

Evaluation de détecteurs de points d'intérêt pour la couleur

Evaluation of point of interest detectors for color images

Valérie Gouet¹

Philippe Montesinos¹

Rachid Deriche²

Danielle Pelé³

¹ Laboratoire de Génie Informatique et Ingénierie de Production

² Institut National de Recherche en Informatique et en Automatique

³ Centre Commun d'Etudes des Télécommunications - France Télécom

¹ EMA/LGI2P - Site EERIE - Parc scientifique Georges Besse - 30035 Nîmes FRANCE

² INRIA - route des Lucioles - 06560 Sophia-Antipolis Cedex FRANCE

³ CCETT - 4 rue du Clos Courtel, BP 59 - 35512 Cesson Sévigné - Rennes FRANCE

{gouet,montesin}@site-eerie.ema.fr , pele@ccett.fr , der@sophia.inria.fr

Résumé

Les travaux présentés dans cet article portent sur la mise en place de deux nouveaux détecteurs de points d'intérêt spécifiques à la couleur. Ces deux détecteurs sont basés sur des approches différentes et travaillent à l'ordre 1 pour l'un et à l'ordre 2 pour l'autre. Pour évaluer ces détecteurs, nous définissons deux critères : la répétabilité et la localisation. L'évaluation est aussi réalisée sur un détecteur de points en niveau de gris qui a fait ses preuves et qui sert donc de référence. Les mesures sont calculées sur plusieurs séquences d'images différant des principales transformations de l'image, comme la rotation 2D, le changement de luminosité et le changement de points de vue. Les résultats obtenus montrent l'apport de l'information couleur dans la détection de points d'intérêt, puisque les détecteurs mis en place ici obtiennent une répétabilité et une localisation la plupart du temps supérieures aux mesures sur le détecteur en niveau de gris. Une comparaison est également réalisée entre les deux détecteurs couleur. Leurs avantages et inconvénients les rendent plus ou moins adaptés à certains types d'applications.

Mots Clef

Images couleur, Détecteurs de points d'intérêt, Répétabilité, Localisation.

Abstract

We present in this paper two new point of interest detectors for color images. These detectors base its argument on different approaches since they work till first order for the first one and till second order for the second one. We define two criteria to value them : the repeatability and the location. We have too to estimate these criteria with a reference detector in gray value images to explain the results obtained. The measurements are realized on images differing from usual image transformation, like 2D rotation, changes of intensity and changes of viewpoint. The results obtained

show the pertinence of our color approach, since the repeatability and the location are most of the time better for the color detectors. The two color detectors are also compared to each other. Their employment depends on the applications considered.

Keywords

Color images, Point of interest detectors, Repeatability, Point location.

1 Introduction

Les travaux menés dans cet article portent sur la mise en place de détecteurs de points d'intérêt spécifiques à la couleur, en vue de réaliser la mise en correspondance d'images stéréoscopiques couleur et non calibrées. Dans nos précédents travaux sur la couleur [12] [8] [7], nous avons mis en place une caractérisation locale des points d'intérêt basée sur les invariants différentiels de Hilbert [10, 11, 6, 14, 13] connus en niveau de gris et qui exploite l'information couleur. La description ainsi obtenue est tout aussi riche que celle en niveau de gris, mais beaucoup plus robuste, car elle n'implique que les dérivées de l'image du premier ordre. L'information couleur s'est avérée très enrichissante ici. Nous avons alors décidé d'exploiter cette donnée supplémentaire pour définir un détecteur de points spécifique à la couleur, sachant qu'à notre connaissance il n'en existe pas pour ce type d'image. Le détecteur mis en place correspond à la généralisation multi-spectrale de la courbure des lignes d'iso-intensité bien connue en niveau de gris. Ce détecteur étant d'ordre 2, il nous aussi a semblé intéressant de proposer un détecteur de points d'intérêt couleur du premier ordre, puisque notre caractérisation locale des points travaille uniquement à l'ordre 1. Cet opérateur correspond à une généralisation du détecteur de Harris [9] aux images couleur. Il est donc ensuite nécessaire de mettre en place une méthode d'évaluation. Le détecteur de Harris précis ayant fait ses preuves

en niveau de gris, nous lui comparons nos deux détecteurs couleur, face aux transformations que nous autorise notre caractérisation locale, à savoir les transformations orthogonales de l'image et les changements affines de luminosité.

Ce papier est décomposé en trois parties. Dans la première section, nous rappelons brièvement notre méthode de caractérisation basée sur les invariants différentiels et sur l'information couleur. Puis nous continuons à exploiter la couleur pour mettre en place dans la section 3 deux détecteurs de points d'intérêt couleur. Ces détecteurs sont ensuite évalués dans la section 4, au travers de différentes transformations de l'image. Enfin nous discutons de leurs avantages et inconvénients dans la conclusion.

2 Caractérisation locale en couleur

Parmi les approches iconiques de mise en correspondance, la corrélation produit d'excellents résultats mais s'avère très coûteuse en temps de calcul et surtout très restrictive quant aux transformations autorisées entre images. Au contraire, les méthodes basées sur les invariants différentiels sont beaucoup plus robustes face aux transformations orthogonales de l'image, et moins coûteuses en temps. Malheureusement, il est souvent nécessaire de considérer les invariants jusqu'à l'ordre 3 [11, 14, 15] pour obtenir une caractérisation suffisante, et ce bien sûr aux dépens de la stabilité des invariants calculés. Nous avons cependant choisi ce mode de caractérisation pour sa robustesse face aux transformations de l'image, et nous avons eu l'idée d'utiliser l'information couleur pour éliminer son principal inconvénient. Une introduction à la théorie des invariants de Hilbert est présentée dans la section 2.1. La section 2.2 démontre les intérêts de notre approche couleur. Enfin nous rendons cette caractérisation invariante aux changements affines de luminosité dans la section 2.3.

2.1 Les invariants différentiels de Hilbert

Comme Hilbert l'a montré dans [10], n'importe quel invariant d'ordre fini peut être exprimé comme un polynôme d'invariants irréductibles. Cet ensemble est bien connu pour ses propriétés du premier et du second ordre [5, 15], et est rendu indépendant de la rotation image s'il est exprimé en coordonnées de *Gauge*, comme suit: $L, L_\eta, L_{\eta\eta}, L_{\xi\eta}, L_{\xi\xi}$ où η est le vecteur unitaire tel que $\eta = \frac{\nabla I}{|\nabla I|}$ et $\xi \perp \eta$. Dans le cadre de la mise en correspondance d'images noir et blanc, il est préférable de considérer d'un point de vue géométrique et/ou numérique une combinaison de ces cinq invariants, comme par exemple :

$$I = I_{\eta\eta} + I_{\xi\xi} = I_\eta \frac{I_{\xi\xi}}{I_\eta} + I_\eta \frac{I_{\xi\eta}}{I_\eta}$$

Considérons maintenant une image couleur $\{R, V, B\}$ ainsi que le groupe des rotations spécifié par un pa-

ramètre (l'angle de la rotation). Les invariants différentiels du premier et du second ordre sont donc au nombre de $6 \times 3 \pm 1 = 17$. Ceci correspond à nos cinq invariants par plan couleur et donc à deux invariants supplémentaires spécifiques à la couleur et pris dans l'ensemble suivant :

$$\nabla R \cdot \nabla G \quad \nabla R \cdot \nabla B \quad \nabla G \cdot \nabla B$$

Remarquons que ces derniers invariants n'impliquent que les dérivées du premier ordre.

2.2 Une caractérisation couleur robuste

Notre approche consiste à n'utiliser que les invariants de Hilbert à l'ordre 1. La caractérisation qui serait alors obtenue pour des images noir et blanc serait inutilisable, puisque réduite à deux invariants. Par contre l'approche couleur permet d'obtenir un ensemble de $2 \times 3 + 2 = 8$ invariants, que nous appelons \vec{v}_{col} .

$$\vec{v}_{col}(\vec{x}, \sigma) = \left(R, |\nabla R|^2, V, |\nabla V|^2, B, |\nabla B|^2, \nabla R \cdot \nabla V, \nabla R \cdot \nabla B \right) \quad (1)$$

Cette simplification que nous permet la couleur présente des avantages certains: la caractérisation obtenue est robuste au bruit et beaucoup moins complexe à implémenter, puisqu'elle n'implique que les dérivées de l'image d'ordre 1. Notons enfin qu'il aurait fallu aller jusqu'à l'ordre 3 en noir et blanc pour obtenir un vecteur d'invariants de dimension similaire (9 invariants exactement). Nous montrons que l'invariant "perdu" est largement compensé par la robustesse de la méthode. Pour plus de détails et des résultats comparatifs concernant cette caractérisation, le lecteur peut consulter [12].

2.3 Changements affines de luminosité

Notre modèle de caractérisation est invariant aux transformations orthogonales de l'image. Pour rendre la caractérisation encore plus robuste, nous avons choisi de la rendre invariante aux changements de luminosité. En premier lieu, il est nécessaire de modéliser ce type de transformation. Conformément aux récents travaux de Finlayson sur la constance des couleurs [4], nous utilisons le modèle diagonal avec un vecteur de translation additionnel. Nous obtenons le modèle à six degrés de liberté suivant :

$$\vec{x}' = \mathcal{D}\vec{x} + \mathcal{T} \quad (2)$$

où $\vec{x} = (r, v, b)^T$ représente le pixel couleur et $\vec{x}' = (r', g', v')^T$ sa transformation. La modélisation obtenue correspond à un changement affine de luminosité. Des modèles plus complexes existent mais celui-ci détient le meilleur ratio qualité/complexité. Ce modèle possède six degrés de liberté, donc rendre notre caractérisation \vec{v}_{rot} invariante à ce type de transformation reviendrait à ne considérer plus que $8 \pm 6 = 2$ invariants, ce qui bien sûr représente une caractérisation trop pauvre. Nous préférons donc procéder à un pré-traitement de l'image, qui consiste à normaliser chacun des pixels. Pour chaque plan couleur, chacun des niveaux de gris est recalculé dans un intervalle donné, $[0..1]$ par exemple. Ce processus rend l'image

indépendante des paramètres du modèle et permet donc d'appliquer notre caractérisation. Ajoutons enfin que pour préserver les propriétés locales des pixels, il est nécessaire d'implanter la méthode localement dans une fenêtre circulaire centrée sur le pixel à normaliser. Une description complète de la méthode peut être consultée dans [8].

Nous venons de mettre en place une caractérisation de points robuste. Il reste donc maintenant à déterminer un détecteur de points d'intérêt. Au vu des résultats obtenus pour la caractérisation à l'aide de l'information couleur, nous avons choisi de continuer dans cette voie en mettant en place deux détecteurs de points d'intérêt couleur.

3 Extraction de points d'intérêt

A notre connaissance il n'existe aucun détecteur spécifique à la couleur. Nous présentons donc dans cette section deux nouveaux détecteurs qui s'appuient sur l'information couleur $\{R, V, B\}$. Le premier détecteur est présenté dans la section 3.1 ; c'est une adaptation du détecteur de Harris et Stephen précis à la couleur. Puis dans la section 3.2, nous présentons une généralisation au cas multi-spectral de la courbure des lignes d'iso-intensité ou opérateur de Kitchen et Rosenfeld.

3.1 Le détecteur Harris couleur

Le détecteur de Harris et Stephen [9] dans sa version précise [15] semble particulièrement adapté à la mise en correspondance de points. En effet, il est démontré dans [1] que c'est le détecteur de points le plus répétable en noir et blanc face aux transformations usuelles de l'image. Sa stabilité provient principalement du fait que son implantation n'utilise que les dérivées de l'image d'ordre 1. Cette particularité s'avère d'ailleurs fort intéressante dans le cadre de notre étude puisqu'elle est en parfait accord avec la caractérisation présentée à la section précédente. Puisque nous travaillons avec des images couleur, notre contribution a donc consisté à introduire l'information couleur dans le détecteur du premier ordre déjà existant. Soit la matrice M :

$$M = \begin{bmatrix} S_\sigma(R_x^2 + G_x^2 + B_x^2) & S_\sigma(R_x R_y + G_x G_y + B_x B_y) \\ S_\sigma(R_x R_y + G_x G_y + B_x B_y) & S_\sigma(R_y^2 + G_y^2 + B_y^2) \end{bmatrix}$$

où $S_\sigma()$ représente un lissage gaussien. Le détecteur est alors défini en considérant les maxima locaux de l'opérateur suivant :

$$Det(M) \perp K Trace^2(M) \text{ avec } K = 0.04$$

La figure 1 montre un exemple des points obtenus en utilisant ce détecteur. Le filtre utilisé est un filtre gaussien sub-pixel. Il est à noter que les points détectés sont des coins pour la plupart mais pas exclusivement.

3.2 Un détecteur de coins couleur

De la même façon que l'opérateur de Di-Zenzo généralise le gradient d'une image scalaire au cas multi-spectral, nous allons ici généraliser la courbure des

lignes d'iso-intensité définie jusqu'à présent en niveau de gris. Considérons sur le plan image une courbe dont la tangente présente avec l'axe des x l'orientation θ_τ (au sens de Di-Zenzo). La courbure associée CMS peut être exprimée comme suit :

$$CMS = \frac{d\theta_\tau}{ds}$$

Elle peut être décomposée en fonction de l'orientation θ du gradient multi-spectral de la manière suivante :

$$CMS = \frac{\partial\theta}{\partial x} \frac{dx}{ds} + \frac{\partial\theta}{\partial y} \frac{dy}{ds}$$

avec :

$$\frac{dx}{ds} = -\sin\theta \quad \text{et} \quad \frac{dy}{ds} = \cos\theta$$

A titre informatif, nous donnons à l'équation 3 l'expression de $\frac{\partial\theta}{\partial x}$.

$$\frac{\partial\theta}{\partial x} = \frac{\left(\begin{aligned} &AB_x B_{xy} + BB_{xx} B_y - 2G_x G_{xy} G_y^2 + 2B_x^2 B_{xx} B_y + \\ &2B_x B_{xy} B_y^2 - 2B_x B_y R_x R_{xx} + AR_x R_{xy} - 2B_x B_{xx} R_x R_y + \\ &2B_{xy} B_y R_x R_y + AR_{xx} R_y - 2R_x^2 R_{xx} R_y + 2B_x B_y R_{xy} R_y + \\ &2R_x R_{xy} R_y^2 - 2B_x B_y G_x X_{xx} - 2R_x R_y G_x X_{xx} + AG_x G_{xy} - \\ &2B_x B_{xx} G_x G_y + 2R_x R_{xx} G_x G_y + 2B_{xy} B_y G_x G_y - AR_{xx} G_y - \\ &2G_x^2 X_{xx} G_y + 2R_{xy} R_y G_x G_y + 2B_x B_y G_{xy} G_y + 2R_x R_y G_{xy} G_y \end{aligned} \right)}{A^2 + 4(B_x B_y + R_x R_y + G_x G_y)^2} \quad (3)$$

où $A = B_x^2 - B_y^2 + R_x^2 - R_y^2 + G_x^2 - G_y^2$.

La détection des coins couleur se fait donc selon le processus suivant :

- Extraction des contours couleur à l'aide de l'opérateur de Di-Zenzo (en maximisant d'abord la plus grande valeur propre λ_{max} du tenseur multi-spectral selon la direction du vecteur propre associé, puis en procédant à un seuillage par hystérésis),
- Calcul de la courbure couleur $K_{couleur} = CMS \times \lambda_{max}$,
- Extraction des points ayant la plus forte courbure en maximisant $K_{couleur}$ dans la direction de θ : K_{max} ,
- Recherche des points de contour parmi K_{max} ,
- Enfin les coins couleur sont déterminés par seuillage parmi les maxima locaux des points trouvés.

La figure 2 montre un exemple de coins couleur détectés (détection sub-pixel).

Il aurait aussi été possible de définir l'orientation des lignes d'iso-intensité en considérant une fonction scalaire des courbures dans chacun des plans, comme par exemple :

$$f(K_R, K_V, K_B) = Max\{K_R, K_V, K_B\}$$

Mais compte tenu du peu de stabilité de l'opérateur Max pour la détection de contours en précision sub-pixel, nous lui avons préféré l'opérateur CMS .

En comparant les figures 1 et 2, nous pouvons constater de visu que les deux détecteurs mis en oeuvre ici

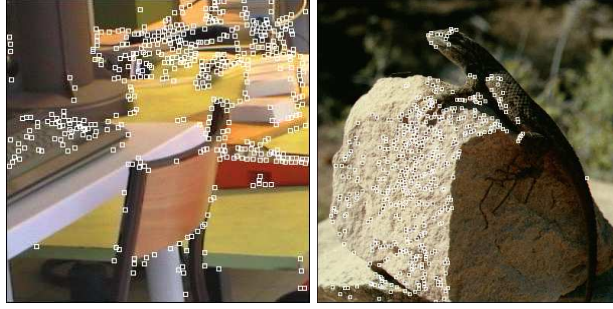


FIG. 1 – Un exemple de points Harris couleur sur deux images : l'image “Lézard” est caractérisée par beaucoup de texture (472 points détectés) alors que l'image “Ordi” présente un grand nombre de contours (493 points détectés).



FIG. 2 – Un exemple de coins couleur sur les mêmes images qu'à la figure 1. 457 coins sont détectés sur “Ordi” et 438 sur “Lézard”.

sont plus ou moins adaptés à certains types d'images. Ainsi le détecteur de coins couleur sera plus approprié pour des images contenant beaucoup de contours comme “Ordi”. Le détecteur Harris couleur quant à lui détecte un point là où le signal est riche en information, ce qui ne limite pas la détection aux coins, comme on peut notamment le voir avec l'image “Lézard” qui contient beaucoup de textures. Après une première évaluation rapide et visuelle des résultats, nous allons mettre en oeuvre dans la section suivante des critères d'évaluation plus rigoureux et spécifiques aux types d'applications envisagées.

4 Évaluation des détecteurs couleur

Il existe beaucoup de travaux sur les détecteurs de points d'intérêt, la méthode d'évaluation la plus répandue étant de vérifier visuellement la qualité des résultats. Ceci fournit une évaluation totalement subjective dans la mesure où l'observateur compare le résultat avec ce qu'il estime comme étant un point d'intérêt et non pas avec ce qui constitue le lieu où le signal est riche en information. En outre, l'évaluation visuelle ne peut pas fournir d'appréciation exploitable quant à la précision de la détection.

D'une manière générale, un détecteur est caractérisé par sa *répétabilité* et sa *localisation*. La répétabilité signifie qu'un même point est détecté dans une suite

d'images. La localisation, par contre, signifie qu'un point détecté dans l'image correspond de façon précise à un point de la scène tridimensionnelle. Plusieurs travaux dont [3] ont mis en évidence que le lissage améliorerait la répétabilité du détecteur mais en revanche détériorait sa localisation. En outre, les critères d'évaluation varient selon l'application envisagée. Une bonne répétabilité sera importante dans des applications comme l'appariement, l'estimation de la géométrie épipolaire ou encore l'indexation. En revanche, pour la reconstruction 3D ou le calibrage, plus généralement pour toutes les applications qui font intervenir la 3D, il faudra en plus tenir compte de la localisation du détecteur.

Nous présentons dans les deux sections suivantes les solutions que nous avons adoptées pour ces deux critères d'évaluation. Une méthode de mesure de ces critères est ensuite détaillée dans la section 4.3, puis mise en oeuvre dans la section 4.4 sur les différents détecteurs introduits.

4.1 Mesure de la répétabilité

Quelques travaux ont été réalisés sur la répétabilité des détecteurs de points. On peut citer notamment [15] qui évalue plusieurs détecteurs en niveau de gris en tenant compte d'une mesure de répétabilité. Soient deux images I_1 et I_2 d'une même scène 3D et M_1 et M_2 les matrices de projection correspondantes. La détection des points images p_1 et p_2 appartenant respectivement à I_1 et I_2 est parfaitement répétable si et seulement si il existe un point P de la scène tel que :

$$p_1 = M_1 P \quad \text{et} \quad p_2 = M_2 P$$

Pour mesurer la répétabilité d'un détecteur, une solution consiste à établir une relation entre p_1 et p_2 . Il n'en existe pas dans le cas général, mais si I_1 et I_2 sont les images d'une scène plane, alors il existe une homographie H_{12} les reliant. Dans le cas d'une répétabilité parfaite, elle est donnée par la relation suivante :

$$p_2 = H_{12} p_1 \quad (4)$$

Dans des conditions réelles, un point détecté n'a pas nécessairement d'homologue dans l'autre image. Il convient donc de calculer la répétabilité sur les parties communes des images, en éliminant du calcul les points dont l'image par H_{12} n'appartient pas à l'image correspondante. En outre, le support de l'image étant discret et la répétabilité d'un détecteur pouvant ne pas être parfaite, l'équation 4 n'a guère de réalité physique. Dans la pratique, nous considérerons donc la répétabilité du détecteur dans un certain voisinage. Pour établir la mesure de répétabilité du détecteur, nous allons donc étudier la distance qui sépare l'image par H_{ij} d'un point détecté dans I_i des points détectés dans I_j .

Soit $C_{ij}(p_k)$ le point de I_j en correspondance homographique avec le point détecté p_k de I_i . C'est le pixel le plus proche de son correspondant exact $H_{ij}.p_k$. Nous lui associons la mesure $D_{ij}(p_k)$ qui traduit la distance euclidienne entre ce correspondant prédit $C_{ij}(p_k)$ et le point le plus proche parmi tous les points détectés p_l de I_j :

$$D_{ij}(p_k) = \min_l \{ \text{dist}(C_{ij}(p_k), p_l) \} \quad (5)$$

La mesure ci-dessus a l'inconvénient de ne pas distinguer les points dont le correspondant dans l'autre image est proche d'un point détecté, des points qui n'ont pas de correspondant. On calcule en effet pour ces derniers une distance relativement grande qui n'a aucun intérêt et qui fait augmenter la moyenne donc dégrader la qualité du détecteur testé. Il faut donc être capable de déterminer si un point correspond à un point de l'autre image en tenant compte de l'erreur de localisation générée par le détecteur, ou bien s'il n'a aucun correspondant dans l'autre image. Ceci est fait en considérant uniquement les points situés en dessous d'un voisinage de rayon ε . Pour une distance supérieure ou égale à ε , le point est considéré comme n'ayant pas de correspondant. La distance définie à l'équation 5 est adaptée pour prendre en compte ce critère :

$$D_{ij}^\varepsilon(p_k) = \begin{cases} D_{ij}(p_k) & \text{si } D_{ij}(p_k) < \varepsilon \\ \varepsilon & \text{sinon} \end{cases} \quad (6)$$

Si N_i est le nombre de points détectés dans l'image I_i ayant un correspondant potentiel dans I_j et n_{ij} le nombre de points effectivement répétés dans le voisinage ε (avec $n_{ij} \leq N_i$), alors nous pouvons définir la mesure de répétabilité R_{ij}^ε de l'image I_i vers l'image I_j telle que :

$$R_{ij}^\varepsilon = \frac{\text{moy}_k(D_{ij}^\varepsilon(p_k))}{\varepsilon.(n_{ij} + 1)} \quad (7)$$

Afin d'obtenir une mesure symétrique entre les deux images I_i et I_j quel que soit le nombre de points détectés, nous définissons enfin la mesure de répétabilité R^ε telle que :

$$R^\varepsilon = \text{moy}(R_{ij}^\varepsilon, R_{ji}^\varepsilon) \quad (8)$$

Notons quelques aspects de la mesure de répétabilité R^ε mise en place ici :

- Tenir compte d'une distance minimale dans le calcul de $D_{ij}^\varepsilon(p_k)$ permet de mettre en valeur le détecteur le plus précis parmi plusieurs détecteurs possédant une bonne répétabilité dans un voisinage ε donné.
- Affecter aux $(N_i \perp n_{ij})$ points non répétés la distance ε (cf. l'équation 6) permet de tenir compte du pourcentage de points répétés par rapport au nombre total de points testés.
- Pondérer le critère R_{ij}^ε par n_{ij} privilégie un détecteur très répétable mais imprécis, à un détecteur peu répétable mais précis.
- Le choix de rendre symétrique le critère R_{ij}^ε le rend plus fiable dans le cas d'un système mal conditionné, lorsque $N_i \ll N_j$ par exemple.
- Le coefficient $\frac{1}{\varepsilon}$ de l'équation 7 permet de normaliser la mesure entre 0 et 1, le détecteur ayant la plus petite mesure de répétabilité étant considéré comme le meilleur.

Signalons enfin que le choix du voisinage ε dépend de la précision choisie pour la détection. En effet, si α est la précision sub-pixel utilisée, alors pour obtenir une mesure de répétabilité dans un voisinage de i pixels (en connexité-8), il est nécessaire de respecter la contrainte suivante :

$$\alpha i \sqrt{2} \leq \varepsilon < \alpha (i + 1) \sqrt{2} \quad (9)$$

Dans le cadre de notre étude, les détecteurs seront implémentés en précision demi-pixel, soit $\alpha = \frac{1}{2}$. Si l'on souhaite chercher le point détecté parmi les voisins directs du pixel prédit, il faudra alors choisir ε dans l'intervalle $[\frac{\sqrt{2}}{2}, \sqrt{2}]$.

4.2 Mesure de la localisation

La mesure de localisation d'un détecteur consiste à vérifier si les points détectés ont une réalité physique dans la scène tridimensionnelle. La qualité de cette mesure est par conséquent très importante dans les applications liées à la reconstruction 3D ou encore au calibrage de caméras. Plusieurs travaux ont étudié la localisation de détecteurs en niveau de gris. Citons notamment [2] qui utilise l'alignement, la reconstruction 3D et l'estimation de la géométrie épipolaire pour établir une mesure de localisation. Dans ce papier, nous avons choisi de nous baser sur deux critères de localisation dont l'importance dépendra des applications envisagées : le critère de *reconstruction projective* et celui de *reconstruction euclidienne*.

La reconstruction projective. Supposons connue la position exacte d'un ensemble de points P de la scène 3D. Soient P_{proj} l'ensemble des projections (en coordonnées réelles et non discrètes pour une plus grande précision) de P sur le plan rétinien d'une caméra, et P_{det} l'ensemble des points détectés sur l'image de la scène vue par la caméra. Il suffit alors de rechercher pour chaque point détecté P_{det}^i le point projeté P_{proj}^j le plus proche, puis de considérer la distance euclidienne qui les sépare. Nous pouvons alors définir un premier critère de localisation $L_{\text{projectif}}$ formalisé par :

$$L_{\text{projectif}} = \text{moy}_i \{ \min_j \{ \text{dist}(P_{\text{det}}^i, P_{\text{proj}}^j) \} \} \quad (10)$$

La mesure obtenue est précieuse dans le cadre du calibrage d'une caméra. Si on ne tient pas compte de la fiabilité de la méthode employée pour l'étalonnage lui-même, on peut dire que plus $L_{\text{projectif}}$ est petit, plus la calibration obtenue sera précise. Le choix des points 3D pour le calcul de cette mesure est détaillé dans section 4.3.

La reconstruction euclidienne. Le critère ci-dessus est insuffisant lorsqu'on souhaite procéder à la reconstruction tridimensionnelle de la scène. Supposons alors que nous ayons à notre disposition au moins deux vues de la scène ainsi qu'une méthode de mise en correspondance des points. Pour obtenir une mesure de localisation spécifique à la 3D, il suffit de mettre en correspondance les points détectés puis de les reconstruire, et enfin de les comparer aux points P^j de la scène. De manière similaire, la mesure de localisation est alors donnée par :

$$L_{\text{euclidien}} = \text{moy}_i \{ \min_j \{ \text{dist}(P_{\text{reconstruits}}^i, P^j) \} \} \quad (11)$$

Le choix des points 3D utilisés ainsi que les méthodes de mise en correspondance et de reconstruction sont détaillés dans la section suivante.

4.3 Cadre de l'évaluation

La méthode d'évaluation mise en oeuvre ici nécessite la donnée d'images couleur d'une scène plane, ainsi que la connaissance de l'homographie H_{12} entre les projections de cette scène. H_{12} doit être déterminée de manière très précise et indépendamment des points détectés, afin de ne pas favoriser le détecteur étudié. Pour ce faire, nous avons choisi de modéliser une scène 3D contenant une texture (une image couleur) que nous plaquons sur un parallélépipède rectangle, comme le montre la figure 3. Dans cette scène, nous positionnons plusieurs caméras modélisées selon le modèle pin-hole. Les paramètres intrinsèques et extrinsèques de ces caméras étant alors connus, la texture 3D peut alors être projetée sur leur plan rétinien par un lancer de rayon. Notons que l'algorithme de projection se doit d'éliminer le crénelage afin de ne pas détecter par la suite des points qui n'ont pas de réalité physique dans la scène 3D.

Bien que les images soient générées synthétiquement, nous les considérons comme réalistes dans la mesure où la texture plaquée est une image issue d'une scène

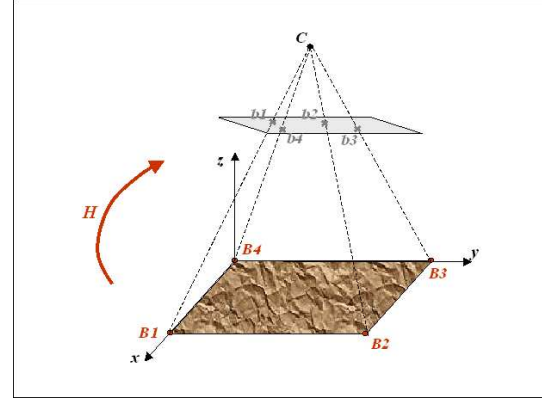


FIG. 3 – Exemple de scène modélisée avec une texture plan et une caméra. La base projective 3D ($B1, B2, B3, B4$) est projetée sur les différentes images I_i en ($b1_i, b2_i, b3_i, b4_i$). L'homographie H est donc calculée de façon très précise.

réelle. Les conditions de notre évaluation sont par conséquent très proches des conditions naturelles de prise de vue. Nous avons en plus à notre disposition la calibration complète des caméras.

L'homographie associée à la texture est donc obtenue de façon exacte et très simple : il suffit de considérer par exemple les 4 sommets de la texture et de les projeter dans les deux images (en n'oubliant pas de conserver les coordonnées réelles et non discrètes des projections pour calculer une homographie aussi précise que souhaitée).

En ce qui concerne la mesure de localisation, la connaissance exacte de la position de chacun des pixels de la texture et de leurs projections (encore en coordonnées réelles) nous est très précieuse. En effet, prenons comme texture une grille de calibration dont la position de certains points est connue avec exactitude. Pour la mesure basée sur la reconstruction projective, il suffit alors de comparer la projection exacte de ces points avec les points détectés dans l'image de la texture. En ce qui concerne la reconstruction euclidienne, il s'agit tout d'abord de mettre en correspondance les points détectés sur deux images. Une solution est de mettre à profit la connaissance complète que nous avons de la calibration : nous appariions les points les plus répétables dans un petit voisinage. Cette solution se révèle être une bonne technique de mise en correspondance, et permet aussi de faire intervenir le moins possible les erreurs de répétabilité dans cette deuxième mesure de la localisation. La reconstruction est ensuite réalisée par triangulation des points appariés.

Enfin la mesure de localisation est obtenue à partir des distances des points reconstruits à la texture.

4.4 Mise en oeuvre et résultats

Les mesures présentées ci-dessus vont être utilisées pour évaluer les deux détecteurs couleur présentés à la section 3. Afin d'être à même de classer les résultats obtenus, nous devons aussi appliquer notre évaluateur à un détecteur qui a fait ses preuves et qui nous ser-

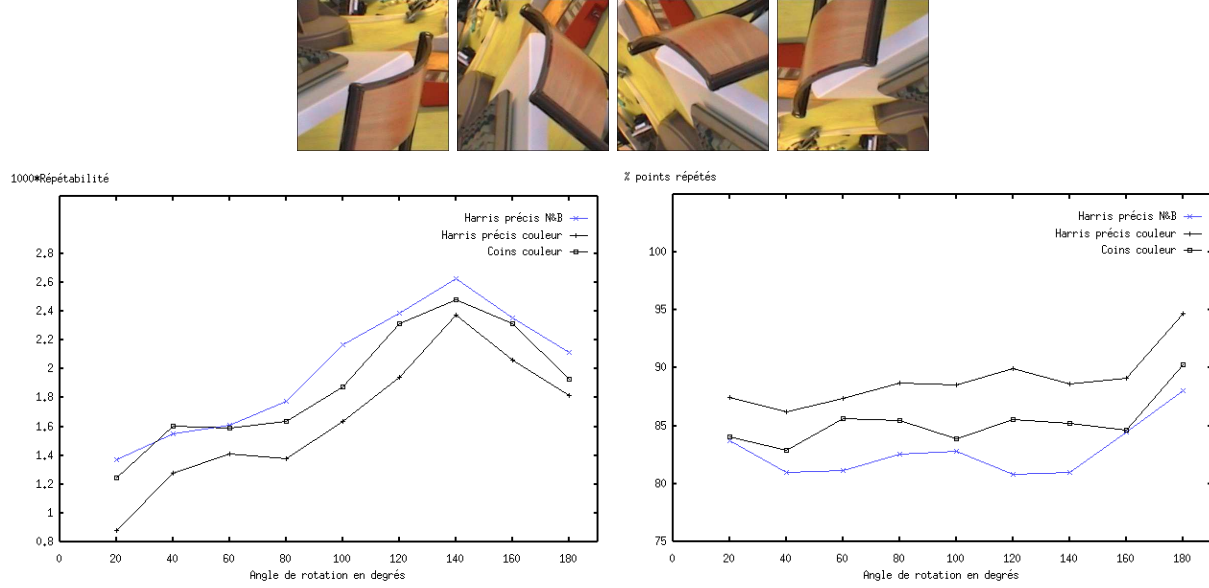


FIG. 4 – Mesure de répétabilité R^1 et pourcentage de points répétés suivant 9 rotations image. Les 4 premières images représentent l'image de référence suivie de 3 images ayant respectivement subi une rotation de 60° , 120° et 180° .

vira donc de référence. Nous avons choisi le détecteur de Harris et Stephen dans sa version précise car il a été élu le plus stable et le plus répétable des détecteurs de points d'intérêt en niveau de gris [15]. L'évaluation sera alors effectuée par rapport à différentes transformations de l'image, notamment la rotation image, le changement affine de luminosité présenté en 2.3 ainsi que le changement de point de vue. Deux textures de nature différente sont utilisées pour la scène 3D : la première provient d'une scène d'extérieur et donc contient beaucoup de texture, alors que la seconde représente une scène d'intérieur avec donc essentiellement des contours. Les filtres utilisés par les opérateurs de détection sont des filtres gaussiens utilisés à une précision sub-pixel. Ajoutons pour finir que dans toutes les expériences menées et pour tous les détecteurs, nous nous sommes attachés à travailler avec un nombre de points compris entre 400 et 500, les images rétinienne étant de dimension (300×300) .

Rotation image. Pour obtenir une séquence de rotations image, nous avons fait pivoter la caméra virtuelle autour de son axe optique, en lui appliquant 9 rotations de 20° jusqu'à 180° . La figure 4 montre trois des rotations obtenues à partir de l'image de référence placée en premier sur la figure. Nous avons choisi une erreur de localisation ε de 1, ce qui indique, si l'on se réfère à l'équation 9, que le point détecté peut être trouvé chez l'un des pixels voisins du pixel prédit dans le cas d'un lissage en précision $1/2$ pixel. Le premier graphique de la figure 4 correspond à la mesure de répétabilité R^1 , multipliée par un facteur 1000 pour une plus grande lisibilité. Le second représente uniquement le pourcentage de points répétés, soit la mesure $100 \times \frac{n_{ij}}{N_i}$. Le premier graphique montre clairement

que le détecteur Harris couleur est le plus répétable face aux rotations image. Plus généralement, les deux détecteurs couleur sont sensiblement plus performants que le détecteur Harris précis en niveau de gris. En effet, moins de points semblent répétés pour ce dernier, comme le montre le second graphique. Notons aussi que les résultats se dégradent globalement lorsque l'angle de rotation augmente, sauf pour la rotation de 180° . Le second graphique nous indique alors que cette dégradation semble plutôt due à une diminution de la précision des détections, puisque les pourcentages de points répétés restent sensiblement les mêmes. La zone de recherche se limitant ici aux premiers voisins ($\varepsilon = 1$), cela indique que d'avantage de points sont détectés dans le voisinage du point prédit.

Changement de luminosité. Nous avons réalisé synthétiquement 7 changements affines de luminosité en référence au modèle d'illumination présenté à la section 2.3. La figure 5 montre sur la première ligne l'image de référence "Ordi" suivie de 3 de ces changements. Le premier graphique montre la mesure de répétabilité R^1 obtenue et le second le pourcentage de points répétés. Au vu des mesures de répétabilité obtenues beaucoup plus faibles en comparaison avec les mesures obtenues pour les rotations image de la figure 4, nous pouvons d'ores et déjà noter que les 3 détecteurs ont une répétabilité excellente face aux changements affines de luminosité. Le deux graphes montrent que les détecteurs sont plus répétables pour les transformations de luminosité les plus faibles (vues 1 et 7). Plus particulièrement, le second graphique traduit que pour les transformations plus importantes (vues 2 à 5), beaucoup moins de points sont répétés. Encore une fois, nous observons la supériorité des détecteurs couleur, le détecteur de coins couleur paraissant le plus

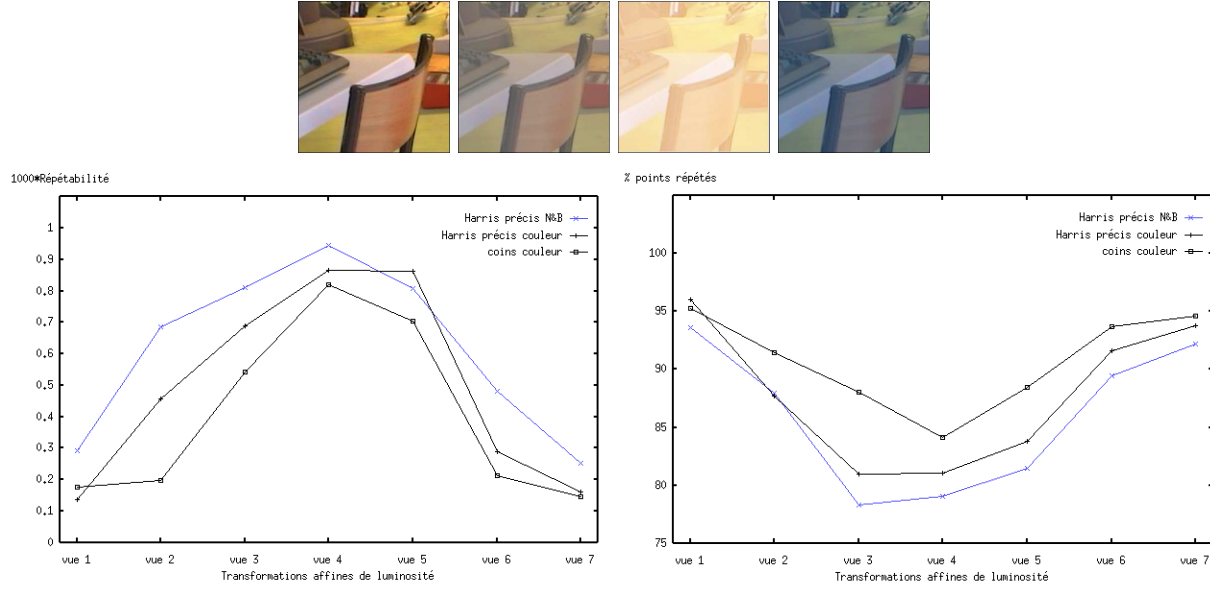


FIG. 5 – Mesure de répétabilité R^1 et pourcentage de points répétés suivant 7 changements affines de luminosité.

répérable.

Changement de point de vue. Pour mesurer la répétabilité des détecteurs face à un changement de point de vue, la caméra a été déplacée 7 fois selon une rotation autour de la texture, en veillant à viser approximativement son centre de façon à en voir la plus grande partie. Le lecteur peut voir quelques exemple des images obtenues dans la figure 6. Sur les graphiques, la vue 1 correspond à la plus petite rotation appliquée à la caméra et la vue 7 la plus grande rotation. Les deux graphes correspondent à nouveau à la mesure de répétabilité R^1 et au pourcentage des points répétés. Il est clair que les résultats se dégradent pour les 3 détecteurs au fur et à mesure que la déplacement augmente. Nous constatons que le détecteur Harris précis couleur obtient les meilleurs résultats, alors que le détecteur de coins couleur les moins bons. Cela s'explique par le fait que les coins couleur sont exclusivement des points de contour, donc des points pour lesquels le voisinage se modifie avec le changement de point de vue (par exemple les pixels correspondant à l'arrière plan de la scène se déplacent plus vite que ceux du premier plan). Nous pouvons donc en conclure que les coins (qu'ils soient en niveau de gris ou en couleur) sont moins bien adaptés au changement de points de vue. L'expérience montre néanmoins encore une fois l'intérêt de l'utilisation de la couleur pour le détecteur Harris précis. Nous constatons au vu du premier graphe que le détecteur de Harris précis est meilleur dans sa version couleur. Il est alors intéressant de noter qu'il semble plus précis que son homologue en niveau de gris, puisque le second graphe nous montre que le pourcentage de points répétés est globalement le même pour les deux détecteurs de Harris. Cela amène à penser que d'avantage de points sont exactement répétés pour le Harris Couleur. Au contraire, le détecteur Harris en niveau de gris trouve plus de points dans le

voisinage direct des point prédits.

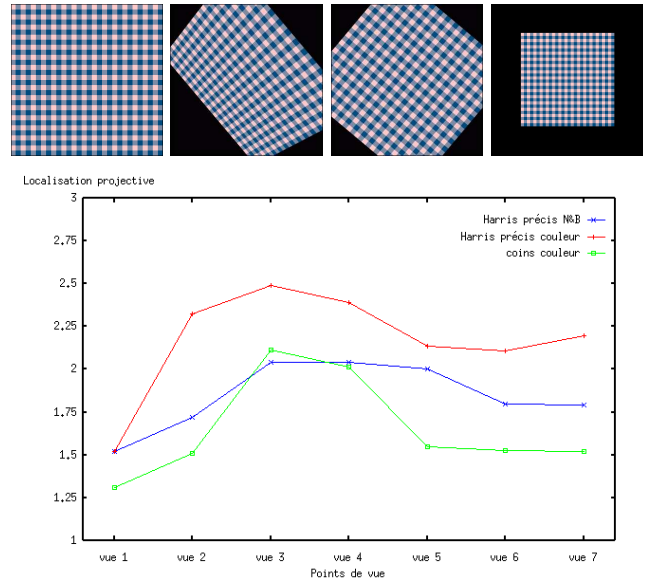


FIG. 7 – Mesures de localisation projective sur des images de la grille prises sous 7 points de vue différents.

Mesures de localisation. Afin d'effectuer la mesure de localisation projective, nous utilisons en guise de texture 3D une grille de calibration pour laquelle la position des points déterminants (ceux qui ont toutes les chances d'être détectés dans la suite du traitement par chacun des détecteurs) est connue avec précision. Le lecteur peut voir à la figure 7 la grille utilisée sous

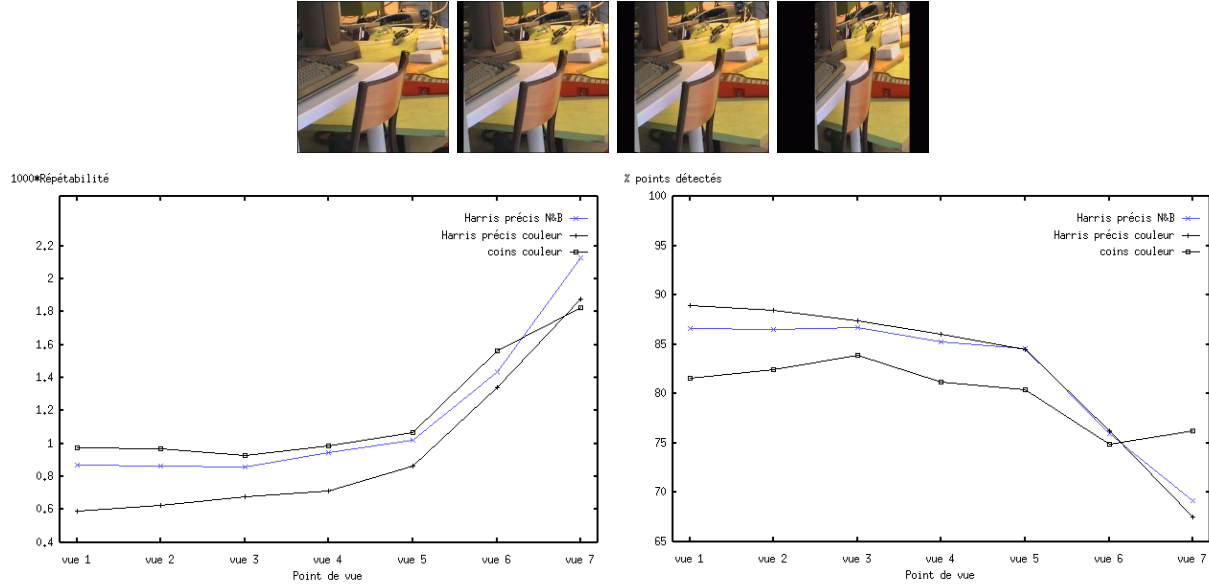


FIG. 6 – Mesure de répétabilité R^1 et pourcentage de points répétés suivant 7 points de vue.

quelques points de vue. Le graphe associé montre les mesures de $L_{\text{projectif}}$ obtenues. Le détecteur de coins couleur présente visiblement les meilleurs résultats, alors que le détecteur Harris couleur se révèle plutôt décevant par rapport aux autres. L'expérience montre donc que c'est le détecteur de coins couleur qui détecte les points le plus précisément. Par conséquent, nous préconisons son utilisation dans le cadre du calibrage des caméras. Pour calculer la mesure de localisation euclidienne, nous avons utilisé la séquence d'images qui a servi à la mesure de la répétabilité face aux différents points de vue (quelques vues sont présentées à la figure 6). Les 7 vues ont été mises en correspondance avec la vue de référence en exploitant la mesure de répétabilité (la méthode est détaillée à la fin de la section 4.3). La reconstruction euclidienne a été réalisée suivant deux erreurs de répétabilité ε de 0.5 et de 1.5. La figure 8 montre les résultats obtenus. D'un point de vue général, nous constatons que quels que soient les détecteurs employés, les résultats s'améliorent nettement plus le point de vue est éloigné du point de vue référence. Ce résultat est tout à fait normal dans la mesure où la précision de la triangulation augmente avec l'écart de points de vue. Ainsi, pour la vue 1, la localisation euclidienne est très mauvaise ($L_{\text{euclidien}} \geq 5000$) alors qu'elle est plus de 100 fois meilleure pour la vue 7 ($L_{\text{euclidien}} \simeq 50$).

Le calcul avec les deux seuils de répétabilité ε montre également que la répétabilité d'un détecteur intervient dans la précision de la reconstruction 3D. Ainsi nous obtenons de meilleurs résultats pour l'erreur ε la plus petite. Nous constatons enfin que le détecteur de coins couleur ne donne pas les meilleurs résultats, alors qu'il était le plus fiable pour la reconstruction projective. Il faut justement attribuer à ce résultat la relative mauvaise répétabilité de ce détecteur face aux changements de points de vue (se référer à la figure 6). Quant aux

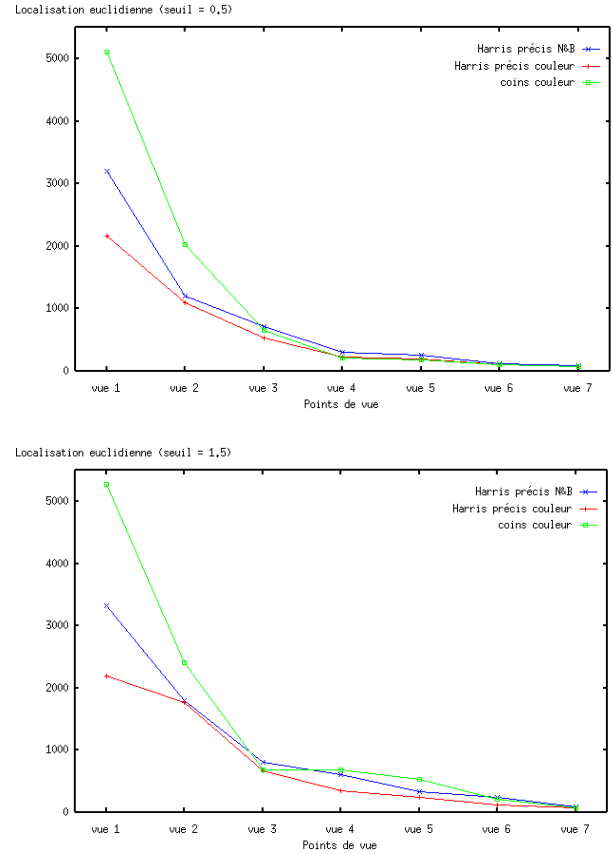


FIG. 8 – Mesures de localisation euclidienne sur des images prises sous des points de vue différents. ε vaut 0.5 pour la première figure et 1.5 pour la seconde.

détecteurs Harris précis, les résultats semblent globalement meilleurs lorsqu'on tient compte de l'information couleur.

5 Conclusion

Nous avons dans ce papier tenté d'évaluer les deux détecteurs de points que nous avons mis en place pour la couleur. Les résultats obtenus sont globalement probants, puisque à chaque expérience, les mesures de répétabilité et de localisation sont meilleures que celles obtenues avec le détecteur noir et blanc pour au moins l'un des deux détecteurs couleur. L'information couleur est visiblement plus riche que les seuls niveaux de gris. Par contre, la comparaison entre les deux détecteurs couleur doit être beaucoup plus nuancée. En effet, le détecteur Harris précis couleur semble mieux adapté aux images texturées, dans la mesure où les points détectés ne sont pas nécessairement des coins. En revanche, le détecteur de coins couleur se révèle plus adapté aux images d'intérieur qui contiennent en général beaucoup de contours. Les expériences menées ont néanmoins permis de montrer que ces deux détecteurs étaient équivalents face aux rotations image. Le détecteur de coins semble meilleur face aux changements affines de niveaux de gris, mais moins répétable en ce qui concerne les changements de points de vue. Nous sommes également arrivés à la conclusion que le choix entre ces détecteurs dépend aussi du type d'application envisagé. Le détecteur de coins possède une meilleure localisation projective, donc semble mieux adapté pour le calibrage de caméras par exemple. En revanche, le détecteur Harris précis couleur possède une meilleure localisation euclidienne (en partie parce que sa répétabilité est aussi meilleure face aux changements de points de vue) et sera donc préconisé pour la reconstruction 3D.

References

- [1] C. Bauckhage and C. Schmid. Evaluation of keypoint detectors. Technical report, INRIA, 1996.
- [2] P. Brand and R. Mohr. Accuracy in image measure. *SPIE Videometrics*, 3:218–228, 1994.
- [3] J.F. Canny. A computational approach to edge detection. *PAMI*, pages 679–698, 1986.
- [4] G.D. Finlayson, M.S. Drew and B. Funt. Color constancy: Generalized diagonal transforms suffice. *Journal of the Optical Society of America A*, 11(11):3011–3019, november 1994.
- [5] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink and M.A. Viergever. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4:171–187, 1994.
- [6] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern and Machine Intelligence*, 13(9):891–906, 1991.
- [7] V. Gouet, P. Montesinos and D. Pelé. A fast matching method for color uncalibrated images using differential invariants. In *Proceedings of the British Machine Vision Conference, Southampton, UK*, 1998.
- [8] V. Gouet, P. Montesinos and D. Pelé. Stereo matching of color images using differential invariants. In *Proceedings of the International Conference on Image Processing, Chicago, USA*, 1998.
- [9] C. Harris and M. Stephens. A combined corner and edge detector. *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.
- [10] D. Hilbert. *Theory of Algebraic Invariants*. Cambridge Mathematica Library, Cambridge University Press, 1890.
- [11] J.J. Koenderink and A.J. Van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.
- [12] P. Montesinos, V. Gouet and R. Deriche. Differential Invariants for Color Images. *Proceedings of 14th International Conference on Pattern Recognition, Brisbane, Australia*, 1998.
- [13] R.P.N. Rao and D.H. Ballard. Object indexing using an Iconic Sparse Distributed Memory. *Proceedings of 5th International Conference on Computer Vision*, pages 24–31, 1995.
- [14] A.H. Salden, B.M. ter Haar Romeny, L.M.J. Florack, M.A. Viergever and J.J. Koenderink. A complete and irreducible set of local orthogonally invariant features of 2-dimensional images. *Proceedings of 11th International Conference on Pattern Recognition*, pages 180–184, 1992.
- [15] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris - Application à l'indexation d'une base d'objets*. PhD thesis, INPG, France, 1996.