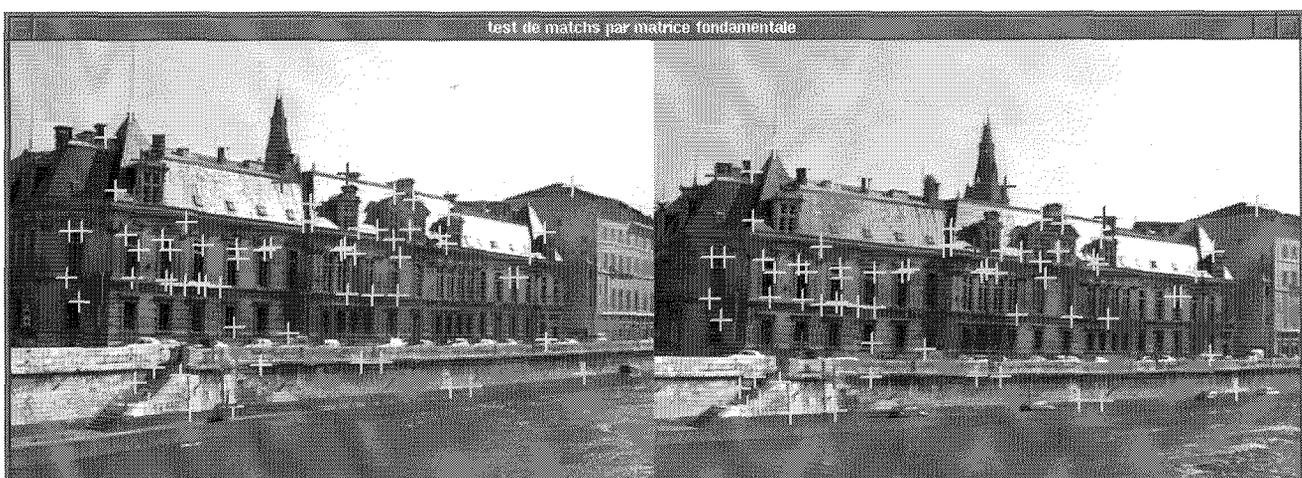


Figure 14. – 84.34 % d'appariement corrects dans le cas d'une scène d'extérieur.

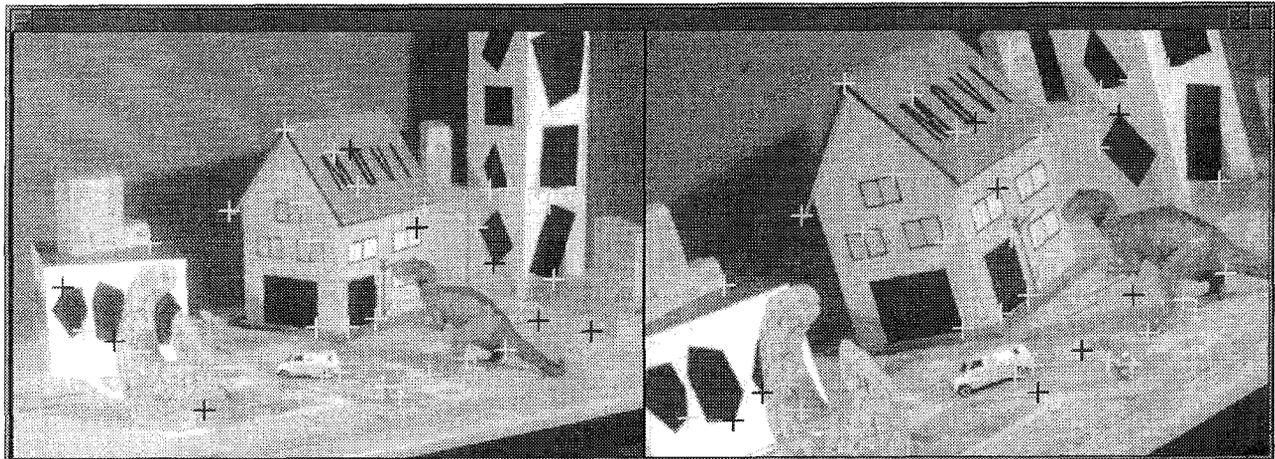


Figure 15. – 80.0 % d'appariements corrects dans le cas d'une scène 3D.

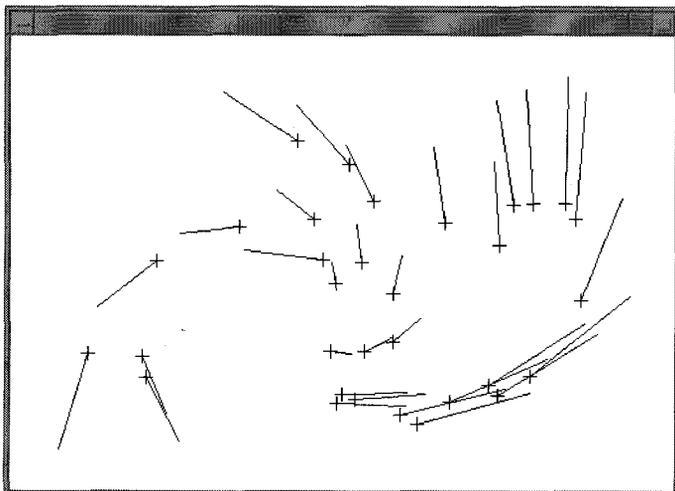


Figure 16. – Vecteurs de déplacement pour les appariements corrects de la figure 18.

Ces deux expériences justifient l'utilisation des invariants de troisième ordre. Ces invariants s'avèrent suffisamment stables et apportent une information non négligeable.

5. conclusion et discussion

Les invariants différentiels sont utilisables pour caractériser des points dans des images. Si les travaux antérieurs ont caractérisé

ces invariants, il restait à déterminer comment ils pouvaient être calculés de façon à être utilisables pratiquement. En outre, nous avons montré l'utilité pratique des invariants d'ordre troisième. En particulier, les formules fournissant des invariants pour différentes échelles ne sont pas utilisables car elles supposent que les calculs ne sont pas influencés par le facteur d'échelle; pour cette raison nous n'avons pas introduit ces derniers ici.

Sans hypothèse sur la déformation de l'image par une similitude, la méthode proposée permet de mettre en correspondance des points d'intérêt avec un taux de mise en correspondance qui est supérieur à 40%, et souvent plus de deux fois meilleur. Les mesures faites sont opérées sur des fenêtres gaussiennes avec un σ de 3. Il s'agit donc bien de mesures locales qui sont de ce fait insensibles à l'occultation d'une partie des formes.

Les expérimentations illustrent bien la qualité des appariements obtenus, et ce en opérant de façon rudimentaire. Nous avons également vu l'influence du détecteur de points d'intérêt utilisé.

Il reste à imaginer comment une telle approche peut s'opérer si l'on autorise des déformations affines. Là encore, le problème n'est pas d'ordre mathématique : les invariants affines sont relativement simples à calculer; le problème est plutôt d'ordre technique : comment réussir à calculer des dérivées sur des supports dont on autorise une variation affine.

Remerciement

Le travail doctoral de Cordelia Schmid a été réalisé dans le cadre d'une bourse européenne CHM. Nous remercions la société ISTAR d'avoir mis à notre disposition les images de la ville de Marseille, et Stéphane d'avoir bien voulu nous prêter son dinosaure favori.

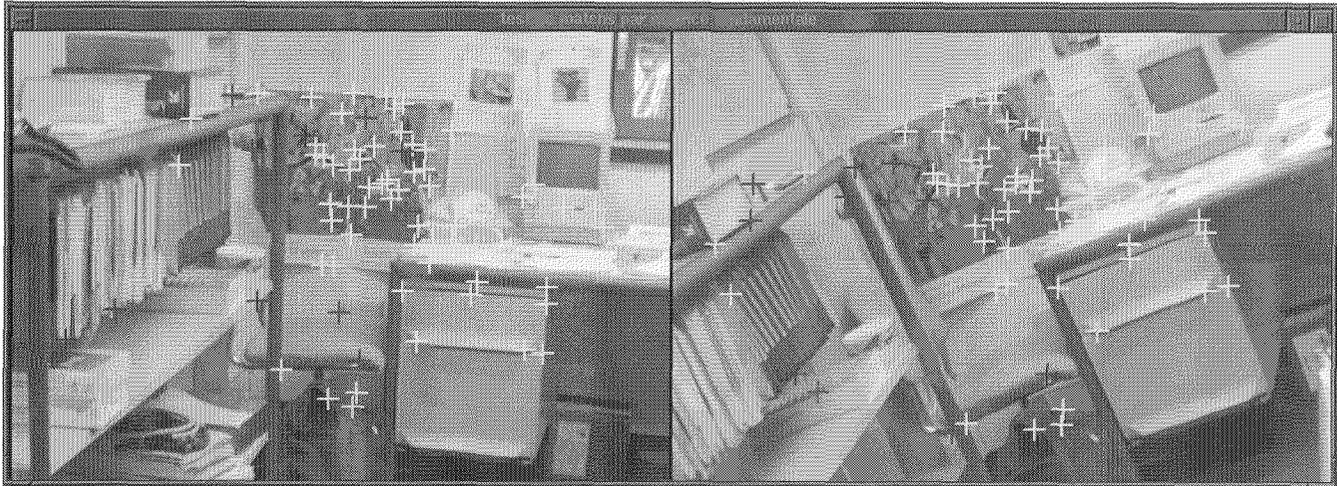


Figure 17. – 84.9% d'appariements corrects dans le cas d'une scène 3D.

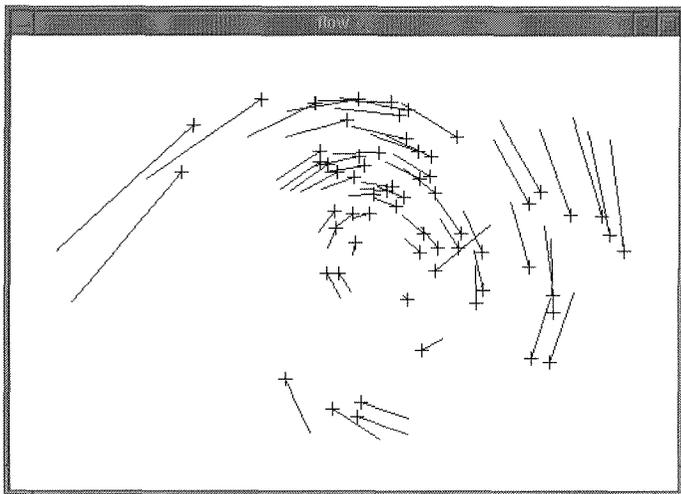


Figure 18. – Vecteurs de déplacement pour les appariements corrects de la figure 20.

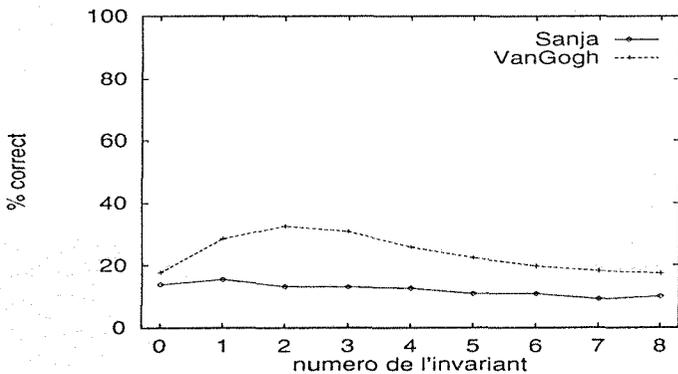


Figure 19. – Taux d'appariement en utilisant une composante du vecteur de caractéristiques.

BIBLIOGRAPHIE

- [Ack 84] F. Akermann. Digital image correlation : Performance and potential application photogrammetry. *Photogrammetric Record*, 64 (11) : 429-439, October 1984.
- [Der 84] R. Deriche, Z. Zhang, Q.T. Luong, and O. Faugeras. Robust recovery of the epipolar geometry for uncalibrated stereo rig. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer vision, Stockholm, Sweden*, volume 800 of *lecture Notes in Computer Science*, pages 567-576. Springer-Verlag, May 1994.
- [Fau 92] O. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat, and Z. Zhang. Quantitative and qualitative comparisons of some area and feature-based stereo algorithms. In W. Forstner, R. Haralick, and B. Radig, editors, *Robust Computer Vision*, page 1-26, February 1992. Comparisons of area and feature-based stereo algorithms.
- [Fle 91] D.J. Fleet, A.D. Jepson, and M.R.M Jenkin. Phase-base disparity measurement. *Computer Vision graphics and Image Processing*, 53(2) : 198-210, 1991.
- [Flo 93] L. Florack. *The Syntactical Structure of Scalar Images*. PhD thesis, universiteit Utrecht, m11 1993.
- [Flo 94] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4 : 171-187, 1994.
- [Fre 91] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 13(9) : 891-906, 1991.
- [Har 88] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147-151, 1988.
- [Hil 93] D. Hilbert, Ueber die vollen Invariantensystemen. *Math. Annalen*, 42 : 313-373, 1893.
- [Hor 89] R. Horaud and Th. Skordas. Stereo correspondence through feature grouping and maximal cliques. *IEEE Transaction on Pattern Analysis and machine Intelligence*, 11 (11) : 1168-1180, 1989.
- [Hor 90] R. Horaud, Th. Skordas, and F. Veillon. Finding geometric and relational structures in an image. In *Proceedings of the 1st European Conference on Computer Vision, Antibes, - France*, Lecture Notes in Computer Science, pages 374-384. Springer-Verlag, April 1990.
- [Hor 95] R. Horaud, R. Mohr, F. Dornaika, and B. Boufama. The advantage of mounting a camera onto a robot arm. In *Europe-China Work-shop on Geometrical Modelling and invariants for computer Vision, Xian, China*, pages 206-213, April 1995.

Mise en correspondance par invariants locaux

- [Hu 94] X. Hu and N. ahuja. Feature extraction and matching as signal detection. *International Journal of Pattern Recognition and Artificial Intelligence*, 8(6) : 1343-1379, 1994.
- [Koe 84] J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50 : 363-369, 1984.
- [Koe 87] J.J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55 : 367-375, 1987.
- [Lin 94] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
- [Lon 86] P. Long and G. Giraudon. Stereo matching based on contextual line-region primitives. in *Proceedings of the 8th International Conference on Pattern Recognition, Paris, France*, pages 974-977, 1986.
- [Lot 94] J.L. Lotti and G. Giraudon. Adaptive window algorithm for aerial image stereo. In *Proceedings of the 12th International Conference on Pattern Recognition Jerusalem, Israel*, pages 701-703, October 1994.
- [Rem 94] P.K. Remagnino, P. Brand, and R. Mohr. Correlation techniques in adaptive template matching with uncalibrated cameras. In *Vision Geometry III, SPIE's international symposium on photonic sensors & control commercial applications*, volume 2356, pages 252-253, October 1994;
- [Rom 94a] B.M. ter Haar Romeny. *Geometry-Driven Diffusion in computer Vision*. Kluwer Academic Publishers, 1994.
- [Rom 94b] B.M. ter Haar Romeny, L.M.J. Florack, A.H. Salden, and M.A. Viergever. Higher order differential structure of images. *Image and Vision Computing*, 12(6) : 317-325, 1994.
- [Rot 92] C.A. Rothwell, A. Zisserman, D.A. Forsyth, and J.L. Mund. Canonical frames for planar object recognition. in G. Sandini, editor *Proceeding of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 757-772. Springer-verlag, May 1992.
- [Sal 92] A.H. Salden, B.M. ter Haar Romeny, L.M.J. Florack, M.A. Viergever, and J.J. Koenderink. A complete and irreducible set of local orthogonally invariant feature of 2-dimensional images. In *Proceedings of the 11th International Conference on Pattern Recognition, The Hague, Netherland*, pages 180-184, 1992.
- [San 88] T. Sanger. Stereo disparity computation using gabor filters. *Biological Cybernetics*, 25(59) : 405-418, 1988.
- [Sch 90] B.G. Schunck. Robust computational vision. In *Proceeding of the International Workshop on Robust Computer Vision, Seattle, Washington, USA*, pages 1-19, October 1990.
- [Sch 96] C. Schmid. *Appariement d'images par invariants locaux de niveaux gris*. Thèse de doctorat, GRAVIR – IMAG – INRIA Rhône-Alpes, July 1996.
- [Sem 52] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford Science Publication, 1952.
- [Tor 86] V. Torre and T.A. Poggio. On edge detection. *IEEE Transactions on Pattern Analysis and machine Intelligence*, 8 (2) : 147-163, 1986.
- [vD 94] M. van Diest, I. van Gool, T. Moons, and E. Pauwels. Projective invariants for planar contour recognition. in J.O. Eklund, editor, *Proceeding of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 527-534. Springer-Verlag, 1994.
- [Wei 91] I. Weiss. Noise-resistance invariant of curves. In *proceeding of the DARPA-ESPRIT workshop on applications of invariants in Computer Vision, Reykjavik, Iceland*, pages 319-344, 1991.
- [We92] C.J. Westelius. *Preattentive gaze control for robot vision*. Ph D thesis, Linköping University, Sweden, June 1992.
- [Wit 83] A.P. Witkin. Scale-space filtering. In *International Joint conference on Artificial Intelligence*, pages 1019-1023, 1983.
- [Wu 94] X. Wu and B. Bhanu. Target recognition using multi-scale gabor filters. in *ARPA Image Understanding Workshop Monterey, CA*, Novembre 1994.
- [Wu 95] X. Wu and B. Bhanu. Gabor wavelets for 3D object recognition. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 537-542, 1995.
- [Zab 94] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondance. IN *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 151-158. Springer-Verlag, May 1994.
- [Zha 89] J. Zhao. *extraction d'information tri-dimensionnelle par sérovision*. Thèse de doctorat, Université Paul Sabatier, Toulouse, Juillet 1989.
- [Zha 94] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Rapport de recherche 2273, INRIA May 1994.

Manuscrit reçu le 21 mai 1996.

LES AUTEURS

Cordélia SCHMID



Diplômée en informatique de l'université de Karlsruhe, elle a obtenu un doctorat de l'INPG en 1996, puis elle a effectué un post-doctorat à l'université d'Oxford. Depuis mai 1997 elle est Chargée de Recherche à l'INRIA Rhones-Alpes. Ses travaux de recherche concernent la recherche d'information dans de grandes bases d'images, la caractérisation d'images, l'appariement entre images et la reconstruction d'images aériennes.

Roger MOHR



Professeur à l'INPG depuis 1988, il dirige le projet de modélisation et vision par ordinateur associé au laboratoire GRAVIR et à l'INRIA. Ancien élève de l'ENS de Cachan et docteur d'Etat de l'université de Nancy, ses intérêts portent sur la géométrie de la vision tridimensionnelle et sur l'indexation d'images. Voir <http://www.inrialpes.fr/movi/people/Mohr/index.html> pour plus de détails.