Appariement d'images à des échelles différentes

Matching Images at Different Scales

Y. Dufournaud[1] C. Schmid[2] R. Horaud[2]

[1] Aerospatiale,
2, rue Béranger
92323 Châtillon

[2] INRIA Rhône-Alpes
655 av. de l'Europe
38330 Montbonnot, France

Yves.Dufournaud@inrialpes.fr

Résumé

Cet article présente une méthode nouvelle pour apparier des images qui diffèrent d'un facteur d'échelle important, et qui est invariante aux rotations et robuste aux déformations perspectives. L'appariement est basé sur l'extraction de points d'intérêt sur les images et la comparaison de ces points caractérisés par des invariants différentiels en niveaux de gris. Nous montrons comment adapter simultanément cette détection et cette caractérisation aux changements d'échelle. Des résultats expérimentaux étendus valident la faisabilité d'une telle approche et un algorithme d'appariement multi-échelles est proposé. Celui-ci donne d'excellents résultats pour des changements d'échelles allant de 1 à 6.

Mots Clef

mise en correspondance image–image, extraction de primitives.

Abstract

This paper presents a new method for image matching in the presence of large scale changes, which is invariant to image rotations and robust to perspective déformations. Matching is based on extracting interest points in images and on comparing these extracted parts based on local greyvalue invariants. It is derived how to adapt both interest point extraction and local invariant to scale changes. Extensive experimental results validate the feasibility of this approach and a multi-scale matching algorithm is proposed. Results for matching image pairs are extremely good for scale changes varying between 1 and 6.

Keywords

image–image matching, feature extraction.

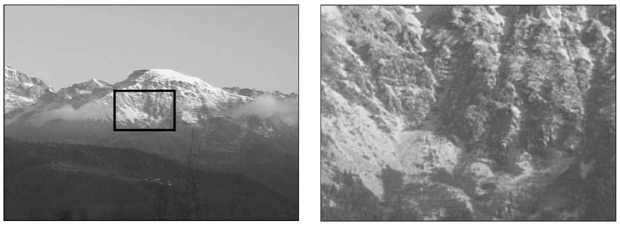

FIG. 1: *Exemple d'une paire d'images comportant un grand changement d'échelle dû à l'utilisation d'un zoom. Le facteur d'échelle entre les images est de 6. La partie commune est encadrée par un rectangle dans la vue de gauche et représente moins de 17% de l'image.*

1 Introduction

L'appariement automatique de deux images reste un sujet de recherche actif en vision artificielle. La grande majorité des méthodes existantes se base sur l'extraction de primitives qu'elles tentent ensuite de mettre en correspondance sur la base de contraintes géométriques (par exemple à l'aide d'un modèle approximatif du mouvement entre les images) et de mesures de similarités photométriques entre les caractéristiques. Deux situations ont particulièrement été étudiées : (i) le cas d'une paire stéréo dans laquelle chaque image correspond à la même scène observée d'un point de vue légèrement différent, et (ii) le cas d'une caméra mobile où chaque caractéristique est suivie dans deux images consécutives. Dans ce dernier cas encore, le changement entre chaque image est léger.
Ici nous abordons un problème différent. Nous cherchons à mettre en correspondance deux prises de vue d'une même scène entre lesquelles soit la caméra s'est rapprochée, soit la focale a changé. A cause de ce mouvement les objets composant la scène apparaissent à des tailles différentes dans chacune des deux images. Un exemple d'une telle situation est donné dans la Figures 1. L'image de gauche correspond à la vue initiale,

l'image de droite est prise en augmentant la focale de l'objectif. Ainsi seule une petite région de l'image initiale reste visible. Dans cet exemple où le changement d'échelle est d'un facteur 6, la zone commune représente moins de 17% de la vue initiale.

L'appariement de ces images pose deux difficultés majeures. La première est la prise en compte du changement de résolution entre les images (c'est à dire le rapport $\frac{f}{Z}$ où f correspond à la focale de la caméra et Z représente la distance moyenne à la scène ; cf. Figure 2) et son impact sur les données photométriques, en particulier sur l'extraction de primitives. Deuxièmement, dans une telle situation, la partie commune des images est si réduite que les possibilités de faux appariements sont augmentés, particulièrement lorsque l'on utilise une description locale du signal pour réaliser cet appariement.

Récemment plusieurs auteurs ont abordé le problème de la mise en correspondance d'images prisent de points de vue très différents [2, 8, 12]. Cependant aucun d'eux n'a considéré le cas particulier d'un rapprochement important de l'observateur conduisant à un changement d'échelle important. Les approches hiérarchiques [3, 6, 9], bien qu'utilisant une représentation de la même image à plusieurs échelles, sont incapables de mettre en correspondance des images d'une scène prises à des résolutions différentes. Leur seul but est de stabiliser ou d'accélérer la mise en correspondance d'images prises de points de vue identiques. De ce fait le problème que nous abordons – la mise en correspondance d'images où une scène est observée à des résolutions différentes – est différent de celui résolu par les approches de mise en correspondance par stéréo hiérarchique.

Notre approche est une extension des travaux menés dans [10]. Elle repose sur l'utilisation d'un espace d'échelles pour modéliser le changement de résolution entre les images, et d'un algorithme de mise en correspondance permettant de tolérer de nombreux faux appariements. Cet algorithme utilise une description locale, pour mettre en correspondance des points d'intérêt d'une image à l'autre. Par rapport aux travaux précédents nous apportons trois contributions. D'abord nous montrons comment adapter aux changements de résolution la détection des points d'intérêt. Jusqu'à présent ces points ont été considérés comme stables aux changements d'échelle, mais des expérimentations ont montré que ce n'est plus le cas au delà d'un facteur 2. Cette adaptation est fondamentale car c'est sur ces points que les caractérisations sont calculées et elles n'ont de sens que si on est capable de les calculer sur la projection d'un même point d'une image à l'autre. Nous avons validé expérimentalement cette adaptation et nous montrons son impact sur la répétabilité de détection à différentes échelles. Enfin nous proposons un algorithme capable de mettre en correspondance des

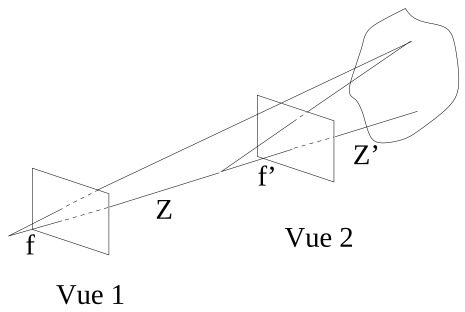

FIG. 2: *Relation entre la résolution* $r = \frac{f}{Z}$ *d'une image et la notion de facteur d'échelle entre deux image* $s = r_1/r_2$ *dans le cas d'une scène 3D où soit la caméra se rapproche de la scène, soit la focale change.*

images dans le cas d'un grand changement d'échelle avec rotation, translation et légère déformation perspective. En particulier pour cela nous utilisons le calcul robuste d'une contrainte globale pour rechercher le changement de points de vue.

1.1 Organisation

La section 2 introduit brièvement le cadre multi-échelles. Il est appliqué dans la section 3 pour la détection des points d'intérêt et le calcul de la caractérisation de ces points. La section 4 présente l'algorithme d'appariement et montre des résultats de mise en correspondance pour des images prisent à différentes échelles. Nous présentons aussi des résultats combinant rotation et changement d'échelle ainsi que de légères déformations perspectives. La section 5 présente des résultats expérimentaux pour la détection des points d'intérêt et leur caractérisation qui valident l'approche multi-échelles sur des images réelles jusqu'à un facteur 6.

2 Cadre Multi-échelles

Dans cette section, nous rappelons la définitions des espaces d'échelles. Nous nous plaçons dans le cas particulier de la Figure 2 où soit la caméra s'est rapprochée de la scène, soit la focale de la caméra a changé. Dans ces conditions, on peut définir le changement d'échelle entre deux images comme le rapport de leur résolution $r = \frac{f}{Z}$ et notamment parler d'image haute résolution (HR) et d'image basse résolution (BR) [1]. Dans ce qui suit, nous supposons de plus que la zone de la scène observable dans l'image HR est plane et parallèle au plan image.

2.1 Définitions et Notations

L'approche multi-échelles est maintenant bien connue [7, 13]. Nous l'expliquons brièvement dans ce qui suit.

1. Sur la Figure 1, l'image de gauche est l'image BR et l'image de droite est l'image HR.

Soient I^1 et I^2 deux fonctions images liées par un facteur d'échelle s :

$$I^1(\mathbf{x}) = I^2(\mathbf{x}'), \text{ avec } \mathbf{x} = s\mathbf{x}' + \mathbf{c}$$

où $\mathbf{x}$ un vecteur de dimension 2 (un point image)

Si on note I_i les dérivées de I par rapport à i ($i \in \{x, y\}$), les dérivées de I^1 et de I^2 sont reliées par :

$$I^1_{i_1 \ldots i_n}(\mathbf{x}) = s^n I^2_{i_1 \ldots i_n}(\mathbf{x}')$$

Dans le contexte d'une représentation discrète, comme c'est le cas pour les images, les dérivées sont calculées par convolution avec les dérivées gaussiennes :

$$I^1(\mathbf{x}) * G_{i_1 \ldots i_n}(\sigma) = s^n I^2(\mathbf{x}') * G_{i_1 \ldots i_n}(s\sigma) \qquad (1)$$

où G représente la Gaussienne et $G_{i_1 \ldots i_n}$ ses dérivées suivant $i_1 \ldots i_n$

La convolution avec les dérivées gaussiennes est dans la suite abrégée en $L_{i_1 \ldots i_n}(\sigma)$. L'équation 1 s'écrit alors :

$$L^1_{i_1 \ldots i_n}(\mathbf{x}, \sigma) = s^n L^2_{i_1 \ldots i_n}(\mathbf{x}', s\sigma) \qquad (2)$$

À partir de cette équation, on peut adapter le calcul des dérivées à condition de connaître le facteur d'échelle s. Comme en général ce facteur est inconnu, on aura recours à une approche multi-échelles dans laquelle les dérivées seront calculées à plusieurs échelles s_i parmi lesquelles s sera sélectionnée (cf. section 4).

3 Caractéristiques et Descripteurs Multi-échelles

Notre algorithme d'appariement est basé sur des points d'intérêt caractérisés par des invariants en niveaux de gris. La détection des points d'intérêt et le calcul de leur descripteur doivent tous les deux être intégrés dans un cadre multi-échelles pour pouvoir absorber les changements d'échelle. Une telle intégration a été proposée précédemment dans le cas des descripteurs seuls (voir par exemple [10]). Cependant, les points d'intérêt ont été supposés robustes aux changements d'échelle. En fait ceci n'est le cas que jusqu'à un facteur 2.

3.1 Points d'intérêt Multi-échelles

Dans le contexte de l'appariement, la détection de caractéristiques doit être répétable, c'est-à-dire stable sous diverses transformations. Des travaux précédents sur les points d'intérêt [11] on montré que le détecteur de Harris [4] était le plus répétable. Ils ont aussi montré que cette répétabilité se dégrade rapidement en fonction du changement d'échelle. Dans cette section nous présentons ce détecteur et nous montrons ensuite comment l'adapter aux changements d'échelle.
L'idée de base de ce détecteur est d'utiliser la fonction d'auto-corrélation pour déterminer les positions où le signal change dans deux directions simultanément. En prenant en compte les dérivées premières du signal sur une fenêtre, une matrice liée à cette fonction d'auto-corrélation est calculée :

$$G(\tilde{\sigma}) \otimes \begin{bmatrix} L_x^2(\sigma) & L_x L_y(\sigma) \\ L_x L_y(\sigma) & L_y^2(\sigma) \end{bmatrix} \qquad (3)$$

Les vecteurs propres de cette matrice sont les courbures principales de la fonction d'auto-corrélation. Deux valeurs significatives indiquent la présence d'un point d'intérêt. Notez que les facteurs de lissage utilisés pour le calcul des dérivées et pour le fenêtrage ne sont pas forcément égaux ($\tilde{\sigma}$ et σ peuvent être différents).
En présence d'un changement d'échelle s entre les deux images I^1 et I^2, le détecteur de points d'intérêt doit être adapté pour obtenir des résultats répétables. Les deux facteurs de lissage $\tilde{\sigma}$ et σ doivent être multipliés par s (cf. équation (2)). La matrice résultante est alors multipliée par s^2 pour avoir des valeurs propres comparables. Si on utilise l'équation (3) pour détecter les points dans I^1, alors la matrice utilisée pour la détection dans I^2 est :

$$s^2 G(s\tilde{\sigma}) \otimes \begin{bmatrix} L_x^2(s\sigma) & L_x L_y(s\sigma) \\ L_x L_y(s\sigma) & L_y^2(s\sigma) \end{bmatrix}$$

Comme dans la section 2, une approche multi-échelles doit être utilisée si le facteur d'échelle est inconnu, et les points d'intérêt doivent être calculés à plusieurs échelles s_i.

3.2 Invariants multi-échelles

L'intégration des invariants en niveaux de gris dans un cadre multi-échelles a déjà été étudiée précédemment [10]. Dans ce qui suit, nous décrivons ces invariants et montrons comment les intégrer.
Les invariants en niveaux de gris sont des combinaisons de dérivées sur le signal image qui sont invariantes aux rotations [5]. Ces dérivées sont calculées par convolution avec les dérivées gaussiennes. L'ensemble des invariants utilisés dans ce travail est donné dans l'équation (4) en notation tensorielle – la convention de sommation d'Einstein. Cet ensemble est limité à l'ordre 3. La première composante représente la moyenne de la luminance, la seconde le carré de la norme du gradient et la quatrième le laplacien.

$$\begin{bmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ji} \\ \varepsilon_{ij}(L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{iij} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ -\varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix} \qquad (4)$$

où L_i est défini par l'équation (2) et ε_{ij} est le tenseur epsilon 2D antisymétrique défini par $\varepsilon_{12} = -\varepsilon_{21} = 1$ et $\varepsilon_{11} = \varepsilon_{22} = 0$.

Quand on intègre ces invariants dans un cadre multi-échelles, les $L_{i_1 \ldots i_n}$ sont calculés à différentes échelles (cf. équation (2)). Pour un facteur d'échelle s entre I^1 et I^2, on obtient par exemple $L_i^1(\sigma)L_i^1(\sigma) = s^2 L_i^2(s\sigma)L_i^2(s\sigma)$ (cf. annexe A pour une adaptation numériquement stable du calcul de ces invariants).

Une autre approche est d'utiliser l'équation 2 en remarquant que l'on peut obtenir des invariants aux changements d'échelle à partir du quotient de deux dérivées :

$$\frac{[L_{i_1 \ldots i_n}(\sigma)]^{\frac{k}{n}}}{L_{i_1 \ldots i_k}(\sigma)} = \frac{[L_{i_1 \ldots i_n}(s\sigma)]^{\frac{k}{n}}}{L_{i_1 \ldots i_k}(s\sigma)}$$

Cependant des expérimentations ont montré que ces invariants n'apportent pas de stabilité supplémentaire par rapport à ceux invariants aux rotations : en particulier le support de calcul σ de ces dérivées doit, lui aussi, être adapté aux changements d'échelle et l'utilisation d'une méthode multi-échelles apparaît de toute façon nécessaire.

4 Algorithme Multi-échelles

Cette section décrit notre algorithme d'appariement d'images à des échelles différentes. La difficulté ici est de trouver l'échelle correcte pour la détection et la caractérisation. Nous commençons d'abord par décrire l'approche sur laquelle se base l'algorithme, puis l'algorithme lui même et ensuite présentons des résultats d'appariement pour quelques paires d'images comprenant un changement d'échelle important.

Comme nous le montrons dans la section 3, il est possible d'adapter la détection des points d'intérêt et le calcul de leur descripteur aux changements d'échelles. Mais cette adaptation requiert la connaissance du facteur d'échelle existant entre les images pour fonctionner. Pour pouvoir se passer de cette connaissance, nous avons recours à un algorithme multi-échelles dont le but est de trouver simultanément le facteur d'échelle $\hat{s}$ (et ainsi permettre la détection et la caractérisation correctes des points) et d'autre part d'apparier correctement ces points.
Pour cela nous supposons que nous savons au moins quelle est l'image HR (si ce n'est pas le cas, il suffit d'appliquer l'algorithme deux fois en permutant les images, puis de comparer les solutions). Nous construisons un espace d'échelles sur l'image HR (cf. Figure 3) et nous cherchons à mettre en correspondance les caractéristiques détectées sur l'image BR. avec celles détectées pour une sélection d'échelle $S = \{1, 1.5, 2, \ldots, 7\}$ Une condition nécessaire pour que cet appariement soit correct est de trouver $\hat{s} \in S$ proche de $S_{\text{réel}}$ le facteur d'échelle existant entre les deux images. Car alors l'adaptation présentée section 3 est correcte et l'algorithme d'appariement présenté plus loin peut fonctionner. Pour trouver $\hat{s}$ parmi les $s_i \in S$, nous définissons un critère $C(s_i)$ qui est maximum lorsque l'appariement des points entre l'image BR et l'échelle s_i est correct.

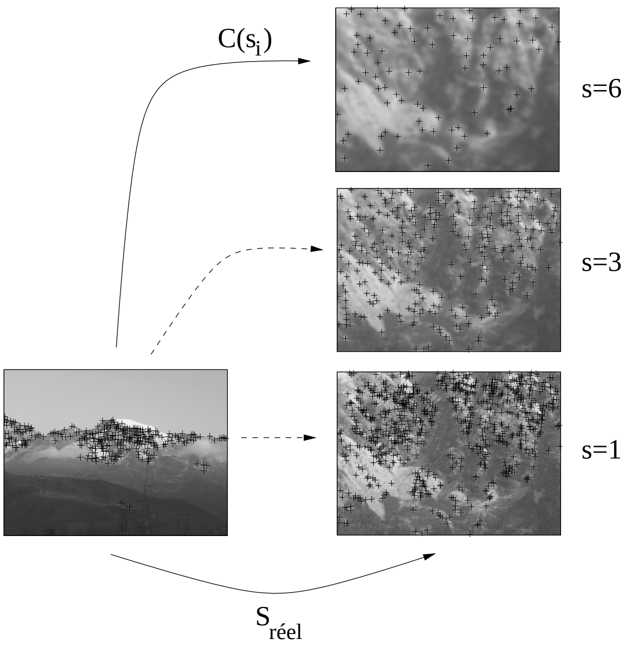

FIG. 3: *Les caractéristiques détectées sur l'image* BR *(à gauche) sont mises en correspondances avec celles détectés dans l'espace d'échelles construit sur l'image* HR *(à droite).*

$C(s_i)$ peut être calculé directement comme le nombre de points appariés correctement par l'algorithme de mise en correspondance. Pour chaque appariement réalisé entre l'image BR et l'image HR à l'échelle s_i, ce nombre est déterminé automatiquement par une estimation robuste de la transformation entre les deux images : les points satisfaisants cette transformation sont les points appariés correctement.

4.1 Evaluation Robuste

L'évaluation robuste de la transformation se fait à partir des correspondances trouvées par l'algorithme d'appariement entre les caractéristiques extraites de l'image BR notée $L^1(\mathbf{x}, \sigma)$ dans la suite et celles extraitent de l'image HR à l'échelle s_i (notée $L^2(\mathbf{x}', s_i\sigma)$). En pratique nous utilisons RANSAC[1] pour estimer une transformation affine entre les images. Ce modèle affine est bien adapté aux modifications que nous prenons en compte (cf. Figure 2), qui sont un rapprochement de la caméra où une modification de la focale.
Cette vérification par application d'une contrainte globale est d'autant plus importante que l'algorithme d'appariement se base sur des informations locales pour la mise en correspondance. En particulier elle permet d'accepter où de rejeter les appariements à partir d'une contrainte physique réelle : le mouvement de la caméra.

4.2 Algorithme d'Appariement

Cet algorithme a pour but de mettre en correspondance les points de $L^1(\mathbf{x}, \sigma)$ avec ceux de $L^2(\mathbf{x}', s_i\sigma)$. L'association de deux points $p^1 \in L^1(\mathbf{x}, \sigma)$ et $p^2 \in L^2(\mathbf{x}', s_i\sigma)$ est faite par vérification croisée : on recherche $p^2 \in L^2(\mathbf{x}', s_i\sigma)$ le points le plus ressemblant à p^1 au sens d'un critère de ressemblance R, puis on cherche $\hat{p}^1 \in L^1(\mathbf{x}, \sigma)$ par rapport à p^2 selon le même critère. Si $p^1 = \hat{p}^1$ alors les points p^1 et p^2 sont appariés.

4.3 Critère d'Appariement

La définition du critère d'appariement est importante car elle conditionne la qualité des résultats. Idéalement, si les invariants utilisés pour décrire les points étaient complètement discriminant, seule la distance de Mahalanobis entre les vecteurs d'invariants suffirait et permettrait l'appariement des points. Mais ce n'est pas le cas et une image peut comporter de nombreux points ayant une caractérisation similaire. Ces ambiguïtés sont la source de nombreux faux appariements.

La prise en compte du voisinage local des points est nécessaire pour lever ces ambiguïtés. Et cette prise en compte doit être faite d'autant plus tôt que les étapes suivantes de l'algorithme ne créent pas d'appariement correct, mais sont chargées au contraire d'éliminer les faux appariements. Pour cette raison nous intégrons la prise en compte de ce voisinage dans le critère de comparaison des points.

Ce critère est donc défini en prenant en compte non seulement la distance de Mahalanobis entre les points composant les couples (c'est le critère principal), mais aussi la cohérence du voisinage, mesurée en vérifiant qu'une mise en correspondance des voisins à l'aide de la distance de Mahalanobis conserve les angles entres les voisins et leur rapports de distances.

4.4 Résultats expérimentaux

Dans cette section nous présentons les résultats de l'algorithme d'appariement pour un ensemble de paires d'images et notamment les séquences Laptop, Van Gogh et Astérix (cf. figure 5) qui sont calibrées et sur lesquelles le changement d'échelle est connu avec précision. Cette connaissance, qui n'est pas nécessaire à l'algorithme, nous permet une vérification de la justesse des résultats obtenus. Toutes les paires d'images présentées dans cet article sont des images réelles. En particulier les rotations et les changements d'échelle sont dues au déplacement de l'objectif et au changement de la focale.

La Figure 4 montre le résultats de l'application de l'algorithme sur la paire d'images présentée en exemple. Comme on le voit les résultats sont excellents. La Figure 6 montre les résultats de l'appariement dans le cas ou il existe en plus une rotation image. La figure 7

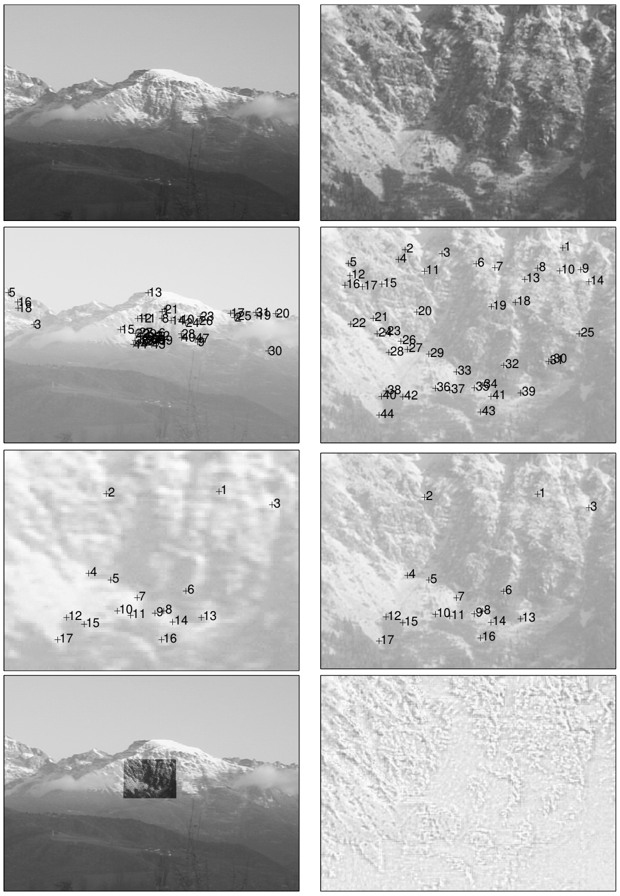

FIG. 4: *Exemple d'appariement obtenu par l'algorithme multi-échelles. Le facteur d'échelle est de 5.8. De haut en bas apparaissent : les images originales, les points appariés sans l'utilisation de la contrainte globale (44), les points correctement appariés après filtrage par la contrainte globale (17 soit 61% d'outliers), sur cette ligne la vue de gauche est agrandie pour permettre la visualisation des appariements. La dernière ligne montre : (à gauche) l'image* HR *projetée dans l'image* BR *par la transformation affine calculée à partir des appariements filtrés (le contraste entre les deux images à été augmenté) ; (à droite) la projection inverse. Dans ce dernier cas c'est la différence entre les deux images qui est affichées (blanc =* 0*). Malgré le fort taux d'outliers, la transformation est estimée correctement et peut être validée grâce à la différence des deux images.*

est un exemple de scène 3D avec rotation et légère déformation perspective due au décalage du point de vue. Le facteur d'échelle est d'environ 3.3. Le fort taux d'outliers (74%) s'explique par le caractère 3D de la scène. En effet dans ce cas le changement d'échelle entre les deux images n'est pas le même pour tout les points ce qui perturbe l'algorithme. D'autre part il est évident que dans le cas d'une scène ayant un relief prononcé, une homographie n'est pas une contrainte globale adaptée ce qui conduit à rejeter certains bons appariements. Néammoins l'algorithme obtient une solutions correcte.

Ces appariements prennent moins de 1 minute sur une Ultra Sparc10 pour un découpage de l'espace d'échelle de l'image HR en 20 niveaux avec une moyenne de 100 caractéristiques par niveau.

Les résultats concernant les séquences calibrées sont résumés dans le tableau 1. La dernière image de chacune des séquences est appariée successivement avec les précédentes. La colonnes de gauche indique le facteur d'échelle réel existant entre les images. On trouve ensuite le nombre de points total appariés lors de l'étape multi-échelles, puis le nombre de ceux qui sont reconnus comme corrects après application de la contrainte globale estimée par la méthode robuste. La dernière colonne montre le pourcentage que ces points représentent par rapport à la deuxième colonne. Il faut noter qu'à partir du moment où la méthode robuste estime correctement la transformation globale, les inliers sont nécessairement appariés correctement. De plus cette estimation est toujours correcte si le taux d'outliers est inférieur à 50%.

Sur la séquence Laptop, les résultats sont excellents sur toute la séquence (soit un facteur d'échelle de 6 pour les paires d'images extrêmes). Comparativement, les séquences Astérix et Van Gogh donnent de moins bons résultats, pour des facteurs d'échelles allant jusqu'à 4. Ceci s'explique par le fait que dans la séquence Laptop le fond de la scène est très différent de la partie commune aux images (et donc la possibilité de fausses mise en correspondance avec cette partie est quasi nulle). D'autre part, les séquences Astérix et Van Gogh ont été prises à l'aide d'un zoom présentant un défaut de tirage optique. A cause de cela les images de fin de séquences sont légèrement flou, ce qui nuit à la détection des points d'intérêts. Pour ces deux séquences, les résultats sont excellents jusqu'à un facteur 4. Au delà ils se dégradent rapidement mais restent exploitables à condition de tolérer un taux d'outliers supérieur à 50%. La séquence Astérix donne de meilleurs résultats car la structure de lignes apparaissant dans cette séquence facilite la répétabilité des points d'intérêt.

5 Validations expérimentales

Dans cette section nous validons expérimentalement l'approche multi-échelles pour la détection et la caractérisation (cf. section 3). Nous présentons d'abord les expérimentations réalisées, puis les résultats pour les points d'intérêt et les invariants en niveaux de gris.

5.1 Expérimentations

Pour nos expérimentations, nous avons utilisé les trois séquences de la Figure 5. Pour pouvoir évaluer les résultats, la transformation entre les images doit être connue. Dans le cas de scènes planaires, les images

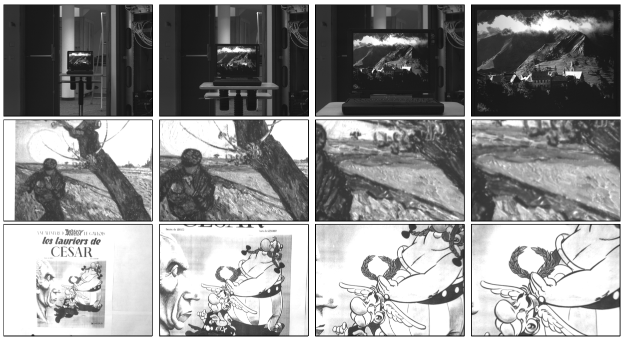

FIG. 5: *Quelques images des séquences Laptop, Van Gogh et Astérix. L'image de gauche est utilisée comme image de référence. Les échelles des images de gauche à droite sont respectivement 1, 2.13, 4.16 et 6.25 pour Laptop, 1, 1.52, 2.97 et 4.12 pour Van Gogh et 1, 2.05, 3.11 et 4.17 pour Astérix.*

échelle	Points appariés		
	initialement	nb inliers pour H	% inliers
Séquence Laptop			
1.52	23	23	100%
2.07	19	19	100%
2.57	20	20	100%
2.96	22	21	95%
3.55	20	20	100%
3.83	21	21	100%
4.43	18	17	94%
4.85	19	19	100%
5.20	21	20	95%
6.26	19	18	95%
Séquence Van Gogh			
2.05	29	28	97%
2.33	31	30	97%
2.72	30	28	93%
2.97	29	27	93%
3.39	24	21	87%
3.70	26	18	69%
4.12	26	17	65%
4.55	21	7	33%
Séquence Astérix			
2.05	27	27	100%
2.31	23	23	100%
2.86	25	25	100%
3.11	25	24	96%
3.68	17	15	88%
4.17	17	13	75%
4.70	13	8	62%

TAB. 1: *Résultats obtenus par l'algorithme multi-échelles sur les séquences de la Figure 5. L'échelle indiquée a été déterminée par calibration. Le nombre total de points appariés sans la contrainte globale est donné ainsi que la quantité de points corrects (déterminé par la méthode robuste) et le pourcentage qu'ils représentent par rapport aux appariements initiaux.*

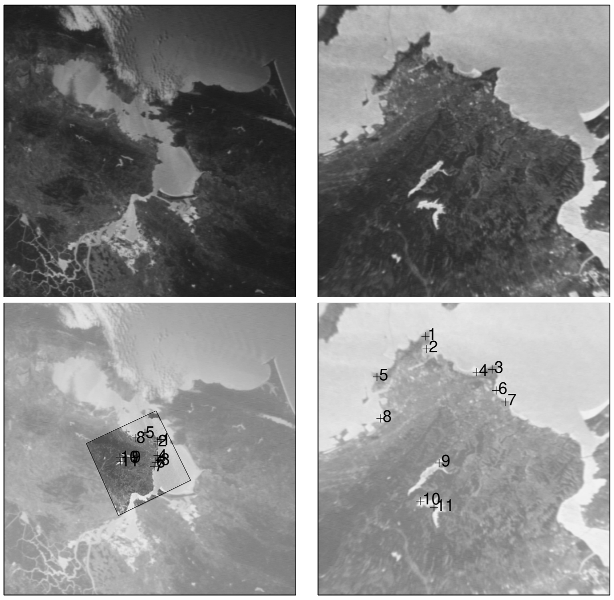

FIG. 6: *Appariement d'une paire avec changement d'échelle et rotation (en haut). Le facteur d'échelle est de de 3.9. Il y a 11 points sur 22 appariés après filtrage (en bas), soit 50% d'outliers. Le contraste a été augmenté pour faire resortir la superposition des deux images en utilisant l'homographie calculée à partir des appariements.*

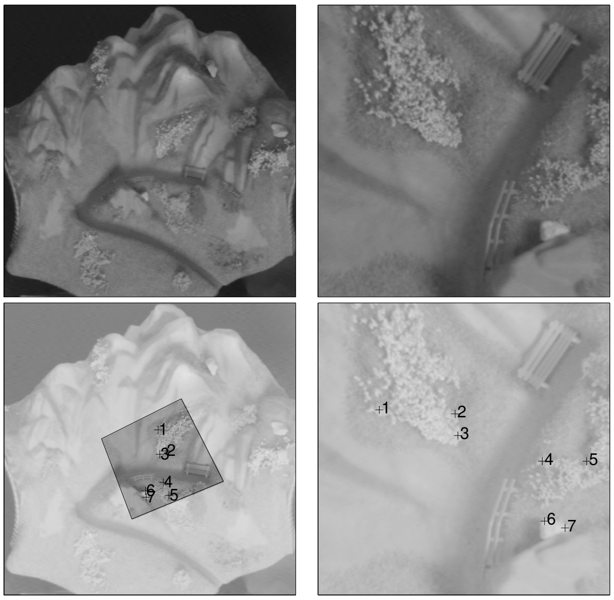

FIG. 7: *Appariement d'une paire d'image pour une scène 3D avec changement d'échelle et rotation (en haut). Le facteur d'échelle est de 3.3. Il y a 7 points sur 27 d'appariés, soit 74% d'outliers. Le contraste a été augmenté pour faire resortir le résultat de la superposition des deux images en utilisant l'homographie calculée à partir des appariements.*

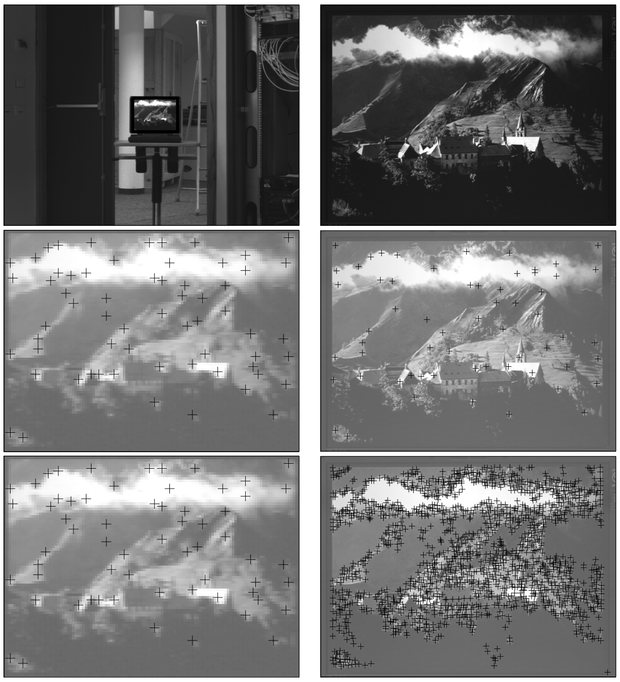

FIG. 8: *Résultats d'adaptation du détecteur de points d'intérêt sur la séquence Laptop. On peut voir la paire d'images utilisées (en haut), les points détectés par la version adaptée (au milieu), les points détectés par la version standard (en bas). Les images de gauche (milieu et bas) correspondent à la partie encadrée dans l'image* BR*, agrandie pour une meilleur visualisation. Le facteur d'échelle est de 6.25. On constate que pour la version adaptée, la plupart des points sont détectés aux mêmes endroits dans les images* HR *et* BR.

sont liées par une homographie. Une mire de calibration nous permet de calculer cette homographie précisément (voir [11] pour plus de détails). En particulier les homographies permettent dans ce cas précis de déterminer les facteurs d'échelle.

5.2 Points d'intérêt

Dans cette section nous supposons que le facteur d'échelle est connu et nous évaluons le résultat de l'adaptation du détecteur à ce changement. Nous présentons d'abord un exemple pour une paire d'images (cf. Figure 8) et donnons ensuite les résultats d'une évaluation systématique.

Nous avons systématiquement évalué la stabilité des points d'intérêt aux changements d'échelle en utilisant un critère de similarité semblable à celui utilisé par [11]. Ce taux mesure le pourcentage de points qui sont répétés entre deux images. Les points ne sont généralement pas détectés à la position exacte de leur projection, mais dans un voisinage. On note δ la taille de ce voisinage. Les points qui se trouvent sur des parties qui ne sont observables que sur une des images sont exclus. Le taux de similarité $S_i(\delta)$ entre les images I_1

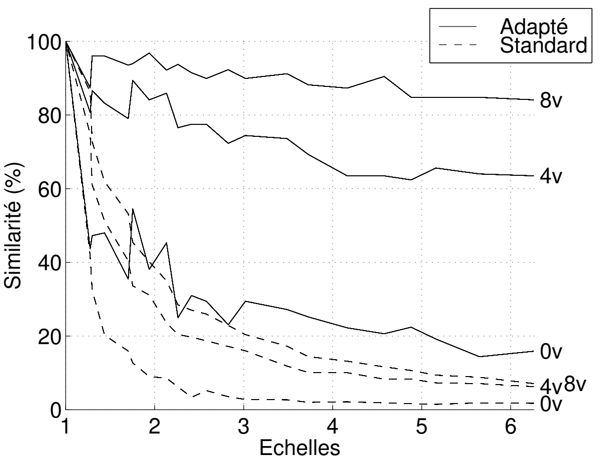

FIG. 9: *Comparaison de la version standard et de la version adaptée du détecteur de Harris. Le taux de similarité est donnée au pixel près (0v), pour un 4-voisinage (4v) et un 8-voisinage (8v) dans l'image* BR.

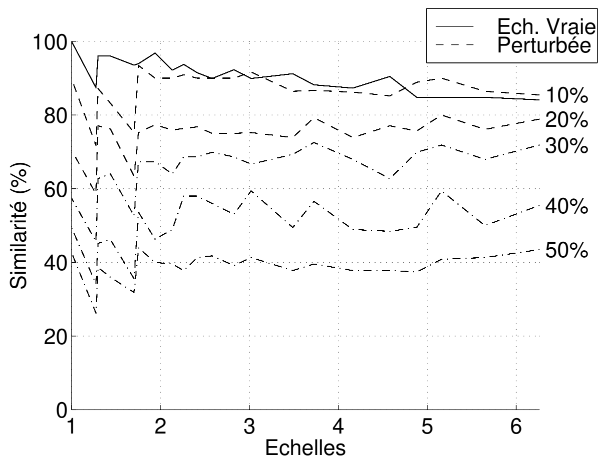

FIG. 10: *Observation de l'impact sur le taux de similarité d'une perturbation de l'échelle d'adaptation. Cette expérimentation a été faite sur la séquence Laptop, les résultats sont montrés sur la courbe correspondant au 8-voisinage de la Figure 9. Comme on le voit, jusqu'à 20% la similarité est très peu affecté, au dela de 40% le taux passe en dessous des 50%.*

et I_i est alors :

$$S_i(\delta) = \frac{|C(\delta)|}{\text{moy}(n_i, n_1)}$$

où $C(\delta)$ sont les points qui se correspondent dans un δ-voisinage et n_1, n_i le nombre de points détectés dans les images I_1, I_i

Comme on le voit sur la figure 9, la version adaptée du détecteur se maintient largement au dessus de la version standard pour une correspondance exacte. L'adaptation à l'échelle fonctionne très bien, puisque jusqu'à un facteur 6, plus de 60% des points sont détectés dans un 4-voisinages par rapport à leur position réelle, et plus de 80% dans un 8-voisinages.

Ces mesures ont été faites dans le cas où l'adaptation à l'échelle correspondait précisément au facteur existant entre les images. Pour étudier la robustesse de cette adaptation, nous avons bruité le facteur d'adaptation et mesuré son impact sur le taux de similarité (cf. Figure 10). Les résultats ne sont présentés que pour le 8-voisinage, les résultats étant proportionnels pour les deux autres voisinages. Pour une perturbation de 10 à 20% l'impact est faible, au delà de 30% seule la répétabilité dans un 8-voisinage donne des résultats exploitables.

5.3 Invariants différentiels en niveaux de gris

Dans cette section nous validons l'adaptation de l'échelle aux invariants différentiels en niveaux de gris. Le facteur d'échelle est supposé connu. Pour être indépendant des erreurs dues à l'extraction des points, nous détectons d'abord les points d'intérêt dans la première image de la séquence, puis nous propageons ces points dans les images suivantes à l'aide des homographies.

La distance entre deux vecteurs d'invariants $\mathbf{v}_1$ et $\mathbf{v}_2$ est mesurée en utilisant la distance de Mahalanobis

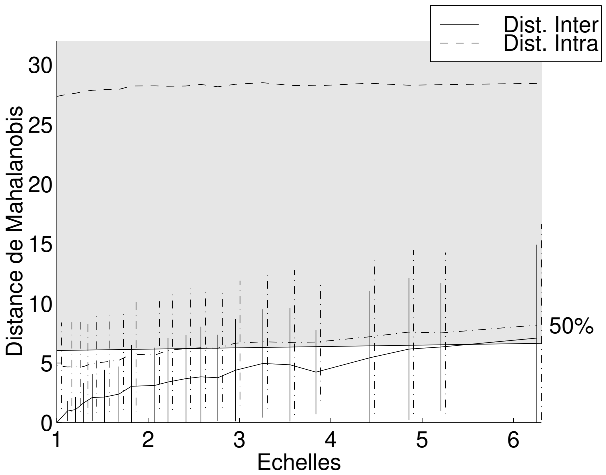

FIG. 12: *Observation de l'impact d'une perturbation de l'échelle d'adaptation sur la distance entre invariant. Cette expérimentation a été faite sur la séquence Laptop. Comme on le voit, même une perturbation de 50% de l'échelle d'adaptation ne dégrade pas excessivement la distance entre les correspondances.*

dist_M :

$$\text{dist}_M(\mathbf{v}_1, \mathbf{v}_2) = \sqrt{(\mathbf{v}_1 - \mathbf{v}_2)^T \Lambda^{-1} (\mathbf{v}_1 - \mathbf{v}_2)} \qquad (5)$$

Pour pouvoir obtenir des résultats exploitables pour cette distance, il est important d'avoir une matrice de covariance représentative qui tienne compte du bruit, des variations de luminance ainsi que de l'imprécision de localisation des points d'intérêt. Comme tout calcul théorique paraît impossible, cette matrice est ici esti-

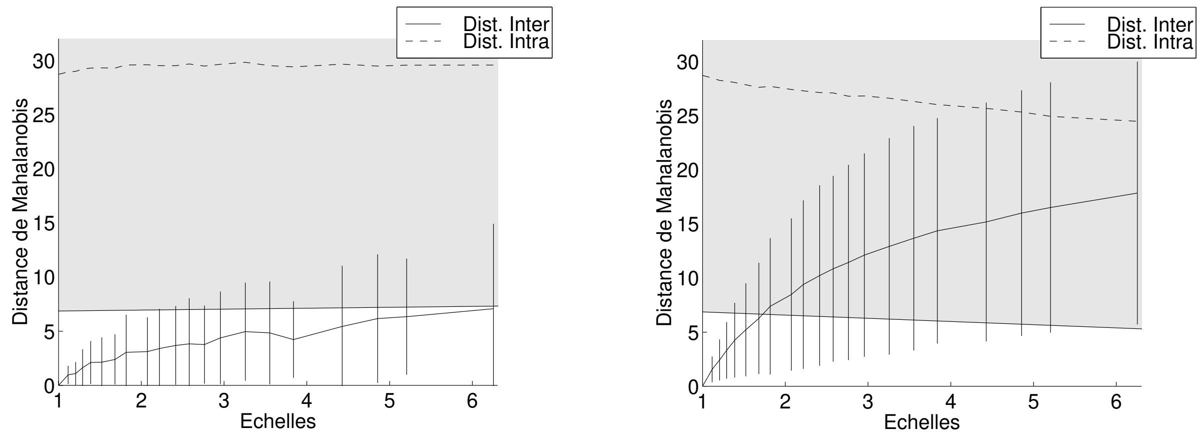

FIG. 11: *Comparaison des invariants standards et adaptés. Le graphe de gauche montre les résultats de l'adaptation à l'échelle des invariants, celui de droite les invariants standards. La courbe "intra" montre la distance moyenne de Mahalanobis entre les points se correspondant. La courbe "inter" représente la distance moyenne entre les paires ne se correspondant pas. L'écart type est représenté par des barres verticales pour la distance "intra" et par la partie grisée pour la distance "inter". Plus la distance "intra" et son écart type sont loin de la distance "inter", meilleur est la caractérisation des invariants et moindre est la probabilité de fausse mise en correspondance avec cette caractérisation.*

mée statistiquement en traquant des points d'intérêt dans les images de la séquence.

La Figure 11 compare les invariants standards et adaptés. On constate encore une fois que la version adaptée donne de bien meilleurs résultats. De la même façon que pour les points d'intérêt, nous avons mesuré l'impact d'une perturbation de l'échelle sur la caractérisation. Comme on peut le voir dans la Figure 12, une perturbation de 50% de l'échelle affecte très peu la caractérisation.

De manière à mesurer l'impact de la répétabilité des points d'intérêt sur notre algorithme d'appariement, nous avons aussi étudié l'impact d'une répétabilité sur un 4-voisinage et un 8-voisinage sur la caractérisation (cf. Figure 13). Celle-ci n'est pas excessivement perturbée, même lorsque l'on est proche d'un facteur 6. Ceci montre que les points d'intérêts répétables dans un 8-voisinage sont exploitables pour l'appariement et explique les bons résultats de l'algorithme.

6 Conclusion et discussion

Nous avons présenté une méthode d'adaptation aux changements d'échelle de la détection des points d'intérêts et de leur caractérisation par invariants différentiels. Les nombreuses expérimentations ont montré l'importance de cette adaptation et ont validé cette approche.

Nous avons proposé un algorithme d'appariement incorporant dans un cadre multi-échelles cette adaptation. Celui-ci permet sans avoir de connaissance préalable sur l'échelle d'apparier deux images. Cette algorithme donne d'excellents résultats et il a ainsi été possible de mettre en correspondance des images jusqu'à un facteur d'échelle 6. L'appariement d'images à un tel facteur reste cependant un problème ambitieux et difficile, même pour un être humain. Notre méthode est de plus invariante aux rotations images et dans une moindre mesure robuste aux déformations perspectives.

Plusieurs extensions sont possibles à ce travail. En particulier, la recherche d'un critère optimal pour le choix de l'espacement des calculs sur l'espace d'échelles afin de réduire la complexité de l'appariement. Un problème plus ambitieux est l'estimation de l'échelle directement à partir du signal ainsi que la prise en compte des déformations perspectives.

A Adaptation stable des dérivées

Pour pouvoir comparer deux images à des résolutions différentes, les valeurs d'une des images doivent être adaptée en fonction du facteur d'échelle. Ce qui suit montre qu'il est préférable d'adapter les valeurs calculées sur l'image BR, plutôt que sur l'image HR. En effet, en présence de bruit, l'équation 1 peut être écrite comme :

$$(L^1_{i_1 \ldots i_n}(\mathbf{x}, \sigma) + \epsilon^1) \approx s^n (L^2_{i_1 \ldots i_n}(s\mathbf{x}, s\sigma) + \epsilon^2)$$

où ϵ^1 et ϵ^2 correspondent aux erreurs de mesures

La comparaison des dérivées calculées sur les deux images s'exprime alors comme :

$$|L^1_{i_1 \ldots i_n}(\mathbf{x}, \sigma) - s^n L^2_{i_1 \ldots i_n}(s\mathbf{x}, s\sigma)| \leq s^n \epsilon^2 + \epsilon^1 \qquad (6)$$

On a le choix d'adapter soit les valeurs de l'image HR soit de l'image BR :

– Si on adapte le facteur de lissage de l'image HR,

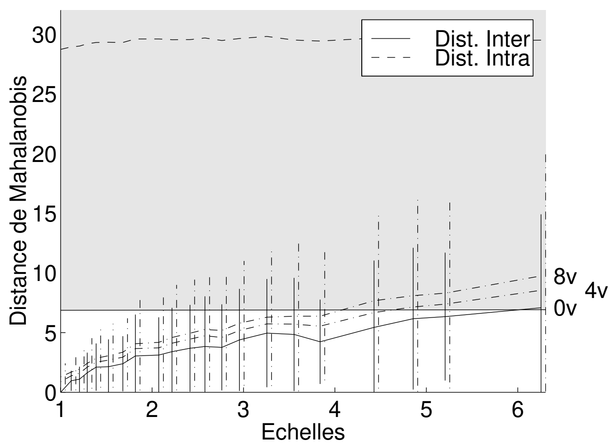

FIG. 13: *Observation de l'impact du calcul sur un 4- et un 8-voisinage sur la distance entre invariants. Sur cette figure 0v signifie que l'invariant à été calculée sur le pixel le plus proche (ceci correspond à la figure 11), 4v correspond à la moyenne des distances pour un calcul sur les voisins dans un 4-voisinage, sans tenir compte des points 0v. De même 8v correspond à la moyenne évalué sur un 8-voisinage sans tenir compte ni des points 4v ni 0v. Ceci montre que les points d'intérêt détectés en 8-voisinages peuvent être caractérisés correctement par les invariants (cf figure 9).*

on a $s > 1$ dans l'équation 6. L'erreur commise sur la différence est alors $s^n\epsilon^2 + \epsilon^1 \approx s^n\epsilon^2$ ($s > 1$, $n \geq 1$). Dans le cas d'un grand changement d'échelle cette erreur deviendra majoritaire, surtout pour le calcul de dérivées d'ordre élevé : dans nos exemples $s^n = 6^3 = 216$.

– Si au contraire on adapte ce facteur sur l'image BR, on a $s < 1$ dans l'équation 6. L'erreur commise sur la différence est alors $s^n\epsilon^2 + \epsilon^1 \approx \epsilon^2 + \epsilon^1$ ($s < 1$, $n \geq 1$). Cette erreur est indépendante du facteur d'échelle.

Lorsqu'on utilise des vecteurs de combinaisons de dérivées (cf. section 3.2) et que la comparaison de ces vecteurs utilise une distance statistique (cf. équation 5), l'utilisation de la deuxième solution est indispensable pour pouvoir comparer deux vecteurs sur la base d'un même seuil d'erreurs. Car toute normalisation à posteriori est alors impossible. Ceci a été confirmé lors de nos expérimentations.

Références

[1] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381 – 395, June 1981.

[2] N. Georgis, M. Petrou, and J. Kittler. On the correspondence problem for wide angular separation of non-coplanar points. *Image and Vision Computing*, 16:35–41, 1998.

[3] F. Glazer, G. Reynolds, and P. Anandan. Scene matching by hierarchical correlation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Washington, DC, USA*, pages 432–441, 1983.

[4] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[5] J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[6] M. S. Lew and T. S. Huang. Optimal multi-scale matching. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, volume 2, pages 88–93, June 1999.

[7] T. Lindeberg. *Scale-Space Theory in Computer Vision.* Kluwer Academic Publishers, 1994.

[8] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 754–760. IEEE Computer Society Press, January 1998.

[9] L. H. Quam. Hierarchical warp stereo. In *Reading in computer Vision*, pages 80–86. Morgan Kaufman, 1987.

[10] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–534, May 1997.

[11] C. Schmid, R. Mohr, and Ch. Bauckhage. Comparing and evaluating interest points. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 230–235, January 1998.

[12] T. Tuytelaars, L. Van Gool, L. D'haene, and R. Koch. Matching of affinely invariant regions for visual servoing. In *Proceedings of* IEEE *International Conference on Robotics and Automation*, 1999. to appear.

[13] A.P. Witkin. Scale-space filtering. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, Germany*, pages 1019–1023, 1983.