

Autocorrélation basée sur les transformations pour la détection de régions affines covariantes

S. Khoualed

A. Bartoli

T. Collins

Clermont Université, France

{khoualed.samir, adrien.bartoli, toby.collins}@gmail.com

Résumé

Dans cette contribution nous introduisons une nouvelle approche pour détecter les éléments saillants dans des images. Elle est basée sur ce que nous avons appelé *warp-based autocorrelation*. Nous combinons cette méthode avec une paramétrisation à base de points de contrôle, afin de détecter les régions covariantes. L'algorithme commence par une recherche exhaustive sur un ensemble de paramètres affines. Afin de rejeter les régions instables, un seuillage est appliqué. Les résultats obtenus montrent que notre approche surpasse les autres méthodes en terme de répétabilité et de précision.

1 Introduction

La détection des éléments saillants (c.-à-d., caractéristiques)¹ est une étape préliminaire pour beaucoup de tâches en vision par ordinateur. Par conséquent, différents détecteurs d'éléments saillants ont été proposés, qui diffèrent largement par le type d'éléments détectés. Ils peuvent être divisés en différentes catégories. Les détecteurs de régions covariantes surpassent les autres détecteurs en terme de répétabilité. Parmi les détecteurs existants, les détecteurs Harris-Affine, Hessian-Affine [1, 2] et MSER [3] sont les meilleurs en terme de répétabilité. Harris-Affine et Hessian-Affine sont deux méthodes similaires basées sur le principe du détecteur de coin de Harris-Stephen [4]. Les deux approches sont basées sur l'adaptation de forme affine (*affine shape adaptation*). Leurs algorithmes peuvent être résumés en deux étapes : 1- identification de la localisation spatiale initiale des régions à l'aide de détecteur de Scale-Invariant Harris-Laplace (pour Harris-Affine) et à base de la matrice Hessienne pour Hessian-Affine [5], 2- pour chaque point initial, une normalisation itérative à base de l'adaptation de forme affine est effectuée afin de rendre la région invariante par rapport à la transformation affine. La méthode MSER (Maximally Stable Extremal Regions) est basée sur une approche différente. Elle repose sur l'extraction de composantes connexes en utilisant un seuillage de l'image.

1. Il n'existe pas une définition standard de l'élément saillant. Souvent il dépend du contexte de problème ou du type d'application. Cependant, la définition générale est qu'un élément saillant est une partie saillante d'une image : les arêtes, coins/points d'intérêts, gouttes/régions d'intérêts, etc.

Les régions extrémales de MSER, sont constituées de tous pixels dont l'intensité est plus (ou moins) élevée que celle des pixels de son extérieur. Une autre méthode exploitant les arêtes présentes dans l'image a été proposée dans [6, 7]. Le concept d'EBR est que les arêtes sont des éléments caractéristiques stables et qu'elles peuvent être facilement détectées malgré de nombreuses transformations (échelle, angle de vue, luminosité, etc). Les auteurs de [6, 7] ont également proposé une autre approche, appelée IBR. L'algorithme IBR commence par la détection des extrêmes de l'intensité à plusieurs échelles. Le voisinage de chaque extrémum est ensuite inspecté de manière radicale afin de marquer le contour de la région. Les régions avec des formes arbitraires sont ensuite remplacées par des ellipses. Le problème avec les détecteurs cités ci-dessus est constaté par les auteurs de [2]. Ils ont effectué une comparaison complète entre ces méthodes. Cette comparaison a montré qu'il n'existe pas de détecteur qui surpasse absolument les autres détecteurs pour tous types de déformations. Par exemple, MSER surpasse Harris-Affine et Hessian-Affine par rapport aux changements d'angle de vue. En revanche, pour le changement d'échelle, il vient en second rang après le détecteur de Hessian-Affine. Hessian-Affine et Harris-Affine ont les scores les plus élevés de répétabilités par rapport au flou. Toutefois, MSER est le plus sensible à ce type de changement. Par conséquent, il semble utile de développer un détecteur de régions qui garantie explicitement une bonne robustesse par rapport à tous les types de déformation (échelle, angle de vue, illumination, etc). Ceci est l'objectif de notre contribution, qui généralise l'approche de Harris et Stephen [4] pour inclure non seulement la détection des points d'intérêts mais aussi la détection des régions, pour aboutir à une meilleure résistance aux différents types de déformations.

Notre approche est basée sur la détection des transformations locales qui ont des maxima locaux sur la fonction de saillance. La fonction de saillance est l'optimisation de ce que nous avons nommé le *warp-based autocorrelation*². Les régions d'images qui maximisent la fonction de saillance sont qualifiées de covariantes.

Contenu. §2 présente un aperçu sur la méthode de Harris et

2. *warp-based autocorrelation* ou **autocorrélation basée sur les transformations pour la détection de régions affines covariantes**

Stephens qui est la base de notre approche. §3 décrit notre approche pour la détection de régions affines covariantes. §4 présente des résultats expérimentaux.

Notation. $g(\cdot, \sigma)$ désigne une Gaussienne centrée de covariance σ . Dans la suite, σ_0 désigne une matrice de covariance isotrope. Le vecteur $\gamma = (s_c \ h_1 \ h_2)^\top$ est le vecteur de paramètres affines : l'échelle s_c , le cisaillement horizontal h_1 et le cisaillement vertical h_2 (le paramètre de rotation est omis car il est indéterminée par la transformation d'une ellipse vers un cercle). Les notations I_0 , I et \tilde{I} désignent l'image d'entrée (de référence), l'image transformée de la référence par $A(\gamma)$ et la version filtrée de l'image transformée avec la Gaussienne g respectivement. La matrice affine $A(\gamma)$ est définie comme la transformation locale d'une ellipse définie dans l'image de référence I_0 vers un cercle fixe C défini dans l'image transformée I . Sauf indication contraire, nous supposons que l'image est continue et nous ignorons les effets d'échantillonnage.

2 Travaux antérieurs

La méthode de Harris et Stephens [4] est une approche communément utilisée pour détecter les points d'intérêts dans l'image. Elle sert de base à beaucoup d'autres approches. Plusieurs de ces méthodes se basent sur la maximisation de la fonction d'autocorrélation locale proposée par Harris et Stephens.

Le principe de la méthode de Harris et Stephen est basé sur la maximisation de l'autocorrélation locale utilisant un décalage local à 2-dimensions d'une région (patch) pondérée isotropiquement. Ensuite, une recherche exhaustive à 2-dimensions est lancée sur la position afin de sélectionner les maxima locaux de l'autocorrélation locale. Considérons une région circulaire C dans l'image I , centrée sur le point $q \in \mathbb{R}^2$. La variation en intensité de pixel correspondante à une variation élémentaire δq autour de q dans C peut être écrite comme :

$$c(q; \delta q) = \int_C g(\tau; \sigma_0) \|I(\tau + q + \delta q) - I(\tau + q)\|^2 d\tau, \quad (1)$$

où τ est la coordonnée locale du pixel appartenant à la région circulaire centrée C . $g(\cdot; \sigma_0)$ définit une gaussienne circulaire isotrope dans C , centrée sur q et avec une covariance isotrope σ_0 . Pour la variation élémentaire δq autour de q , $I(\tau + q + \delta q)$ peut être approché avec une approximation de Gauss-Newton (c.-à-d., expansion de Taylor d'ordre 1) comme suit :

$$I(\tau + q + \delta q) \approx I(\tau + q) + \left(\frac{\partial I}{\partial q}(\tau + q, \cdot) \right)^\top \delta q. \quad (2)$$

Ceci mène à l'approximation :

$$c(q; \delta q) \approx \delta q^\top \left(\int_C g(\tau; \sigma_0) \nabla I(\tau + q) \nabla I^\top(\tau + q) d\tau \right) \delta q, \quad (3)$$

ce qui donne en notation matricielle :

$$c(q; \delta q) \approx \delta q^\top M(q) \delta q, \quad (4)$$

où $M(q)$ est la matrice d'autocorrélation :

$$M(q) = \int_C g(\tau; \sigma_0) \nabla I(\tau + q) \nabla I^\top(\tau + q) d\tau. \quad (5)$$

La fonction de saillance est définie par :

$$s : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad s(q) = \min_{\delta q \in \mathbb{R}^2} c(q; \delta q).$$

En se basant sur l'analyse spectrale, le minimum de $c(q, \delta q)$ sur δq sous la contrainte $\delta q^\top \delta q = 1$ est donné par la plus petite valeur propre de M : $s(q) = \lambda_{\min}(M(q))$. Enfin, les points saillants (c.-à-d., les points d'intérêts) \tilde{q} sont sélectionnés avec :

$$\tilde{q} = \arg \text{local max}_{q \in I} s(q). \quad (7)$$

3 Détection basée sur les transformations

Notre approche est basée sur ce que nous nommons la *warp-based autocorrelation*, ainsi définit :

$$c_I(q, \gamma; \delta q) = \int_C g(\tau; \sigma_0) \left\| \tilde{I}(q + \tau + A(\gamma)^{-1} \delta q) - \tilde{I}(q + \tau) \right\|^2 d\tau. \quad (8)$$

Nous rappelons que la matrice affine $A(\gamma)$ est définie comme la transformation locale d'une ellipse définie dans l'image de référence I_0 vers un cercle fixe C défini dans l'image transformée I et centrée en q . \tilde{I} est le résultat de la convolution de I avec $g(\cdot; \sigma_0)$:

$$\tilde{I}(q + \tau) = \int_C g(\alpha; \sigma_0) I(q + \tau + \alpha) d\alpha. \quad (9)$$

Dans sa définition générale proposée dans (8), la valeur de $c(q, \gamma; \delta q)$ est la variation de l'intensité de l'image au point q vis-à-vis d'une variation locale δq de la région circulaire isotrope définie par $g(\cdot; \sigma_0)$. L'image référence I_0 est reliée au point q de l'image transformée par le matrice affine A (dans le reste de l'article $A \equiv A(\gamma)$) :

$$I(q + \tau + \alpha) = I_0(A^{-1}q + A^{-1}\tau + A^{-1}\alpha). \quad (10)$$

La fonction optimisée de $c_I(q, \gamma; \delta q)$ sur δq mesure la saillance au point $(q, \gamma)^\top$, notée $s_I(q, \gamma)$:

$$s_I(q, \gamma) = \min_{\delta q \in \mathbb{R}^2} c_I(q, \gamma; \delta q). \quad (11)$$

L'idée sous-jacente de notre approche *warp-based autocorrelation*, est que si l'image I_0 est transformée localement au point \tilde{q}_0 par une transformation affine \tilde{A} , et si son image transformée I répond avec des maxima locaux au point $\tilde{q} = A\tilde{q}_0$ sur la fonction de saillance (11), alors le vecteur $(\tilde{q}, \tilde{\gamma})^\top$ correspond au point spatiale \tilde{q}_0 et à la transformation \tilde{A} est qualifiée *saillant*. Cependant, la région circulaire C , centrée au point $\tilde{q} = A\tilde{q}_0$ et définie dans I , est transformée en région elliptique E , centrée en \tilde{q}_0 et définie dans I_0 . Cette région est qualifiée de *covariante* :

$$E = \left\{ q_0 \in I_0 : (q_0 - \tilde{q}_0)^\top \tilde{A}^\top \tilde{A} (q_0 - \tilde{q}_0) \leq r^2 \right\}. \quad (12)$$

Ici, r est une constante correspondant au diamètre de C . Comme l'image donnée est I_0 , ceci nous amène à dériver l'expression de la fonction de saillance $s(q_0, \gamma)$ correspondante à l'image I_0 , c.-à-d., écrire $c(\cdot, \cdot; \cdot)$ en fonction de I_0 , q_0 , γ et δq_0 . Commençons à partir des équations (9) et (10), on peut écrire :

$$\tilde{I}(q + \tau) = \int_C g(\alpha; \sigma_0) I_0(A^{-1}q + A^{-1}\tau + A^{-1}\alpha) d\alpha. \quad (13)$$

La définition de la région circulaire c'est :

$$C = \{p \in I : (p - q)^\top (p - q) \leq r^2\}. \quad (14)$$

Le changement de variables $\alpha_0 = A^{-1}\alpha$ dans (13) conduit à :

$$\tilde{I}(q + \tau) = \int_E g(\alpha_0; \sigma) I_0(q_0 + A^{-1}\tau + \alpha_0) d\alpha_0. \quad (15)$$

Avec $q_0 = A^{-1}q$, $\sigma = A^{-1}\sigma_0 A^{-\top}$ est une covariance Gaussienne anisotrope et E une région elliptique définie dans I_0 :

$$E = \{p_0 \in I_0 : (p_0 - q_0)^\top A^\top A (p_0 - q_0) \leq r^2\}. \quad (16)$$

Le résultat obtenu dans (15) montre que la convolution d'une image transformée avec une Gaussienne isotrope et définie dans une région circulaire, est équivalente, à la convolution de l'image origine avec une Gaussienne transformée avec la même transformation, c.-à-d., une Gaussienne anisotrope définie sur une région elliptique.

Approximation de Gauss-Newton. L'approximation de Gauss-Newton, permet d'écrire :

$$\tilde{I}(q + \tau + A^{-1}\delta q) - \tilde{I}(q + \tau) = (\nabla \tilde{I}(q + \tau))^\top A^{-1}\delta q. \quad (17)$$

Ici $\nabla \tilde{I} = \frac{\partial \tilde{I}}{\partial q}$. En accord avec la propriété de commutativité de la convolution³, la dérivée peut être obtenue d'une façon équivalente, par la convolution de l'image originale avec la dérivée de la Gaussienne. Cela, conduit à :

$$\nabla \tilde{I}(q + \tau) = A^{-1}(\nabla g(\cdot; \sigma) * I_0)(q_0 + A^{-1}\tau). \quad (18)$$

Ici $*$ est l'opérateur de convolution. $\nabla g(\cdot; \sigma)$ est la dérivée de $g(\cdot; \sigma)$. La théorie liée à l'espace d'échelle et la dérivée normalisée de la Gaussienne, permet d'écrire :

$$\nabla_{\zeta_{-norm}} g(\cdot; \sigma) = \sigma^\zeta \nabla g(\cdot; \sigma). \quad (19)$$

Ici $\nabla_{\zeta_{-norm}}$ est l'opérateur de la dérivée normalisée. $\zeta \in [0, 1]$ est un paramètre lié à la dimensionnalité des éléments saillants dans l'image. Ainsi, la dérivée normalisée de l'image \tilde{I} est donnée par :

$$\nabla \tilde{I}(q + \tau) = A^{-1} \sigma^\zeta (\nabla g(\cdot; \sigma) * I_0)(q_0 + A^{-1}\tau). \quad (20)$$

3. $\nabla(f_a * f_b) = (\nabla f_a) * f_b = f_a * (\nabla f_b)$.

La substitution de l'expression de $\nabla \tilde{I}(q + \tau)$ obtenue à partir de (20) dans (8), conduit à écrire : $c_I(q, \gamma; \delta q) =$

$$\int_C g(\tau; \sigma) \left\| \left(A^{-1} \sigma^\zeta (\nabla g(\cdot; \sigma) * I_0)(q_0 + A^{-1}\tau) \right)^\top A^{-1} \delta q \right\|^2 d\tau. \quad (21)$$

En substituant τ dans (21) avec $\tau_0 = A^{-1}\tau$, on obtient que :

$$c_I(q, \gamma; \delta q) = \int_E g(\tau_0; \sigma) \left\| \left(A^{-1} \sigma^\zeta ((\nabla g(\cdot; \sigma)) * I_0)(q_0 + \tau_0) \right)^\top A^{-1} \delta q \right\|^2 d\tau_0. \quad (22)$$

Sachant que $\delta q_0 = A^{-1}\delta q$, la dernière intégrale (22) peut être réarrangée comme suit :

$$c_I(q, \gamma; \delta q) \approx \delta q_0^\top A^{-\top} M(q_0, \gamma) A^{-1} \delta q_0, \quad (23)$$

avec $M(q_0, \gamma) =$

$$\sigma^\zeta \left(g(\cdot; \sigma) * \left((\nabla g(\cdot; \sigma) * I_0) (\nabla g(\cdot; \sigma) * I_0)^\top \right) \right) (q_0) \sigma^\zeta^\top. \quad (24)$$

Par conséquent, la définition du *warp-based autocorrelation* en fonction de l'image originale I_0 , est donnée par :

$$c(q_0, \gamma; \delta q_0) \approx \delta q_0^\top A^{-\top} M(q_0, \gamma) A^{-1} \delta q_0. \quad (25)$$

En considérant que le principe de détection correspondant est :

$$(\tilde{q}_0, \tilde{\gamma}) = \arg \max_{(q_0, \gamma) \in I_0 \times \mathbb{R}^3} \text{local } s(q_0, \gamma), \quad (26)$$

$$\text{avec : } s(q_0, \gamma) = \min_{\delta q_0 \in \mathbb{R}^2} c(q_0, \gamma; \delta q_0),$$

l'expression de $c(q_0, \gamma; \delta q_0)$ est la forme quadratique correspondant à la matrice : $A^{-\top} M(q_0, \gamma) A^{-1}$. Comme dans le cas de Harris et Stephen's, la fonction de saillance est :

$$s(q_0, \gamma) = \lambda_{\min} \left(A^{-\top} M(q_0, \gamma) A^{-1} \right). \quad (27)$$

En fait, le critère approprié en terme de répétabilité, est établi ainsi :

$$s(q_0, \gamma) = \frac{\det(A^{-\top} M(q_0, \gamma) A^{-1})}{\text{trace}(A^{-\top} M(q_0, \gamma) A^{-1})}. \quad (28)$$

Ici, \det et trace , désignent le déterminant et la trace respectivement.

3.1 Covariance du warp-based autocorrelation

En considérant que I_0 et I'_0 sont deux images liées par la transformation affine $w(q, \gamma) = Bq + t$ tel que $I_0(q) = I'_0(Bq + t)$. La transformation affine $w(q, \gamma)$ transforme la région elliptique E définie dans I_0 (équation (16)) en région elliptique E' dans I'_0 , centrée en $q'_0 = Bq_0 + t$ et définie par la matrice affine $A' = AB^{-1}$:

$$E' = \{p_0 \in I'_0 : (p_0 - q'_0)^\top A'^\top A' (p_0 - q'_0) \leq r^2\}. \quad (29)$$

En utilisant les notations $s(q_0, A)$ et $c(q_0, A; \delta q_0)$ à la place de $s(q_0, \gamma)$ et $c(q_0, \gamma; \delta q_0)$ respectivement, ainsi la condition de covariance est définie par :

$$\begin{aligned} (\tilde{q}_0, \tilde{A}) &= \arg \text{local max}_{(q_0, A) \in I_0 \times \mathbb{R}^3} s(q_0, A) \iff \\ (\tilde{q}'_0, \tilde{A}') &= \arg \text{local max}_{(q'_0, A') \in I'_0 \times \mathbb{R}^3} s'(q'_0, A') \quad (30) \\ \text{tel que : } \tilde{q}'_0 &= B\tilde{q}_0 + t, \quad \tilde{A}' = \tilde{A}B^{-1}. \end{aligned}$$

$s(\cdot, \cdot)$ et $s'(\cdot, \cdot)$ sont les fonctions de saillances correspondantes à I et I' respectivement.

L'expression de $c(\cdot, \cdot, \cdot)$ correspondante à la fonction $s(\cdot, \cdot)$ est donnée par l'équation (25) :

$$c(q_0, A; \delta q_0) \approx \delta q_0^\top A^{-\top} M(q_0, A) A^{-1} \delta q_0. \quad (31)$$

Il est facile de montrer que $M(\cdot, \cdot)$ peut être écrite comme :

$$M(q_0, A) = B^\top M'(Bq_0 + t, AB^{-1})B \quad (32)$$

cela donne :

$$c(q_0, A; \delta q_0) \approx \delta q_0^\top A^{-\top} B^\top M'(Bq_0 + t, AB^{-1})BA^{-1} \delta q_0. \quad (33)$$

Après le réarrangement, on obtient :

$$c(q_0, A; \delta q_0) \approx \delta q_0^\top (AB^{-1})^{-\top} M'(Bq_0 + t, AB^{-1}) (AB^{-1})^{-1} \delta q_0. \quad (34)$$

Les équations $q'_0 = Bq_0 + t$ et $A' = AB^{-1}$ donnent :

$$c(q_0, A; \delta q_0) \approx \delta q_0^\top A'^{-\top} M'(q'_0, A') A'^{-1} \delta q_0. \quad (35)$$

En fin on obtient :

$$c(q_0, A; \delta q_0) \approx c'(q'_0, A'; \delta q_0), \quad (36)$$

ce qui implique

$$s(q_0, A) \approx s'(q'_0, A') \quad (37)$$

Ceci mène à l'approximation :

$$\begin{aligned} (\tilde{q}_0, \tilde{A}) &= \arg \text{local max}_{(q_0, A) \in I_0 \times \mathbb{R}^3} s(q_0, A) \iff \\ (\tilde{q}'_0, \tilde{A}') &= \arg \text{local max}_{(q'_0, A') \in I'_0 \times \mathbb{R}^3} s'(q'_0, A') \quad (38) \\ \text{avec : } \tilde{q}'_0 &\approx B\tilde{q}_0 + t, \quad \tilde{A}' \approx \tilde{A}B^{-1}. \end{aligned}$$

Nous concluons que la fonction *warp-based autocorrelation* est approximativement covariante vis à vis des transformations affines.

3.2 Paramétrisation par points de contrôle

Des paramètres avec différentes unités vont générer des matrices de transformation $A(\gamma)$ non normalisées, rendant à leur tour les matrices M , liées la fonction de *warp-based autocorrelation*, non normalisées. Notre approche pour résoudre ce problème utilise une paramétrisation à base de points de contrôle comme suit :

1. Les points d'entrées sont paramétrés comme suit :

$$m_0^i = \begin{pmatrix} x \\ y \end{pmatrix}, m_1^i = \begin{pmatrix} x + s_c + h_1 \\ y + h_2 \end{pmatrix}, m_2^i = \begin{pmatrix} x + h_2 \\ y + s_c - h_1 \end{pmatrix}, \quad (39)$$

2. Tandis que, les points de sorties sont fixes :

$$m_0^o = \begin{pmatrix} x \\ y \end{pmatrix}, m_1^o = \begin{pmatrix} x + r \\ y \end{pmatrix}, m_2^o = \begin{pmatrix} x \\ y + r \end{pmatrix}, \quad (40)$$

3. Notant M^i et M^o deux ensembles d'entrée et de sortie respectivement :

$$M^i = \begin{pmatrix} m_0^i & m_1^i & m_2^i \end{pmatrix}, M^o = \begin{pmatrix} m_0^o & m_1^o & m_2^o \end{pmatrix}. \quad (41)$$

La matrice de transformation A est calculée pour satisfaire la relation suivante : $M^o = AM^i(\gamma)$.

Ici, $q_0 = (x \ y)^\top$ sont les coordonnées spatiales d'un point dans l'image originale. r est une valeur fixée. En fait, la transformation affine $A(\gamma)$ est définie comme la transformation d'une ellipse vers un cercle fixe. L'avantage de cette approche est d'utiliser les mêmes unités (pixels) pour tous les paramètres affines $q_0 = (x \ y)^\top$ et $\gamma = (s_c \ h_1 \ h_2)^\top$. Cela conduit à une structure de tenseur M normalisée.

4 Résultats expérimentaux

Les expériences sont effectuées sur des séquences d'images réelles⁴, obtenues à partir de la base de données de [5, 2]. Trois différents types de détecteurs sont in-



Figure 1 – Quelques échantillons d'images de test. Colonne (1) : 2 images d'une séquence de changement d'échelle + rotation (1ère et 6ème séquences). Colonne (2) : 2 images d'une séquence de changement d'angle de vue (1ère et 6ème séquences). Colonne (3) : 2 images d'une séquence de changement d'illumination (1ère et 6ème séquences).

clus dans notre comparaison : Harris-Affine [1, 8], Hessian-Affine [1] et MSER [3]. Nous avons utilisé les implémentations originales des auteurs en ajustant les valeurs des paramètres par défaut recommandés par les auteurs. Les paramètres liés à notre détecteur comprennent : r le rayon du cercle prédéfini, σ_0 la variance de la Gaussienne isotrope $g(\cdot, \sigma_0)$, k la taille de l'espace affine échantillonné

4. <http://www.robots.ox.ac.uk/~vgg/research/affine/>
<https://lear.inrialpes.fr/people/Mikolajczyk/Database>

(dans lequel, la recherche exhaustive est effectuée, *e.g.*, nombre d'échelles à tester), th le seuil de saillance (comme pour le détecteur de coins de Harris) pour éliminer les régions instables. Les valeurs expérimentales sont :

$$r = 3, k = 100, \zeta = 0.25, \sigma_0 = 2, th = 0.1$$

4.1 Évaluation par l'erreur de chevauchement

Nous comparons les performances de notre détecteur à celles obtenues par les détecteurs ci-dessus. En fait, ces détecteurs (*i.e.*, Harris-Affine [1, 8], Hessian-Affine [1] et MSER [3]) outrepassent les autres approches existantes. La comparaison est basée sur la mesure de répétabilité [2]. La répétabilité pour une paire d'images est calculée comme le rapport entre le nombre de correspondances point-à-points et le nombre minimum de points détectés dans les deux images [5]. Deux régions sont considérées en correspondance si l'erreur de chevauchement liée à la surface commune couverte par les deux régions est $\epsilon < 0.4$:

$$\epsilon = 1 - \frac{R(E_1) \cap R(B^T E_2 B)}{R(E_1) \cup R(B^T E_2 B)}. \quad (42)$$

Ici $R(E_1)$ et $R(E_2)$ sont deux régions elliptiques définies par : $q^T E_1 q = 1$ et $q^T E_2 q = 1$ respectivement, \cap et \cup correspondent aux opérateurs d'intersection et d'union respectivement. B est la transformation affine entre les deux images.

4.2 Répétabilité

Pour tous les détecteurs inclus (sauf indication contraire), nous avons fixé le seuil de l'erreur de chevauchement à 40%. La taille normalisée des régions est ajustée à 30 pixels [2]. Nous évaluons la répétabilité de chaque détecteur vis à vis des changements d'échelle, changement d'angle de vue, changement d'éclairage et changement de flou. Des échantillons de séquences d'images utilisées sont affichées dans la figure 1. Le test de répétabilité permet d'évaluer comment le nombre de correspondances change entre l'image de référence et la séquence, en fonction du changement des transformations. Pour chaque image de la séquence, le nombre relatif et total des régions détectées sont enregistrées. Le détecteur idéal a un score de répétabilité élevé et un grand nombre de correspondances.

Changement d'échelle. Les résultats de ces évaluations sont reportés dans les figures 2(a) et 2(b). Bien que simple (notre algorithme est effectué en une seule étape), notre détecteur dispose d'un score de répétabilité meilleurs que ceux obtenus par les meilleures approches existantes. La figure 2(b) montre que notre détecteur a un nombre de correspondances supérieur à celui obtenu avec MSER, et inférieur à ceux obtenus avec Harris-Affine et Hessian-Affine. Cela signifie que nos régions détectées sont plus stables que celles obtenues par les autres détecteurs. Généralement, cela est lié à la propriété de covariance de la fonction *warp-based autocorrelation* par rapport aux transformations affines.

Changement d'angle de vue. Les figures 2(c) et 2(d) montrent les résultats de répétabilité et du nombre de correspondances en fonction des transformations liées au changement d'angle de vue. Le meilleur taux de répétabilité est obtenu avec le détecteur MSER suivi de notre détecteur (figure 2(c)). Cela est dû à la précision élevée de MSER, en particulier pour les régions homogènes avec des limites distinctives [2]. Le plus grand nombre de régions correspondantes est donné par Hessian-Affine (≈ 3000) suivi par Harris-Affine (≈ 2700) (voir Fig. 2(d)). Notre détecteur donne ≈ 500 régions et MSER est dernier avec ≈ 250 régions.

Changement d'éclairage. Les figures 2(e) et 2(f) affichent les résultats liés au changement d'éclairage. Il est facile de voir que notre détecteur et MSER obtiennent les taux les plus élevés de répétabilité. Les courbes de répétabilité liées à notre détecteur et MSER sont approximativement horizontales, soit un taux de répétabilité presque constante. Ceci montre qu'ils ont une bonne robustesse au changement d'éclairage. Le nombre de régions détectées avec notre détecteur est beaucoup plus grand que ceux obtenus avec les autres détecteurs (voir Fig. 2(f)). Cela montre clairement que notre détecteur répond à différents types d'éléments saillants, à savoir coins, tâches, etc.

Changement de flou. Les figures 2(g) et 2(h) montrent les résultats liés au changement croissant de flou. Notre détecteur est le plus performant, suivi par Hessian-Affine et Harris-Affine. La courbe de répétabilité pour notre détecteur est presque horizontale (figure 2(g)). Cela prouve que notre détecteur à un très haut niveau de robustesse par rapport à la variation de flou. Le détecteur MSER est nettement le plus sensible au flou parce que les frontières des régions deviennent lisses, et le processus de segmentation est moins précis [2]. Le Hessian-Affine donne le plus grand nombre de régions suivies par notre détecteur. Le nombre de régions détectées avec MSER est très faible, aussi parce que les frontières de régions deviennent lisses.

La figure 3 montre des régions générées par, notre détecteur WBA, Hessian-Affine et MSER sur des sous-parties en correspondances de la première et la quatrième (angle de vue 40°) images de la séquence graffiti (mur), figure 1(c.-à-d., changement d'angle de vue). Les régions détectées sont en jaune et les régions correspondantes projetées à partir de la référence sont en bleu. Les ellipses sont à un facteur de 0.5 de la taille originale détectée.

5 Conclusion

Dans cet article, nous avons introduit une approche simple pour détecter les régions affines covariantes dans les images. Elle est basée sur ce que nous appelons la fonction *warp-based autocorrelation*. La fonction *warp-based autocorrelation* est covariante par rapport aux transformations affines des régions. En combinant la fonction *warp-based autocorrelation* avec un paramétrage à base de points de contrôle, on obtient un nouveau détecteur, répétable et précis vis à vis des transformations affines. Notre méthode est simple et efficace. Elle surpasse nettement les méthodes

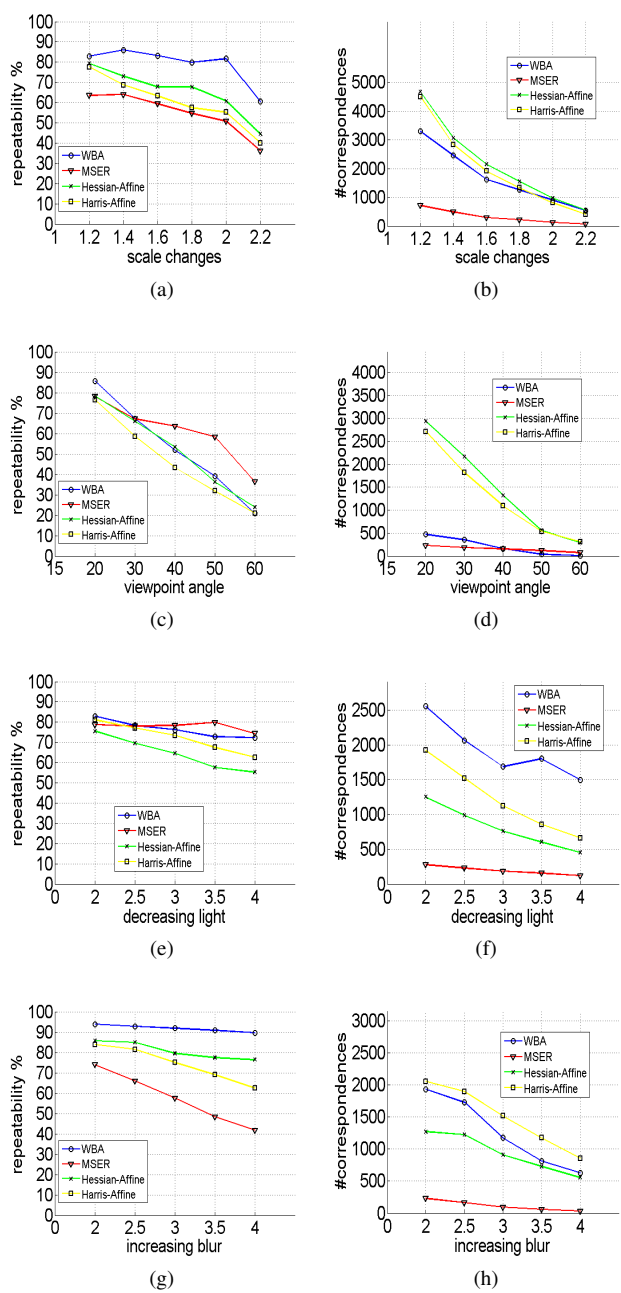


Figure 2 – Répétabilité et nombres de régions détectées (erreur de chevauchement = 40%, taille normalisée = 30 pixels) par rapport au changement d'échelle+rotation (la séquence Bateau, Fig. 1), changement d'angle de vue (la séquence Mur, Fig. 1), changement d'éclairage (la séquence Leuven, Fig. 1) et changement de flou (la séquence Motos). (a) Taux de répétabilité par rapport au changement d'échelle. (b) Nombre de régions détectées par rapport au changement d'échelle. (c) Taux de répétabilité par rapport au changement d'angle de vue. (d) Nombre de régions détectées par rapport au changement d'angle de vue. (e) Taux de répétabilité par rapport au changement d'éclairage. (f) Nombre de régions détectées par rapport au changement d'éclairage. (g) Taux de répétabilité par rapport au changement de flou. (h) Nombre de régions détectées par rapport au changement de flou.

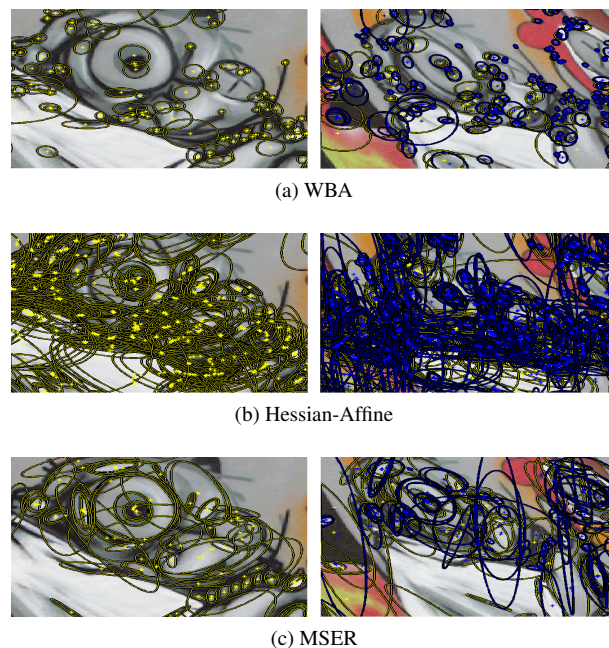


Figure 3 – Exemple de régions détectées par : notre détecteur WBA, Hessian-Affine et MSER sur des sous-parties en correspondances de la 1ère (la référence) et la 4ème (angle de vue 40°) images de la séquence graffiti (mur), figure 1 (c.-à-d., changement d'angle de vue). Les régions détectées sont en jaune et les régions correspondantes projetées à partir de la référence sont en bleu. Les ellipses sont à un facteur de 0.5 de la taille originale détectée.

existantes. Nos expériences ont montré l'efficacité de notre approche. Dans l'avenir, nous allons étendre notre approche vers les images 3D.

Références

- [1] K. Mikolajczyk et C. Schmid. An affine invariant interest point detector. *ECCV*, 2002.
- [2] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, et L.V. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1) :43–72, 2005.
- [3] J. Matas, O. Chum, M. Urban, et T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10) :761–767, 2004.
- [4] C. Harris et M. Stephens. A combined corner and edge detector. *Alvey Vision Conference*, 1988.
- [5] K. Mikolajczyk et C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1) :63–86, 2004.
- [6] T. Tuytelaars et L. Van Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1) :61–85, 2004.
- [7] T. Tuytelaars et L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. *BMVC*, 2002.
- [8] F. Schaffalitzky et A. Zisserman. Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". *ECCV*, 2002.