

*Utilisation de la couleur pour l'appariement et  
l'indexation d'images.*

P. Gros, G. Mclean, R. Delon, R. Mohr, C. Schmid et G. Mistler

**N° 3269**

1<sup>er</sup> septembre 1997

THÈME 3



*Rapport  
de recherche*



## Utilisation de la couleur pour l'appariement et l'indexation d'images.

P. Gros, G. Mclean, R. Delon, R. Mohr, C. Schmid et G. Mistler

Thème 3 — Interaction homme-machine,  
images, données, connaissances  
Projet Movi

Rapport de recherche n° 3269 — 1<sup>er</sup> septembre 1997 — 41 pages

**Résumé :** De nombreux travaux récents ont montré que l'utilisation directe du signal des images en niveaux de gris permet de développer des systèmes puissants d'appariement, d'indexation et de reconnaissance d'images. Le présent document explore l'utilisation d'images en couleur pour ces mêmes objectifs. Les principaux points en sont le choix du modèle de variation de la couleur, l'utilisation d'invariants colorimétriques, et la normalisation des images.

**Mots-clé :** couleur, appariement, indexation, invariants.

*(Abstract: pto)*

## Using Color for Image Matching and Indexing.

**Abstract:** Recent works has demonstrated that the direct use of grey levels for image matching and indexing allows to build very powerful systems. The present report tries to enlarge these results to the case of color images. The main points are color illumination model evaluation, color invariant computation and image normalization.

**Key-words:** color, indexing, matching, invariants.

## 1 Introduction

La puissance toujours plus grande des ordinateurs disponibles sur le marché permet de traiter des informations de taille croissante. Dans le cadre du traitement des images, pour réduire la quantité d'information à traiter, l'utilisation des contours était classiquement considérée comme le meilleur compromis : elle permettait une grosse réduction, tout en gardant un contenu informatif important. De très nombreux algorithmes sont donc basés sur l'emploi de contours ou d'approximations de ceux-ci par des segments ou des courbes algébriques, des B-splines cubiques par exemple.

Désormais, il est possible de traiter les images en niveau de gris directement, et ce, en gardant des temps de calculs raisonnables. De nombreux travaux ont commencé à explorer l'utilisation de cette information pour l'appariement, l'indexation ou la reconnaissance des images. Les résultats sont très prometteurs [Sch96, SM97]. Le signal des niveaux de gris se révèle très informatif, permet le calcul de caractéristiques très discriminantes, et permet donc la construction de systèmes très puissants.

Qu'en est-il de la couleur ? Le signal couleur peut être décomposé de diverses manières en trois composantes. Faut-il alors considérer une image en couleur comme la simple superposition d'images en niveau de gris, ou y a-t-il une information supplémentaire ? La couleur ouvre-t-elle de nouvelles possibilités, ou peut-on en rester aux niveaux de gris ?

Le présent rapport a pour but d'explorer ces questions, en fournissant quelques résultats préliminaires et ouvrant diverses pistes de travail futur. La première partie est un état de l'art sur l'utilisation de la couleur pour l'appariement et l'indexation des images, la deuxième partie s'intéresse à l'évaluation de modèles de variation d'illumination, la troisième part des idées de Finlayson sur les variations d'illumination et propose divers schémas de normalisation d'images, la dernière concerne l'utilisation d'invariants locaux.

## 2 État de l'art

Ce paragraphe présente un état de l'art des méthodes utilisant la couleur pour indexer des images. Son but est d'identifier les techniques qui pourraient s'avérer intéressantes pour prolonger aux cas des images couleur les techniques déjà utilisées pour les images en niveaux de gris, par exemple les invariants locaux développés par Schmid [SM97].

Dans ce travail, le paradigme de base est semblable à celui du hachage géométrique de Wolfson [Wol90] : des points d'intérêts ont tout d'abord détectés, puis caractérisés par des descripteurs locaux calculés autour de ces points. Chacun de ces points fournit donc une clé possible pour rechercher une image dans une base d'images. D'une manière générale, une requête peut porter sur plusieurs clés simultanément. Cette redondance permet des recherches mêmes lors d'occultations partielles.

Dans le présent document, nous nous intéressons à la recherche de descripteurs basés sur l'information de couleur. Nous imposons à un tel descripteur d'être invariant pour le groupe des similitudes entre les images. De plus, ils doivent être aussi invariants pour les changements d'illumination, et de capteurs.

## 2.1 Modélisation des variations d'illumination

Le concept de constance de la couleur vient de l'observation que la perception humaine de la couleur est indépendante du produit de l'illumination par la réflectance, en étant plus fortement corrélée avec la réflectance seule [LM71]. La théorie originale du « Retinex » proposée par Land a été l'objet d'un travail considérable, alors que se poursuit la recherche d'un algorithme général qui pourrait annuler les effets de la variation spatiale de l'illumination dans une scène. Cette recherche n'a pour le moment donné que des résultats modestes, malgré la complication des algorithmes utilisés [BFF96].

Bien que ce ne soit pas équivalent à la constance de la couleur, obtenir une invariance à l'illumination est indispensable pour les algorithmes d'indexation qui doivent marcher pour des objets vus dans des conditions d'illumination changeantes. Brill a proposé un modèle général selon lequel deux surfaces identiques observées dans des conditions d'illumination différentes sont reliées de manière linéaire [Bri78]. Un travail plus récent suggère qu'un modèle diagonal de transformation suffit pour obtenir l'invariance à l'illumination [FDF94] :

$$\mathcal{D}^{ij} p^{i,x} = p^{j,x} \quad (1)$$

où  $p^{i,x}$  et  $p^{j,x}$  représentent deux vecteurs de couleur à la position  $x$  sous les conditions d'illumination  $i$  et  $j$ , et où  $\mathcal{D}^{ij}$  est une matrice diagonale qui lie les deux observations. Le modèle diagonal de variation de l'illumination est valide si la matrice  $\mathcal{D}^{ij}$  est la même pour toute les positions  $x$ .

## 2.2 Appariement global utilisant la couleur.

La majorité des travaux sur l'indexation des images qui utilisent la couleur sont basés sur une approche *globale* de l'image. L'image entière est décrite au moyen d'un unique vecteur de caractéristiques. Ces travaux peuvent être classés en trois catégories: la comparaison d'histogrammes, le regroupement (clustering en anglais), et les moments statistiques. Chacun des trois fait l'objet d'un des paragraphes suivant. Leur adaptation à des techniques plus locales est étudiée au paragraphe 2.4.

### 2.2.1 Mesures de ressemblance entre histogrammes

L'histogramme d'une image représente l'énumération de la densité de chaque couleur présente dans cette image. En pratique, le nombre de couleurs possibles est très grand, aussi les histogrammes sont-ils fait en regroupant les couleurs en paquets. Le calcul de l'histogramme associé à une image est facile, mais il mène à une représentation dans une espace de très grande dimension (par exemple, si l'on divise chaque composante de la couleur en 16 intervalles, cela mène à un espace de dimension 4096), ce qui nécessite l'emploi de techniques d'appariement de forte complexité.

Si l'on ordonne les indices de l'histogramme de manière régulière, la valeur de l'histogramme de l'image  $I$  pour un paquet  $j$ , que l'on note  $h_j^I$ , est simplement le nombre de pixels de l'image  $I$  qui ont un vecteur de couleurs compris dans le paquet  $j$ . Un tel histogramme

peut être normalisé en divisant chacune des valeurs précédentes par le nombre total de pixels de l'images, de telle sorte que l'intégrale de l'histogramme soit égale à 1.

On peut alors définir « l'histogramme cumulatif » comme une primitive de l'histogramme précédent :

$$c_j^I = \sum_{i=1}^j h_i^I$$

**Intersections d'histogrammes.** Le premier travail et sans doute aussi le plus référencé sur l'indexation utilisant la couleur est la méthode d'intersection d'histogrammes proposée par Swain et Ballard [SB91]. Un histogramme 3D est construit pour chaque image de la base d'images. Les images de cette base semblable à une image requête sont trouvées en calculant l'intersection de l'histogramme de cette image avec ceux présents dans la base :

$$\text{Inters}(I, H) = \frac{\sum_k \min(h_k^I, h_k^H)}{\sum_k h_k^H} \quad (2)$$

Des valeurs de la fonction d'intersection proches de 1 indiquent une bonne ressemblance, cette fonction comptant le nombre de pixels du modèle ayant un correspondant direct dans l'image requête. L'intersection des histogrammes n'est pas invariante aux changements d'illumination : les images doivent donc avoir été traitées au préalable pour corriger les effets de ces changements, avant leur utilisation comme requête, ou leur introduction dans la base.

Funt et al. [FF95] utilisent le modèle diagonal de variation de l'illumination pour obtenir des mesures de couleur invariantes. Avec une illumination à peu près constante, le rapport de la réponse du capteur dans deux positions de l'image est donné par :

$$\frac{p_k^{x_1}}{p_k^{x_2}} = \frac{e_k r_k^{x_1}}{e_k r_k^{x_2}} \quad (3)$$

où  $e_k$  est l'illumination de couleur  $k$  et  $r_k^x$  la réflectance à la position  $x$ . En considérant le logarithme de l'image et la différence entre les logarithmes de deux pixels de l'image, on peut calculer un invariant à l'illumination :

$$\log P_k^1 - \log P_k^2 = \log r_k^1 - \log r_k^2 \quad (4)$$

La différence de logarithmes du premier membre peut être utilisé pour calculer des histogrammes, à partir desquels on peut calculer des intersections d'histogrammes. Leur auteur notent que ceci revient à effectuer de manière grossière une détection de contour sur le logarithme de l'image.

**Mesures de distance.** Chaque histogramme permet de représenter une image par un point dans un espace de dimension  $N_I$ . Diverses méthodes de calcul de distance dans cet espace ont été proposées.

Stricker [SO95] a remarqué la similitude existant entre l'intersection des histogrammes et la norme  $L_1$  utilisée pour mesurer la distance entre deux histogrammes dans leur espace de grande dimension, et il utilise ces deux mesures comme si elles étaient équivalentes. Cela a aussi été remarqué par Mehtre [MKN95]:

$$d_{L_1}(I, H) = \sum_j |h_j^I - h_j^H| \quad (5)$$

La distance euclidienne a été utilisée par Stricker [SO95], Niblack [HSE<sup>+</sup>95] et Zhang [ZGLS95] utilisant pour leur part une distance euclidienne généralisée. Bien que l'emploi d'une matrice de pondération, qui permet d'inclure dans le calcul la ressemblance entre les couleurs, a été proposée, dans la pratique, seule la matrice identité a été utilisée. La distance euclidienne généralisée est définie par la formule suivante, où  $A$  est la matrice de pondération :

$$d_{L_2}(I, H) = \sqrt{(H - I)A(H - I)^T} \quad (6)$$

**Tests d'hypothèses.** Un histogramme de couleur peut être considéré comme la réalisation d'une variable aléatoire donnant les couleurs dans une image. À partir de là, la comparaison d'histogrammes peut être ramenée à un problème de tests d'hypothèses, dans lequel il faut déterminer si deux réalisations (c'est à dire deux histogrammes) peuvent provenir de la même distribution. Cette vue des choses a été introduite simultanément par deux équipes [SP95, SO95]. Toutes deux ont montré, sur un ensemble de données expérimentales réduit, que cette méthode conduit à de meilleurs résultats que l'intersection des histogrammes ou la distance euclidienne généralisée.

Dans [SP95], une méthode de détection des cuts vidéos est proposée, et les tests d'hypothèses sont donc appliqués pour voir si deux images sont différentes (les auteurs ne cherchaient donc pas une mesure de ressemblance, mais de dissemblance). Les deux premiers tests utilisés sont les suivants.

#### Report de ressemblance de Yakimovsky

$$y(I, H) = \frac{(\sigma_o^2)^{N_I + N_H}}{(\sigma_I^2)^{N_I} (\sigma_H^2)^{N_H}} \quad (7)$$

où  $\sigma_o^2$  est la variance de la population ayant formé les deux histogrammes,  $\sigma_j^2$  est la variance d'un histogramme et  $N_I$  et  $N_H$  sont les tailles de chacune des populations.

**Test du  $\chi^2$**  Le test du  $\chi^2$  est utilisé pour déterminer la similitude entre deux distributions discrètes, et est donné par :

$$\chi^2 = \sum_j \frac{(h_j^I - h_j^H)^2}{(h_j^I + h_j^H)^2} \quad (8)$$

Ce test est basé sur l'hypothèse que la ou les distributions présentes sont gaussiennes.

Dans [SO95], des analogues robustes des normes  $L_1$  et  $L_2$  sont proposées, et une norme  $L_\infty$  est introduite. Ces mesures conduisent à des tests non paramétriques. Comme ceux-ci sont basés sur les distributions cumulatives, ils ont tendance à être robustes aux changements locaux de forme et aux petits décalages dans l'image. Les trois tests sont les suivants.

$L_1$

$$d_{L_1}(I, H) = \sum_{j=1}^N |c_j^I - c_j^H| \quad (9)$$

$L_2$  (Test de Cramer - von Mises)

$$d_{L_2}(I, H) = \sqrt{\sum_{j=1}^N (c_j^I - c_j^H)^2} \quad (10)$$

$L_\infty$  (Test de Kolmogorov - Smirnov)

$$d_{L_\infty}(I, H) = \max |c_j^I - c_j^H| \quad (11)$$

Le test de Kolmogorov - Smirnov a été utilisé par Sethi [SP95]. Il peut être utilisé pour détecter des changements en utilisant de petits échantillons extraits localement des images.

Stricker montre que le test de Kolmogorov - Smirnov fournit de meilleurs résultats que les deux autres tests et que la mesure de distance euclidienne appliquée aux histogrammes. Sethi rapporte lui que le test du  $\chi^2$  fournit de meilleurs résultats avec le test de Kolmogorov - Smirnov venant en second. Ceci est intéressant, car le test du  $\chi^2$  fait une hypothèse gaussienne et est appliqué sur l'histogramme présumé bruité, et non sur l'histogramme cumulatif. Il faut toutefois bien remarquer que les deux batteries de tests proposées ont peu de choses en commun.

### 2.2.2 Clustering

Le regroupement implicites des couleurs en paquets pendant la construction de l'histogramme est une forme grossière de regroupement des distributions de couleur en un ensemble plus restreint de couleurs. Une simple réduction du nombre des couleurs utilisées pour calculer un histogramme mène à des performances d'indexation très instables [ZGLS95].

Diverses variations sur la distance euclidienne ont été proposées, dans le but de réduire le coût de la comparaison en diminuant la dimension de l'espace de représentation. Jain [JV96] fait la somme des distances euclidiennes des trois histogrammes unidimensionnels, un pour chaque canal de couleur, et Kankanhalli [KMW96] calcule la distance euclidienne entre les moyennes des trois couleurs.

Hafner [HSE<sup>+</sup>95] considère le problème du choix des  $K$  meilleurs vecteurs de couleurs pour l'appariement en utilisant la distance euclidienne et a développé un moyen efficace de faire une « comparaison rapide d'index » comme prétraitement pour l'appariement des histogrammes.

Dans [MKN95], ou bien [HSE<sup>+</sup>95], une « table de référence des couleurs » est utilisée, qui représente le meilleur ensemble de couleurs pour l'ensemble de la base. Les couleurs de chaque image sont alors réduites à cet espace, qui permet donc une espace de moindre dimension que l'ensemble total des couleurs, et la distance euclidienne est alors utilisée. Un problème avec ces approches est qu'un seul vecteur de petite dimension est utilisé pour représenter toutes les images de la base, ce qui rend difficile l'adjonction ou la suppression des images dans la base.

Une méthode pour éviter ce problème est proposée dans [KMW96]. Le regroupement des couleurs est effectué pour chaque image indépendamment, de telle sorte que chaque image est représentée par sa propre table de couleurs et la densité de chacune de ces couleurs. La ressemblance entre images est alors calculée en appariant tout d'abord les deux tables de référence de couleurs, puis en calculant la distance euclidienne entre les densités. Pour deux images  $I$  et  $J$ , si l'on note  $p(i)$  la relation entre les tables de couleurs, la distance entre les images s'écrit alors :

$$D(I, J) = \sum_{i=1}^{\min(N_I, N_J)} \omega_i \sqrt{(h_i^I - h_{p(i)}^J)^2} + \sum_{i=1}^{\min(N_I, N_J)} \text{cdist}(c_i^I, c_{p(i)}^J) \quad (12)$$

où  $\text{cdist}(c_i^I, c_{p(i)}^J)$  est la distance entre les couleurs qui se correspondent dans les deux tables.

### 2.2.3 Moments statistiques

Reconnaître que l'histogramme d'une image est en réalité une fonction de densité de probabilité mène naturellement à l'idée d'utiliser des moments statistiques pour indexer des images. Cela diffère des méthodes de comparaisons d'histogrammes et de regroupement de couleurs en ce que ce ne sont points des « super-couleurs » qui sont utilisées, mais des caractéristiques de la *forme* de la distribution de couleur.

Stricker [SO95] introduit les moments comme un moyen d'échapper au effets de la discrétisation de l'espace des couleurs qui est intrinsèque à l'utilisation des histogrammes. Une somme pondérée de la moyenne, de la variance et du moment du troisième ordre est calculée pour chaque canal de couleur, pour fournir un nombre unique utilisé pour indexer. Si  $p_{ij}$  est la valeur du pixel  $j$  pour le canal  $i$ , ces moments sont définis par :

$$\mu_i = \frac{1}{N} \sum_{j=1}^N p_{ij} \quad (13)$$

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^2} \quad (14)$$

$$s_i = \left( \frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^2 \right)^{\frac{1}{3}} \quad (15)$$

La distance entre deux images est alors définie comme une somme pondérée entre ces quantités calculées pour chacun des canaux :

$$d_{mom}(I, H) = \sum_{i=1}^3 w_{i1} |\mu_i^I - \mu_i^H| + w_{i2} |\sigma_i^I - \sigma_i^H| + w_{i3} |s_i^I - s_i^H| \quad (16)$$

Dans [FCF96], l'utilisation des moments est motivée par la volonté d'aborder le problème du changement de couleur en utilisant un ensemble de caractéristiques qui soient invariantes aux changements d'illumination. Les auteurs font trois remarques importantes quant à l'utilisation des moments :

- en ce qui concerne l'information de couleur, les moments d'ordre peu élevés résument la plupart de l'information utile ;
- malheureusement, ces moments d'ordre peu élevés ne sont pas invariants aux changements d'illumination ;
- des caractéristiques invariantes peuvent être calculées à partir des moments classiques (comme dans [TC92]), mais à partir des moments d'ordre élevés seulement. Les moments d'ordre faible sont perdus dès que l'on recherche l'invariance.

Faisant l'hypothèse de la validité du modèle diagonal, les auteurs proposent alors un moyen de standardiser les statistiques qui préserve les covariances entre les canaux (c'est à dire qui conservent les quantités  $\sigma_r \sigma_b$ ,  $\sigma_r \sigma_g$ , et  $\sigma_b \sigma_g$ ). Si une image est représentée par trois vecteurs de couleur, un vecteur pour chaque canal de couleur, une dimension pour chaque pixel, alors l'effet d'un changement d'illumination est un changement de la norme de ces vecteurs, sans changement de leur direction. Ainsi, entre deux vues de la même scène, les angles entre ces vecteurs restent constants. D'où le nom « colour angle indexing », c'est à dire indexation par les angles de couleur.

Le principal avantage de cette méthode des angles de couleur est que ces angles conservent la covariance entre canaux, d'une manière invariante aux changements d'illumination. Toutefois, dans leurs expérimentations, les auteurs introduisent un deuxième ensemble de caractéristiques basé sur les angles de couleurs entre distributions des contours en couleur.

### 2.3 Indexation locale utilisant la couleur

Toutes les techniques dont il a été question précédemment considèrent l'image globalement. Seuls trois articles ont, jusqu'à présent, tentés de proposer des méthodes utilisant localement les images pour une indexation basée sur la couleur.

Le premier travail de Swain et Ballard incluait une méthode de localisation d'objets dans les images utilisant un algorithme appelé *rétro-projection d'histogrammes*, dans lequel

les couleurs associées à l'objet recherché sont renforcées dans l'image, produisant un effet similaire à la réponse d'un filtre d'appariement [SB91]. Malheureusement, il est nécessaire que l'objet recherché soit complètement séparé du fond sur lequel il apparaît pour que la méthode de rétro-projection marche.

Ennesser et Médioni [EM95] ont remarqué que la méthode de rétro-projection des histogrammes fournissait des résultats très bruités, et ont donc proposé l'utilisation d'histogrammes locaux pour les améliorer. Des histogrammes locaux sont construits partout dans l'image considérée, puis sont intersectés avec l'histogramme de l'objet recherché.

Le dernier des travaux utilisant la couleur de manière locale se trouve dans [ZGLS95]. Chaque image est segmentée en neuf régions ne s'intersectant pas, et des histogrammes sont calculés pour chacune de ces régions. Les résultats de l'indexation sont alors meilleurs que lors de l'utilisation d'un histogramme unique.

## 2.4 Critique

L'indexation des images utilisant les informations de couleur a pour le moment été dominée par les méthodes comparant les populations représentant les distributions de couleurs dans les images (intersection d'histogrammes, mesures de distance, tests d'hypothèses). L'utilisation du regroupement de couleurs permet une avancée, par la réduction qu'elle offre de la dimension de l'espace de représentation et par l'introduction de mesures de distances plus réalistes dans un contexte de classification. Quant à elles, les méthodes basées sur les moments statistiques font un pas de plus en s'affranchissant de la discrétisation de l'espace des couleurs, et en ne caractérisant que la forme des distributions comparées.

Bien que l'on puisse trouver des résultats expérimentaux montrant les bons résultats de chacune de ces méthodes par rapports aux autres, ces diverses expérimentations ne sont pas toujours cohérentes entre elles, et on ne trouve pas de résultats synthétiques testant toutes ces méthodes sur un jeu unique de données. En dépit de cela, on peut, à partir de ces tests, classer grossièrement les méthodes par ordre d'utilité décroissante: moments statistiques, regroupement, tests d'hypothèses, mesures de distance, intersection d'histogrammes.

Seulement deux des méthodes présentées, qui sont toutes deux issues de la même équipe de recherche, incluent la notion d'invariance aux changements d'illumination. Dans ces deux méthodes, une sorte de contours en couleurs est calculée et sert de base de l'appariement. Toutes les méthodes proclament une invariance aux changements d'échelle, aux rotations, et aux translations, mais aucune données expérimentales ne permettent de vérifier ces assertions pour des catalogues d'images réalistes.

Toutes ces méthodes reposent sur la comparaison de distributions de couleurs entre les deux images à comparer. Cela entraîne deux problèmes important pour leur application à une indexation locale.

- 1° Lorsque calculées localement, les distributions de couleur ne contiennent qu'un faible nombre de pixels. Même des régions d'intérêt relativement grandes produisent des distributions sur un nombre de pixels de l'ordre de  $10^2$ , alors que toutes les expérimentations faites portent sur des populations de l'ordre de  $10^4$  pixels ou plus. Il

n'est pas assuré que les méthodes puissent marcher correctement avec de si petites populations.

- 2° Si la technique d'appariement est basée sur l'utilisation de la distribution locale des couleurs, elle risque de n'être que peu robuste aux occultations, puisque le fond sur lequel apparaît les objets peut compter pour une part importante dans la distribution.

Les deux méthodes qui essaient de travailler à un niveau plus local sont intéressantes, mais elles ne permettent pas vraiment de travailler dans un cadre similaire à celui de Schmid et Mohr [SM97] (détection de points d'intérêt, description de ces points par des invariants locaux aux changements d'illumination, vérification d'une cohérence géométrique locale). En bref, il est peu probable qu'une des méthodes présentées puisse marcher lorsqu'utilisée localement.

## 2.5 Quelles directions possibles ?

### 2.5.1 Extension aux images colorées du travail de Schmid et Mohr

Si l'on accepte la validité du modèle diagonal, il n'y a guère de raison que le travail de Schmid et Mohr [SM97] ne puisse être étendu au cas de l'indexation des images en couleur. Chaque canal de couleur est supposé subir une simple homothétie, et une image de couleur peut donc être décomposée en trois images en niveaux de gris qui peuvent être indexées. Cela revient principalement à tripler le nombre de primitives utilisables pour indexer chaque image, en gardant exactement le même algorithme général.

Le jet local calculé sur les trois canaux de couleur jusqu'à l'ordre 3 fournit 30 fonctions à partir desquelles des invariants peuvent être calculés, et les transformations que l'on considère n'ont que 4 paramètres (un angle de rotation commun aux trois canaux et 3 facteurs d'homothétie). On devrait donc pouvoir trouver 26 invariants. La translation est éliminée en utilisant des systèmes locaux de coordonnées. Ceci sera étudié au paragraphe 5.

### 2.5.2 Appariements de distributions locales

Bien que qu'il ne soit pas assuré que les méthodes de comparaison de distributions de couleur fonctionnent convenablement lorsqu'utilisées avec des données locales seulement, il serait toutefois intéressant de déterminer leur limite de fonctionnement. Il serait donc intéressant de rechercher un argument de nature statistique pour définir la borne inférieure à la taille des populations à considérer pour utiliser ces méthodes avec succès.

Une piste possible pour étendre ces méthodes d'appariement de distributions serait de considérer la notions d'appariement robuste de distributions, dans laquelle seule une partie des distributions et donc des histogrammes est considérée. De cette manière, le fond sur lequel un objet apparaît pourrait varier, la méthode cherchant correctement la ressemblance entre les deux parties d'histogrammes correspondant à l'objet recherché.

### 2.5.3 Hachage des couleurs

Au vue de toute l'étude faite, plusieurs points apparaissent :

- 1° l'utilisation des dérivées peut poser un problème de sensibilité au bruit ;
- 2° l'emploi de petites populations rend suspect l'utilisation de méthodes statistiques ;
- 3° les occultations sont toujours possibles et l'emploi de méthodes robustes est indispensable.

Le paradigme du « hachage géométrique » de Wolfson [Wol90] semble donc intéressant dans ce cadre : il n'est pas basé sur des distributions statistiques, il n'utilise pas de dérivées, et il est extrêmement robuste, dû la forte quantité de votes impliquée. On pourrait alors l'étendre à un hachage en couleur.

Supposons valide le modèle diagonal, c'est à dire :

$$\mathcal{D}^{ij} p^{i,x} = p^{j,x} \quad (17)$$

qui signifie que le rapport de deux vecteurs de couleurs en deux points est invariant aux changements d'illumination. On suppose bien évidemment que le changement d'illumination est le même pour ces deux points.

Le vecteur de couleur en un point est utilisé pour définir un « vecteur de couleur de base ». Les vecteurs de couleur en tous les autres points peuvent alors être définis en termes de coordonnées de couleur par rapport à ce vecteur de base. Le modèle diagonal ramène cela à un calcul de rapports entre les vecteurs de couleur des points et celui du point de base.

Un point étant donné, nous pouvons donc exprimer les couleurs de tous les autres en termes de rapport avec le vecteur de couleur en ce point. La matrice diagonale  $\mathcal{D}^{ij}$  peut donc être calculée pour tout point  $j$ , dès que l'on a choisit le point de base  $i$ . On peut alors construire une table de hachage sur les trois couleurs. Chaque entrée de la table pointe sur toutes les paires de points d'une image de la base qui a les rapports de couleur concernés. Ce calcul est cher, mais hors ligne.

Lors d'une requête, on part d'une image ou d'une région d'une image. Toutes les paires de points sont considérées, et les rapports entre vecteurs de couleurs correspondants sont utilisés comme index dans la table de hachage. L'image de la base qui est désignée le plus souvent par la table est alors choisie comme candidate potentielle pour l'appariement.

Un tel algorithme apparaît intéressant sur plusieurs plans :

- il marche à plusieurs échelle (aucun fenêtrage de l'image n'est utilisé), tant que le détecteur de points d'intérêt est stable ;
- il n'utilise aucun outil statistique ;
- il est invariant aux changements d'illumination dans la mesure où le modèle diagonal est valide ;
- il est redondant, et devrait donc être robuste au bruit et aux occultations ;

- la stratégie d'appariement initiale ne dépend d'aucune information géométrique et devrait donc être invariante à la rotation, à la translation... dans la mesure où le détecteur de points d'intérêt est aussi stable pour ces transformations.

### 3 Modèles de variation de l'illumination

L'état de l'art qui vient d'être fait souligne l'importance, pour obtenir une caractérisation robuste d'une image en fonction de la couleur, d'utiliser des quantités invariantes aux changements d'illumination. Cela exige donc, en préalable, de disposer d'un modèle de ces changements.

On peut distinguer deux sources de changements. La première est la variation de la lumière émise, soit en intensité, soit en couleur. Un tel changement sera appelé changement interne de la source de lumière. La deuxième source de changement est le déplacement de la ou des sources de lumières éclairant la scène, elle sera appelée changement externe de la source, et fait l'objet du paragraphe 3.2.

#### 3.1 Modèles pour les changements internes de la source

##### 3.1.1 Quels modèles et comment les évaluer ?

Pour un changement interne de la source, divers modèles ont déjà été proposés : ces modèles décrivent comment est transformé le vecteur  ${}^t(r, g, b)$  des trois composantes de la couleur en chaque pixel de l'image. De nombreux travaux, dont ceux de Finlayson reposent sur le modèle diagonal, mais sans pour autant justifier ce choix, ni comparer ce modèle à d'autres modèles possibles.

Dans la présente étude, nous avons testé divers modèles. Pour chacun d'eux, on indique comment un pixel  $\mathbf{p} = {}^t(r, g, b)$  est transformé en un nouveau vecteur  $\mathbf{p}' = {}^t(r', g', b')$ . On note  $\mathbf{t} = {}^t(t, t, t)$  où  $t$  est un scalaire,  $\mathbf{T} = {}^t(t_x, t_y, t_z)$  un vecteur quelconque de translation,  $\mathbf{D} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}$  une matrice diagonale, et  $\mathbf{M}$  une matrice de taille  $3 \times 3$  quelconque.

Modèle n° 1 :	$\mathbf{p}' = \mathbf{p}$	identité
Modèle n° 2 :	$\mathbf{p}' = \mathbf{p} + \mathbf{t}$	translation simple
Modèle n° 3 :	$\mathbf{p}' = \mathbf{p} + \mathbf{T}$	translation
Modèle n° 4 :	$\mathbf{p}' = \alpha \mathbf{p}$	scalaire
Modèle n° 5 :	$\mathbf{p}' = \alpha \mathbf{p} + \mathbf{t}$	scalaire avec translation simple
Modèle n° 6 :	$\mathbf{p}' = \alpha \mathbf{p} + \mathbf{T}$	scalaire avec translation
Modèle n° 7 :	$\mathbf{p}' = \mathbf{Dp}$	diagonal
Modèle n° 8 :	$\mathbf{p}' = \mathbf{Dp} + \mathbf{t}$	diagonal avec translation simple
Modèle n° 9 :	$\mathbf{p}' = \mathbf{Dp} + \mathbf{T}$	diagonal avec translation
Modèle n° 10 :	$\mathbf{p}' = \mathbf{Mp}$	linéaire
Modèle n° 11 :	$\mathbf{p}' = \mathbf{Mp} + \mathbf{t}$	linéaire avec translation simple
Modèle n° 12 :	$\mathbf{p}' = \mathbf{Mp} + \mathbf{T}$	linéaire avec translation

Le but de l'expérimentation est premièrement de classer ces modèles en fonction des résultats qu'ils fournissent sur des images réelles. Il est sûr que les modèles qui ont plus de paramètres fournissent de meilleures approximations, mais un tel test permet toutefois de décider, par exemple, si un modèle linéaire est plus ou moins réaliste qu'un modèle diagonal avec translation, ce qui n'est pas évident de prime abord, un de ces modèles n'étant pas un simple cas particulier de l'autre.

Pour affiner cette étude, nous avons ensuite cherché à déterminer, pour chacun de ces modèles, quels paramètres sont vraiment nécessaires. Pour cela, on détermine si chaque paramètre a une valeur statistiquement différente de zéro, ou si, au contraire, zéro est une aussi bonne approximation de la valeur théorique que celle fournit par le calcul d'estimation. Cela permet de juger si les paramètres estimés sont caractéristiques du signal ou bien du bruit.

**Méthode d'évaluation des modèles** Pour évaluer chaque modèle, nous avons pris un ensemble d'images présentant une même scène: la caméra et la source de lumière était fixe, mais nous avons fait varier l'intensité de cette source. La figure 1 présente, par exemple, une série de 6 images de la même scène plane avec une simple variation de l'intensité de l'illumination.

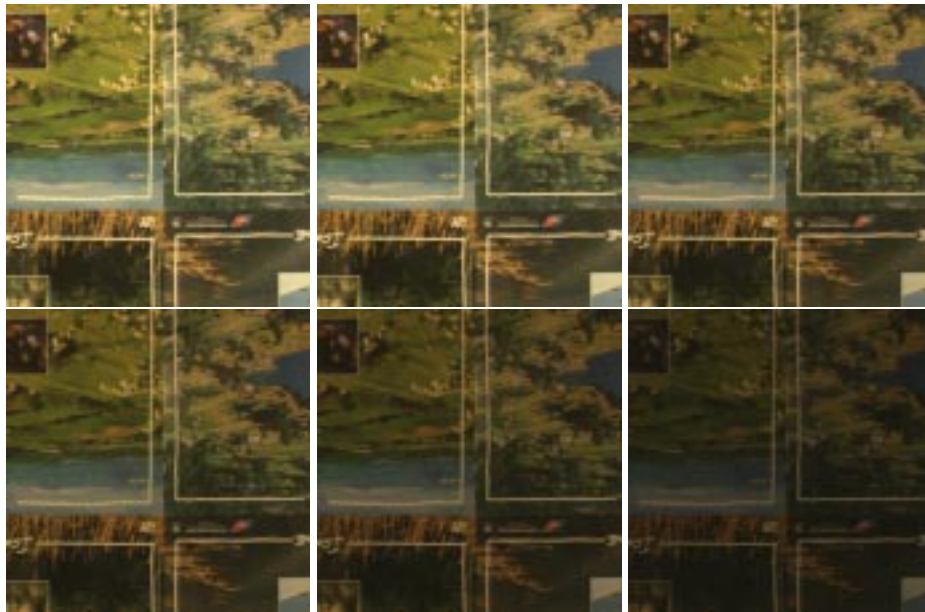


FIG. 1: 6 images avec variation de l'intensité lumineuse.

Nous avons alors calculé les paramètres du modèles de diverses manières. Dans chaque image, nous avons soit utilisé tous les pixels (APS = avec pixels saturés), soit seulement ceux qui ne sont pas saturés (c'est à dire dont chaque composante est incluse dans [2, 253]) (SPS = sans pixels saturés). Par ailleurs, nous avons utilisés deux algorithmes d'estimation : la méthode de décomposition en valeurs singulières, qui fournit une estimation linéaire aux moindres carrés des paramètres des modèles (MC), soit une version robuste de cet algorithme, dite moindres carrés médians (MCM). On a alors calculé la moyenne, le maximum et le médian des erreurs entre les pixels d'une image et ceux de l'autre image corrigée par le modèle estimé.

**Test des paramètres estimés.** Il est prévisible, et cela se vérifie dans les résultats, que les modèles ayant plus de paramètres conduisent à des erreurs plus faibles. La question est de savoir si ces paramètres captent bien de l'information et pas seulement du bruit. Pour cela, nous avons utilisé la méthode proposée par Florou [FM96]. (Il en existe d'autres, cf. [Tor97].)

Cette méthode est basée sur un test statistique. On suppose que les erreurs ont une moyenne nulle. On calcule alors des régions de confiance pour chaque paramètres, c'est à dire des intervalles autour des paramètres estimés, tels que l'on soit sûr à 90% ou 95% qu'ils contiennent la valeur réelle du paramètre.

La formule donnant le rayon de l'intervalle de confiance, pour un taux de confiance de  $\alpha\%$  est :

$$\sigma^2(p_i) = \sqrt{\chi^2(\alpha, m)} \sqrt{\frac{f}{n}} \sqrt{C_{ii}}$$

Dans cette formule :

- 1°  $\chi^2(\alpha, m)$  est le quantile de la distribution  $\chi^2$  pour un taux de confiance de  $\alpha\%$  et  $m$  degrés de liberté; ce nombre est le nombre de paramètres du modèle estimé;
- 2°  $f$  est la somme des erreurs sur tous les pixels, et  $n$  est le nombre de pixels de chaque image;
- 3°  $C_{ii}$  est la variance du paramètre estimé  $p_i$ .

Le calcul de cette dernière quantité est simple : les deux algorithmes SVD et Levenberg Marquardt, qui sont très utilisés pour l'estimation linéaire et non linéaire des paramètres fournissent cette variance comme sous produit. Dans notre cas, une fois la décomposition SVD de la matrice des données faite, sous forme  $\mathbf{A} = \mathbf{U}\mathbf{w}^t\mathbf{V}$ , la variance du paramètre  $p_i$  est simplement :

$$C_{ii} = \sum_{j=1}^m \left( \frac{V_{ij}}{w_j} \right)^2$$

Si un paramètre  $p_i$  est compris entre  $-\sigma^2(p_i)$  et  $\sigma^2(p_i)$ , on peut alors estimer que 0, qui appartient à la région de confiance aurait fourni une estimation aussi valable, et le paramètre  $p_i$  est jugé non significatif. Dans le cas contraire, il est jugé significatif.

Au delà de ce test binaire, le rapport  $p_i/\sqrt{\sigma^2(p_i)}$  fourni un indicateur du degré de signification des tous les paramètres estimés.

### 3.1.2 Quelques expérimentations

**Comparaison des méthodes de calcul des modèles sur deux images.** Dans un premier temps, nous n'avons utilisé que les deux premières images de la série de 6 présentée à la figure 1. Nous avons testé les 4 algorithmes (APS + MC, APS + MCM, SPS + MC, SPS + MCM) pour les 12 modèles proposés. Pour chaque calcul, nous avons calculés la moyenne, le médian et le maximum des erreurs sur les pixels, nous avons estimés les paramètres et vérifié s'ils sont significatifs pour des taux de confiance de 95%.

Modèle	Méthode	Moyenne Médian Maximum	Paramètres $p_i$	$\sigma^2(p_i)$
1		13.6156 13.3041 44.2041		
2	APS + MC	7.63405 7.09989 37.3716	$t=-6.90471$	0.00512899
	SPS + MC	7.59494 7.06818 37.2926	$t=-6.85907$	0.00610538
	APS + MCM	7.63405 7.09989 37.3716	$t=-6.90471$	0.00610538
	SPS + MCM	7.59494 7.06818 37.2926	$t=-6.85907$	0.00612177
3	APS + MC	7.63405	$t_x=-7.53739$	0.0148631
		7.09989	$t_y=-7.81116$	0.0148631
		37.3716	$t_z=-5.36557$	0.0148631
	SPS + MC	7.59494	$t_x=-7.5462$	0.0148973
		7.06818	$t_y=-7.73259$	0.0148973
		37.2926	$t_z=-5.29844$	0.0148973
	APS + MCM	7.41492	$t_x=-7.53739$	0.0148631
		6.83917	$t_y=-7.81116$	0.0148631
37.4608		$t_z=-5.36557$	0.0148631	
SPS + MCM	7.37127	$t_x=-7.5462$	0.0148973	
	6.8071	$t_y=-7.73259$	0.0148973	
	37.3836	$t_z=-5.29844$	0.0148973	
4	APS + MC	6.28593	$\alpha=0.930704$	5.17043e-05

Modèle	Méthode	Moyenne Médian Maximum	Paramètres $p_i$	$\sigma^2(p_i)$
		5.75006 33.1615		
	SPS + MC	6.2232 5.72137 33.3195	$\alpha=0.929811$	5.25978e-05
5	APS + MC	6.22393 5.68836 33.3881	$\alpha=0.941218$ $t=-1.25613$	0.000145241 0.0155626
	SPS + MC	6.17188 5.66493 33.4773	$\alpha=0.9389$ $t=-1.06369$	0.000150302 0.015835
6	APS + MC	6.16254 5.61692 33.591	$\alpha=0.9358$ $t_x=0.0305361$ $t_y=-0.868263$ $t_z=-1.36889$	0.000211725 0.0290863 0.027338 0.0199204
	SPS + MC	6.10906 5.59438 33.6832	$\alpha=0.932691$ $t_x=0.290919$ $t_y=-0.541783$ $t_z=-1.17335$	0.000150302 0.0297677 0.0279523 0.020176
7	APS + MC	6.13422 5.5932 33.5231	$\alpha=0.937683$ $\beta=0.9283$ $\gamma=0.913997$	0.000106157 0.000116876 0.000194164
	SPS + MC	6.08232 5.568 33.643	$\alpha=0.93611$ $\beta=0.928035$ $\gamma=0.913476$	0.000107991 0.000118921 0.000198494
8	APS + MC	6.12493 5.58598 33.599	$\alpha=0.941708$ $\beta=0.932776$ $\gamma=0.921109$ $t=-0.553835$	0.000196537 0.000217789 0.000351607 0.021732
	SPS + MC	6.07727 5.56259 33.677	$\alpha=0.938552$ $\beta=0.930752$ $\gamma=0.917819$ $t=-0.329332$	0.000203719 0.000225877 0.000366471 0.0223
9	APS + MC	6.11995 5.58589 33.5543	$\alpha=0.947401$ $\beta=0.931968$ $\gamma=0.914737$ $t_x=-1.33695$ $t_y=-0.453836$ $t_z=-0.0576278$	0.000355988 0.000418015 0.000550085 0.0453338 0.0483505 0.0382997

Modèle	Méthode	Moyenne Médian Maximum	Paramètres $p_i$	$\sigma^2(p_i)$
	SPS + MC	6.07466 5.56175 33.6486	$\alpha=0.941923$ $\beta=0.931121$ $\gamma=0.913186$ $t_x=-0.784023$ $t_y=-0.374063$ $t_z=0.0220638$	0.000370861 0.000438366 0.000575203 0.0464724 0.0498828 0.0392142
10	APS + MC	6.12315 5.58288 33.4262	$a_{11}=0.930345$ $a_{12}=0.026941$ $a_{13}=-0.005546$ $a_{21}=0.0072958$ $a_{22}=0.893035$ $a_{23}=0.0114265$ $a_{31}=0.0014726$ $a_{32}=0.0100464$ $a_{33}=0.905289$	0.00137736 0.00137736 0.00137736 0.00186245 0.00186245 0.00186245 0.00114226 0.00114226 0.00114226
	SPS + MC	6.07104 5.55503 33.5326	$a_{11}=0.928524$ $a_{12}=0.0264167$ $a_{13}=-0.006095$ $a_{21}=0.0111645$ $a_{22}=0.893896$ $a_{23}=0.0123249$ $a_{31}=-0.004829$ $a_{32}=0.0091549$ $a_{33}=0.904241$	0.00137539 0.00137539 0.00137539 0.00137539 0.00137539 0.00137539 0.00114597 0.00114597 0.00114597
11	APS + MC	6.10746 5.5735 33.4911	$a_{11}=0.92865$ $a_{12}=0.0252459$ $a_{13}=-0.007241$ $a_{21}=0.0166187$ $a_{22}=0.902358$ $a_{23}=0.0207494$ $a_{31}=-8.63728e-05$ $a_{32}=0.0084874$ $a_{33}=0.90373$ $t=-0.812391$	0.00143318 0.00143318 0.00143318 0.00197407 0.00197407 0.00197407 0.00118888 0.00118888 0.00118888 0.033807
	SPS + MC	6.06153 5.5455 33.5603	$a_{11}=0.927537$ $a_{12}=0.02543$ $a_{13}=-0.00708184$ $a_{21}=0.0171308$ $a_{22}=0.899862$	0.0014315 0.0014315 0.0014315 0.00197233 0.00197233

Modèle	Méthode	Moyenne Médian Maximum	Paramètres $p_i$	$\sigma^2(p_i)$
			$a_{23}=0.0182912$ $a_{31}=-0.00554763$ $a_{32}=0.00843657$ $a_{33}=0.903523$ $t=-0.542308$	0.00197233 0.00119245 0.00119245 0.00119245 0.034802
12	APS + MC	6.10165 5.56994 33.4797	$a_{11}=0.92687$ $a_{12}=0.0261926$ $a_{13}=-0.00640779$ $a_{21}=0.0264103$ $a_{22}=0.897151$ $a_{23}=0.0161648$ $a_{31}=-0.00172377$ $a_{32}=0.00935812$ $a_{33}=0.904496$ $t_x=-1.66562.$ $t_y=-0.35866$ $t_z=-0.412892$	0.00143318 0.00143318 0.00143318 0.00197407 0.00197407 0.00197407 0.00118888 0.00118888 0.00118888 0.033807 0.033807 0.033807
	SPS + MC	6.06153 5.5455 33.5603	$a_{11}=0.926659$ $a_{12}=0.0258903$ $a_{13}=-0.00666419$ $a_{21}=0.0224396$ $a_{22}=0.897079$ $a_{23}=0.0157657$ $a_{31}=-0.00618681$ $a_{32}=0.00877168$ $a_{33}=0.903827$ $t_x=-1.02486$ $t_y=-0.289313$ $t_z=-0.312755$	0.00153612 0.00153612 0.00153612 0.0021905 0.0021905 0.0021905 0.00127901 0.00127901 0.00127901 0.0645579 0.0645579 0.0645579

Plusieurs enseignements peuvent être tirés de ces données. Tout d'abord, l'apport de la méthode des moindres carrés médians est du même ordre que la suppression des points saturés, avec toutefois un léger avantage pour la première. L'élimination des points saturés ne supprime en effet qu'une des causes d'aberration dans les données. À l'inverse, une méthode robuste peut en considérer d'autres comme la défectuosité d'un CCD qui renverrait un signal aléatoire ou constant. C'est pourquoi, nous n'avons mis les résultats avec les quatre méthodes que pour les modèles 2 et 3. Il est sûr que la méthode avec moindres carrés médians sans points saturés fournit les meilleurs résultats.

Le test de pertinence des modèles n'apporte que peu d'information : tous les paramètres sont jugés significatifs. Cela est dû au nombre de points utilisés pour calculés ces paramètres. Lorsque l'on utilise tous les points d'une image  $512 \times 512$ , cela en fait 262 144, et l'on conçoit bien qu'estimer 12 paramètres sur autant de données fournit des résultats précis. Il pourrait ne pas en être de même avec une image fortement saturée, où le nombre de points utilisables serait réduit. On peut toutefois voir, avec le modèle 12, que les paramètres non diagonaux sont moins significatifs que les paramètres de translation.

Pour la comparaison des modèles, certains choix apparaissent clairement : le modèle 4 (scalaire) fournit de plus petites erreurs que le modèle 3 (translation), le modèle 7 (diagonal) en fournit de plus petites que le modèle 6 (scalaire avec translation). Par contre, le choix entre les modèles 9 (diagonale avec translation) et 10 (linéaire) n'est pas évident, et dépend de la considération ou non des points saturés.

**Comparaison des modèles sur 6 images.** Pour compléter cette première série d'expériences, nous avons calculés les modèles entre la première image et chacune des 5 autres. Nous ne donnons que les résultats obtenus avec la méthode SPS + MCM qui fournit les meilleurs résultats. Le tableau qui suit reprend les colonnes 1 et 3 du tableau précédent, en n'utilisant que la méthode SPS + MCM, mais pour 5 paires d'images différentes.

Modèle	Moyenne, médian et maximum des erreurs entre deux images				
	Images 1 et 2	Images 1 et 3	Images 1 et 4	Images 1 et 5	Image 1 et 6
1	13.5362	38.6435	60.5171	83.4653	108.794
	13.1909	38.3275	60.1664	83.1204	107.75
	44.2041	94.3769	141.128	186.786	247.348
2	7.59494	15.3892	24.0688	35.0397	49.4871
	7.06818	14.4009	22.6872	33.1939	46.6429
	37.2926	57.3137	83.0569	107.255	145.529
3	7.37127	13.8359	20.8556	29.5884	40.9449
	6.8071	12.3306	18.3124	25.94	36.1986
	37.3836	56.8309	81.829	104.864	140.55
4	6.2232	6.98909	7.44423	7.29016	6.86809
	5.72137	6.57889	7.028	6.79211	6.10553
	33.3195	35.3157	34.5089	35.3656	39.7315
5	6.17188	6.61222	7.00892	7.17807	6.5142
	5.66493	6.15432	6.56571	6.66285	5.89518
	33.4773	35.4115	35.1389	35.4988	35.1536
6	6.10906	6.26885	6.53342	6.7874	6.35143
	5.59438	5.78016	6.05672	6.282	5.76897
	33.6832	34.9332	33.0012	33.7218	33.7024
7	6.08232	6.14477	6.27132	6.39796	6.81627
	5.568	5.67147	5.80588	5.88565	6.04758
	33.643	34.8631	31.5682	33.1802	39.7918

Modèle	Moyenne, médian et maximum des erreurs entre deux images				
	Images 1 et 2	Images 1 et 3	Images 1 et 4	Images 1 et 5	Image 1 et 6
8	6.07727	6.07263	6.20241	6.39816	6.19346
	5.56259	5.58556	5.72385	5.89675	5.62311
	33.677	34.1654	32.1378	33.0248	33.5508
9	6.07466	6.0483	6.13661	6.2572	6.09289
	5.56175	5.56269	5.65976	5.73528	5.51266
	33.6486	33.9358	32.1776	33.223	33.3612
10	6.07104	6.09724	6.20927	6.29894	6.63859
	5.55503	5.62512	5.74105	5.78087	5.82957
	33.5326	34.5064	31.6982	34.1667	39.7582
11	6.06153	6.03857	6.15636	6.29847	6.0787
	5.5455	5.553	5.67788	5.79797	5.51805
	33.5603	33.7033	32.4158	33.964	33.0951
12	6.05874	6.02175	6.10728	6.20183	6.01641
	5.54143	5.53818	5.62793	5.68712	5.45086
	33.5563	33.5722	32.1895	33.8315	32.9997

L'avantage de cette deuxième série de données est qu'elle comporte des images plus contrastées : les images les moins lumineuses ont des pixels à valeur plus rapprochées. D'autre part, on peut tester la dégradation des résultats quand la différence d'illumination croît.

Première constatation, les méthodes n'ayant qu'une translation ne fournissent pas de bons résultats, et ces résultats se dégradent nettement avec des images très différentes. Il n'en est pas de même avec les autres, même avec la simple méthode 4 (scalaire) qui n'a qu'un coefficient. Nous éliminons ces modèles (1, 2 et 3) pour la suite de l'analyse.

Deuxième point, dans chacun des groupes de modèles scalaires (4, 5 et 6), diagonaux (7, 8 et 9) ou linéaires (10, 11 et 12), l'importance des termes de translation croît avec la différence entre images. Pour les images 1 et 2, la présence de coefficients de translation n'a guère d'influence sur les erreurs, alors que cette influence est nette entre les images 1 et 6.

Si l'on considère maintenant les groupes de modèles ayant les mêmes paramètres de translation, c'est à dire les groupes  $\{4, 7, 10\}$ ,  $\{5, 8, 11\}$  et  $\{6, 9, 12\}$ , on peut étudier l'influence du nombre de paramètres dans la partie linéaire. Comme précédemment, cette influence devient plus nette avec des images plus différentes. Les modèles à un seul paramètres connaissent une dégradation de performances, et leur différence avec les autres modèles croît quand la différence entre les images croît.

Par contre, s'il y a bien une différence entre modèles diagonaux et modèles linéaires, cette différence est faible (plus faible par exemple que celle existant entre modèles diagonaux et modèles scalaires), et reste constante quelle que soit la différence entre image.

Dernière série de comparaison, vaut-il mieux avoir des paramètres de translation et moins de paramètres dans la partie linéaire, ou l'inverse ? Le modèle 9 (diagonal avec translation) donne de meilleurs résultats que le modèle 10 (linéaire), sauf pour les image 1 et 2. Il donne des résultats équivalents au modèle 11 (linéaire avec translation simple), avec un léger

avantage au modèle 9 pour les paires d'images les plus différentes. L'avantage du modèle 12 (linéaire avec translation) est peu surprenante, mais ne concerne que la deuxième décimale des erreurs.

En conclusion, nous pouvons retenir que le modèle 9 (diagonal avec translation) fournit un excellent ratio qualité / complexité. Il faut ajouter 6 paramètres supplémentaires, dont certains ne sont pas significatifs sur de petites images, pour obtenir de meilleurs résultats. On peut donc retenir ce modèle 9 comme étalon, et c'est lui dont nous nous servirons au paragraphe 5.

Par rapport à cet étalon, le modèle préconisé par Finlayson apparaît bon, tant que les images ne sont pas trop différentes.

**Dégénérescence des modèles.** La dernière série d'expériences réalisée met en évidence la dégénérescence des modèles. Nous avons noté, dans les premières expériences, que tous les paramètres estimés pour tous les modèles sont significatifs à cause de la grande quantité de données sur lesquels ils sont estimés. Que se passe-t-il quand le nombre de données diminue?

Dans cette expérience, nous avons considéré les deux images extrêmes de la séquence de la figure 1 et nous avons extrait de ces deux images des sous-images de taille croissante. Pour chaque paire de sous-images, nous avons estimé le modèle 12 par la méthode MCM + SPS, puis nous avons testé la signification des paramètres estimés.

Dans le tableau suivant, nous indiquons, pour chaque taille de sous-image, les paramètres qui sont significatifs (S) et ceux qui ne le sont pas (N).

Taille	$a_{11}$	$a_{12}$	$a_{13}$	$a_{21}$	$a_{22}$	$a_{23}$	$a_{31}$	$a_{32}$	$a_{33}$	$t_x$	$t_y$	$t_z$
5	S	N	N	N	N	N	N	N	N	N	N	N
10	S	N	N	N	N	N	N	N	N	N	N	N
15	S	N	N	N	N	N	N	S	N	N	S	S
20	S	N	N	S	S	N	N	S	N	N	S	S
25	S	S	N	N	S	N	N	N	N	N	S	S
30	S	N	N	N	S	N	N	N	S	S	S	S
35	S	N	S	N	S	S	N	N	S	S	S	S
40	S	N	N	N	S	S	N	N	S	S	S	S
45	S	N	N	N	S	N	N	N	S	S	S	S
50	S	N	N	N	S	S	N	N	S	S	S	S
60	S	S	N	S	S	N	N	N	S	S	S	S
70	S	S	N	S	S	S	N	N	S	S	S	S
80	S	S	N	S	S	N	N	N	S	S	S	S
90	S	N	N	N	S	N	S	S	S	S	S	S
100	S	N	N	S	S	S	S	S	S	S	S	S
125	S	S	S	S	S	S	S	S	S	S	S	S
150	S	S	S	S	S	S	S	S	S	S	S	S

Dans ces résultats, il ne faut pas négliger le fait que nous n'avons travaillé qu'avec deux images particulières: la composition de ces images, en particulier la présence prépondérante d'une des trois couleurs peut expliquer la dissymétrie observée entre les paramètres relatifs aux divers canaux.

Le résultat qui apparaît le plus clairement est que les paramètres diagonaux et ceux de translation sont bien plus significatifs que les paramètres non diagonaux. Le calcul exact et pertinent d'un modèle linéaire n'est donc possible que lorsque l'on dispose de très grandes images. Dans le cas où l'on calcule des invariants, et où l'on ne s'intéresse qu'à de petites régions de l'image (cf. le paragraphe 5), il n'est donc pas nécessaire d'utiliser un modèle linéaire.

En ce qui concerne les paramètres de translation, ceux-ci apparaissent significatifs même pour de petites sous-images. Cela ajoute une réserve supplémentaire par rapport au modèle préconisé par Finlayson, qui n'apparaît donc justifié que pour de très petites images.

### 3.2 Changement de la position et de l'orientation de la source de lumière

Le paragraphe précédent était consacré à l'étude des modèles prenant en compte le cas où la source de lumière changeait d'intensité ou de couleur. Un autre changement possible est celui de la position et de l'orientation de cette source de lumière, que nous avons appelé changement externe. On peut soit supposer que seules la position et l'orientation changent, ou que ce changement est couplé à un changement d'intensité ou de couleur.

Dans le cas du changement interne, Finlayson utilise le modèle diagonal: chaque composante  $r$ ,  $g$ ,  $b$  du signal est indépendamment multipliée par un facteur. Un pixel  $(r, g, b)$  est transformé en  $(\alpha r, \beta g, \gamma b)$ . Pour trouver le modèle de transformation entre ces deux images, il suffit donc de trouver un coefficient multiplicatif pour chacun des trois canaux des images.

Dans le cas d'un déplacement de la source, il suppose que chaque pixel est multiplié par un facteur. Ainsi, un pixel  $(r, g, b)$  est transformé en  $\delta(r, g, b)$ . Le facteur multiplicatif dépend de la position et de l'orientation de la source de lumière par rapport aux deux positions successives de cette source et à celles de la caméra, et varie donc de pixel en pixel.

Le modèle de transformation entre les deux images est donc composé, dans ce cas, d'un coefficient multiplicatif pour chaque pixel de l'image.

## 4 Normalisation des images

Il y a deux façons d'exploiter les modèles présentés à la section précédente. Soit on calcule des descripteurs des images qui sont invariants aux transformations utilisées pour le modèle, c'est ce qui est fait dans la section suivante, soit on normalise les images. Cette normalisation consiste à transformer chaque image de telle sorte que le résultat de cette transformation soit indépendant des paramètres du modèle étudié.

Appelons images similaires toutes les images qui peuvent se déduire l'une de l'autre par un modèle de transformation tel que ceux étudiés précédemment. Une telle relation de simi-

larité entre les images est une relation d'équivalence entre images. Calculer un descripteur invariant pour une image revient à calculer un descripteur non par pour cette image, mais pour la classe d'équivalence de cette image. Si ce descripteur peut être calculé de la même manière pour toutes les images, et varie d'une classe à l'autre, il permet de retrouver à quelle classe appartient une image.

Dans ce cadre, normaliser une image consiste à choisir une image particulière dans chaque classe comme représentante de cette classe d'équivalence pour la relation de similarité. Le choix de ce représentant doit être fait de sorte qu'on puisse le retrouver ou le calculer à partir de toute image de sa classe.

Par exemple, si l'on étudie le problème du déplacement de la source, on peut calculer une image normalisée en multipliant chaque pixel par un facteur tel que le résultat vérifie  $r + g + b = 3$ . L'image normalisée obtenue peut alors être calculée de manière analogue à partir de toute autre image prise dans les mêmes conditions que la première, sauf un éventuel changement de position de la source de lumière, dans la mesure où le modèle utilisé est correct.

Pour un changement interne de la source, il faut normaliser chaque canal. Supposons que l'image soit constituée d'un ensemble de  $n$  pixels  $(r_i, g_i, b_i)$ . On peut alors multiplier toutes les composantes rouges des pixels de l'image par l'unique facteur  $f_r = n / \sum_i r_i$ . De même pour les composantes vertes et bleues. On obtient alors une image qui ne dépend plus de la couleur et de l'intensité de la source de lumière, dans la mesure où le modèle utilisé est juste bien entendu.

Un problème est de réaliser une normalisation conjointe pour les deux modèles à la fois. Il s'agit donc, à partir d'une image  $(r_i, g_i, b_i)$ , de calculer une image  $(r'_i, g'_i, b'_i)$  telle que :

$$\forall i \quad r'_i + g'_i + b'_i = 3 \quad \text{et} \quad \sum_i r'_i = \sum_i g'_i = \sum_i b'_i = n$$

#### 4.1 Quelques méthodes de normalisation

**Méthode itérative de Finlayson.** Finlayson utilise une méthode itérative. Chaque itération comporte deux étapes. Dans la première, il normalise l'image pour le changement interne de la source, et dans la deuxième il normalise le résultat de l'étape précédente pour le changement externe. Il répète alors le processus jusqu'à convergence. Le processus est assez rapide et converge en quelques itérations.

Soit  $(r_i^{(p)}, g_i^{(p)}, b_i^{(p)})$  un pixel de l'image à l'étape  $p$ . La valeur du pixel à l'étape  $p+1$  est :

$$r_i^{(p+1)} = \frac{nr_i^{(p)}}{r_i^{(p)} + g_i^{(p)} + b_i^{(p)}} \left( \sum_{j=1}^n \frac{r_j^{(p)}}{r_j^{(p)} + g_j^{(p)} + b_j^{(p)}} \right)^{-1} \quad (18)$$

$$g_i^{(p+1)} = \frac{ng_i^{(p)}}{r_i^{(p)} + g_i^{(p)} + b_i^{(p)}} \left( \sum_{j=1}^n \frac{g_j^{(p)}}{r_j^{(p)} + g_j^{(p)} + b_j^{(p)}} \right)^{-1} \quad (19)$$

$$b_i^{(p+1)} = \frac{nb_i^{(p)}}{r_i^{(p)} + g_i^{(p)} + b_i^{(p)}} \left( \sum_{j=1}^n \frac{b_j^{(p)}}{r_j^{(p)} + g_j^{(p)} + b_j^{(p)}} \right)^{-1} \quad (20)$$

**Méthode directe pour deux pixels.** On peut trouver une méthode directe pour réaliser cette même normalisation. Plaçons nous dans le cas d'une image de deux pixels. La transformation subie par cette image lors d'un changement interne et externe de la source peut être modélisée par l'équation suivante :

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} r_1 & g_1 & b_1 \\ r_2 & g_2 & b_2 \end{pmatrix} \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} = \begin{pmatrix} r'_1 & g'_1 & b'_1 \\ r'_2 & g'_2 & b'_2 \end{pmatrix}$$

Un argument de comptage simple permet de voir qu'il existe deux invariants indépendants pour cette transformation : l'image a six paramètres indépendants, la transformation a 5 paramètres dont seuls 4 sont indépendants, d'où le nombre d'invariants indépendants. On peut trouver facilement trois invariants :

$$A = \frac{r_1 g_2}{r_2 g_1} \quad B = \frac{r_1 b_2}{r_2 b_1} \quad C = \frac{g_1 b_2}{g_2 b_1}$$

L'image normalisée recherchée est donc :

$$\begin{pmatrix} r'_1 & g'_1 & b'_1 \\ r'_2 & g'_2 & b'_2 \end{pmatrix} \text{ telle que } \begin{cases} r'_1 + g'_1 + b'_1 = 3 \\ r'_1 + g'_2 + b'_2 = 3 \\ r'_1 + r'_2 = 2 \\ g'_1 + g'_2 = 2 \\ b'_1 + b'_2 = 2 \\ \frac{r'_1 g'_2}{r'_2 g'_1} = A \\ \frac{r'_1 b'_2}{r'_2 b'_1} = B \end{cases}$$

Les contraintes fournissent un ensemble d'équations qui définissent complètement les paramètres cherchés. Pour les trouver, on peut faire le changement de variables suivant :

$$x = r'_1 - 1 \quad y = g'_1 - 1 \quad z = b'_1 - 1$$

On est alors ramené à résoudre un système de trois équations :

$$\begin{cases} x + y + z = 0 \\ \frac{1+x}{1-x} \frac{1-y}{1+y} = A \\ \frac{1+x}{1-x} \frac{1-z}{1+z} = B \end{cases}$$

On peut éliminer  $z$  par  $z = -x - y$ , puis  $y$  grâce à :  $y = \frac{1-A+x(1+A)}{1+A+x(1-A)}$ . Cela mène à une unique équation du troisième degré en  $x$  :

$$2 - 2AB + x(5 + A + 5AB + B) + x^2(4 - 4AB) + x^3(1 - A - B + AB) = 0$$

Cette équation a deux solutions complexes conjuguées et une solution réelle qui est celle du problème posé. On reportant cette valeur dans les équations précédentes, on trouve l'image normalisée.

**Méthode directe pour  $n$  pixels.** Dans le cas, le plus courant dans la réalité, où l'image comporte plus de deux pixels, les équations de la transformation deviennent :

$$\begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix} \begin{pmatrix} r_1 & g_1 & b_1 \\ r_2 & g_2 & b_2 \\ \vdots & \vdots & \vdots \\ r_n & g_n & b_n \end{pmatrix} \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} = \begin{pmatrix} r'_1 & g'_1 & b'_1 \\ r'_2 & g'_2 & b'_2 \\ \vdots & \vdots & \vdots \\ r'_n & g'_n & b'_n \end{pmatrix}$$

Dans ce cas, l'image a  $3n$  paramètres, la transformation en a  $n+3$  dont  $n+2$  indépendants, ce qui mène à  $3n - n - 2 = 2n - 2$  invariants indépendants. Ces invariants peuvent être choisis parmi l'ensemble des invariants suivants :

$$A_i = \frac{r_1 g_i}{r_i g_1} \quad B_i = \frac{r_1 b_i}{r_i b_1} \quad C_i = \frac{g_1 b_i}{g_i b_1}$$

Pour la résolution, on part donc d'un ensemble de  $3(n-1)$  équations. Mais notre choix d'invariants permet de se ramener à la résolution de  $n-1$  problèmes à 2 pixels (après changement de variables), que l'on peut résoudre de la manière précédente. C'est là la limite de la méthode directe: elle mène à des calculs assez complexes, une équation du troisième degré dont la résolution théorique est simple, mais dont la solution a une stabilité probablement assez faible, et elle traite les pixels les uns après les autres sans permettre de calcul global.

**Normalisation locale.** La normalisation pour les changements externes de la source de lumière est purement locale: chaque pixel est traité successivement. Par contre, en ce qui concerne les changements internes, la normalisation est appliquée sur chaque canal, mais

globalement à toute l'image. Cette normalisation est donc sensible à tout changement de composition de l'image. Cela peut être un problème lorsque l'on étudie des propriétés locales.

Pour éviter ce problème, on peut utiliser une méthode plus locale. Pour chaque pixel de l'image, on considère une imagerie centrée autour de ce point. On normalise cette imagerie. La valeur centrale de l'imagerie normalisée est alors retenue comme valeur finale pour le pixel. Cette méthode admet un seul paramètre: la taille des images à considérer.

## 4.2 Quelques résultats

Voici quelques résultats de normalisation. On part des quatre images de la figure 2. Les trois premières diffèrent par un changement d'illumination de la source de lumière, les trois et quatrième par un déplacement de cette source. Ces images présentent des difficultés: la source n'est pas unique, le fond de la scène étant éclairé par une autre source. D'autre part, il y a une couleur fortement dominante: le jaune.

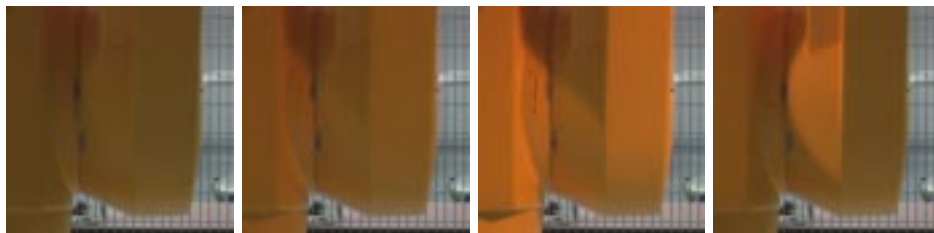


FIG. 2: 4 images avec variation de l'intensité et de la position de la source de lumière.

La matrice des médianes des distances entre les pixels de ces images est la suivante:

$$\begin{pmatrix} 0,0 & 17,54 & 43,70 & 12,20 \\ 17,54 & 0,0 & 29,27 & 14,62 \\ 43,70 & 29,27 & 0,0 & 42,26 \\ 12,20 & 14,62 & 42,26 & 0,0 \end{pmatrix}$$

Nous présentons maintenant la normalisation de ces images selon diverses méthodes et les matrices des médianes des distances entre pixels.

**Normalisation au changement externe de la source.** Les images sont sur la figure 3.

La matrice des médianes après normalisation est:

$$\begin{pmatrix} 0,0 & 15,13 & 27,36 & 12,84 \\ 15,13 & 0,0 & 13,74 & 12,32 \\ 27,36 & 13,74 & 0,0 & 19,10 \\ 12,84 & 12,32 & 19,10 & 0,0 \end{pmatrix}$$

Cette normalisation donne à chaque pixel la même intensité. Les variations observées entre les trois premières images sont dues au fait que le changement de la source était interne entre les prises de vues correspondantes. La normalisation appliquée n'a pas pour but de supprimer ces différences. Il n'en est pas de même entre les deux dernières, pour lesquelles le changement de la source de lumière était purement externe. Les différences qui subsistent montrent que le modèle n'est bien qu'une approximation de la vérité.

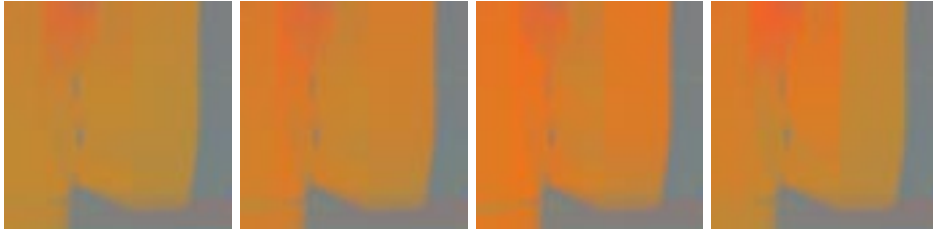


FIG. 3: Les 4 images normalisées vis à vis d'un changement externe de la source de lumière.

#### Normalisation globale au changement interne de la source de lumière.

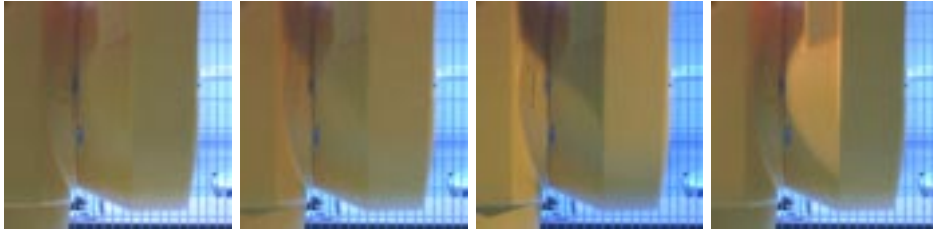


FIG. 4: Les 4 images normalisées globalement vis à vis d'un changement interne de la source de lumière.

Les images sont montrées sur la figure 4. Le caractère global de la normalisation est clair : l'objet jaune rend le fond bleu, pour que les quantités de rouge, vert et bleu soient égales dans l'image. La matrice des médianes est, dans ce cas :

$$\begin{pmatrix} 0,0 & 20,39 & 37,16 & 21,67 \\ 20,39 & 0,0 & 23,68 & 21,40 \\ 37,16 & 23,68 & 0,0 & 35,86 \\ 21,67 & 21,40 & 35,86 & 0,0 \end{pmatrix}$$

#### Normalisation globale au changement interne et externe de la source de lumière.

On a appliqué cette fois une normalisation pour les deux changements, par la méthode

itérative de Finlayson. Voir les images sur la figure 5. L'effet global joue encore pleinement. La matrice des médianes est :

$$\begin{pmatrix} 0,0 & 13,49 & 21,23 & 13,63 \\ 13,49 & 0,0 & 13,37 & 14,69 \\ 21,23 & 13,37 & 0,0 & 18,54 \\ 13,63 & 14,69 & 18,54 & 0,0 \end{pmatrix}$$

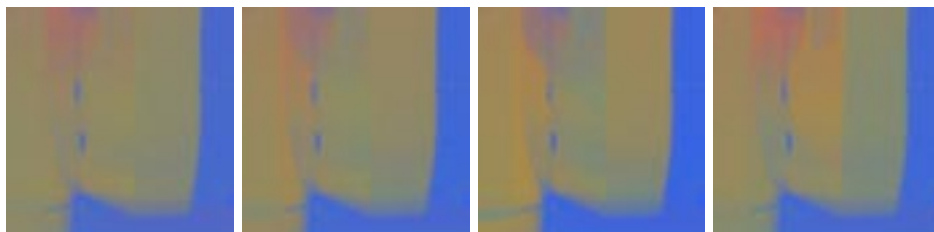


FIG. 5: Les 4 images normalisées globalement vis à vis d'un changement interne et externe de la source de lumière.

#### Normalisation locale pour les changements internes de la source.

Dans ce cas, nous avons pris des imagerie de taille  $21 \times 21$  pixels. Les images obtenues sont montrées sur la figure 6. Elles paraissent grises : en mettant localement les niveaux de rouge, vert et bleu au même niveau, on perd forcément l'information de couleur. Par contre, on voit que ce sont les formes qui ressortent, et il n'y a plus de différence d'illumination. L'invariance au changement externe de la source, même si non recherché au départ, semble donc atteint aussi.

Cela peut paraître frustrant : on perd la couleur des images, mais ce n'est pas surprenant : pour être invariant à un changement de couleur de la source de lumière, il ne faut pas dépendre de la couleur. Par contre, et ce fait est très important, cela permet de s'affranchir de la position de source de lumière, et donc des ombres, ce qui n'était pas possible avec des images en niveaux de gris (sauf à se ramener à une image uniforme et donc sans intérêt).

La matrice des médianes des distances entre pixels est :

$$\begin{pmatrix} 0,0 & 13,34 & 13,03 & 13,34 \\ 13,34 & 0,0 & 12,56 & 13,03 \\ 13,03 & 12,56 & 0,0 & 12,72 \\ 13,34 & 13,03 & 12,72 & 0,0 \end{pmatrix}$$

Cette matrice est la meilleure obtenue par normalisation (ses coefficients sont les plus petits).

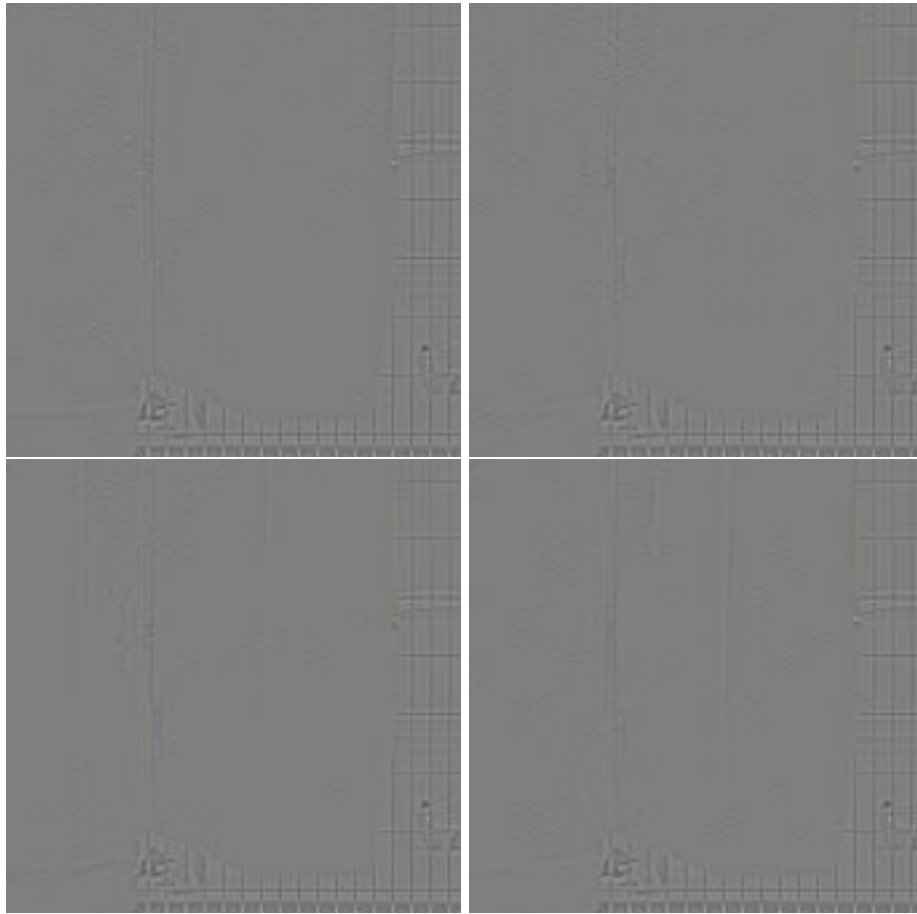


FIG. 6: Les 4 images normalisées localement vis à vis d'un changement interne de la source de lumière.

Ces résultats ne sont que préliminaires. D'autres seront utilisés dans la section suivante. Ils montrent toutefois que la normalisation est un outil plus délicat à utiliser qu'une simple estimation de modèle. Elle ouvre toutefois des possibilités intéressantes dans le problème de l'indexation des images où on ne connaît pas, avant que la requête ne soit formulée, les deux images utilisées.

## 5 Calcul d'invariants locaux

Cette section concerne l'expérimentation de divers invariants locaux basés sur l'information de couleur. La normalisation, telle qu'appliquée dans la section précédente, qu'elle soit d'ailleurs calculée de manière globale ou locale, mène à une comparaison globale des images. D'autre part, la normalisation ne permet pas d'obtenir une invariance à la rotation ou à la translation.

Lorsqu'on s'intéresse à des parties d'images seulement, par exemple pour reconnaître un objet dans une image, il est nécessaire d'utiliser des outils plus locaux : les invariants apparaissent alors naturellement. La normalisation ne perd pas tout son intérêt : elle peut être utilisée comme un prétraitement, et permet d'obtenir des résultats intéressants, comme montré ultérieurement.

La base de ce travail est la thèse de Cordelia Schmid [Sch96]. Dans ce travail est présenté un système d'appariement et d'indexation d'images basé sur le calcul de descripteurs locaux des images.

Pour chaque image concernée, on utilise tout d'abord un détecteur de points d'intérêt. Ces points sont les points les plus informatifs de l'image. Le détecteur utilisé est une amélioration de celui de Harris, qui est basé sur la recherche des minima de la fonction d'autocorrelation spatiale du signal image. Les points d'intérêt sont, dans ce cadre, les points qui ressemblent le moins à leurs voisins.

En chacun de ces points, on décompose le signal image autour de ce point sur une base de fonctions. Ces fonctions sont choisies de telle sorte que le résultat de la convolution entre ces fonctions et le signal soit invariant par rotation de l'image, et éventuellement par un changement des conditions d'illumination.

Les vecteurs de descripteurs obtenus, ou vecteurs d'invariants, sont alors comparés avec ceux obtenus à partir d'une autre image pour apparier ces images, ou ils peuvent être stockés dans une base pour indexer l'image.

Ce même schéma peut être repris avec des images couleur. Dans ce cas, on dispose de trois informations par pixels au lieu d'une seule, mais la même démarche est possible.

**Calcul des dérivées.** On commence par convoluer chacun des canaux avec une gaussienne :

$$R = \text{Gaus} * r \quad G = \text{Gauss} * g \quad B = \text{Gauss} * b$$

On dérive alors les signaux obtenus jusqu'à l'ordre 3. On notera, par exemple,  $R_{xy}$  la dérivée seconde de  $R$  par rapport à  $x$  et à  $y$ . On obtient ainsi 3 fonctions non dérivées, 6 dérivées à l'ordre 1, 9 dérivées à l'ordre 2, et 12 dérivées à l'ordre 3, soit 30 fonctions en tout.

**Invariants à la rotation.** La rotation est définie par un seul paramètre : son angle. On suppose que l'on s'affranchit de la translation par un changement de paramètres local. On doit donc obtenir 29 invariants indépendants. Pour les images en niveaux de gris, on en

disposait de 9. On peut reprendre ces 9 invariants, calculés sur chacun des canaux cette fois, puis trouver 2 invariants supplémentaires mélangeant les canaux.

Voici les invariants pour chaque canal (ici écrit dans le cas du rouge) :

$$\left\{ \begin{array}{l} R \\ R_x R_x + R_y R_y \\ R_{xx} R_x R_x + 2R_{xy} R_x R_y + R_{yy} R_y R_y \\ R_{xx} + R_{yy} \\ R_{xx} R_{xx} + 2R_{xy} R_{xy} + R_{yy} R_{yy} \\ R_{xxx} R_y R_y R_y + 3R_{xyy} R_y R_x R_x - 3R_{xxy} R_x R_y R_y - R_{yyy} R_x R_x R_x \\ R_{xxx} R_x R_y R_y + R_{xxy} (-2R_x R_x R_y + R_y R_y R_y) + R_{xyy} (-2R_x R_y R_y + R_x R_x R_x) + R_{yyy} R_x R_x R_y \\ R_{xxy} (-R_x R_x R_x + 2R_x R_y R_y) + R_{xyy} (-2R_y R_x R_x + R_y R_y R_y) - R_{yyy} R_x R_y R_y + R_{xxx} R_x R_x R_y \\ R_{xxx} R_x R_x R_x + 3R_{xyy} R_y R_x R_x + 3R_{xxy} R_x R_y R_y + R_{yyy} R_y R_y R_y \end{array} \right.$$

Pour compléter l'ensemble, on peut choisir deux invariants supplémentaires parmi les trois suivants :

$$R_x G_x + R_y G_y \quad G_x B_x + G_y B_y \quad B_x R_x + B_y R_y$$

**Invariants à la rotation et au changement d'illumination.** Pour obtenir des invariants, non seulement à la rotation, mais aussi au changement d'illumination, il faut choisir le modèle de changement d'illumination que l'on souhaite prendre en compte. Dans le cadre de la présente étude, on a choisit le modèle numéroté 9 au paragraphe 3 : il s'agit du modèle diagonal avec translation.

Pour trouver des invariants pour ce modèle, il faut donc éliminer 6 paramètres entre les invariants à la rotations trouvés précédemment. Pour éliminer les paramètres de translation, il suffit de supprimer les trois invariants non différentiels  $R$ ,  $G$  et  $B$ .

Pour éliminer les trois paramètres diagonaux, on peut diviser chaque invariant par la puissance adéquate de la norme du gradient, puis on supprime ces normes de la liste des invariants. On obtient ainsi les 23 invariants recherchés.

Notons  $\mathcal{R}_i$  les 9 invariants à la rotation obtenus précédemment pour le canal rouge. À partir de ces invariants, on peut donc calculer 7 invariants pour le changement d'illumination, qui sont :

$$\frac{\mathcal{R}_3}{\mathcal{R}_2^{3/2}} \quad \frac{\mathcal{R}_4}{\mathcal{R}_2^{1/2}} \quad \frac{\mathcal{R}_5}{\mathcal{R}_2} \quad \frac{\mathcal{R}_6}{\mathcal{R}_2^2} \quad \frac{\mathcal{R}_7}{\mathcal{R}_2^2} \quad \frac{\mathcal{R}_8}{\mathcal{R}_2^2} \quad \frac{\mathcal{R}_9}{\mathcal{R}_2^2}$$

Pour les invariants mélangeant les canaux, on peut faire de même, et choisir deux invariants parmi les trois suivants :

$$\frac{R_x G_x + R_y G_y}{(R_x R_x + R_y R_y)^{1/2} (G_x G_x + G_y G_y)^{1/2}}$$

$$\frac{G_x B_x + G_y B_y}{(G_x G_x + G_y G_y)^{1/2} (B_x B_x + B_y B_y)^{1/2}}$$

$$\frac{B_x R_x + B_y R_y}{(B_x B_x + B_y B_y)^{1/2} (R_x R_x + R_y R_y)^{1/2}}$$

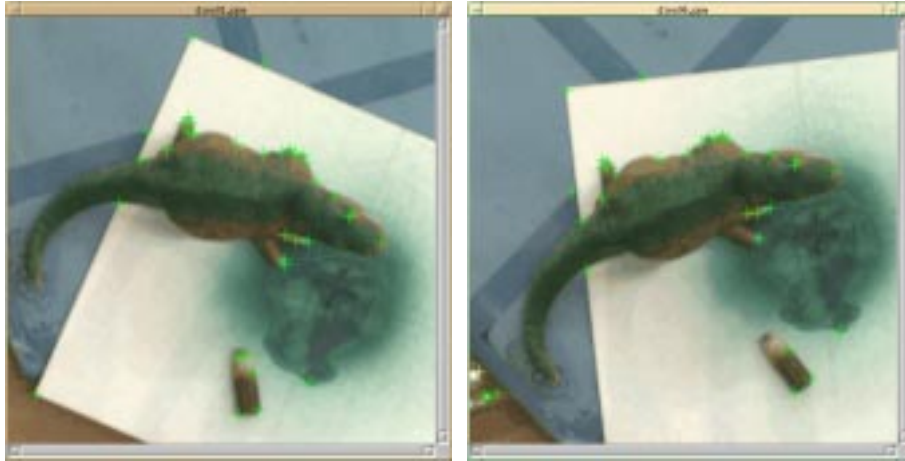


FIG. 7: 2 images d'une séquence prise avec rotation de la scène devant la caméra.

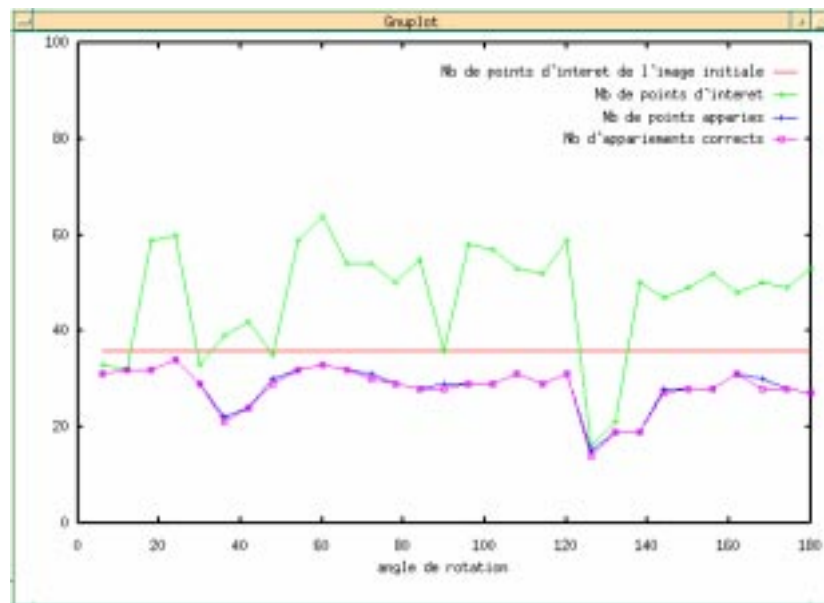


FIG. 8: Résultats de l'appariement sur la séquence du dinosaure.

**Quelques résultats pour une rotation de la scène.** On a tout d'abord testé ces invariants lors d'une rotation de la scène devant la caméra, avec l'axe de la rotation parallèle à l'axe optique. La figure 7 montre deux images d'une séquence montrant un dinosaure jouet. La séquence complète comporte une trentaine d'images prise lors d'une rotation globale de 180 degrés.

De chaque image ont été extraits une quarantaine de points d'intérêts, auxquels on a associé des invariants à la rotation (les points d'intérêt sont montrés sur les images). Ces invariants ont alors été appariés et les résultats sont montrés sur la figure 8. Ces résultats montrent que tous les appariements faits sont corrects. Le fait que tous les points ne sont pas appariés est dû tant à des faux négatifs, points qui auraient dus être appariés et qui ne le sont pas, qu'au fait que ce ne sont pas exactement les mêmes ensembles de points qui sont extraits de chaque image. Ces points sont peu nombreux et les résultats peuvent donc être déclarés très bons.

La même opération a été effectuée sur une autre séquence, plus plane que la première, mais avec plus de points d'intérêt. Deux images de la séquence sont montrées sur la figure 9, et les résultats sur la figure 10. Là aussi, les résultats sont très bons.

**Résultats pour un changement d'intensité de la source de lumière.** La série d'expérimentation suivante a été obtenue pour tester les invariants lors d'un changement interne de la source de lumière. On a pris une séquence de 10 images, montrant la couverture d'un magazine, dont les deux extrêmes sont montrées sur la figure 11. On a alors utilisé, d'une part les invariants à la rotation seule, dont les performances sont montrées sur la figure 12, d'autre part les invariants à la rotation et au changement interne de la source de lumière. Les résultats obtenus avec ces derniers sont montrés sur la figure 13.

Ces derniers invariants permettent de bien meilleurs résultats, tout au moins tant que les images ne sont pas trop saturées.

**Résultats pour un changement de position de la source de lumière.** La dernière série d'expérimentations réalisées teste la résistance de la deuxième famille d'invariants à un changement de la position de la source de lumière. 7 images ont été prises en faisant tourner la source de lumière autour de la scène. Deux de ces images sont montrées sur la figure 14. On a alors calculé et apparié les invariants comme précédemment. Les résultats sont montrés sur la figure 15. Les invariants étant utilisés hors de leur zone de validité, les résultats sont décevants dès que la source de lumière a trop varié.

Pour ne pas rester sur un échec, quand bien même il était prévisible, on a essayé de normaliser ces images à l'aide de la technique locale exposée précédemment. On a alors extraits les points de ces images normalisées, puis on les a appariés. Ceux des images normalisées sont montrées sur la figure 16, et les résultats sont sur la figure 17.

On a effectué la normalisation avec de grandes imageries, de taille  $41 \times 41$  pixels. Cette normalisation a eu pour premier effet une plus grande répétabilité du détecteur de points d'intérêt : les divers ensembles de points extraits sont beaucoup plus semblables que précédemment. Ensuite, il y a beaucoup moins de faux appariements, et beaucoup plus d'ap-



FIG. 9: 2 images d'une autre séquence prise avec rotation de la scène devant la caméra.

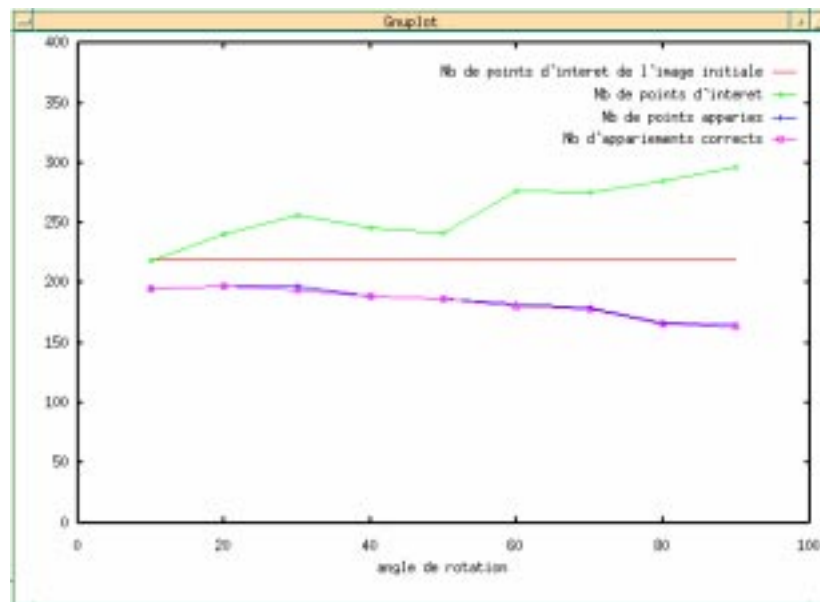


FIG. 10: Résultats de l'appariement sur la séquence de la boîte de jeu.



FIG. 11: 2 images d'une séquence prise avec changement de l'intensité de la source d'illumination.

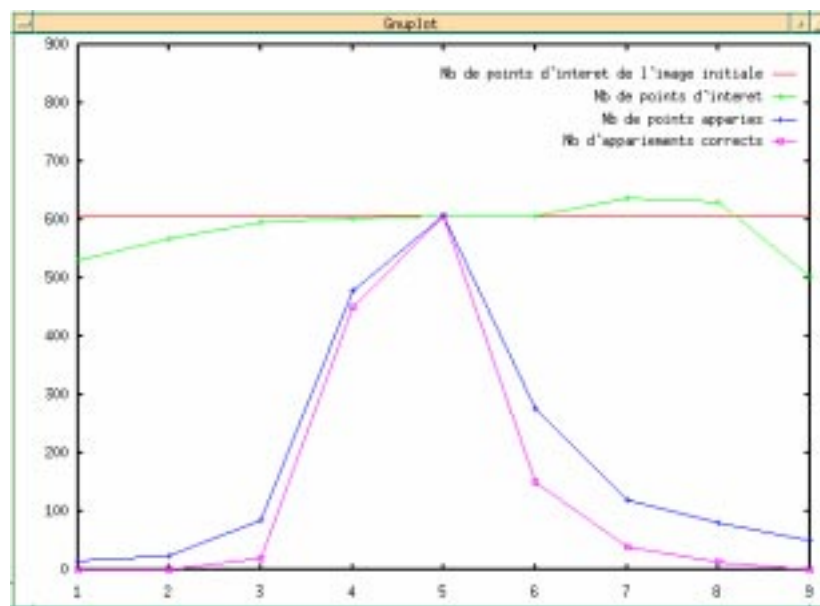


FIG. 12: Résultats de l'appariement sur la séquence du magazine, en utilisant les invariants à la seule rotation.

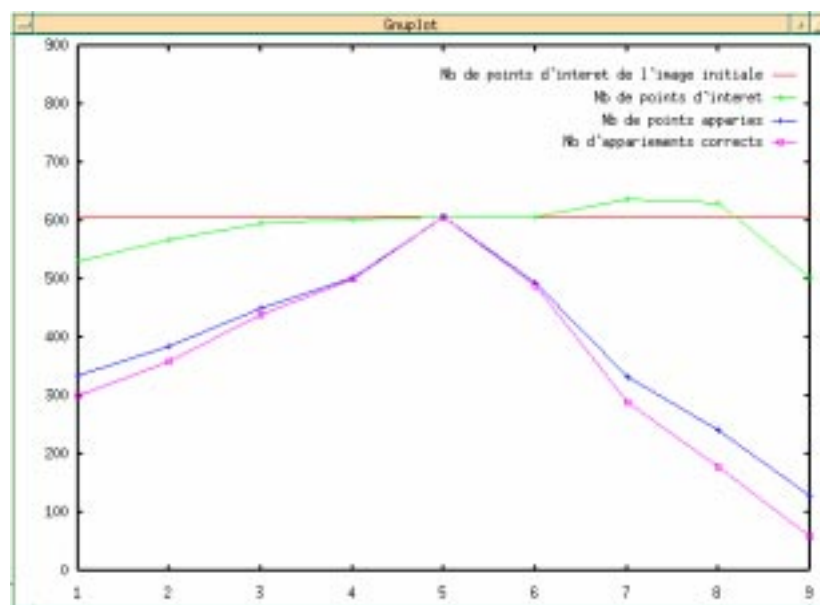


FIG. 13: Résultats de l'appariement sur la séquence du magazine, en utilisant les invariants à la rotation et au changement d'illumination.

pariements corrects (de l'ordre de 45 au lieu d'une quinzaine). Cela montre l'intérêt de la normalisation dans un cas concret.

## 6 Conclusion

Dans ce rapport, nous avons exploré quelques voies pour l'utilisation de la couleur pour l'indexation et l'appariement des images. Ont été testés divers modèles de variation de la couleur au changement des conditions d'illumination d'une scène, des techniques de normalisation d'images pour s'affranchir de ces changements, et enfin l'utilisation d'invariants pour appairer ou indexer des images, en suivant la méthodologie proposée par C. Schmid dans sa thèse.

Les résultats obtenus sont bons, mêmes si divers réglages restent à faire, et d'autres d'expérimentations à mener. La couleur ouvre, par exemple, la possibilité de s'affranchir des ombres, ce qui est une révolution par rapport aux images en niveau de gris, où ces ombres représentent un bruit très fort et presque incontournable. Il y a donc là des nouvelles possibilités à explorer.



FIG. 14: 2 images d'une séquence prise avec déplacement de la source d'illumination.

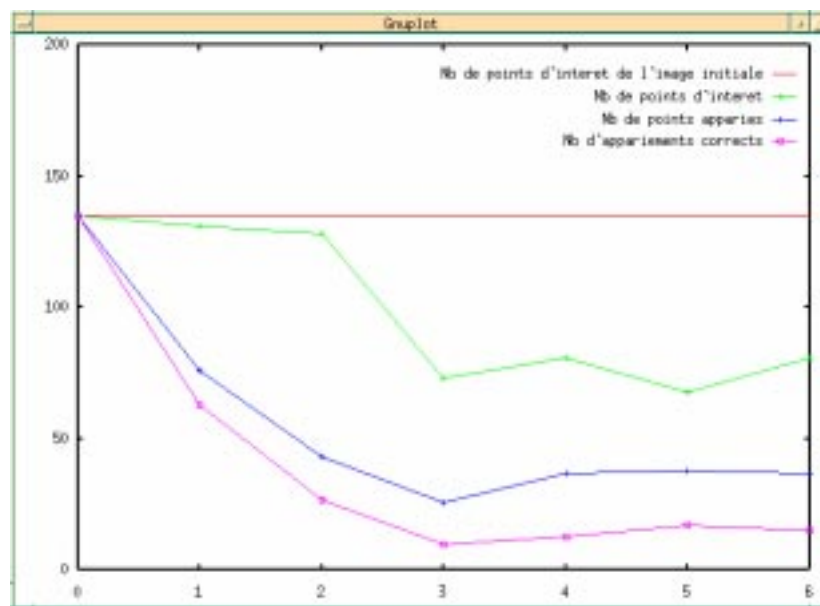


FIG. 15: Résultats de l'appariement sur la séquence de la maisonnette, en utilisant les invariants à la rotation et au changement interne de la source.

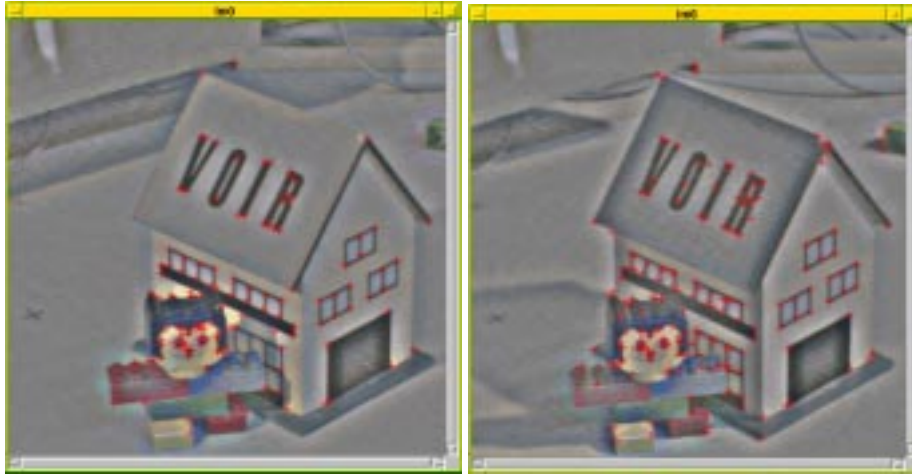


FIG. 16: 2 images de la séquence précédente, après normalisation locale.

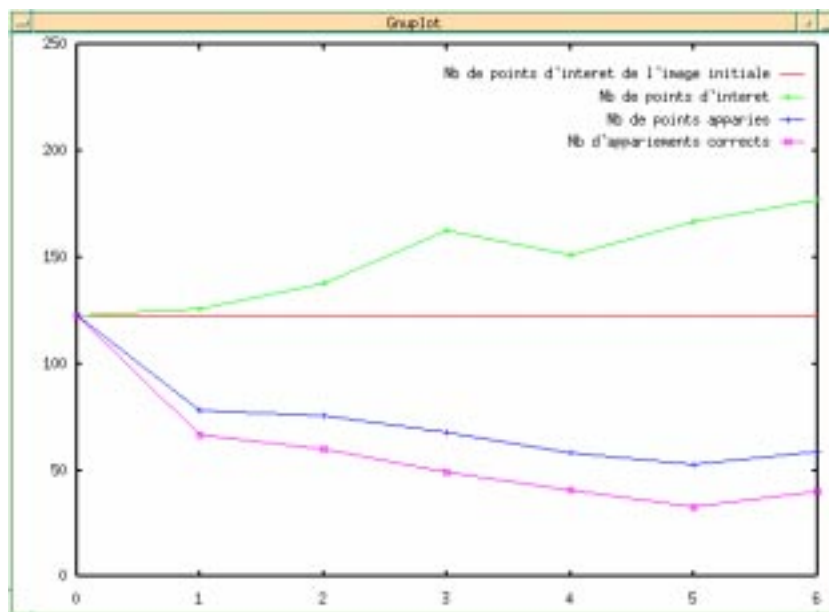


FIG. 17: Résultats de l'appariement sur la séquence de la maisonnette avec normalisation locale, en utilisant les invariants à la rotation et au changement interne de la source.

## Références

- [BFF96] K. Barnard, G. Finlayson, et B. Funt. Colour constancy for scenes with varying illumination. Dans *Proceedings of the 4th European Conference on Computer Vision, Cambridge, Angleterre*, pages 3–15, juillet 1996.
- [Bri78] M.H. Brill. A device performing illuminant-invariant assessment of chromatic relations. *Journal of Theoretical Biology*, pages 473–478, 1978.
- [EM95] R. Ennesser et G. Medioni. Finding waldo, or focus of attention using local color information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):805–809, août 1995.
- [FCF96] G. Finlayson, S. Chatterjee, et B. Funt. Color angular indexing. Dans *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, pages 16–27, 1996.
- [FDF94] G.D. Finlayson, M.S. Drew, et B. Funt. Color constancy: Generalized diagonal transforms suffice. *Journal of the Optical Society of America A*, 11(11):3011–3019, novembre 1994.
- [FF95] B.V. Funt et G.D. Finlayson. Color constant color indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.
- [FM96] G. Florou et R. Mohr. What accuracy for 3D measurements with cameras? Dans *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 354–358. IEEE Computer Society Press, 1996.
- [HSE<sup>+</sup>95] J. Hafner, H.S. Sawhney, W. Equitz, M. Flickner, et W. Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, July 1995.
- [JV96] A.K. Jain et A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29:1233–1244, 1996.
- [KMW96] M.S. Kankanhalli, B.M. Mehtre, et J.K. Wu. Cluster based colour matching for image retrieval. *Pattern Recognition*, 29(4):701–708, 1996.
- [LM71] E.H. Land et J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America A*, 61(1):1–11, janvier 1971.
- [MKN95] B.M. Mehtre, M.S. Kankanhalli, et A.D. Narasimhalu. Color matching for image retrieval. *Pattern Recognition Letters*, 16:325–331, 1995.
- [SB91] M.J. Swain et D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

- 
- [Sch96] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris*. Thèse de doctorat, GRAVIR – IMAG – INRIA Rhône–Alpes, juillet 1996.
- [SM97] C. Schmid et R. Mohr. Object recognition using local characterization and semi-local constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–534, May 1997.
- [SO95] M.A. Stricker et M. Orengo. Similarity of color images. Dans *SPIE Conference on Storage and Retrieval for Image and Video Databases V*, volume SPIE-2420, pages 381–392, 1995.
- [SP95] I.S. Sethi et N. Patel. A statistical approach to scene change detection. Dans *SPIE Conference on Storage and Retrieval for Image and Video Databases*, volume SPIE-2420, pages 329–337, 1995.
- [TC92] G. Taubin et D.B. Cooper. Object recognition based on moment (or algebraic) invariants. Dans J.L. Mundy et A. Zisserman, éditeurs, *Geometric Invariance in Computer Vision*. The MIT Press, Cambridge, MA, USA, 1992.
- [Tor97] P.H.S. Torr. An assessment of information criteria for motion model selection. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 47–52, juin 1997.
- [Wol90] H.J. Wolfson. Model-based object recognition by geometric hashing. Dans O. Faugeras, éditeur, *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 526–536. Springer-Verlag, avril 1990.
- [ZGLS95] H. Zhang, Y. Gong, C.Y. Low, et S.W. Smoliar. Image retrieval based on color features: an evaluation study. Dans *SPIE Conference on Storage and Retrieval for Image and Video Databases*, volume SPIE-2606, pages 381–392, 1995.



---

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,  
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY  
Unité de recherche INRIA Rennes, Irista, Campus universitaire de Beaulieu, 35042 RENNES Cedex  
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN  
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex  
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

---

Éditeur  
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399