Indexation à l'aide de points d'intérêt invariants à l'échelle

Krystian Mikolajczyk, Cordelia Schmid

Indexation à l'aide de points d'intérêt invariants à l'échelle

Krystian MIKOLAJCZYK Cordelia SCHMID

INRIA Rhône-Alpes
655 av. de l'Europe
38330 Montbonnot, France
Krystian.Mikolajczyk@inrialpes.fr, Cordelia.Schmid@inrialpes.fr

Résumé

Cet article propose une méthode innovante de détection de points d'intérêt invariants aux changements d'échelle. La méthode est basée sur deux résultats utilisant l'espace d'échelle : 1) Les points d'intérêt peuvent être adaptés à l'échelle et donnent des résultats répétables (géométriquement stables). 2) Les extrema locaux de dérivées normalisées à l'échelle indiquent la présence de structures locales caractéristiques. D'abord, on extrait des points d'intérêt à plusieurs échelles avec le détecteur de Harris. Ensuite on sélectionne les points où la mesure locale (Laplacien) donne une réponse maximale dans l'espace d'échelle. Ceci permet de choisir un ensemble de points discriminants pour lesquels les échelles locales sont connues. Ces points sont invariants aux changements d'échelle, à la rotation et à la translation.

Ces points détectés permettent d'indexer des images ayant subi une rotation, une translation et un changement d'échelle important. Les images sont caractérisées par un ensemble de points. L'échelle associée à chaque point permet de calculer des descripteurs invariants aux changements d'échelle. Les résultats expérimentaux pour l'indexation montrent une excellente performance de la méthode jusqu'à un facteur d'échelle de 4.5 pour une base de 800 images.

Mots clés : sélection d'échelle, multi-échelles, points d'intérêt, appariement, indexation.

1 Introduction

La difficulté dans l'indexation d'objets est de déterminer l'identité d'un objet en présence de changements des conditions de prise de vue ou d'occultation. La caractérisation locale est bien adaptée à ce problème. Les petites régions caractéristiques autour des points d'intérêt sont robustes aux occultations et aux changements de fond. Pour obtenir l'invariance à des transformations plus complexes, il faut appliquer des descripteurs robustes à ces transformations. Les méthodes d'indexation proposées récemment dans la littérature utilisent différents types de descripteurs.

Schmid et Mohr [10] détectent des points d'intérêt qui sont ensuite caractérisés par des invariants à la rotation. La robustesse à l'échelle est obtenue en calculant les descripteurs sur plusieurs échelles. Pour obtenir l'invariance à l'échelle, Lowe [8] cherche des maxima dans un espace d'échelle. Les niveaux successifs de l'espace sont calculés à partir des résultats de soustractions de filtres gaussiens. Tuytelaars et Van Gool [13] ont proposé un descripteur invariant aux changements affines. Ils cherchent des régions caractéristiques en utilisant simultanément des points d'intérêt et des contours. Les régions sont caractérisées par des invariants affines basés sur la couleur. Au lieu d'utiliser un ensemble de points d'intérêt,

Chomat [2] détecte des échelles caractéristiques pour chaque point de l'image et calcule des descripteurs à ces échelles. Un objet est représenté par un histogramme de ces desripteurs. Toutes les méthodes mentionnées sont limitées à un facteur d'échelle de 2.

D'autres approches ont été proposées dans le contexte d'appariement d'images prises sous des angles de vue très différents [1, 3, 5, 9, 12]. Prichett et Zisserman [9] apparient des régions limitées par quatre segments de droites. Ensuite, ils utilisent les régions correspondantes pour calculer une homographie. De telles approches sont difficiles à adapter dans le contexte de l'indexation. Deux publications concernant l'appariement d'images prises sous des points de vue très différents, abordent le problème du changement d'échelle. Afin de trouver la correspondance entre deux images, Hansen [5] a proposé une méthode de corrélation de traces d'échelle construites par des filtres gaussiens calculés à plusieurs échelles. Une trace d'échelle est un ensemble des valeurs en un point sur des niveaux de résolution consécutifs. Dufournaud [3] a proposé une approche multi-échelle pour apparier des images à des échelles très différentes. Des points et des descripteurs sont calculés à plusieurs niveaux d'échelle. Un algorithme d'appariement robuste permet ensuite de sélectionner l'échelle correcte. Les deux approches [3, 5] ne sont toutefois pas applicables dans un contexte d'indexation, puisqu'elles nécessitent une comparaison *image à image*. Des descripteurs discriminants, directement accessibles sont nécessaires dans un contexte d'indexation. Le stockage de plusieurs niveaux d'échelle est une solution partielle, puisqu'elle augmente la possibilité de faux appariements.

Dans cet article, nous proposons une approche qui permet l'indexation d'images ayant subi des changements d'échelle importants. La réussite de la méthode repose sur l'utilisation d'un détecteur de points répétables et discriminants. La méthode de détection est fondée sur deux résultats dans un contexte de changement d'échelle : 1) Les points d'intérêt peuvent être adaptés à l'échelle et donnent des résultats répétables [3]. 2) Les extrema locaux des dérivées normalisées dans la dimension d'échelle indiquent la présence de structures locales caractéristiques [7]. Dans un premier temps on calcule des points d'intérêt à plusieurs niveaux d'échelle. Ensuite, on sélectionne les points où la mesure locale (Laplacien) donne une réponse maximale dans la dimension d'échelle. Cela nous permet de rejeter les points les moins discriminants détectés dans le premier stade de l'algorithme. Cette méthode permet de choisir un ensemble de points discriminants pour lesquels les échelles locales sont connues. Ainsi, on obtient un ensemble de points d'intérêt qui signalent les endroits où le signal est le plus informatif. Les points sont invariants aux changements d'échelle, à la rotation et à la translation. Nous montrons par la suite, que la répétabilité de cette méthode est meilleure que celle des approches proposées récemment dans la littérature. Ainsi, nous pouvons obtenir de meilleurs résultats d'indexation. La deuxième contribution de cet article est la qualité des résultats d'appariement et d'indexation.

Organisation Cet article est organisé de la manière suivante : La section 2 introduit brièvement le cadre multi-échelle. Nous décrivons notre détecteur de points d'intérêt dans la section 3. La section 4 présente l'algorithme d'appariement et d'indexation. Les résultats expérimentaux sont présentés dans la section 5.

2. Représentation multi-échelle

Les propriétés d'échelles caractéristiques ont été étudiées par [7]. Les points caractéristiques sont indiqués par des maxima dans l'espace d'échelle construit à partir de dérivées normalisées. Plusieurs fonctions ont été proposées pour construire des représentations d'images en échelle. Les fonctions dépendent du type d'information que l'on veut extraire d'une image (i.e. blobs, contours). La représentation en échelle est un ensemble d'images d'une scène représentée à plusieurs niveaux de résolution. Nous pouvons représenter le contenu d'une image (i.e. contours, coins) à différents niveaux de résolution en utilisant une fonction adéquate. La fonction doit être normalisée par rapport au niveau d'échelle. Ainsi,

on obtient une pyramide représentant une image. Soient I et I' deux images à des échelles différentes liées par la relation suivante :

$$I(\mathbf{x}) = I'(\mathbf{x}'),\ où\ \mathbf{x}' = s\mathbf{x} + c,\ \mathbf{x} = (x, y) \tag{1}$$

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

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

Les dérivées de l'image sont calculées par convolution avec des dérivées gaussiennes.

$$L_{i_1 \ldots i_n}(\mathbf{x}) = I(\mathbf{x}) * G_{i_1 \ldots i_n}(\sigma) = s^n I'(\mathbf{x}') * G_{i_1 \ldots i_n}(s\sigma) \tag{3}$$

Afin de maintenir un changement d'information uniforme entre deux niveaux successifs, l'espacement du facteur d'échelle doit être exponentiel. Un ensemble de réponses calculées pour un point donné $\mathbf{x}$ dans une image I est représenté par

$$F(I, \mathbf{x}, s_n),\ s_n = s_0^n \tag{4}$$

où s_0 est le facteur d'échelle initiale correspondant à la plus haute résolution d'image, et s_n dénote les niveaux consécutifs dans l'espace d'échelle. F représente la fonction qui construit les niveaux de résolution. Pour un point donné nous pouvons calculer des réponses de telles fonctions pour plusieurs facteurs s. Ainsi, on construit la représentation en échelle pour le voisinage de ce point. Dans l'ensemble des réponses, on cherche une échelle caractéristique relativement indépendante de la résolution de l'image. Cette échelle caractéristique est déterminée par un maximum local de la fonction (figure 1). Le quotient des échelles trouvées pour deux points correspondants est égal au facteur d'échelle réelle entre les deux images. Nous pouvons appliquer des fonctions différentes pour construire la représentation d'une

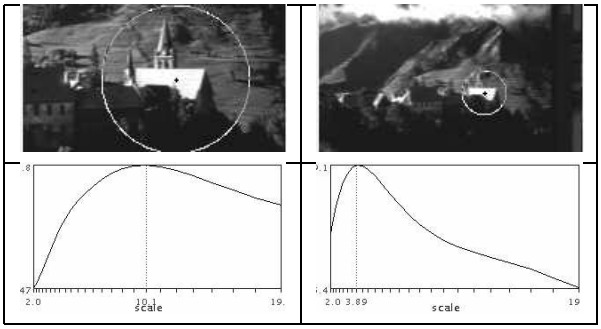

FIG. 1 – *Ensemble de réponses $F(I, \mathbf{x}, s_n)$ pour la fonction du Laplacien (traces d'échelle).*

image en échelle. Les fonctions doivent être au moins invariantes à la rotation. L'invariance aux conditions d'éclairage est moins nécessaire car nous cherchons des extrema. Des variations d'éclairage ne changent pas les profils de traces d'échelle mais leurs valeurs absolues qui sont moins importantes dans la sélection des extrema. Dans nos expérimentations nous avons utilisé les expressions différentielles suivantes :

– carré du gradient
$$s_n^2(L_x^2(\mathbf{x}, s_n\sigma) + L_y^2(\mathbf{x}, s_n\sigma)) \tag{5}$$

– Laplacien
$$|s_n^2(L_{xx}(\mathbf{x}, s_n\sigma) + L_{yy}(\mathbf{x}, s_n\sigma))| \tag{6}$$

– fonction proposée par Lowe
$$|I(\mathbf{x}) * G(s_{n-1}\sigma) - I(\mathbf{x}) * G(s_n\sigma)| \tag{7}$$

– fonction de Harris adaptée à l'échelle
$$\mathbf{C}(\mathbf{x}, \sigma, \tilde{\sigma}) = s_n^2 G(\mathbf{x}, s_n\tilde{\sigma}) \star \begin{bmatrix} L_x^2(\mathbf{x}, s_n\sigma) & L_xL_y(\mathbf{x}, s_n\sigma) \\ L_xL_y(\mathbf{x}, s_n\sigma) & L_y^2(\mathbf{x}, s_n\sigma) \end{bmatrix} \tag{8}$$

Chomat et al. [2] ont montré que la fonction du gradient peut être utilisée pour détecter l'échelle caractéristique d'un point. L'amplitude du gradient est naturellement invariante à la rotation. La même propriété est assurée par le Laplacien qui est particulièrement bien adapté pour la détection de blobs [7]. La fonction proposée par Lowe [8] est très similaire au Laplacien mais elle permet d'accélérer le calcul de l'espace d'échelle. Des travaux précédents sur les points d'intérêt [11] ont montré que le détecteur de Harris était le plus répétable. Néanmoins, cette répétabilité se dégrade rapidement en fonction du changement de la résolution. Pour pouvoir aborder ce problème il faut adapter le détecteur de Harris aux changements d'échelle [3]. Un point d'intérêt $\mathbf{x}$ est sélectionné si le déterminant et la trace de la matrice (8) vérifient la condition suivante :

$$det(\mathbf{C}) - \alpha trace^2(\mathbf{C}) > seuil_h \quad (9)$$

où $seuil_h$ est un seuil et α un paramètre choisi expérimentalement. L'invariance à la rotation est assurée par la symétrie de la matrice $\mathbf{C}$.

Pour illustrer le comportement de la sélection automatique d'échelle, nous avons effectué une sélection d'échelle caractéristique en tous points d'images réelles. Ensuite, nous avons établi des correspondances de points à l'aide de la matrice de transformation entre deux images. Nous avons observé le comportement des fonctions (5, 6, 7, 9) en présence de changements d'échelle. En cherchant l'échelle caractéristique d'un point d'intérêt nous devons considérer que la taille de la région autour du point doit être limitée. Cela signifie que la haute fréquence contenue dans la région est plus importante. La haute fréquence est aussi moins sensible aux changements d'éclairage. De cette façon, des détecteurs basés sur des dérivées d'un ordre plus élevé sont capables de détecter de petits changements locaux du signal. Par contre, les dérivées d'ordre petit sont moins sensibles au bruit et à l'erreur de la localisation d'un point d'intérêt. La figure 2 montre une configuration théorique du signal où le Laplacien atteint un maximum et le gradient non.

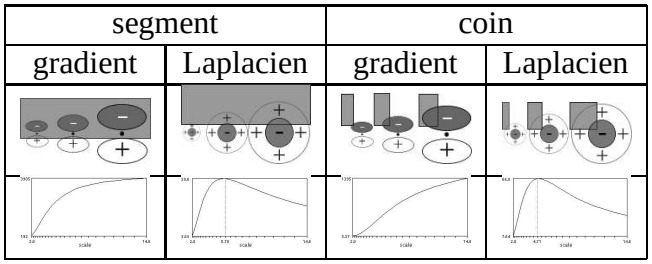

FIG. 2 – *Le gradient vis-à-vis du Laplacien.*

Pendant nos expérimentations nous avons remarqué que les extrema locaux sont détectés dans le voisinage des changements du signal où la réponse du gradient est forte. Par contre les extrema sont rarement atteints exactement aux points où le gradient atteint un maximum dans la surface d'image (contours, coins, figure 3). La gamme d'échelles utilisée pour créer la représentation en échelle doit être limitée et être la même pour toutes les images si on n'a pas la connaissance a priori du facteur d'échelle entre les images considérées. Parfois, des nouveaux changements du signal peuvent apparaître dans l'image de haute résolution. Ceci peut provoquer un changement du profil de la trace de l'échelle et fausser la détection de l'échelle caractéristique. On peut supposer que cette échelle est trouvée près de la limite haute de s_n, dans l'image de basse résolution. Dans ce cas, il est très probable que le point correspondant dans l'image de haute résolution se trouve trop loin du changement du signal. Le gradient ou le Laplacien ne peuvent pas capturer ce changement dans les limites d'échelle appliquées. Nos résultats ont montré que les points détectés en cherchant des maxima uniquement dans la direction d'échelle sont trop sensibles à cet effet. D'autre part, on ne peut pas appliquer une gamme d'échelles trop grande, sinon on perd le caractère local des points d'intérêt et l'effet de bord devient trop important.

Dans la figure 3, nous comparons le pourcentage de points avec des échelles correctement estimées, par rapport aux points pour lesquels les extrema ont été détectés. Les échelles de deux points correspondants sont correctement estimées si leur quotient est proche du facteur d'échelle entre deux images. La

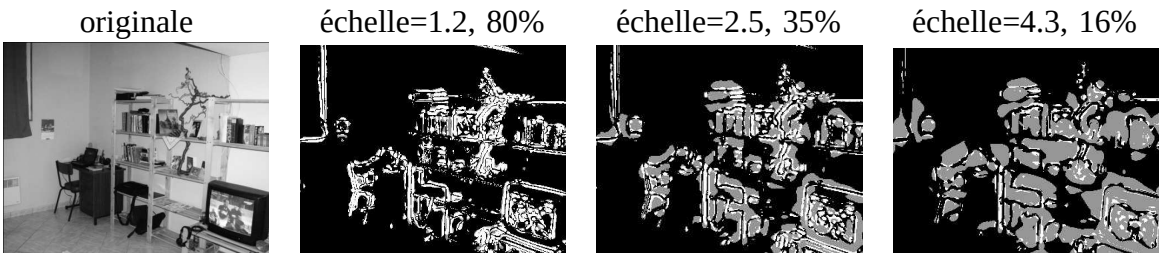

FIG. 3 – *Les échelles caractéristiques des points. En gris : les échelles détectées, En blanc : les échelles détectées et correctement appariées. Afin d'augmenter la visibilité, les images de basse résolution ont été élargies :* $échelle = \frac{image\,originale}{image\,transformée}$.

détection d'échelle caractéristique a été appliquée pour tous les points des images d'échelles différentes. Les points ainsi détectés ne sont plus stables au-delà d'un facteur d'échelle de 2.5. Le pourcentage des échelles correctement estimées est inférieur à 50%. Nous avons aussi remarqué que les fonctions de Harris et du gradient détectent relativement moins de points caractéristiques par image. Cela signifie que ces fonctions sont beaucoup plus sélectives dans la recherche de points d'intérêts.

3. Points d'intérêt invariants à l'échelle

Afin de trouver un ensemble de points plus stables, nous cherchons les maxima dans la représentation en échelle (tri-dimensionnelle) d'une image. Un point dans l'espace est considéré comme un point d'intérêt s'il est un maximum dans le voisinage le plus proche et si, de plus, sa valeur est plus grande qu'un certain seuil (figure 4, équation 10).

$$\forall \mathbf{x_w} \in W \;\; seuil_a < F(I, \mathbf{x}, s_n) > F(I, \mathbf{x_w}, s_m),\; m \in \{n-1, n+1\} \tag{10}$$

où W est le voisinage immédiat du point $\mathbf{x}$.

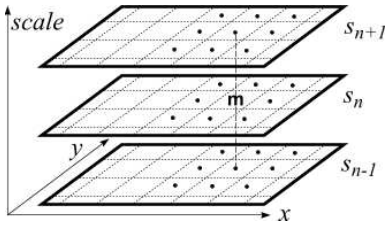

FIG. 4 – *Le maximum de la fonction* $F(I, \mathbf{x}, s_n)$ *dans l'espace d'échelle.*

La fonction de Harris atteint rarement des maxima dans l'espace 3D. Si trop peu de points sont détectés, la représentation de l'image n'est pas robuste aux différentes transformations. Pour résoudre ce problème, deux représentations en échelle sont construites pour une image donnée. La première est créée par la fonction de Harris adaptée. On obtient un ensemble de points en détectant à chaque niveau de la pyramide des maxima dans la surface d'image (équation 11). La majorité des points (x, y) sont détectés à des endroits très proches sur plusieurs niveaux d'échelle consécutifs.

$$\forall \mathbf{x_w} \in W \;\; seuil_h < F(I, \mathbf{x}, s_n) > F(I, \mathbf{x_w}, s_n) \tag{11}$$

Afin d'obtenir une représentation plus compacte, on vérifie pour chaque point détecté précédemment s'il constitue un maximum dans la dimension d'échelle dans la deuxième représentation (équation 12). La deuxième représentation est construite avec le Laplacien

$$F(I, \mathbf{x}, s_{n-1}) < F(I, \mathbf{x}, s_n) > F(I, \mathbf{x}, s_{n+1}) \;\cap\; F(I, \mathbf{x}, s_n) > seuil_l \tag{12}$$

avec $seuil_h$ et $seuil_l$ choisis expérimentalement. Le détecteur de Harris modifié détermine la localisation de points d'intérêt dans la surface d'image pour un niveau de résolution donné. La fonction Laplacien

détermine l'échelle caractéristique d'un point d'intérêt. Le choix de la fonction de Harris a été motivé par les résultats de la comparaison de différents détecteurs menée par [11], où le détecteur de Harris s'est avéré le plus fiable. Une extension de ces travaux proposée par [3] a montré la possiblité d'adapter ce détecteur aux changements d'échelle. Le choix de la fonction de recherche de maxima dans la direction d'échelle est moins critique. Nous proposons d'utiliser l'opérateur Laplacien. La figure 5 montre la représentation en échelle de deux images réelles avec des points détectés par la méthode de Harris-Laplacien. Pour les deux images présentant le même objet nous montrons les points d'intérêt sur des niveaux d'échelle consécutifs pour lesquels les points ont été détectés. Il y a beaucoup de points qui se correspondent sur les niveaux (indiqué par les flèches) pour lesquels la relation des échelles correspond au facteur d'échelle réel. Les points sont alors caractéristiques sur la surface d'image et dans la dimension d'échelle.

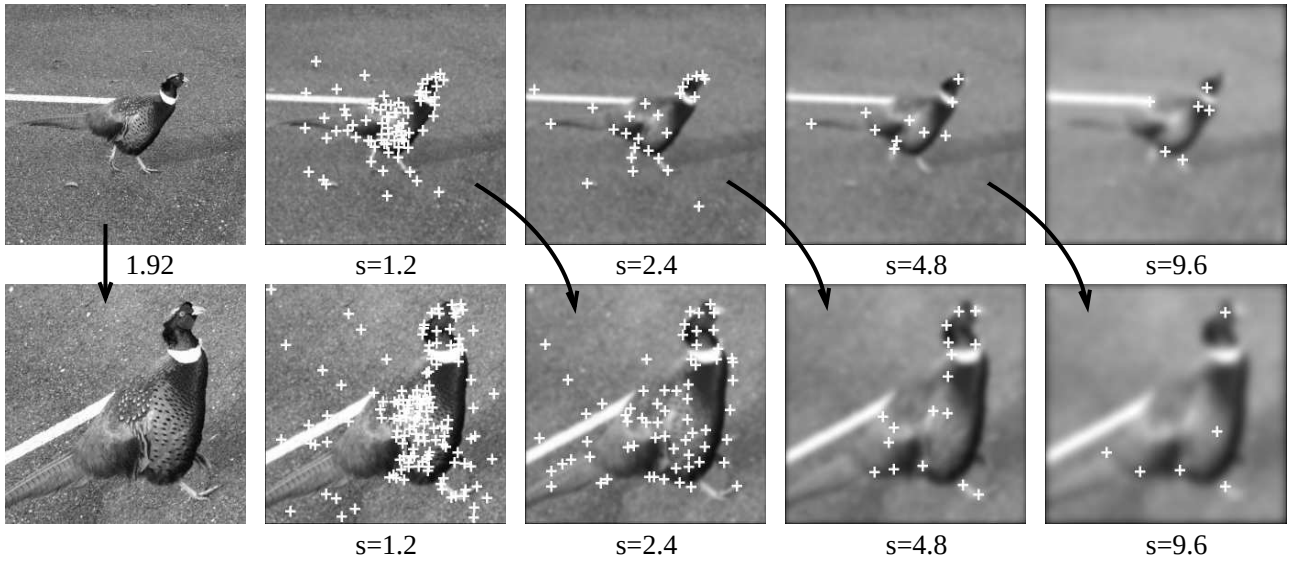

FIG. 5 – *Points détectés sur différents niveaux de résolution par la méthode de Harris-Laplacien.*

Nous allons évaluer les performances des fonctions basées sur les dérivées premières et secondes. Nous cherchons des points d'intérêt et leurs échelles caractéristiques dans l'espace tridimensionnel construit pour chacune des fonctions-gradient, Laplacien, fonction de Lowe, et détecteur de Harris modifié. Afin de pouvoir mesurer le profit d'adaptation à l'échelle nous présentons aussi des résultats pour la version du détecteur de Harris non adapté aux changements d'échelle (`Harris2D_na`). Nous avons évalué la stabilité des points d'intérêt en utilisant un critère introduit par [11]. Ce taux mesure le pourcentage de points qui sont répétés entre deux images, par rapport au nombre moyen de points détectés dans les deux images $S_{1,2} = \frac{C(I_1,I_2)}{moy(m_1,m_2)}$, où $C(I_1, I_2)$ est le nombre de points mis en correspondance et $m_1,\ m_2$ est le nombre de points détectés dans les deux images. Deux points se correspondent si l'erreur de localisation dans l'image de basse résolution ne dépasse pas $1.5\ pixel$ et si la différence relative entre le facteur d'échelle réelle et le quotient des échelles caractéristiques est inférieure à 20%. La figure 6 présente le taux de répétabilité calculé pour les séquences du test. Les meilleurs résultats sont obtenus pour la méthode de Harris-Laplacien (`Harris`). Ils peuvent s'expliquer par une bonne répétabilité du détecteur de Harris adapté et la bonne performance du Laplacien en sélection d'échelle caractéristique. Comme nous l'avons déjà remarqué, un taux légèrement meilleur pour la méthode du Laplacien (`Laplacian`) est dû à un plus grand nombre d'échelles sélectionnées par point.

Pour nos expérimentations, nous avons utilisé 10 séquences d'images réelles. Chaque séquence est constituée d'images ayant subi un changement d'échelle et une rotation. Le facteur d'échelle varie entre 1.2 et 4.5.

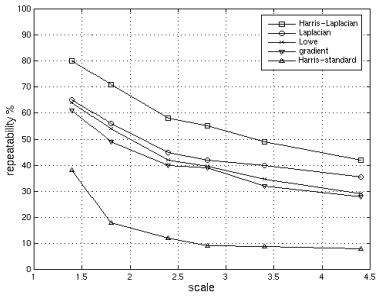

FIG. 6 – *Le taux de répétabilité des détecteurs.*

4. Appariement robuste et indexation

Dans la suite nous décrivons brièvement l'algorithme d'appariement et d'indexation. Les deux sont fondés sur le même principe :

1. Extraction des points d'intérêt par la méthode de Harris-Laplacien (cf. section 3).
2. Calcul de descripteurs invariants à la rotation pour chaque point en son échelle caractéristique. On utilise les filtres orientables [4].
3. Calcul de similarité entre les descripteurs, basée sur la distance de Mahalanobis.

Points d'intérêt Pour détecter des points d'intérêt, nous construisons 17 niveaux de résolution correspondant aux échelles choisies. Nous calculons 17 niveaux d'échelles à $s_n = s^n s_0 = 1.2^n,\ n = 0 \ldots 16$. Le paramètre α est égal à 0.06. Les seuils $seuil_h$ et $seuil_l$ sont choisis expérimentalement et correspondent à des niveaux du bruit pour la fonction de Harris et celle du Laplacien (1000 et 10 respectivement).

Descripteurs Les descripteurs sont basés sur des filtres orientables [4]. Pour caractériser un point d'intérêt nous utilisons des dérivées de niveaux de gris calculées jusqu'à l'ordre 3. Afin d'obtenir des dérivées indépendantes de la rotation existante entre deux images, la direction de calcul des dérivées est rapportée à celle du gradient. Ainsi, on obtient un vecteur de 9 invariants.

Comparaison de descripteurs La ressemblance entre des descripteurs est mesurée à l'aide de la distance de Mahalanobis. La matrice de covariance, nécessaire pour calculer la distance, est estimée statistiquement en suivant les points d'intérêt dans des images d'une séquence.

Appariement robuste Pour effectuer un appariement robuste, dans la première phase (I) nous déterminons des correspondances de points entre deux images. Pour chaque descripteur de point d'une image nous cherchons le descripteur le plus ressemblant dans l'autre image. Si la distance entre deux descripteurs est supérieure à un certain seuil (30), les points appariés sont rejetés. Dans le cas, où plusieurs points ont le même point correspondant on garde la paire la plus ressemblante. Ceci permet d'obtenir un premier ensemble de correspondances. Dans la phase finale (II), une estimation robuste de la transformation entre deux images, fondée sur la méthode *RANSAC*, permet de rejeter les faux appariements. Puisque la transformation entre deux régions correspondantes est un zoom, nous avons estimé une homographie entre les deux images pour identifier les appariements corrects. Un algorithme de sélection de modèle (cf. [6]) peut être employé pour déterminer la transformation la plus adéquate.

Indexation La recherche d'images dans une base est effectuée à l'aide de la technique de vote. Chaque point de la base est comparé à la liste de points extraits de l'image requête. Un vote est ajouté à une entrée d'une table de vote si la distance de Mahalanobis entre le point de la base et un point de la liste est inférieur à un seuil (15). On obtient une table de votes où les meilleurs scores correspondent aux images les plus similaires à l'image requête. Notons qu'un point de la base ne peut voter qu'une fois pour l'image correspondante.

5. Validations expérimentales

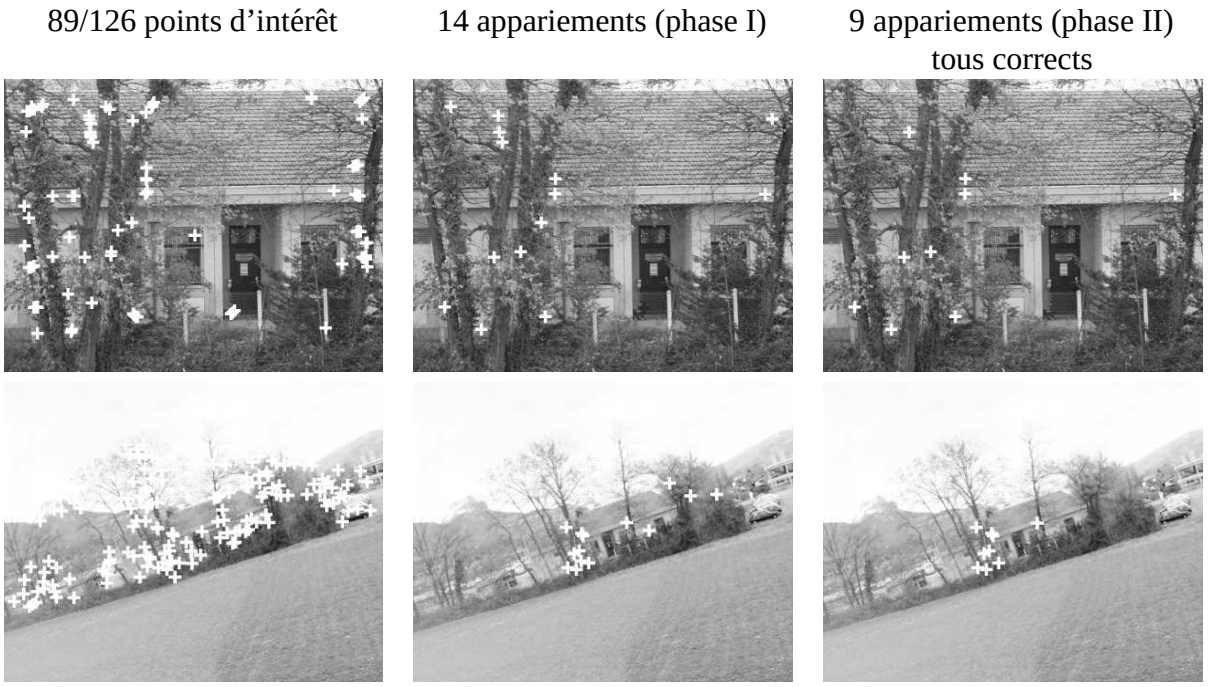

FIG. 7 – *Appariement robuste : Il y a 89 points détectés dans l'image du haut et 126 dans l'image du bas. Après la phase I il y a 14 points appariés. Après la phase II il reste 9 appariements, tous corrects. Le facteur d'échelle estimé est de* 4.9*, et l'angle de rotation de* $19°$.

Nous présentons dans ce paragraphe, deux exemples validant notre algorithme. Nous avons effectué un test d'appariement, et un test d'indexation. Un exemple d'appariement est présenté dans la figure 7. Les points détectés sont présentés dans la colonne de gauche. Il y a respectivement 89 et 126 points détectés dans l'image du haut et du bas. Le nombre de points détectés est équivalent aux résultats obtenus avec un détecteur standard appliqué sur un niveau de résolution. Le détecteur standard, sans sélection de maxima dans la direction d'échelle, détecte environ 2000 points pour l'ensemble des échelles. Ceci montre la sélectivité de notre méthode. Dans la figure 7 (au milieu) nous présentons 14 paires de points appariés après la première phase. Parmi eux, il y a 9 appariements correct après la deuxième phase (à droite). Le facteur d'échelle estimé est égal à 4.9. La figure 9 montrent d'autres exemples d'appariement. Les résultats sont également très bons.

réponse	échelle					
	1.4	1.8	2.4	2.8	3.4	4.4
1	100%	100%	70%	60%	50%	50%
3			100%	90%	80%	70%
6					90%	80%

TAB. 1 – *Résultats d'indexation d'images.*

Dans la suite nous présentons les résultats d'indexation et de recherche dans une base d'images. La base est constituée de 800 images. La moitié de ces images provient de la base "Columbia". La deuxième partie est issue d'une séquence vidéo de journaux télévisés. Il y a 116 045 descripteurs dans la base. Nous avons inclus dans la base une image par séquence de test. La figure 8 (en bas) montre quelques exemples de ces images. Les autres images des séquences ont servi pour évaluer la performance

de la recherche dans la base. La figure 8 (première ligne) montre les images requêtes pour lesquelles les images correspondantes (deuxième ligne) ont été correctement retrouvées. Nous pouvons constater que la méthode donne de très bons résultats jusqu'à un facteur d'échelle de 4.4. Les résultats du test sont présentés dans le tableau 1. Pour un facteur d'échelle de 4.4, 50% des images requêtes ont été correctement retrouvées, 70% se trouvent parmi les 3 premières réponses et 80% parmi les 6 premières réponses.

3.2 1/4.2 4.4 4.1 1/4.1

FIG. 8 – *Résultats d'indexation. Quelques exemples d'images requêtes (en haut). Les images correctement retrouvées sont visibles en bas et les facteurs d'échelle en dessous.*

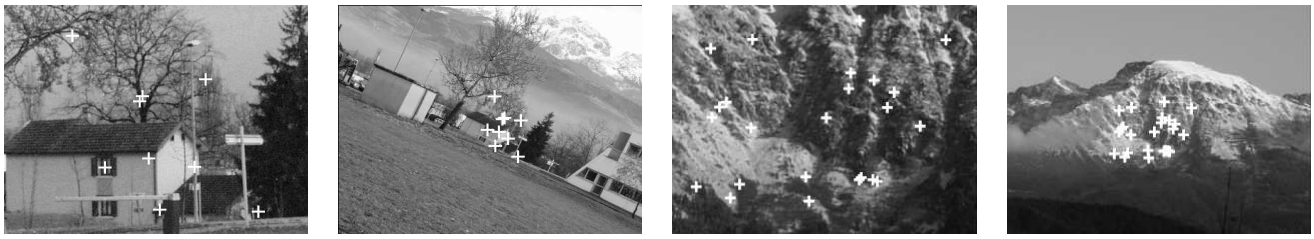

FIG. 9 – *Exemples d'appariement d'images. Pour la paire de gauche : le facteur d'échelle est de* 4.28*, l'angle de la rotation est de* $25.4°$*. 112 points sont détectés à gauche et 89 à droite. 9 appariements restent après la phase II, tous corrects. Pour la paire de droite : le facteur d'échelle est de* 3.25*. 115 points sont détectés à gauche et 128 à droite. 24 appariements restent après la phase II, tous corrects.*

6. Conclusions et perspectives

Nous avons présenté un algorithme de détection de points d'intérêt robuste aux changements d'échelle importants. Les expérimentations ont été menées sur un nombre de données important. Nous avons comparé notre détecteur de points à ceux proposés récemment dans la littérature. Les résultats de la comparaison sont favorables à notre détecteur. L'élément essentiel dans notre approche est la sélection de points dans deux espaces d'échelle construits pour chaque image. Les excellents résultats valident cette approche dans les cas de changements d'échelle jusqu'à un facteur de 4.5. L'indexation d'images à un tel facteur reste un problème ambitieux et rarement abordé dans le domaine de la vision. Nous pensons améliorer les résultats d'indexation avec un descripteur plus robuste, et envisageons une extension de ce travail aux cas des transformations affines et des changements d'éclairage d'une scène, notamment en utilisant l'information de couleur.

Remerciement

Je tiens à remercier le projet RNRT AGIR qui a soutenu ce travail.

Références

[1] A. Baumberg. Reliable feature matching across widely separated views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pages 774–781, juin 2000.

[2] O. Chomat, V. Colin de Verdière, D. Hall, et J. Crowley. Local scale selection for gaussian based description techniques. In *Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland*, pages 117–133, juin 2000.

[3] Y. Dufournaud, C. Schmid, et R. Horaud. Matching images with different resolutions. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pages 612–618, juin 2000.

[4] W.T. Freeman et E.H. Adelson. The design abd use of steerable filters. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.

[5] B. B. Hansen et B. S. Morse. Multiscale image registration using scale trace correlation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, volume 2, pages 202–208, juin 1999.

[6] K. Kanatani. Geometric information criterion for model selection. *International Journal of Computer Vision*, 26(3):171–189, 1998.

[7] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[8] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 1150–1157, septembre 1999.

[9] P. Pritchett et 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, janvier 1998.

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

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

[12] D. Tell et S. Carlsson. Wide baseline point matching using affine invariants computed from intensity profiles. In *Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland*, pages 814–828, juin 2000.

[13] T. Tuytelaars et L. Van Gool. Content-based image retrieval based on local affinely invariant regions. In *Visual'99, Amsterdam, The Netherlands*, pages 493–500, juin 1999.