Mise en correspondance d'images en couleur : Application à la synthèse de vues intermédiaires

Valérie Gouet

Numéro d'identification

ACADEMIE DE MONTPELLIER

UNIVERSITE MONTPELLIER II

SCIENCES ET TECHNIQUES DU LANGUEDOC

THESE

présentée à l'Université Montpellier II Sciences et Techniques du Languedoc
pour obtenir le diplôme de DOCTORAT

Spécialité : INFORMATIQUE

Formation Doctorale : Informatique

Ecole Doctorale : Information, Structures, Systèmes

MISE EN CORRESPONDANCE D'IMAGES EN COULEUR

Application à la synthèse de vues intermédiaires

par

Valérie GOUET

Soutenue le 25 Octobre 2000 devant le jury composé de :

M. HABIB Michel	Professeur, LIRMM, Montpellier II	Président
M. DERICHE Rachid	Directeur de recherche, INRIA Sophia-Antipolis	Rapporteur
M. ROUX Christian	Professeur, ENST Bretagne, Rennes I	Rapporteur
Mme BOUJEMAA Nozha	Chargée de recherche, INRIA Rocquencourt	Examinateur
Mme PELÉ Danielle	Ingénieur de recherche, France Télécom R&D	Examinateur
M. BAJARD Jean-Claude	Professeur, LIRMM, Montpellier II	Directeur de Thèse
M. MONTESINOS Philippe	Maître assistant, LGI2P, Ecole des Mines d'Alès	Encadrant de Thèse

A Anna, ma grand-mère.

Remerciements

Je tiens tout d'abord à exprimer ma reconnaissance aux personnes qui m'ont fait l'honneur de participer au jury de cette thèse.

Monsieur Rachid Deriche, Directeur de recherche à l'INRIA de Sophia Antipolis et Monsieur Christian Roux, Professeur à l'ENST de Bretagne, qui m'ont fait le plaisir d'en être les rapporteurs. Je les remercie pour le temps qu'ils ont consacré à juger ce travail et les précieux conseils qu'ils m'ont donnés pour l'améliorer.

Monsieur Michel Habib, Professeur à l'université Montpellier II et Madame Nozha Boujemaa, Chargée de recherche à l'INRIA de Rocquencourt, pour l'intérêt qu'il ont porté à ce travail en acceptant de faire partie du jury.

Madame Danielle Pelé, Ingénieur de recherche à France Télécom R&D (Rennes), pour les nombreux échanges que nous avons eus tout au long de notre collaboration et pour avoir accepté de faire partie de ce jury.

Monsieur Jean-Claude Bajard, Professeur à l'université Montpellier II qui a assumé la direction de cette thèse, ainsi que Monsieur Mourad Oussalah, Professeur à L'université de Nantes, pour avoir dirigé mon travail pendant les deux premières années de mon doctorat.

Monsieur Philippe Montesinos, Maître assistant au LGI2P, pour m'avoir accueillie au sein de l'équipe Vision du laboratoire et proposé un sujet de thèse passionnant. Je le remercie pour ses idées, son enthousiasme, pour avoir su me guider et pour la confiance qu'il m'a témoignée tout au long de notre collaboration.

Cette thèse a été réalisée au sein du LGI2P, sous la direction de Madame Janine Magnier. Je la remercie de m'avoir accueillie dans son laboratoire. J'adresse également ma sympathie à tous les membres du laboratoire qui ont contribué, de près ou de loin, à l'accomplissement de ce travail dans une ambiance amicale que j'ai particulièrement appréciée.

Toute ma sympathie va notamment aux membres de l'équipe Vision, Samir mais aussi les anciens, pour leur disponibilité et tout ce que j'ai pu apprendre à leur contact.

Je remercie aussi François T. et Nicolas pour leur amitié, leur infinie patience et l'aide technique inestimable qu'ils m'ont apportée, en particulier pour la rédaction de ce mémoire avec Latex !

Egalement un grand merci à François L. et à Vincent pour avoir eu la gentillesse de consacrer du temps à la relecture de ce mémoire et pour leurs remarques pertinentes.

Et bien sûr les incontournables : Sylvie, Christelle, Christophe et Karima pour leurs conseils avisés, mais surtout pour les moments fantastiques que j'ai passés en leur compagnie, que ce soit pendant les pauses café/thé, autour d'une bonne table ou encore devant un bon film !

Enfin, mes parents et mes proches pour leur inconditionnel soutien, leurs encouragements dans les moments difficiles et leur disponibilité malgré la distance qui me sépare d'eux.

Résumé

La mise en correspondance d'images est une étape clé pour un grand nombre d'applications de Vision par Ordinateur. Parmi toutes les approches existantes en niveau de gris, les méthodes iconiques fournissent les résultats les plus probants. Cependant, aucune d'entre elles ne tire profit de la richesse de la couleur, alors que leur essence même est d'exploiter au maximum l'information contenue dans le signal de l'image. La principale contribution des travaux développés dans ce mémoire consiste alors à mettre en place une méthode de mise en correspondance plus robuste que celles rencontrées jusqu'à ce jour, en mettant en jeu l'information supplémentaire que contiennent les images en couleur. Cette approche est innovante à plusieurs niveaux du processus d'appariement. Ainsi, nous présentons dans un premier traitement une classe de détecteurs de points d'intérêt spécifique à la couleur. Deux nouveaux opérateurs sont proposés, évalués et se révèlent plus stables que les détecteurs jusqu'alors connus en niveau de gris. La couleur nous permet également de mettre en place une méthode de caractérisation de ces points, locale et robuste. Celle-ci est basée sur les invariants différentiels de Hilbert calculés seulement à l'ordre un, grâce à l'apport de l'information couleur. Elle est donc invariante aux transformations euclidiennes de l'image ; nous proposons également une méthode originale qui la rend invariante aux changements d'illumination. La forte combinatoire des méthodes d'appariement classiques rendent ces dernières inexploitables avec beaucoup de points. Nos travaux consistent en outre à intégrer les primitives couleur ainsi caractérisées dans un nouveau processus de mise en correspondance, rendu efficace face aux grands ensembles de points par l'utilisation de contraintes géométriques robustes quelles que soient les transformations de l'image.

Mots clés :
Vision par Ordinateur, Images en couleur, Détecteurs de points d'intérêt, Invariants différentiels, Constance des couleurs, Mise en correspondance, Optimisation combinatoire, Géométrie épipolaire, Géométrie projective, Reconstruction tridimensionnelle, Transfert d'images.

Abstract

Image matching is a key step in many Computer Vision applications. Among the existing approaches for gray value images, iconic methods produce the most convincing results. However, none of these take advantage of the richness of color, although the very reason why they were developed is to exploit to the maximum the information contained in the image signal. The main contribution of the work described in this thesis therefore consists in developing an image matching method that is more robust than those used up till now, by bringing into play the additional information contained in color images. This approach is innovative at several steps of the matching process. Indeed, we present, in a first treatment, a class of points of interest detectors specific to color. Two new operators are proposed, and assessed : they prove more stable than the detectors already used for gray value images. Color also allows us to define a new characterization method of these points, which is local and robust. This description is based on the Hilbert differential invariants, which need only to be computed to order one, thanks to the contribution of the color information. So the description is invariant to Euclidean transformations of the image; in addition, we propose an original method to make it invariant to changes in illumination. The excessive combinatory of ordinary matching methods make them unexploitable when there are many points. In addition, the color primitives thus characterized are then integrated into a new matching scheme, which has been made efficient for big sets of point by using geometric constraints that are robust whatever the image transformations are.

Keywords :
Computer Vision, Color images, Points of interest detectors, Differential invariants, Color constancy, Matching, Combinatorial optimization, Epipolar geometry, Projective geometry, 3D reconstruction, Image transfer.

Table des matières

Table des figures

Liste des tableaux

Introduction

Les travaux présentés dans ce mémoire s'inscrivent dans le domaine de *la mise en correspondance* d'images, encore appelée appariement. Il s'agit d'un domaine très vaste de la Vision par Ordinateur, qui a suscité et suscite encore de nombreuses recherches. Il représente une étape clé pour des problèmes variés, comme l'indexation d'images ou encore la reconstruction de scènes tridimensionnelles. Dans ce contexte, les travaux réalisés jusqu'à présent portent essentiellement sur l'analyse et le traitement d'images définies en *niveau de gris*. L'approche qui est proposée dans cette thèse vise à mettre à profit l'information supplémentaire disponible dans les images en *couleur*, dans le but de développer des méthodes d'appariement plus performantes.

Cette introduction est découpée comme suit : nous présentons en premier lieu les motivations qui nous ont amenés à étudier le problème de la mise en correspondance dans un contexte d'images en couleur. Puis l'approche que nous proposons est présentée et située par rapport aux méthodes existantes. Nous donnons ensuite les contributions de notre travail, avant de terminer par un plan détaillé du mémoire.

Motivations

Avec l'engouement croissant pour les techniques du multimédia, les images sont de plus en plus présentes dans notre environnement. Ainsi ces dernières années, de nombreuses applications se sont développées ou ont vu le jour autour de l'imagerie numérique. Citons notamment *la recherche documentaire*, qui consiste à rechercher dans une banque d'images celle qui par exemple illustre un événement politique ; ou encore *la synthèse d'images*, qui, par le biais de la réalité virtuelle, permet à un utilisateur de visiter un musée, une boutique virtuel ou encore d'être immergé dans les mondes imaginaires des jeux. Ces applications prennent une dimension particulière avec les immenses potentialités qu'offrent maintenant les autoroutes de l'information comme Internet.

Toutes ces applications impliquent la résolution du problème de la mise en correspondance d'images. En effet, la recherche documentaire nécessite de trouver parmi plusieurs images celle qui correspond le mieux à un modèle donné. Les applications dérivées de la synthèse d'images requièrent certaines connaissances sur la géométrie des caméras, qui, lorsqu'elle n'est pas disponible est calculable à partir de l'appariement d'images de cette scène.

De nombreuses solutions ont été proposées pour résoudre les différents problèmes liés à la mise en correspondance. Elles présentent cependant de fortes limitations : la plupart ne permettent pas d'apparier deux images dans des conditions générales, comme par exemple lorsque celles-ci diffèrent d'une importante rotation ou d'un changement d'illumination. Elles sont également fortement combinatoires et ne parviennent pas à traiter des données volumineuses en un temps raisonnable. Enfin, l'ensemble des approches proposées travaillent à partir d'images en niveau de gris et par conséquent ne mettent pas à profit toute l'information que peut contenir une image en couleur. L'objet de cette thèse est de proposer une approche innovante par rapport à ces méthodes et à leurs limitations.

Parmi toutes les applications potentielles de la mise en correspondance d'images, nous nous intéressons plus particulièrement dans ce mémoire aux techniques de *transfert d'images* qui consistent à synthétiser une image de la scène tridimensionnelle sous un point de vue donné, à partir d'images acquises sous des points de vue différents. Autrement dit, ces approches cherchent à définir des techniques de traitement d'images et d'analyse de scènes permettant à partir de plusieurs scènes d'un environnement réel, acquises au préalable dans des conditions de prises de vues différentes (angles différents, distances différentes, etc), de recalculer les séquences d'images simulant les scènes que verraient une personne se déplaçant dans cet environnement.

Le transfert d'images a de nombreuses applications. Il permet par exemple de simuler la visite virtuelle de lieux connus uniquement au travers de photos. Dans un futur proche, il pourra également permettre à un téléspectateur de choisir son point de vue lorsqu'il regarde un match de basket à la télévision. La technique peut être enfin vue comme une méthode de compression vidéo. En effet, il suffit de ne stocker que quelques vues d'une séquence d'images et de les accompagner d'une certaine information, le tenseur trifocal par exemple, permettant ainsi de recalculer les images manquantes.

Des solutions au transfert d'images existent déjà. Citons la plus connue du grand public : le *morphing*, qui fournit effectivement des vues intermédiaires, mais qui malheureusement ne respecte pas la géométrie de la scène. Ou encore le logiciel QuickTime VR d'Apple, qui permet de simuler la visite d'un lieu, à partir de la donnée de quelques photos. Signalons pour ce dernier que la technique employée requiert des conditions de prise de vue particulières ainsi que l'aide de l'utilisateur pour certaines étapes du traitement.

Approche proposée

Les techniques dédiées au transfert d'images nécessitent une certaine connaissance de la géométrie des caméras qui ont fourni les images. Parfois les paramètres optiques et la position des caméras sont connus. Néanmoins le plus souvent, on peut vouloir traiter des images qui proviennent d'une caméra ou d'un appareil photo pour lesquels ces paramètres ne sont pas disponibles, tout simplement parce celles-ci proviennent d'un film préexistant ou bien ont été prises par des non spécialistes. Dans tous ces cas, le système est dit *non calibré*, et la seule information géométrique qui puisse être obtenue se traduit par des relations linéaires entre caméras, comme par exemple la géométrie épipolaire lorsque l'on dispose de deux oculaires. C'est à ce niveau qu'interviennent les techniques de mise en correspondance d'images. Il est en effet nécessaire de faire correspondre un certain nombre de points entre les images pour estimer cette géométrie.

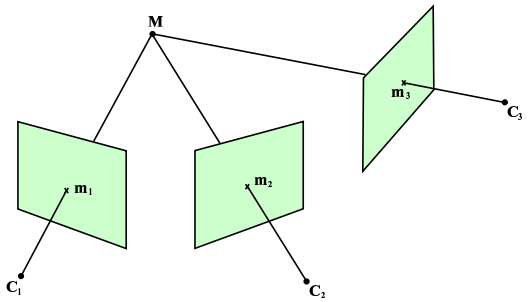


Figure 1 – *Exemple d'appariement trinoculaire* (m_1, m_2, m_3) *issu du point 3D* M.

L'appariement de points d'intérêt[1] a fait l'objet de nombreux travaux. Le lecteur peut se reporter à [Zhang, 1993] pour un état de l'art. Il consiste à déterminer quelles sont les projections qui se correspondent dans les images, c'est-à-dire quels sont les points 2D qui représentent les projections d'un même point 3D, comme l'illustre la figure 1. L'approche que nous avons choisi d'explorer dans ce mémoire repose sur une caractérisation exploitant directement l'information photométrique de l'image. Cette classe de méthodes porte le nom de méthodes *iconiques* et permet de caractériser les points de l'image de manière *locale*. D'autres types d'approches existent, elles sont pour certaines basées sur une description globale et ne sont par conséquent robustes ni aux changements d'arrière-plan, ni aux occultations, ni aux principales transformations de l'image. D'autres méthodes quantifient l'information par des grandeurs géométriques, comme les contours, les segments ou encore les régions. Elles sont toutefois issues de primitives symboliques, qui entraînent inévitablement une perte d'information, et surtout rendent la qualité de la mise en correspondance fortement dépendante de la structure de l'image. Lorsque l'on se place dans le contexte du transfert d'images, l'approche iconique semble donc être la plus robuste et la mieux adaptée à la mise en correspondance.

Toutes les méthodes de mise en correspondance iconiques que nous avons rencontrées ont pour support les images en niveau de gris. A notre grand étonnement, aucune n'essaie de tirer profit de la richesse des images en couleur, alors que leur essence même est d'exploiter au maximum l'information contenue dans le signal de l'image [Luong, 1991]. La principale contribution de nos travaux consiste donc à mettre en place une méthode de mise en correspondance ayant pour support les images en couleur.

Notre méthode suit le schéma suivant :

1. Un traitement de bas niveau nous permet de traiter le signal et d'en extraire des points d'intérêt. De multiples détecteurs de points d'intérêt existent en niveau de gris, mais nous n'en connaissons aucun qui exploite l'information couleur. Nous proposons alors deux types de détecteurs, qui correspondent à une généralisation multi-spectrale des opérateurs de Kitchen et Rosenfeld [Kitchen et Rosenfeld, 1982] et de Harris [Harris et Stephens, 1988; Bauckhage et Schmid, 1996] ;
2. Puis il s'agit de déterminer des grandeurs numériques qui décrivent de manière discriminante les primitives observées. Encore une fois, nous mettons à profit la richesse

[1] Un point d'intérêt représente l'endroit de l'image où le signal est riche en information. Cette notion est définie plus précisément dans la première partie de ce mémoire.

du signal couleur pour développer une méthode de caractérisation iconique robuste. Cette description est basée sur les invariants différentiels de Hilbert [Hilbert, 1890]. Elle étend à la couleur l'approche proposée par Schmid dans sa thèse [Schmid, 1996] et apporte des précisions concernant la robustesse au bruit;

3. Enfin, la phase d'appariement proprement dite consiste à établir une correspondance entre les points ayant des caractérisations similaires entre images. L'ajout de contraintes semi-locales rend la méthode d'appariement plus robuste et son intégration dans un processus incrémental nous permet de traiter efficacement de grands ensembles de points.

A l'issue de cette dernière étape, nous avons à notre disposition un ensemble de points d'intérêt mis en correspondance entre plusieurs images. Pour des scènes rigides, cette donnée permet alors d'estimer la géométrie reliant les caméras entre elles. Dans le cas d'un système binoculaire, il s'agit de la géométrie épipolaire; dans le cas d'un système trinoculaire, on parle de trilinéarités ou encore de tenseur trifocal. C'est seulement lorsque cette information géométrique est connue avec précision que l'on peut envisager de mettre en œuvre les méthodes de transfert d'images.

Contributions

La principale contribution de cette thèse est d'avoir développé une méthode de mise en correspondance plus robuste que celles rencontrées jusqu'à ce jour. Le succès de notre approche s'explique essentiellement par le choix d'exploiter l'information couleur. A plusieurs étapes de notre traitement, celle-ci s'est en effet révélée posséder un contenu informatif plus fertile que les traditionnelles images en niveau de gris. En premier lieu, cette richesse nous a permis de mettre en place deux détecteurs de points d'intérêt plus stables que les meilleurs opérateurs connus en niveau de gris, et ce pour les principales transformations de l'image[2]. Bien que d'approches différentes, ces deux détecteurs se sont révélés équivalents au travers des nombreux tests réalisés. Nous avons cependant dénoté quelques aspects favorisant l'un par rapport à l'autre selon les applications dans lesquelles ils sont intégrés.

La couleur nous a également permis de simplifier considérablement la méthode de caractérisation de points proposée par Schmid et basée sur les invariants différentiels de Hilbert. L'approche traditionnelle impose en effet de calculer ces invariants jusqu'aux dérivées troisièmes de l'image, pour que la description qui en découle soit suffisamment discriminante. Il en résulte des calculs coûteux et plutôt sensibles au bruit. Au contraire, l'information couleur permet de réduire ce calcul aux dérivées d'ordre un. La caractérisation ainsi obtenue est plus stable et bien plus simple à mettre en œuvre. Comme pour les détecteurs, elle est robuste aux principales transformations de l'image.

L'autre point fort de notre approche réside dans la mise en place d'un processus de mise en correspondance robuste aux grands ensembles de points. La forte combinatoire des méthodes d'appariement classiques les rendent en effet vite inexploitables dans ce contexte, alors que pour être efficaces, les méthodes de transfert d'images requièrent justement la donnée d'un grand nombre d'appariements. Notre approche est incrémentale et exploite des contraintes géométriques robustes quelles que soient les transformations de l'image. Elle

[2]Les translations, les rotations, les changements d'illumination, les changements d'échelle, les changements de point de vue et le bruit.

permet d'obtenir, pour des nombres de points de l'ordre du millier, des taux d'appariements corrects de près de 95%, dans des temps de calcul considérablement réduits.

Ajoutons pour terminer que cette étude a également permis de faire une synthèse, que nous espérons complète, des différentes méthodes de reconstruction tridimensionnelle et de transfert d'images existants dans la littérature de la Vision par Ordinateur.

Plan du mémoire

Les travaux menés dans ce mémoire se décomposent en trois grandes parties. La première est consacrée à l'extraction de points d'intérêt dans des images en couleur. Elle se présente sous la forme de deux chapitres.

- *Chapitre 1 : Segmentation couleur*
  Nous commençons dans ce chapitre par faire un bref état de l'art des principaux extracteurs de points d'intérêt que l'on rencontre en Vision par Ordinateur. Ceux-ci sont exclusivement dédiés aux images en niveau de gris. Nous passons ensuite en revue les quelques travaux de segmentation réalisés pour des images en couleur, avant de présenter les deux détecteurs de points d'intérêt que nous avons mis en place et qui exploitent l'information couleur.
- *Chapitre 2 : Evaluation des détecteurs couleur*
  Le chapitre précédent est illustré par de nombreux résultats de détection, obtenus à partir des deux nouveaux extracteurs couleur, ce qui permet au lecteur d'évaluer visuellement la qualité de la détection. Afin d'être à même de juger de leur précision et de les situer par rapport aux détecteurs existants, nous mettons en place dans ce chapitre une méthode d'évaluation basée sur deux critères : la répétabilité du détecteur à travers des séries d'images ainsi que sa localisation. Nos deux détecteurs couleur sont alors évalués aux côtés d'un détecteur de points en niveau de gris qui sert de référence. Pour tester la robustesse de ces opérateurs, l'étude est réalisée dans différents contextes incluant les principales transformations de l'image.

La deuxième partie est consacrée à la mise en correspondance des points d'intérêt entre deux images en couleur. La première étape de ce travail consiste à les caractériser selon une approche locale. La caractérisation obtenue pour chacun des points est ensuite exploitée au sein de techniques de mise en correspondance, à l'issue desquelles un ensemble cohérent de points appariés, c'est-à-dire ayant des caractéristiques similaires, est obtenu.

- *Chapitre 3 : Caractérisation locale des points d'intérêt - Etat de l'art*
  Nous passons en revue dans ce chapitre les différents travaux qui ont donné lieu à des méthodes de caractérisation, aussi bien en niveau de gris qu'en couleur. Un certain nombre de ces caractérisations extrait une information locale autour du point d'intérêt, mais travaille uniquement sur des images en niveau de gris. En parallèle, les quelques méthodes rencontrées exploitant l'information couleur extraient une information beaucoup plus globale, qui se révèle peu adaptée à la caractérisation et à la mise en correspondance de points d'intérêt.
- *Chapitre 4 : Les invariants différentiels couleur*
  C'est ainsi que nous présentons dans ce chapitre une nouvelle approche de caractérisation locale de points d'intérêt, qui, elle, exploite l'information couleur. La

méthode est basée sur les invariants différentiels de Hilbert. Elle permet de mettre en place une caractérisation invariante à la rotation image. Nous montrons que l'information couleur permet de simplifier considérablement la mise en place de cette caractérisation, tout en la rendant plus robuste au bruit. Une technique issue du problème de la constance des couleurs est également présentée pour la rendre invariante aux changements d'illumination.

- *Chapitre 5 : Méthodes d'appariements - Etat de l'art*
  Dans ce chapitre, nous passons en revue les principales méthodes de mise en correspondance qui exploitent les diverses caractérisations présentées au chapitre 3. L'objectif de ces approches est de mettre en place une mesure de similarité permettant de les comparer. Nous verrons également que certaines de ces méthodes exploitent des informations géométriques plus globales pour améliorer l'appariement. Les principales techniques de mise en correspondance dense, consacrées à l'appariement de l'intégralité des images, sont aussi présentées.

- *Chapitre 6 : Une méthode de mise en correspondance robuste*
  Nous avons présenté au chapitre 4 une nouvelle méthode de caractérisation de points. Ce chapitre se consacre en premier lieu à la mise en place d'une métrique permettant de l'exploiter pour la mise en correspondance. Des contraintes géométriques invariantes aux principales transformations de l'image sont également introduites. Puis, nous présentons une méthode de mise en correspondance robuste rendant efficace l'appariement d'un très grand nombre de points. Le processus complet d'appariement est ensuite testé et comparé aux approches traditionnelles.

A l'issue de ces deux parties, nous sommes capables de mettre en correspondance efficacement un grand nombre de points d'intérêt entre deux images en couleur. Nous allons maintenant voir dans la dernière partie de ce mémoire comment il est possible d'utiliser ces appariements dans le cadre d'applications de stéréovision dédiées à la reconstruction tridimensionnelle et au transfert d'images.

- *Chapitre 7 : Géométrie des systèmes binoculaires et trinoculaires non calibrés*
  Dans ce chapitre, nous rappelons au lecteur les principales notions requises pour manipuler les systèmes de vision constitués de deux ou trois caméras, dans un contexte non calibré. Dans un premier temps, les différents modèles de caméra sont passés en revue. Puis nous présentons les entités caractérisant les systèmes binoculaires, notamment la géométrie épipolaire. Ceux-ci représentent la configuration minimale nécessaire pour envisager une reconstruction tridimensionnelle. Nous nous intéressons également aux systèmes trinoculaires, qui manipulent des entités indispensables au transfert d'images. Enfin, nous passons en revue les différentes méthodes permettant l'estimation des diverses entités mises en jeu. Les caméras n'étant pas calibrées, les seules informations exploitables pour ces estimations sont les mises en correspondance de points.

- *Chapitre 8 : Synthèse de vues intermédiaires*
  Ce chapitre est consacré aux méthodes permettant de synthétiser des images d'une scène sous différents points de vue, à partir d'images existantes. Ces approches peuvent être décomposées en deux classes : d'une part les approches procédant à une reconstruction tridimensionnelle de la scène, avant de la reprojeter sur le plan rétinien de l'image virtuelle, et d'autres part les méthodes, dites méthodes de

transfert, qui génèrent l'image virtuelle uniquement à partir de l'information bidimensionnelle contenue dans les images de référence. Nous illustrons ce chapitre par des résultats de reconstruction tridimensionnelle et de transfert projectif d'images, réalisés à partir de points appariés par la méthode présentée dans les deux premières parties de ce mémoire.

En conclusion, nous récapitulons les principales contributions de notre approche, les améliorations à apporter et les perspectives ouvertes par cette thèse.

Première partie

Détection de points d'intérêt pour la couleur

Chapitre 1

Segmentation couleur

Ce chapitre débute par un bref état de l'art des principaux extracteurs de points d'intérêt que l'on rencontre en Vision par Ordinateur. Ceux-ci sont exclusivement dédiés aux images en niveau de gris. Nous passons ensuite en revue les quelques travaux de segmentation réalisés pour des images en couleur, avant de présenter les deux détecteurs de points d'intérêt que nous avons mis en place et qui exploitent l'information couleur.

Sommaire

Nous présentons dans ce chapitre la première étape du processus d'appariement : l'extraction des points d'intérêt. Le choix de ce type de primitive est tout d'abord expliqué et justifié à la section 1.1. L'originalité des travaux présentés ici réside dans le fait que les images utilisées sont en couleur. A notre connaissance, il n'existe aucun détecteur de points d'intérêt exploitant l'information couleur. Nous commençons donc par faire à la section 1.2 un état de l'art sur les différents détecteurs existants en niveau de gris. Nous détaillons particulièrement le détecteur de Kitchen et Rosenfeld et celui de Harris et Stephens dans sa version précise. Le détecteur de Kitchen et Rosenfeld fait partie de la catégorie des détecteurs travaillant à l'ordre deux qui exploitent la magnitude du gradient ainsi que la courbure des contours. Il a été démontré équivalent à plusieurs autres détecteurs de la même catégorie. Il fait l'objet de la section 1.2.1. Le détecteur de Harris et Stephens utilise quant à lui uniquement les dérivées du premier ordre. Il a été évalué comme étant le plus stable, surtout s'il est implémenté dans sa version précise. La méthode est présentée à la section 1.2.2.

Ces deux détecteurs font l'objet d'une présentation plus approfondie car ils apparaissent comme étant les plus représentatifs et/ou performants parmi tous les détecteurs existants et puis parce qu'ils extraient les points directement à partir du signal. Nous montrons alors à la section 1.3 comment il est possible de les généraliser pour qu'ils tiennent compte de l'information couleur. Dans cette section, un état de l'art des quelques méthodes de segmentation qui existent pour la couleur est réalisé. Puis nous présentons à la section 1.3.2.1 l'adaptation couleur du détecteur de Kitchen et Rosenfeld et à la section 1.3.2.2 le détecteur de Harris Précis Couleur. Des exemples de détection de points d'intérêt viennent illustrer la présentation de ces travaux, permettant ainsi une première évaluation visuelle.

Ce chapitre suppose d'avoir à l'esprit quelques notions de filtrage en traitement d'images. Le lecteur peut consulter l'annexe C pour une introduction au filtrage en précisions pixel et sub-pixel, ainsi qu'à la segmentation contours en niveau de gris. Dans toutes nos expérimentations, les dérivées de l'image sont calculées à partir d'un support gaussien, en précision pixel pour certaines et sub-pixel pour d'autres.

1.1 Choix d'une primitive : les points d'intérêt

L'extraction d'indices dans une image est une étape indispensable en Vision par Ordinateur. Ainsi par exemple le calcul de la géométrie épipolaire ou encore le processus de reconstruction tridimensionnelle nécessitent la donnée d'un certain nombre de mises en correspondance entre deux images. Traditionnellement, les méthodes permettant de déterminer ces mises en correspondance peuvent être séparées en deux classes :

- Les approches dites *iconiques* qui sont basées directement sur le signal (l'image et ses dérivées) : elles permettent de mettre en correspondance des points d'intérêt entre images, ou bien tous les pixels (appariement dense), uniquement à partir d'informations photométriques ;
- Les approches *géométriques* basées sur d'autres primitives de l'image de plus haut niveau, comme les contours, segments de droite, courbes ou régions : elles prennent en compte leur géométrie et certaines caractéristiques locales pour procéder à l'ap-

pariement. Citons notamment les travaux sur la couleur de [Jordan et Bovik, 1988; Coutance et al., 1989] en ce qui concerne l'appariement de segments, ainsi que ceux de [Brockelbank et Yang, 1989] pour l'appariement de contours. Le lecteur peut se référer à [Jones, 1997] pour un état de l'art.

Nous ne passerons pas en revue dans ce chapitre ce dernier type d'approches car les primitives géométriques s'avèrent trop instables d'une image à l'autre, notamment en présence d'occultations. De plus, les grandeurs symboliques les caractérisant sont des approximations plus ou moins précises de la réalité. Il n'est donc pas souhaitable de les utiliser pour une estimation robuste de la calibration ou de la géométrie épipolaire, lorsque les images ne sont pas suffisamment structurées.

Dans ce contexte, la détection de points d'intérêt directement à partir du signal semble au contraire s'avérer la technique la plus précise. On peut par exemple citer les travaux de Zhang [Zhang et al., 1995b] qui s'appuient sur ce type de primitives pour estimer de façon robuste la géométrie épipolaire.

D'autre part, la détection directement à partir du signal se révèle être la solution la plus adaptée à la méthode de caractérisation que nous développons dans le chapitre 4. En effet, les points d'intérêt sont détectés aux endroits de l'image où se trouve l'information la plus discriminante. Les positions trouvées permettent donc d'exploiter au mieux ce type de caractérisation.

Enfin la détection des points d'intérêt s'avère fiable quelles que soient les images traitées. Elle peut en effet être envisagée de la même façon sur des images très texturées comme sur des images contenant beaucoup de contours, contrairement aux méthodes basées sur les caractéristiques symboliques dont l'efficacité dépend fortement du contenu géométrique de la scène.

1.2 Etat de l'art en niveau de gris

Un grand nombre de travaux ont été réalisés concernant la détection iconique de points d'intérêt dans des images en niveau de gris. Les plus anciens sont ceux de Moravec [Moravec, 1977], qui extrait des coins de l'image à partir de la fonction d'auto-corrélation du signal. C'est le premier a avoir développé l'idée de "points d'intérêt". Selon lui, un point d'intérêt est défini lorsque les valeurs de l'intensité varient beaucoup dans toutes les directions. De nombreux travaux proposent des améliorations de cette première définition. Citons notamment le détecteur mis en place par Beaudet [Beaudet, 1978] lié à la courbure gaussienne de l'image, ensuite amélioré par Dreschler et Nagel dans [Dreschler et Nagel, 1982] ou encore celui développé dans [Zuniga et Haralick, 1983]. Il a d'ailleurs été démontré dans [Nagel, 1983; Shah et Jain, 1984; Noble, 1988] que ces deux derniers détecteurs, avec celui de Kitchen et Rosenfeld présenté plus loin, sont similaires, puisqu'ils exploitent tous les trois le produit de la magnitude du gradient avec la courbure des isophotes [1]. Deriche et Giraudon proposent également dans [Deriche et Giraudon, 1991] une méthode permettant de localiser précisément le point comme l'endroit où le laplacien s'annule sur la droite définie par les points détectés par l'approche de Beaudet implémentée à différentes échelles.

[1] ou lignes d'iso-intensité

Förstner propose quant à lui dans [Förstner et Gülch, 1987; Forstner, 1994] une approche basée sur la statistique locale de l'image.

Dans les deux sections qui suivent, nous passons en revue le détecteur de Kitchen et Rosenfeld [Kitchen et Rosenfeld, 1982] ainsi que celui de Harris [Harris et Stephens, 1988; Bauckhage et Schmid, 1996]. Ceux-ci font l'objet d'une section chacun car c'est à partir de ces deux détecteurs que nos travaux se sont orientés vers la détection de points d'intérêt en couleur.

1.2.1 Les coins de Kitchen et Rosenfeld

L'approche de Kitchen et Rosenfeld est une des approches les plus représentatives de toutes celles qui travaillent à l'ordre deux et qui exploitent la magnitude du gradient ainsi que la courbure des contours. Elle consiste plus précisément à maximiser la courbure des isophotes multipliée par la norme du gradient. La mesure obtenue donne des réponses fortes lorsque les valeurs de courbure et de gradient sont simultanément élevées. Comme pour les contours, cette phase s'appuie sur une étape de filtrage, ses résultats dépendent donc fortement de l'estimation des dérivées premières et secondes de l'image par rapport à x et à y et notées I_x, I_y, I_{xx}, I_{xy} et I_{yy} (cf. annexe C).

La courbure des isophotes peut s'exprimer par l'opérateur suivant :

$$K = \frac{I_{xx}I_y^2 - 2\,I_{xy}I_xI_y + I_{yy}I_x^2}{\| \vec{\nabla} I \|^3} \tag{1.1}$$

Localement, l'orientation des isophotes par rapport à l'axe des x peut s'exprimer de la manière suivante :

$$\theta = \arctan\left(\frac{-I_x}{I_y}\right) \tag{1.2}$$

L'opérateur de Kitchen et Rosenfeld s'écrit donc :

$$KR = \frac{I_{xx}I_y^2 - 2\,I_{xy}I_xI_y + I_{yy}I_x^2}{\| \vec{\nabla} I \|^2} \tag{1.3}$$

Nous présentons ci-dessous l'algorithme de détection :

- Détection des contours selon le processus présenté à la section C.2 de l'annexe C (avec un seuillage par hystérésis très bas de manière à obtenir de nombreux contours) ;
- Calcul de l'opérateur de Kitchen et Rosenfeld KR;
- Calcul de l'orientation des isophotes ;
- Calcul des maxima locaux de KR dans la direction de ces isophotes ;
- Multiplication de l'image de ces maxima avec l'image des contours. On obtient un ensemble de coins ;
- Calcul des maxima locaux et seuillage de ces coins.

1.2.2 Le détecteur de Harris et Stephens

Le détecteur de Harris et Stephens [Harris et Stephens, 1988] est aussi connu sous le nom de détecteur de points caractéristiques de Plessey. Il a été conçu sur le même principe que le détecteur de Moravec, à la différence que la mesure de l'auto-corrélation du signal $\mathcal{A}$ (cf. annexe A) est ici estimée à partir des dérivées de l'image du premier ordre.

1.2.2.1 Stabilisation du détecteur

Dans sa version précise [Bauckhage et Schmid, 1996], les calculs de dérivation utilisent les dérivées d'une fonction de lissage gaussien, les rendant ainsi plus stables.

L'idée de base de ce détecteur est d'utiliser la fonction d'auto-corrélation pour déterminer les positions où le signal change dans deux directions simultanément. Les dérivées du signal sont calculées sur un support gaussien de taille σ (cf. annexe C). En prenant en compte les dérivées premières sur une fenêtre $\widetilde{\sigma}$, une matrice M liée à cette fonction d'auto-corrélation est calculée :

$$M = G(\widetilde{\sigma}) \otimes \begin{bmatrix} I_x^2(\sigma) & I_x(\sigma)I_y(\sigma) \\ I_x(\sigma)I_y(\sigma) & I_y^2(\sigma) \end{bmatrix} \tag{1.4}$$

où $G(\widetilde{\sigma})$ représente un lissage gaussien. Notons que le facteur de lissage σ utilisé pour le calcul des dérivées et celui pour le fenêtrage $\widetilde{\sigma}$ ne sont pas obligatoirement égaux.

Les valeurs propres de M sont les courbures principales de la fonction d'auto-corrélation. Deux valeurs suffisamment grandes indiquent la présence d'un point d'intérêt. Pour ne pas avoir à extraire ces valeurs propres, Harris utilise une mesure reposant sur le déterminant et la trace de la matrice M. Les coins sont alors extraits à partir de l'opérateur suivant :

$$Det(M) - k.Trace^2(M) \quad \text{avec} \quad k = 0.04 \tag{1.5}$$

où k est un paramètre permettant de combiner l'information de contour donnée par la trace de M avec l'information d'angularité donnée par le déterminant de M. L'algorithme de détection est présenté ci-après :

- Calcul des dérivées premières de l'image avec un lissage gaussien de paramètre σ ;
- Calcul de la matrice M avec lissage gaussien dans la fenêtre $\widetilde{\sigma}$;
- Calcul des maxima locaux de l'opérateur défini à l'équation 1.5.

Notons que ce détecteur extrait des points là où l'image est riche en information, ce qui ne limite pas l'extraction aux seuls coins de l'image.

Cet opérateur est en fait une version légèrement modifiée de l'opérateur de Plessey qui minimise le rapport de la trace par le déterminant de M. Noble a montré dans [Noble, 1988] que la matrice M peut se décomposer en une somme de deux termes. Si les dérivées premières de l'intensité peuvent être approchées par un développement de Taylor au premier ordre et si la fenêtre de calcul est symétrique par rapport à l'origine et par rotation, alors on peut écrire :

$$M = A_g + \alpha^2 He^2 \tag{1.6}$$

avec :

$$A_g = \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \quad He = \begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix} \tag{1.7}$$

He est la matrice Hessienne et reflète la courbure de la surface des intensités. Ces deux termes montrent que M combine les contributions des dérivées premières et secondes. Le paramètre α permet de mêler les informations de gradient et de courbure et correspond à la largeur du filtre gaussien.

Remarquons que développée sous la forme 1.7, la matrice M est utilisée par Nagel dans [Nagel, 1987] pour l'estimation de flots optiques. Elle est également impliquée dans plusieurs autres types de détecteurs. Förstner par exemple définit dans [Förstner et Gülch, 1987] un point comme l'intersection de deux ou plusieurs contours. Si ces derniers sont rectilignes alors l'algorithme de détermination des points d'intérêt consiste à calculer l'intersection des droites les portant, en minimisant les distances orthogonales entre toutes les droites dans un certain voisinage. C'est cette minimisation qui fait intervenir la matrice M. De même, le détecteur mis en place par Rohr dans [Rohr, 1992] extrait les points d'intérêt en maximisant le déterminant de M.

Evaluation

Dans [Schmid, 1996], Schmid évalue la version précise du détecteur de Harris et Stephens avec plusieurs autres détecteurs de points d'intérêt (les détecteurs de Heitger, de Förstner, de Horaud et de Cottier). Le critère d'évaluation utilisé est la répétabilité de la détection lorsque l'image est prise dans des conditions différentes : rotation image, changement d'échelle, changement de luminosité, changement de point de vue et bruit du système de prise de vue. Dans tous les cas, Schmid arrive à la conclusion que le détecteur de Harris précis donne des résultats meilleurs ou équivalents aux autres détecteurs. Les résultats de cette étude montrent que c'est le détecteur de Harris précis qui est le plus répétable en niveau de gris. Sa stabilité provient principalement du fait qu'il utilise les dérivées d'une fonction de lissage mais aussi qu'il travaille seulement à l'ordre un. En ce qui concerne la robustesse face aux changements d'échelle, il est néanmoins nécessaire d'envisager une approche multi-échelle, lorsque le facteur d'échelle n'est pas connu. Cette approche est développée dans la prochaine section.

1.2.2.2 Approche multi-échelle

Un changement d'échelle peut être dû soit au déplacement de la caméra, soit à un changement de la distance focale de celle-ci. Les résolutions des deux images différant de ce changement en sont alors modifiées. Si la résolution est définie par $r = \frac{f}{Z}$ où f correspond à la focale de la caméra et Z à la distance moyenne à la scène, alors le changement d'échelle α entre les deux images vaut $\alpha = \frac{r_1}{r_2}$ où r_i est la résolution de la caméra i.

Soient deux fonctions f_1 et f_2 différant de ce facteur d'échelle α. Ce dernier peut être modélisé par le changement affine suivant :

$$f_1(x) = f_2(u(x)) = f_2(\alpha x + c) \tag{1.8}$$

Dans le contexte d'une représentation discrète comme c'est le cas avec les images, les dérivées sont calculées par convolution avec les dérivées gaussiennes. Si G représente la gaussienne et $G_{i_1...i_n}$ ses dérivées suivant $i_1...i_n$, alors on obtient :

$$I_1(x) * G_{i_1...i_n}(\sigma) = \alpha^n I_2(\alpha x + c) * G_{i_1...i_n}(\alpha\sigma) \tag{1.9}$$

Ce résultat montre que le détecteur de points doit être adapté pour obtenir des résultats répétables. Les deux facteurs de lissage doivent être multipliés par α. La matrice résultante est alors multipliée par α^2 pour avoir des valeurs propres comparables. Si la matrice utilisée pour la première image est celle de l'équation 1.4, alors celle pour la seconde image est :

$$M_2 = \alpha^2 G(\alpha\tilde{\sigma}) \otimes \begin{bmatrix} I_x^2(\alpha\sigma) & I_x(\alpha\sigma)I_y(\alpha\sigma) \\ I_x(\alpha\sigma)I_y(\alpha\sigma) & I_y^2(\alpha\sigma) \end{bmatrix} \tag{1.10}$$

Ainsi les notions de différenciation et de changement d'échelle sont intimement liées. Cette constatation est à l'origine de l'analyse multi-échelle, dont le principe nous vient de Marr [Marr, 1982]. Nous ne considérerons pas ici les approches qui effectuent un sous-échantillonnage de l'image [Murase et Nayar, 1995], car elles ont l'inconvénient d'introduire du bruit de discrétisation ou des distorsions dans l'image traitée. Notre problème repose plutôt sur la notion d'*espace échelle*[2], introduite sous sa forme continue par Witkin [Witkin, 1983] et Koenderink [Koenderink, 1984]. La définition de l'espace échelle pour les signaux 1D continus est la suivante :

Définition 1 (Espace échelle) *Soit $f(x)$ une fonction et $G_\sigma(x)$ la gaussienne d'écart-type σ. On appelle espace échelle le lieu des réalisations de la transformation S définie par :*

$$(Sf)(x,\sigma) = f \star G_\sigma(x), \qquad (x,\sigma) \in \mathcal{R} \times \mathcal{R}^+ \tag{1.11}$$

On appellera S opérateur de changement d'échelle et l'on notera $E = (x,\sigma)$ l'espace échelle.

La transformation énoncée ci-dessus conduit à la représentation d'une fonction sous la forme d'une surface décrite dans E. On peut donc étudier directement cette surface en utilisant des arguments de géométrie différentielle. Dans notre cas, pour obtenir les dérivées gaussiennes, nous simplifions la représentation en dérivant la surface uniquement suivant les variables x et y.

Si le facteur d'échelle est inconnu, ce qui est le cas le plus souvent, alors il faut envisager de calculer la matrice de l'équation 1.10 dans un contexte multi-échelle, c'est-à-dire pour plusieurs valeurs de α. De nombreuses études ont été menées sur la discrétisation de l'espace échelle, on peut notamment consulter [Falzon, 1994] pour un état de l'art ou encore [Armande, 1997] pour une analyse multi-échelle dans le cadre de la détection de réseaux fins. Dans [Schmid, 1996], Schmid propose un espacement linéaire de 0.5. Les résultats obtenus sont alors bons jusqu'à un facteur d'échelle de 3. Dans une version plus récente [Dufournaud et al., 2000], le facteur d'échelle est estimé lors de l'étape de la mise en correspondance, ce qui permet d'obtenir des résultats plus robustes jusqu'à un facteur d'échelle de 6. Cette méthode est plus amplement détaillée dans la section 5.1.4.3 du chapitre 5.

[2]Pour un résumé, consulter par exemple [Lindeberg, 1993; ter Haar Romeny, 1996].

1.3 Segmentation couleur

Comme en niveau de gris, la segmentation couleur s'appuie sur les étapes de filtrage qui peuvent être réalisées en sub-pixel (cf. annexe C). Les contours et les coins sont des généralisations des méthodes définies en niveau de gris. Dans la prochaine section, nous rappelons les principales méthodes existantes permettant d'extraire les contours d'une image couleur, à savoir l'approche basée sur le gradient de Di-Zenzo et celle basée sur le gradient maximal. Jusqu'à ce jour, aucune approche couleur n'a été développée pour l'extraction de points d'intérêt. Nous présentons dans la section 1.3.2 deux approches spécifiques à la couleur, qui correspondent aux extensions des détecteurs de Harris précis et de Kitchen-Rosenfeld. Les travaux concernant ces deux nouveaux détecteurs ont donné lieu à des publications [Montesinos et al., 1998; Gouet et al., 2000; Montesinos et al., 2000].

1.3.1 Segmentation contours

Il existe dans la littérature de nombreuses méthodes permettant d'extraire des informations (contours, coins, régions) d'images couleur [Luong, 1991], certaines travaillent directement dans l'espace RVB comme dans [Machuca et Phillips, 1983; Di Zenzo, 1986; Cumani et al., 1990; Fellah, 1994; Zhu et Yuille, 1996; Sapiro et Ringach, 1996]. D'autres encore travaillent dans des espaces couleur HLS, HSI, CIELUV, CIELAB [Luong, 1991; Schettini, 1993; Perez et Koch, 1994]. Les méthodes directes (RVB) semblent bien adaptées à l'extraction de contours et points caractéristiques [Di Zenzo, 1986; Cumani et al., 1990; Machuca et Phillips, 1983], en revanche de nombreuses méthodes de segmentation région travaillent dans d'autres espaces couleur [Luong, 1991; Schettini, 1993; Perez et Koch, 1994].

Nous présentons dans cette section deux approches qui sont des généralisations de la segmentation contours en niveau de gris au cas multi-spectral et qui travaillent directement dans l'espace RVB. La première méthode est basée sur le gradient multi-spectral de Di-Zenzo [Di Zenzo, 1986; Cumani et al., 1990] et la seconde sur la maximisation du gradient dans les trois plans RVB [Fellah, 1994]. Comme en témoignent les images de contour qui sont présentées, les résultats obtenus à l'aide de ces méthodes sont de bonne qualité, la détection couleur permettant de diviser par trois le rapport signal/bruit par rapport aux images en niveau de gris.

1.3.1.1 Le gradient multi-spectral de Di-Zenzo

Soit une image multi-spectrale $I(x)$. Nous nous intéressons aux fortes variations de cette fonction dans le plan image où un pixel est défini par un vecteur $x = (x^1, x^2)^T$. La fonction image étant elle même vectorielle telle que $I(x) = (R(x), V(x), B(x))^T$, nous exprimons les variations de la norme de ce vecteur, soit :

$$dI^2 = \sum_{h=1}^{2} \sum_{k=1}^{2} g_{hk} dx^h dx^k \tag{1.12}$$

sous la condition :

$$\sum_{h=1}^{2}\sum_{k=1}^{2} dx^h dx^k = 1 \tag{1.13}$$

et où g_{hk} représente la composante (h,k) du tenseur multi-spectral $G_{(2\times 2)}$ telle que :

$$g_{hk} = \vec{v}_h.\vec{v}_k \tag{1.14}$$

avec

$$\vec{v}_1 = \begin{pmatrix} \frac{\partial R(\vec{x})}{\partial x^1} \\ \frac{\partial V(\vec{x})}{\partial x^1} \\ \frac{\partial B(\vec{x})}{\partial x^1} \end{pmatrix} \qquad \vec{v}_2 = \begin{pmatrix} \frac{\partial R(\vec{x})}{\partial x^2} \\ \frac{\partial V(\vec{x})}{\partial x^2} \\ \frac{\partial B(\vec{x})}{\partial x^2} \end{pmatrix} \tag{1.15}$$

La norme du gradient multi-spectral est donnée par la racine carrée de la valeur propre maximale λ_{max} du tenseur multi-spectral G et l'orientation du gradient multi-spectral par rapport à l'axe des x est donnée par le vecteur propre $\vec{P}_{max}$ associé à λ_{max}, soit :

$$\lambda_{max} = \frac{1}{2}\left(g_{11} + g_{22} + \sqrt{\Delta}\right) \tag{1.16}$$

avec

$$\Delta = (g_{11} - g_{22})^2 + 4\,g_{12}^2 \tag{1.17}$$

et

$$\vec{P}_{max} = \frac{1}{\sqrt{4\,g_{12}^2 + \left(g_{11} - g_{22} - \sqrt{\Delta}\right)^2}} \begin{pmatrix} -2\,g_{12} \\ g_{11} - g_{22} - \sqrt{\Delta} \end{pmatrix} \tag{1.18}$$

L'angle du gradient multi-spectral par rapport à l'axe x est donné par la formule suivante :

$$\theta = \arctan\left(-\frac{g_{11} - g_{22} - \sqrt{\Delta}}{2\,g_{12}}\right) \tag{1.19}$$

ou encore en passant par l'angle double :

$$\theta = \frac{1}{2}\arctan\left(\frac{2\,g_{12}}{g_{11} - g_{22}}\right) \tag{1.20}$$

La détection de contours se poursuit de manière analogue aux méthodes en niveau de gris : suppression des non maxima locaux du gradient dans la direction du gradient et seuillage par hystérésis (se reporter à la section C.2 de l'annexe C).

Nous avons implémenté cette méthode. L'image couleur utilisée pour tous les tests de ce chapitre est présentée à la figure 1.1. Nous avons également à notre disposition son équivalent en niveau de gris (cf. annexe B). Nous présentons des exemples de segmentation contour en couleur mais également en niveau de gris. Ceux-ci ont été réalisés en précision pixel à la figure 1.2 et en précision quart de pixel à la figure 1.3, à l'issue d'un seuillage

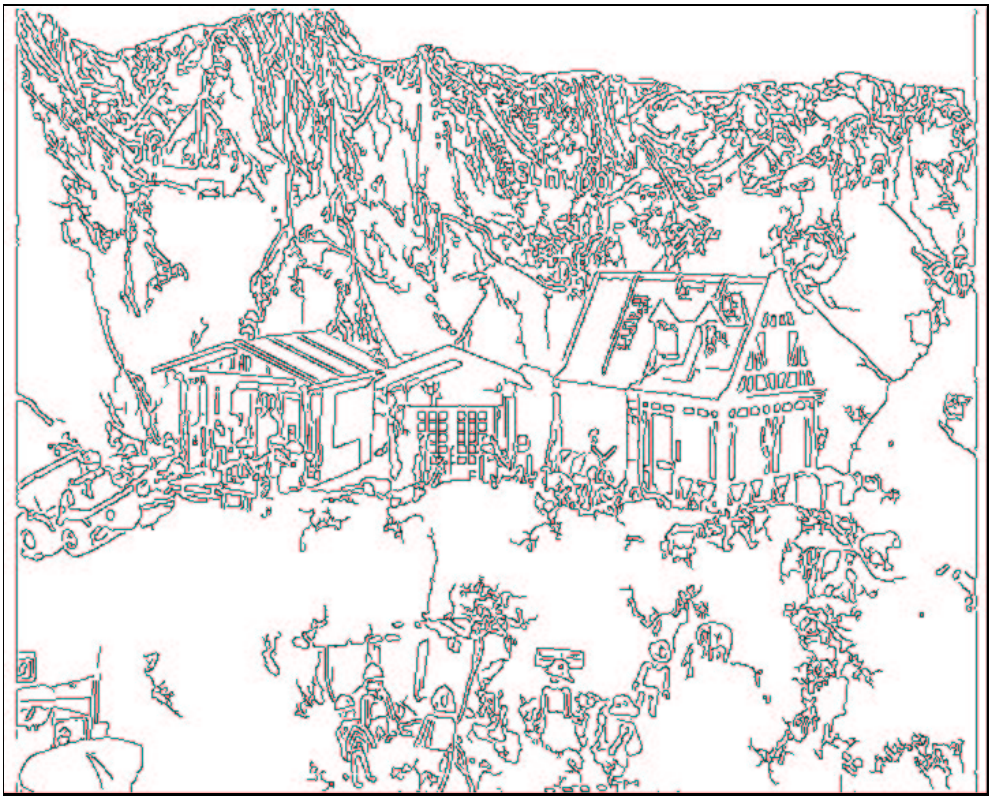
par hystérésis. Le filtre utilisé est un filtre gaussien récursif ($\sigma = 1$). La première ligne de chaque figure montre les résultats de détection obtenus en niveau de gris à partir du gradient mono-spectral, alors que la deuxième ligne présente les contours obtenus à partir du gradient multi-spectral de Di-Zenzo. Le lecteur peut constater une amélioration notable de la détection lorsque l'information couleur est exploitée, par exemple au niveau du toit de la maison la plus à gauche dans les images de la figure 1.2. Des résultats complémentaires sont consultables à la section D.1.1 de l'annexe D.

Figure 1.1 – *Image initiale "Toys".*

1.3.1.2 Gradient maximal

Alors que le gradient multi-spectral de Di-Zenzo utilise les propriétés différentielles exactes des fonctions vectorielles, il est possible de définir de nombreuses autres méthodes pour combiner les gradients provenant des différents plans d'une image couleur. Il s'agit de définir des fonctions scalaires croissantes par rapport à chacune des variables, de R^6 dans R^+ (où plus généralement de R^{2n} dans R^+ dans le cas d'images 2D multi-spectrales à n composantes) . Dans [Fellah, 1994], Fellah construit de telles fonctions à l'aide des

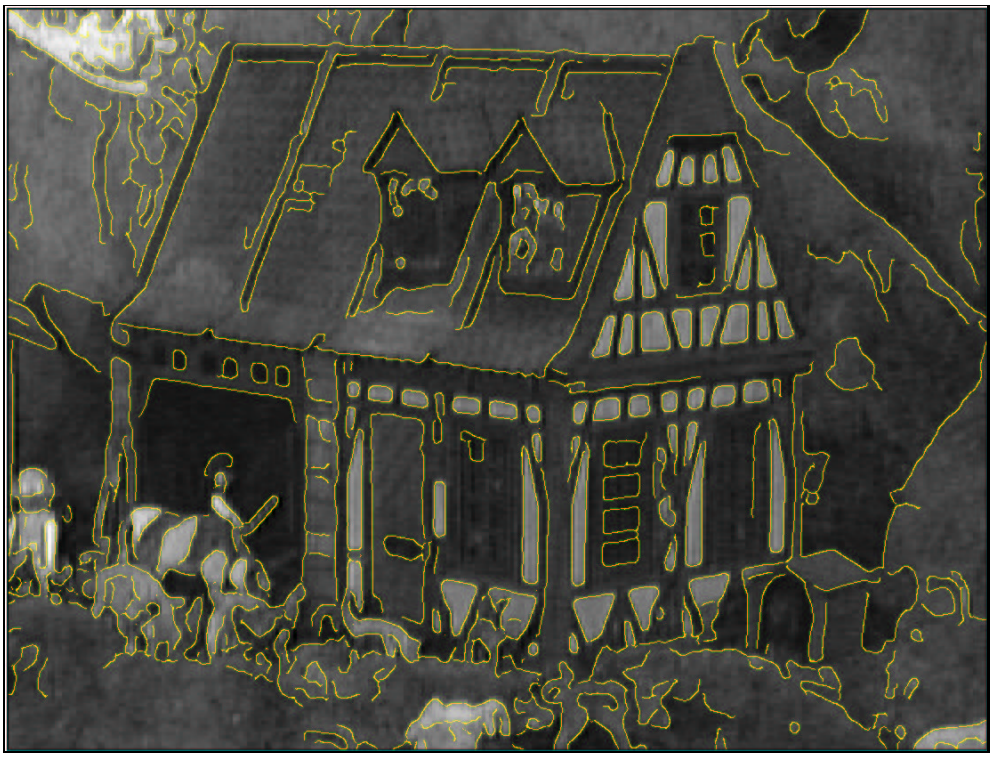
En niveau de gris : contours obtenus à partir du gradient mono-spectral



En couleur : contours obtenus à partir du gradient multi-spectral de Di-Zenzo

Figure 1.2 – *Détection de contours en précision pixel.*

En niveau de gris : contours obtenus à partir du gradient mono-spectral



En couleur : contours obtenus à partir du gradient multi-spectral de Di-Zenzo

Figure 1.3 – *Détection de contours en précision quart de pixel : détail des "maisons".*

normes classiques de R^2 et R^3 :

$$f\left(\frac{\partial R}{\partial x}, \frac{\partial R}{\partial y}, \frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}, \frac{\partial B}{\partial x}, \frac{\partial B}{\partial y}\right) = \| \left(\|\nabla R\|_n, \|\nabla V\|_n, \|\nabla B\|_n\right) \|_m \tag{1.21}$$

Si $n = 2$ et $m = \infty$, on obtient :

$$f\left(\frac{\partial R}{\partial x}, \frac{\partial R}{\partial y}, \frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}, \frac{\partial B}{\partial x}, \frac{\partial B}{\partial y}\right) = \max\left\{\|\nabla R\|_2, \|\nabla V\|_2, \|\nabla B\|_2\right\} \tag{1.22}$$

Dans ce cas, la norme du gradient multi-spectral est la norme infinie des normes quadratiques des gradients dans chacun des plans, l'orientation du gradient multi-spectral est donnée par l'orientation correspondant au maximum des trois gradients. La détection de contours se poursuit de manière analogue au cas en niveau de gris : suppression des non maxima locaux du gradient dans la direction du gradient et seuillage par hystérésis (se reporter à la section C.2 de l'annexe C). Dans la suite des travaux, nous appelons MAX cet opérateur.

Cette méthode a été implémentée et testée. Les résultats obtenus présentent une amélioration significative par rapport à la détection de contours en niveau de gris en précision pixel et semblent aussi meilleurs que ceux obtenus par la méthode de Di-Zenzo en précision pixel, comme l'illustre la figure 1.4, pour laquelle un filtrage gaussien récursif a été utilisé ($\sigma = 1$) ainsi qu'un seuillage par hystérésis. Des résultats complémentaires sont consultables à la section D.1.2 de l'annexe D.

Dédoublement des contours

En revanche, lorsque l'on passe en précision sub-pixel, il semble que cet opérateur détecte aussi les imperfections des capteurs, et notamment des petits décalages sub-pixelliques entre les différents plans de l'image couleur. Ces décalages se traduisent alors par des instabilités et un doublement de certains contours. A partir de la précision quart de pixel, l'opérateur de Di-Zenzo se comporte mieux que l'opérateur MAX et fournit toujours les meilleurs résultats de segmentation. Cette idée est illustrée par les exemples de détection en huitième de pixel à la figure 1.5.

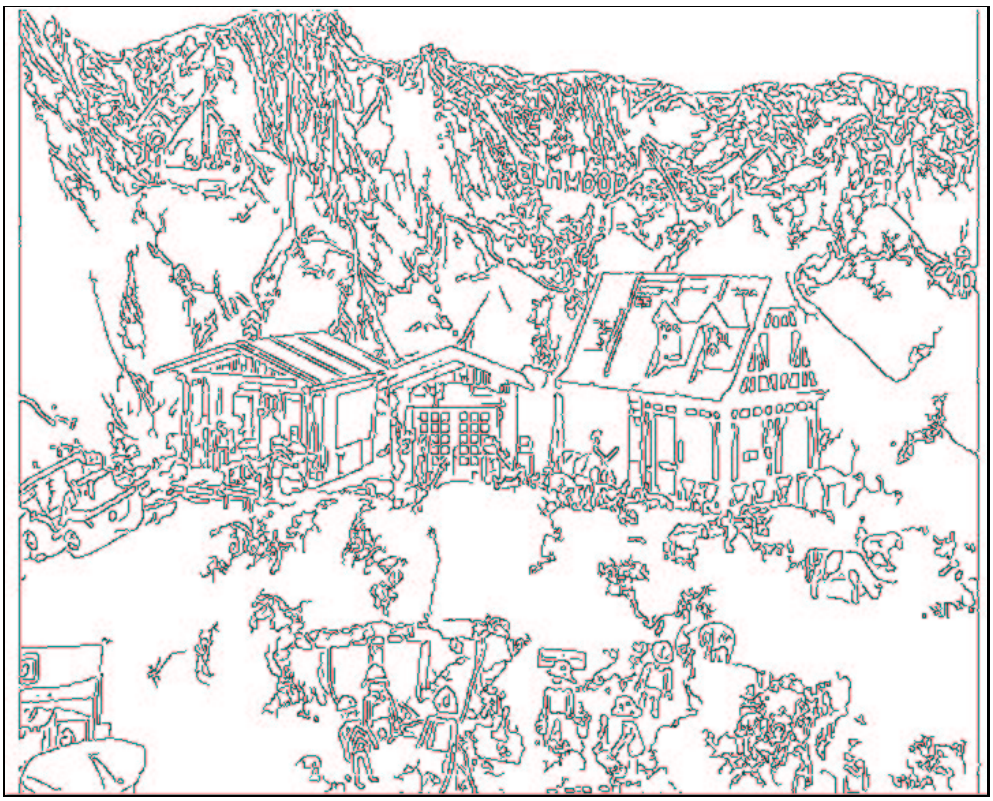
1.3.2 Extraction de points caractéristiques

Pour chacun de ces détecteurs, nous présentons la méthode employée ainsi que quelques résultats obtenus pour différentes images couleur, fournissant une évaluation visuelle de ces deux nouvelles méthodes. Une évaluation plus poussée sera mise en place au prochain chapitre.

1.3.2.1 Les coins

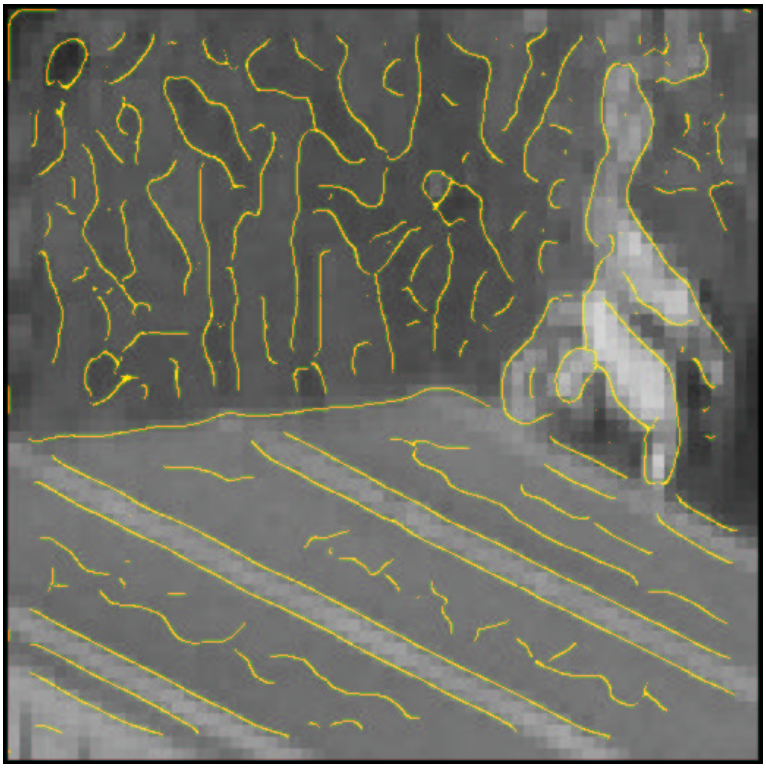
Différentiation d'un champ de vecteur

Opérateur de Di-Zenzo

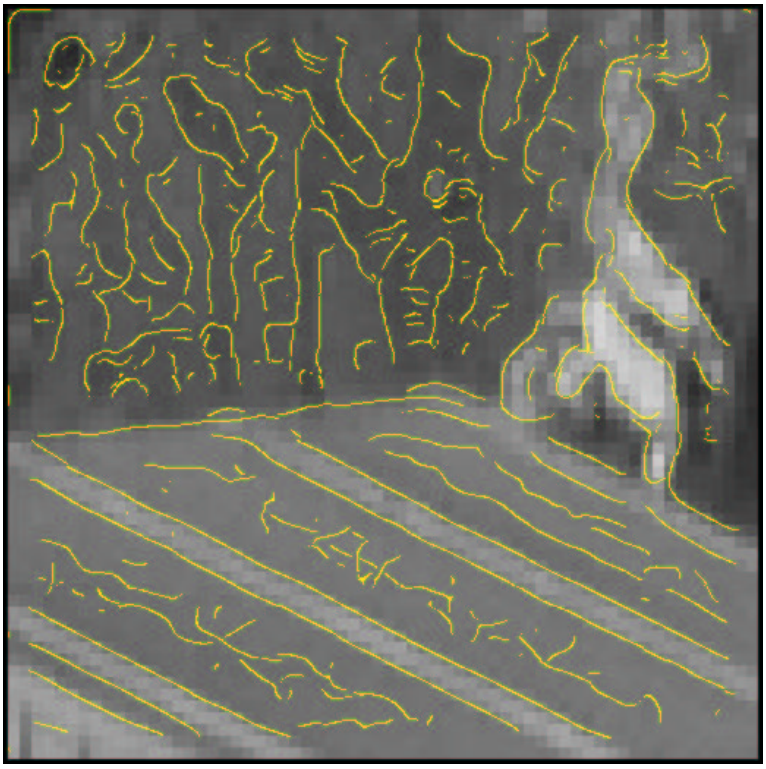


Opérateur MAX

Figure 1.4 – *Détection de contours en précision pixel par les opérateurs de Di-Zenzo et* MAX.



Opérateur de Di-Zenzo



Opérateur MAX

Figure 1.5 – *Détection de contours couleur en précision huitième de pixel (détail d'un "toit de maison").*

De la même façon que l'opérateur de Di-Zenzo généralise le gradient d'une image scalaire au cas des champs de vecteurs, nous généralisons ici la courbure des isophotes au cas multi-spectral $\{R, V, B\}$.

Soit une courbe tracée dans le plan image dont l'orientation de la tangente par rapport à l'axe des x, effectue un angle θ_τ ; la Courbure Multi-Spectrale de cette courbe (notée CMS) s'exprime par :

$$CMS = \frac{\mathrm{d}\theta_\tau}{\mathrm{d}s} \tag{1.23}$$

Ici, θ_τ représente l'orientation de la tangente au contour (au sens de Di-Zenzo) par rapport à l'axe des x, ou encore l'orientation du vecteur propre $\vec{P}_{min}$ correspondant à la valeur propre minimale du tenseur multi-spectral λ_{min} ($\vec{P}_{max}$ et $\vec{P}_{min}$ sont deux vecteurs orthogonaux). Cette expression peut être développée en fonction de l'orientation du gradient multi-spectral θ de la manière suivante :

$$CMS = \frac{\mathrm{d}\theta_\tau}{\mathrm{d}s} = \frac{\partial\theta}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}s} + \frac{\partial\theta}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}s} \tag{1.24}$$

avec

$$\frac{\mathrm{d}x}{\mathrm{d}s} = -\sin\theta \qquad \text{et} \qquad \frac{\mathrm{d}y}{\mathrm{d}s} = \cos\theta \tag{1.25}$$

Aux équations 1.26 et 1.27, nous donnons les expressions de $\frac{\partial\theta}{\partial x}$ et $\frac{\partial\theta}{\partial y}$ en fonction des dérivées de l'image {R,V,B} et d'un paramètre A défini à l'équation 1.28 :

$$\frac{\partial\theta}{\partial x} = \frac{\left(\begin{array}{l} A\,B_x\,B_{xy} + B\,B_{xx}\,B_y - 2\,V_x\,V_{xy}\,{V_y}^2 + 2\,{B_x}^2\,B_{xx}\,B_y + 2\,B_x\,B_{xy}\,{B_y}^2 - \\ 2\,B_x\,B_y\,R_x\,R_{xx} + A\,R_x\,R_{xy} - 2\,B_x\,B_{xx}\,R_x\,R_y + 2\,B_{xy}\,B_y\,R_x\,R_y + \\ A\,R_{xx}\,R_y - 2\,{R_x}^2\,R_{xx}\,R_y + 2\,B_x\,B_y\,R_{xy}\,R_y + 2\,R_x\,R_{xy}\,{R_y}^2 - \\ 2\,B_x\,B_y\,V_x\,X_{xx} - 2\,R_x\,R_y\,V_x\,X_{xx} + A\,V_x\,V_{xy} - 2\,B_x\,B_{xx}\,V_x \\ V_y + 2\,B_{xy}\,B_y\,V_x\,V_y - 2\,R_x\,R_{xx}\,V_x\,V_y + 2\,R_{xy}\,R_y\,V_x\,V_y + \\ A\,X_{xx}\,V_y - 2\,{V_x}^2\,X_{xx}\,V_y + 2\,B_x\,B_y\,V_{xy}\,V_y + 2\,R_x\,R_y\,V_{xy}\,V_y \end{array}\right)}{A^2 + 4\,(B_x\,B_y + R_x\,R_y + V_x\,V_y)^2} \tag{1.26}$$

$$\frac{\partial\theta}{\partial y} = \frac{\left(\begin{array}{l} -2.\,{B_x}^2\,B_{xy}\,B_y + 2.\,B_x\,{B_y}^2\,B_{yy} - 2.\,B_x\,B_y\,R_x\,R_{xy} - 2.\,B_x\,B_{xy}\,R_x\,R_y + \\ 2.\,B_y\,B_{yy}\,R_x\,R_y - 2.\,{R_x}^2\,R_{xy}\,R_y + 2.\,B_x\,B_y\,R_y\,R_{yy} + 2.\,R_x\,{R_y}^2\,R_{yy} - \\ 2.\,B_x\,B_y\,V_x\,V_{xy} - 2.\,R_x\,R_y\,V_x\,V_{xy} - 2.\,B_x\,B_{xy}\,V_x\,V_y + 2.\,B_y\,B_{yy}\,V_x\,V_y - \\ 2.\,R_x\,R_{xy}\,V_x\,V_y + 2.\,R_y\,R_{yy}\,V_x\,V_y - 2.\,{V_x}^2\,V_{xy}\,V_y + \\ B_{xy}\,B_y\,A + B_x\,B_{yy}\,A + R_{xy}\,R_y\,A + R_x\,R_{yy}\,A + V_{xy}\,V_y\,A + 2.\,B_x\,B_y\,V_y\,V_{yy} + \\ 2.\,R_x\,R_y\,V_y\,V_{yy} + 2.\,V_x\,{V_y}^2\,V_{yy} + V_x\,A\,V_{yy} \end{array}\right)}{A^2 + 4\,(B_x\,B_y + R_x\,R_y + V_x\,V_y)^2} \tag{1.27}$$

avec

$$A = {B_x}^2 - {B_y}^2 + {R_x}^2 - {R_y}^2 + {V_x}^2 - {V_y}^2 \tag{1.28}$$

La détection des coins couleur se fait alors selon le processus suivant [Montesinos et al., 1998] :

- 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.

Courbure maximale

De la même manière que pour le gradient maximal, il est possible de définir une fonction scalaire des courbures dans chacun des plans par exemple :

$$f(K_R, K_V, K_B) = \max\{K_R, K_V, K_B\} \tag{1.29}$$

L'orientation des lignes d'iso-intensité peut alors être choisie comme l'orientation correspondant à la courbure maximale (orientation perpendiculaire au gradient). Compte tenu du peu de stabilité de l'opérateur MAX pour la détection de contours en précision sub-pixel, nous n'avons pas testé ici cet opérateur.

Quelques résultats

Nous présentons ici des résultats de détection de coins couleur à partir de l'opérateur CMS (équation 1.24), en précision pixel à la figure 1.6. Les dérivées sont estimées à l'aide de filtres gaussiens récursifs en précision pixel avec $\sigma = 1$. Les paramètres de détection des coins sont de 0.05 pour le seuil et de 9 pixels pour le diamètre de la fenêtre circulaire pour la maximisation des coins. D'autres résultats sont disponibles à la section D.2 de l'annexe D.

1.3.2.2 Le détecteur Harris Précis Couleur

Le détecteur de Harris et Stephens dans sa version précise semble particulièrement adapté à la mise en correspondance de points, de part son excellente répétabilité (cf. section 1.2.2). Sa stabilité provient essentiellement du fait que son implantation n'utilise que les dérivées de l'image d'ordre un. Cette particularité s'avére d'ailleurs fort intéressante pour notre étude car elle est en accord avec la caractérisation que nous présentons au chapitre 4. Notre contribution a donc consisté à introduire l'information couleur $\{R, V, B\}$ dans le détecteur du premier ordre déjà existant. La matrice M utilise alors le tenseur multi-spectral (défini à l'équation 1.14) et devient [Montesinos et al., 1998] :

$$M = G(\widetilde{\sigma}) \otimes \begin{bmatrix} R_x^2(\sigma) + V_x^2(\sigma) + B_x^2(\sigma) & R_x(\sigma)R_y(\sigma) + V_x(\sigma)V_y(\sigma) + B_x(\sigma)B_y(\sigma) \\ R_x(\sigma)R_y(\sigma) + V_x(\sigma)V_y(\sigma) + B_x(\sigma)B_y(\sigma) & R_y^2(\sigma) + V_y^2(\sigma) + B_y^2(\sigma) \end{bmatrix} \tag{1.30}$$

Figure 1.6 – *Détection de coins couleur en précision pixel : généralisation de la courbure des lignes d'iso-intensité au cas multi-spectral.*

Le lecteur peut voir à la figure 1.7 un exemple de points Harris Précis Couleur détectés sur l'image 1.1. Un filtre gaussien ($\sigma = 1$) a été utilisé pour le calcul des dérivées gaussiennes en précision demi-pixel. Le choix de la précision demi-pixel impose de doubler la taille du support pour le second lissage. Il faut donc prendre $\tilde{\sigma} = 2$. La taille de la fenêtre pour le seuillage local est de 15 pixels et la valeur du seuil est 0.05. D'autres résultats sont disponibles à la section D.2 de l'annexe D.

Figure 1.7 – *Détection de points Harris Précis Couleur en précision demi-pixel.*

1.3.2.3 Evaluation visuelle

Considérons un extrait de l'image 1.1 appelé "Maison". Les figures 1.8 et 1.9 montrent les points d'intérêt obtenus respectivement avec le détecteur de Kitchen et Rosenfeld (cf. section 1.2.1) et le détecteur Harris Précis (cf. section 1.2.2) sur l'image en niveau de gris correspondante. Les figures 1.10 et 1.11 quant à elles montrent les résultats obtenus en utilisant les détecteurs correspondants en couleur. Dans tous les cas, la détection a été réalisée à l'aide d'un lissage gaussien de paramètre $\sigma = 1$ en demi-pixel (qui implique $\tilde{\sigma} = 2$ pour le lissage avec Harris). La taille de la fenêtre locale pour le seuillage est de 9

et le seuil varie en fonction des détecteurs, de façon à obtenir à peu près le même nombre de points pour toutes les détections.

Figure 1.8 – *Extrait "Maison" : 121 coins selon Kitchen et Rosenfeld* **en niveau de gris**.

On constate de visu que les quatre ensembles de points caractéristiques extraits semblent d'avoir un grand nombre d'éléments en commun. Cependant, pour certains points (l'épaule du personnage à gauche par exemple ou encore le détail du torchis sur la façade de la maison), le point détecté ne semble pas être localisé exactement au même endroit dans les images contenant les points détectés par les opérateurs de Harris et celles associées aux opérateurs de Kitchen et Rosenfeld. Ceci sous-entend que ces deux classes de détecteurs ne réalisent pas la détection avec la même précision. Il semble que, contrairement aux opérateurs de Kitchen et Rosenfeld, les détecteurs de Harris ne détectent pas exactement

Figure 1.9 – *Extrait "Maison" : 128 points Harris Précis* **en niveau de gris**.

Figure 1.10 – *Extrait "Maison" : 142 coins selon Kitchen et Rosenfeld* **en couleur**.

Figure 1.11 – *Extrait "Maison" : 119 points Harris précis* **en couleur**.

les coins.

On peut également constater quelques différences entre les détections en niveau de gris et les détections correspondantes en couleur pour chaque classe de détecteur (voir le détail du torchis au dessus de la fenêtre la plus à gauche par exemple).

Il est bien évident qu'une évaluation simplement visuelle ne suffit pas ici, il faudrait pouvoir comparer avec précision les points obtenus sur une série d'images, afin de pouvoir émettre un jugement plus "qualitatif". Ceci sera fait dans la suite de ces travaux.

1.4 Conclusion

Dans ce chapitre nous avons présenté un état de l'art des principales méthodes de détection de points d'intérêt. Nous nous sommes particulièrement intéressés aux approches utilisant directement le signal. C'est en effet ce type d'approche qui est susceptible d'exploiter au mieux l'information couleur dont nous voulons tenir compte, en plus de donner les résultats de détection les plus précis et les plus stables. Dans ce contexte, deux de ces méthodes ont été examinées en particulier. Il s'agit du détecteur de Kitchen et Rosenfeld et du détecteur de Harris et Stephens Précis. Notre travail a alors consisté à les généraliser au cas d'images en couleur.

Les résultats de détection qui ont été présentés en niveau de gris et en couleur mettent en avant deux points précis. D'une part, il semble que les deux détecteurs de Harris et ceux basés sur Kitchen et Rosenfeld n'extraient pas exactement les mêmes points. Quelle est alors la classe de détecteur qui offre la meilleure précision ? D'autre part, il semble aussi que des différences apparaissent dans la détection dès que l'on exploite l'information couleur. Est-ce que cette information supplémentaire permet d'améliorer l'extraction des points ? Il a été prouvé que le détecteur de Harris Précis en niveau de gris était plus stable que les principaux autres détecteurs, selon un critère de répétabilité. Il est donc également nécessaire de mettre en place des critères plus rigoureux qu'une simple évaluation visuelle pour juger les deux nouveaux détecteurs dédiées à la couleur que nous venons de présenter. C'est ce travail qui fait l'objet du prochain chapitre. Un état de l'art sur les quelques critères d'évaluation d'extracteurs de points existants y est présenté. Puis deux critères sont proposés et testés sur nos deux nouveaux détecteurs ainsi que sur le détecteur Harris Précis en niveau de gris qui sert de référence.

Chapitre 2

Evaluation des détecteurs couleur

Afin d'être à même de juger de la précision des deux nouveaux détecteurs couleur qui viennent d'être présentés, nous mettons en place dans ce chapitre une méthode d'évaluation basée sur deux critères : la répétabilité du détecteur ainsi que sa localisation. Les deux détecteurs couleur sont alors évalués aux côtés d'un détecteur de points en niveau de gris qui sert de référence. Pour tester la robustesse de ces opérateurs, l'étude est réalisée dans différents contextes incluant les principales transformations de l'image.

Sommaire

En Vision par Ordinateur, beaucoup d'applications telles que l'indexation ou la reconstruction tridimensionnelle ont pour base l'extraction de points d'intérêt. Leur fiabilité dépend donc grandement de la précision des points extraits. Il existe beaucoup de travaux sur les détecteurs de points d'intérêt en niveau de gris, mais la méthode d'évaluation de ces détecteurs la plus répandue consiste à vérifier visuellement la qualité des points trouvés. Cette analyse fournit bien souvent une appréciation 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. Nous avons donc jugé utile de mettre en œuvre dans ce chapitre une méthode d'évaluation pour évaluer avec précision la valeur des deux détecteurs couleur présentés au chapitre précédent. Dans la section 2.1, nous présentons les critères que nous avons choisis pour évaluer ces détecteurs. Dans la section 2.2, nous les estimons sur des séries d'images très différentes les unes des autres, afin d'observer le comportement de nos détecteurs dans plusieurs situations. Pour être à même de comparer les résultats obtenus, les mesures sont aussi réalisées sur un détecteur en niveau de gris qui a fait ses preuves. L'ensemble des résultats est enfin discuté dans la section 4.4. L'ensemble des travaux développés dans ce chapitre a fait l'objet d'une publication [Gouet et al., 2000].

2.1 Méthode d'évaluation

Après avoir passé en revue dans la section 2.1.1 les quelques méthodes d'évaluation qui existent déjà, nous définissons en 2.1.2 et 2.1.3 les critères que nous avons adoptés pour l'évaluation des extracteurs de points couleur. Dans la section 2.1.4, nous présentons le contexte qui a permis d'établir ces mesures.

2.1.1 Etat de l'art

D'un point de vue général, un détecteur de points d'intérêt peut être caractérisé par les deux critères suivants :

- sa *répétabilité* : le détecteur est dit répétable si le même point est détecté dans une suite d'images ;
- sa *localisation* : elle signifie qu'un point détecté dans l'image correspond de façon précise à un point de la scène tridimensionnelle.

Bien que le critère d'appréciation des détecteurs soit très souvent simplement visuel, quelques travaux ont été réalisés sur la répétabilité et la localisation des détecteurs en niveau de gris. Dans sa thèse [Schmid, 1996], Schmid met en place un critère de répétabilité qui comptabilise le nombre de points répétés d'une vue à l'autre dans une séquence d'images. Cette mesure lui permet de mettre en évidence la stabilité du détecteur de Harris précis. Elle ne tient pas compte de la mesure de localisation car les applications visées dans son étude portent sur l'appariement et la reconnaissance d'objets dans une base de données. Dans un article récent [Schmid et al., 2000], elle propose en plus une mesure d'entropie pour déterminer quels sont les détecteurs qui permettent d'extraire le

point là où l'information photométrique est la plus riche. Quelques années plus tôt, Deriche et Giraudon dans [Deriche et Giraudon, 1993] étudient quant à eux la localisation des détecteurs les plus classiques en niveau de gris, à partir de leur modèle de point. Une des conclusions de leur étude est qu'aucun de ces détecteurs ne possède une excellente mesure de localisation. Ils exploitent d'ailleurs cette lacune pour estimer la position exacte du point en lissant plus ou moins l'image. Plus tard, Brand estime dans [Brand, 1995] une mesure de localisation pour comparer plusieurs détecteurs en niveau de gris. L'application visée dans son cas étant la reconstruction tridimensionnelle, la mesure qu'il définit s'appuie sur l'alignement de points, la précision de la reconstruction, la précision du calcul de la géométrie épipolaire et la stabilité des birapports. Dans sa thèse [Blaszka, 1997], Blaszka passe en revue les principaux extracteurs de coins et quantifient leur précision et leur fiabilité. Il travaille à partir d'images synthétiques bruitées d'angles et de jonctions. La détection est réalisée pour plusieurs valeurs de lissage et une erreur de localisation est calculée entre les points extraits et la position réelle du sommet de l'angle ou de la jonction. Heyden et Rohr quant à eux étudient dans [Heyden et Rohr, 1996] la précision de la détection pour cinq extracteurs de points, en utilisant la conservation d'invariants projectifs entre des points de la scène tridimensionnelle et les points détectés. De façon similaire, Coelho et al. développe dans [Coelho et al., 1992] une méthode pour évaluer les performances de trois détecteurs de points, à partir d'invariants projectifs.

Plusieurs travaux [Canny, 1986; Deriche et Giraudon, 1993; Wang et Brady, 1992; Fidrich et Thirion, 1998] ont mis en évidence que le lissage améliore la répétabilité du détecteur mais en revanche détériore sa localisation. Dans le cas général, ces deux critères d'évaluation prennent plus ou moins d'importance selon l'application envisagée. Une bonne répétabilité du détecteur sera précieuse par exemple lorsqu'il s'agit de mise en correspondance, de l'estimation de la géométrie épipolaire ou encore de l'indexation d'images. En revanche, pour la reconstruction tridimensionnelle ou le calibrage, plus généralement pour toutes les applications qui font intervenir la 3D, il faudra en plus tenir compte de la mesure de localisation.

L'objectif de nos travaux étant le transfert d'images ou encore la reconstruction tridimensionnelle, il s'est donc avéré nécessaire de tenir compte des deux critères. Les deux sections qui suivent présentent les solutions adoptées.

2.1.2 Critère de répétabilité

Nous reprenons ici la définition de la répétabilité introduite par Schmid dans [Schmid, 1996]. Soient deux images I_i et I_j d'une même scène 3D et M_i et M_j les matrices de projection correspondantes. La détection des points images p_i et p_j appartenant respectivement à I_i et I_j est parfaitement répétable si et seulement si il existe un point P de la scène tel que :

$$p_i = M_i P \quad \text{et} \quad p_j = M_j P \tag{2.1}$$

Pour mesurer la répétabilité d'un détecteur, une solution consiste à établir une relation entre p_i et p_j. Il n'en existe pas dans le cas général, mais si I_i et I_j sont les images d'une scène plane, alors il existe une homographie H_{ij} du plan projectif $\mathcal{P}^2$ les reliant[1]. Dans le

[1]Le lecteur peut se référer à la section 7.3.3 du chapitre 7 pour la définition de l'homographie du plan

cas d'une répétabilité parfaite, la relation entre p_i et p_j est donc la suivante :

$$p_j = H_{ij}\, p_i \tag{2.2}$$

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_{ij} n'appartient pas à l'image correspondante, et réciproquement. En outre, le support de l'image étant discret et la répétabilité d'un détecteur pouvant être imparfaite, l'équation 2.2 n'a guère de réalité physique. Il s'avérera donc nécessaire dans la pratique de considérer la répétabilité du détecteur dans un certain voisinage.

Soit $C_{ij}(p_{i,k})$ le point de I_j en correspondance homographique avec le point détecté $p_{i,k}$ de I_i. Il représente le point discret de l'image le plus proche de son correspondant homographique exact $H_{ij}\, p_{i,k}$, comme le montre la figure 2.1.

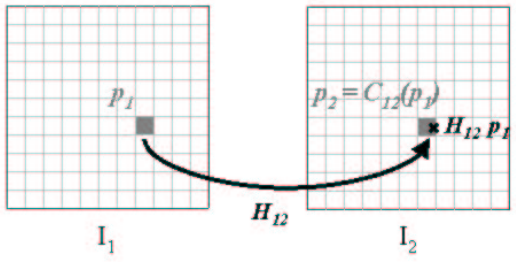


Figure 2.1 – *Discrétisation du correspondant homographique de* p_1 *:* $C_{12}(p_1)$.

Nous lui associons la mesure $D_{ij}(p_{i,k})$ qui traduit la distance euclidienne entre ce correspondant prédit $C_{ij}(p_{i,k})$ et le pixel le plus proche parmi tous les points détectés $p_{j,l}$ de I_j :

$$D_{ij}(p_{i,k}) = \min_l \{dist(C_{ij}(p_{i,k}), p_{j,l})\} \tag{2.3}$$

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, on considère que le point n'a pas de correspondant. La distance définie à l'équation 2.3 est alors adaptée de façon à prendre en compte ce critère :

$$D^{\varepsilon}_{ij}(p_{i,k}) = \begin{cases} D_{ij}(p_{i,k}) & \text{si } D_{ij}(p_{i,k}) < \varepsilon \\ \varepsilon & \text{sinon} \end{cases} \tag{2.4}$$

Cette formule n'élimine pas du calcul les points qui ne sont pas répétés dans le voisinage donné, la distance ε leur étant associée par défaut ; nous verrons plus loin les avantages

projectif.

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

$$R_{ij}^{\varepsilon} = \frac{\text{moy}_k(D_{ij}^{\varepsilon}(p_{i,k}))}{\varepsilon.(n_{ij}^{\varepsilon}+1)} \tag{2.5}$$

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 [Gouet et al., 2000] :

$$R^{\varepsilon} = \text{moy}(R_{ij}^{\varepsilon}, R_{ji}^{\varepsilon}) \tag{2.6}$$

Propriétés de cette mesure de répétabilité

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

- Le coefficient $\frac{1}{\varepsilon}$ de l'équation 2.5 permet de normaliser la mesure entre 0 et 1, le détecteur ayant la plus petite mesure de répétabilité étant considéré comme le meilleur ;
- Tenir compte d'une distance minimale dans le calcul de R_{ij}^{ε} permet de mettre en valeur le détecteur le plus précis parmi plusieurs détecteurs possédant des répétabilités équivalentes dans un voisinage ε donné. Cette idée est illustrée par l'exemple du tableau 2.1, dans lequel nous considérons deux détecteurs de points. k_{max} est le nombre de points détectés dans l'image I_i par chacun d'eux. On a alors par exemple, pour deux distances α_1 et α_2 telles que $\alpha_1 < \alpha_2 < \varepsilon$:

Détecteur	k_{max}	$D_{ij}^{\varepsilon}(p_{i,k})_{1\leq k\leq k_{max}}$								R_{ij}^{ε}
n°1	8	α_1	α_1	α_1	α_1	α_1	α_1	α_1	α_1	$\frac{\alpha_1}{9\varepsilon}$
n°2	8	α_2	α_2	α_2	α_2	α_2	α_2	α_2	α_2	$\frac{\alpha_2}{9\varepsilon}$

Table 2.1 – *Propriétés de la mesure de répétabilité : exemple 1.*

Il est clair que le détecteur le plus précis (celui qui est associé à la distance α_1 dans l'exemple) possède le meilleur score de répétabilité R_{ij}^{ε} ;

- Affecter aux $(N_i - n_{ij}^{\varepsilon})$ points non répétés la distance ε (cf. l'équation 2.4) permet de tenir compte du pourcentage de points répétés par rapport au nombre total de points testés. Cette idée est illustrée par le tableau 2.2 :

Détecteur	k_{max}	$D_{ij}^{\varepsilon}(p_{i,k})_{1\leq k\leq k_{max}}$								R_{ij}^{ε}
n°1	3	$\frac{\varepsilon}{3}$	$\frac{\varepsilon}{3}$	$\frac{\varepsilon}{3}$						$\frac{1}{12} \simeq 0.08$
n°2	8	$\frac{\varepsilon}{3}$	$\frac{\varepsilon}{3}$	$\frac{\varepsilon}{3}$	ε	ε	ε	ε	ε	$\frac{3}{16} \simeq 0.19$

Table 2.2 – *Propriétés de la mesure de répétabilité : exemple 2.*

- Pondérer le critère R_{ij}^{ε} par n_{ij}^{ε} permet de privilégier, dans une certaine mesure, un détecteur très répétable mais imprécis, à un détecteur moins répétable mais plus

précis, comme le montrent les exemples n°1 et n°2 du tableau 2.3. Cela permet également de favoriser les échantillons de points les plus grands, comme le montrent les exemples n°2 et n°3.

Détecteur	k_{max}	$D_{ij}^{\varepsilon}(p_{i,k})_{1\leq k\leq k_{max}}$						R_{ij}^{ε}
n°1	6	$\frac{\varepsilon}{4}$	$\frac{\varepsilon}{4}$	$\frac{\varepsilon}{4}$	$\frac{\varepsilon}{4}$	ε	ε	$\frac{1}{10}$
n°2	6	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{1}{14}$
n°3	5	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$	$\frac{\varepsilon}{2}$		$\frac{1}{12}$

Table 2.3 – *Propriétés de la mesure de répétabilité : exemple 3.*

– 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.

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 :

$$\alpha\, i\, \sqrt{2} \leq \varepsilon < \alpha\, (i+1)\, \sqrt{2} \tag{2.7}$$

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 ($i = 1$), il faudra alors choisir ε dans l'intervalle $[\frac{\sqrt{2}}{2}, \sqrt{2}[$.

2.1.3 Critères de localisation

La mesure de localisation d'un détecteur consiste à vérifier que les points détectés ont une réalité physique dans la scène. La qualité de cette mesure est par conséquent très importante dans les applications liées à la reconstruction tridimensionnelle ou encore au calibrage de caméras. Nous avons choisi ici de nous baser sur deux critères de localisation [Gouet et al., 2000] dont l'importance dépendra des applications envisagées : le critère de *reconstruction projective*, qui met en jeu une distance euclidienne 2D et celui de *reconstruction euclidienne*, qui lui se base sur une distance euclidienne 3D.

2.1.3.1 Localisation projective

Supposons connue la position exacte d'un ensemble de points P de la scène 3D. Soient p_{proj} l'ensemble des projections 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 :

$$L_{\text{projectif}} = \underset{i}{\text{moy}}\{\min_{j}\{dist(p_{\text{det}}^{i}, p_{\text{proj}}^{j})\}\} \tag{2.8}$$

Cette mesure est simple à calculer et peut être rendue encore plus précise en considérant les coordonnées réelles et non discrètes des projections p_{proj}. Toute la difficulté de son estimation réside dans le choix des points 3D et de leurs projections. Ce point est détaillé dans la section 2.1.4.

Ce premier critère se révèle précieux 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.

2.1.3.2 Localisation euclidienne

Le critère ci-dessus s'avère insuffisant lorsque l'on souhaite procéder à la reconstruction tridimensionnelle de la scène ; en effet il ne traduit pas la précision de la reconstruction. Le traitement se faisant à partir de deux vues de la scène au moins, il faut également tenir compte de la mesure de répétabilité du détecteur. Nous avons choisi dans cette partie de définir une mesure qui exprime directement la précision de la reconstruction. Il sera ensuite intéressant de voir comment évoluent les trois mesures ainsi définies les unes par rapport aux autres.

Supposons que nous ayons à notre disposition au moins deux vues de la scène, la calibration des caméras 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, de les reconstruire par triangulation, et enfin de les comparer aux points P^j de la scène. De manière similaire à la mesure projective, la mesure de localisation euclidienne est alors donnée par :

$$L_{\text{euclidien}} = \underset{i}{\text{moy}}\{\min_{j}\{dist(P^i_{\text{reconstruits}}, P^j)\}\} \tag{2.9}$$

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

2.1.4 Cadre de l'évaluation

Dans le cas de la mesure de répétabilité, la méthode d'évaluation mise en œuvre ici nécessite la donnée d'une séquence d'images couleur d'une scène plane, ainsi que la connaissance de l'homographie H_{ij} reliant les projections de cette scène qui diffèrent par leur point de vue. Afin de ne pas favoriser le détecteur étudié, H_{ij} doit être déterminée de manière très précise et indépendamment des points détectés. Certains auteurs, comme [Schmid, 1996] utilisent des cibles circulaires très précisément appariées dans les images et vérifiées manuellement, puis estiment l'homographie à l'aide d'une méthode aux moindres carrés médians, similaire à celle développée pour l'estimation de la matrice fondamentale (cf. section 7.6.1 du chapitre 7).

Le principal inconvénient de cette méthode est que, même si elle s'appuie sur des appariements précis, elles fait intervenir dans l'évaluation le processus d'estimation de l'homographie. Le résultat est un mélange des performances des deux algorithmes. C'est pourquoi 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 2.2. Dans cette scène, nous positionnons plusieurs caméras virtuelles modélisées selon le modèle Sténopé[2]. Les paramètres intrinsèques et extrinsèques de ces caméras étant 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. L'homographie du plan projectif est ensuite calculée de façon précise à partir de la projection exacte de 4 points de la facette, ses sommets par exemple, comme le montre la figure 2.2.

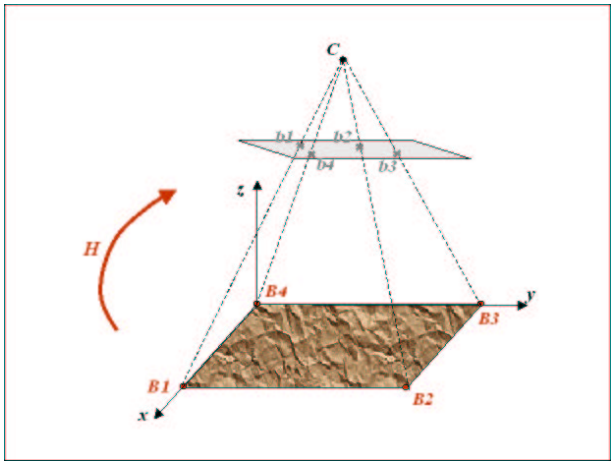


Figure 2.2 – *Modèle de la scène tridimensionnelle pour l'évaluation des détecteurs.*

Pour les séquences d'images prises sous le même point de vue mais n'ayant pas la même luminosité par exemple, ce cadre d'évaluation devient bien sûr inutile, l'homographie du plan projectif étant dans ce cas la matrice identité. La mesure de répétabilité est alors estimée à partir d'images réelles.

Les mesures de localisation nécessitent quant à elles la donnée de la position de points de la scène tridimensionnelle ainsi que de leurs projections dans les images. Nous avons choisi comme texture une grille de calibration dont la position des principaux 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 facette texturée. 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 apparions les points les plus répétables dans un petit voisinage. 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 facette.

Choix du détecteur de référence

Nos deux détecteurs sont implémentés ici en sub-pixel. Afin d'être à même de classifier les résultats obtenus, nous devons aussi appliquer notre évaluateur à un détecteur qui a fait ses preuves et qui nous servira de référence. Nous avons choisi le détecteur de Harris

[2]Le modèle Sténopé est introduit à la section 7.1.1 du chapitre 7.

et Stephens 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 [Schmid et al., 1998].

2.2 Etude comparative des détecteurs couleur

Dans cette partie, nous comparons les deux détecteurs couleur et le détecteur Harris précis, grâce aux critères mis en place dans la section précédente. Les deux détecteurs couleur sont comparés au détecteur en niveau de gris mais également comparés entre eux. Le lecteur peut voir dans la section 2.2.1 des exemples de points extraits de différentes images de référence, à l'aide des deux détecteurs couleur. Dans la section 2.2.2, nous estimons la mesure de répétabilité sur plusieurs séquences d'images, obtenues en appliquant à ces images de référence les principales transformations de l'image telles que la rotation, le changement de luminosité, le changement d'échelle, le changement de point de vue. La stabilité au bruit de la caméra est également testée. Les mesures de localisation sont étudiées dans la section 2.2.4 pour la reconstruction projective et dans la section 2.2.5 pour la reconstruction euclidienne.

2.2.1 Exemples d'images et de détections

Nous présentons dans cette section les images de référence qui vont être utilisées pour la réalisation de notre étude. L'image "Lézard" est caractérisée par beaucoup de texture alors que "Salle robotique" et "Bureau" présentent un grand nombre de contours. La figure 2.3 montre les points Harris couleur obtenus ; 472 points sont extraits sur "Lézard", 493 sur "Salle robotique" et 474 sur "Bureau". Pour les coins couleur (figure 2.4), 457 coins sont extraits sur "Salle robotique", 438 sur "Lézard" et 457 sur "Bureau". Comme cela est expliqué en détail à l'annexe B, ces images sont disponibles en couleur mais aussi en niveau de gris, pour pouvoir détecter les points à partir du détecteur Harris Précis en niveau de gris. Certaines de ces images monochromatiques ont été obtenues en considérant la moyenne des trois plans couleur.

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 rétiniennes ne dépassant pas la dimension (400×400). La détection a été réalisée en précision demi-pixel pour les trois détecteurs (cf. annexe C).

2.2.2 Mesure de la répétabilité

Pour chacun des types de transformation, nous présentons 2 graphes : le premier correspond à la mesure de répétatibilité R^{ε} entre l'image de référence et l'image transformée, et cela pour les 3 détecteurs. Le second représente uniquement le pourcentage de points répétés associé. L'interprétation simultanée de ces 2 graphes permettra par exemple de distinguer un détecteur plus précis dans un voisinage ε, parmi plusieurs détecteurs qui répètent un pourcentage de points équivalent. Ces deux graphes sont toujours accompagnés de quelques unes des vues utilisées, l'image de référence étant à chaque fois entourée d'un cadre plus épais.

Figure 2.3 – *Un exemple de points Harris couleur sur les images "Salle robotique", "Lézard" et "Bureau".*

Figure 2.4 – *Un exemple de coins couleur sur les images "Salle robotique", "Lézard" et "Bureau".*

Les sections suivantes étudient la répétabilité de chacun des détecteurs par rapport aux principales transformations de l'image. La rotation image est traitée à la section 2.2.2.1, le changement de luminosité à la section 2.2.2.2, le changement d'échelle à la section 2.2.2.3 et le changement de point de vue à la section 2.2.2.4.

2.2.2.1 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°. Le lecteur peut voir les résultats de mesure de répétabilité aux figures 2.5 et 2.6. Les images de la première ligne représentent l'image de référence suivie de 4 images ayant subi une rotation image. Nous avons choisi dans les exemples une erreur de localisation ε de 1, ce qui indique, si l'on se réfère à l'équation 2.7, que le point détecté peut être trouvé chez l'un des pixels voisins du pixel prédit dans le cas d'un lissage en précision demi-pixel. Le premier graphe de chaque figure correspond à la mesure de répétatibilité R^1. Le second représente uniquement le pourcentage de points répétés.

Les résultats obtenus sur les deux séquences montrent clairement que le détecteur Harris couleur est le plus répétable face aux rotations image. Le détecteur de coins couleur, quant à lui, obtient des résultats variables d'une séquence à l'autre. Notons aussi que les résultats se dégradent globalement lorsque l'angle de rotation augmente, sauf pour la rotation de 180°. Le second graphe nous indique alors que cette dégradation semble plutôt due à une diminution de la précision des détections, puisque les pourcentages de points répétés restent globalement les mêmes. La zone de recherche se limitant ici aux premiers voisins ($\varepsilon = 1$), nous pouvons en conclure qu'en augmentant l'angle de rotation, de plus en plus de points détectés se retrouvent dans le voisinage direct du point prédit.

2.2.2.2 Changement de luminosité

Dans cette partie, nous passons en revue deux types de changement de luminosité :

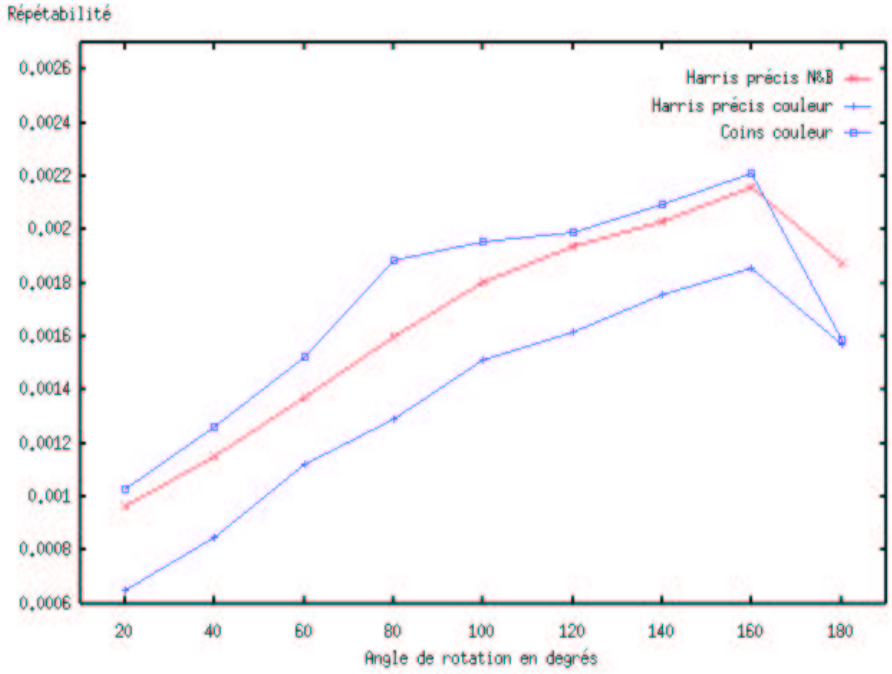
- *Un changement affine de luminosité* selon le modèle à 6 paramètres de Finlayson [Finlayson et al., 1994]. Les images "Lézard" et "Salle robotique" ont été modifiées synthétiquement à 10 reprises, en augmentant inégalement les niveaux de gris de chaque plan {R,V,B}. L'équation 2.10 montre les 6 paramètres affines utilisés ici :

$$\begin{cases} R'_i = 0.2R + 0.08i & \forall i = 1..10 \\ V'_i = 0.8V + 0.02i & \forall i = 1..10 \\ B'_i = 0.5B + 0.05i & \forall i = 1..10 \end{cases} \tag{2.10}$$

  Les images originales suivie de quelques unes de ces transformations sont visibles aux figures 2.7 et 2.8. Pour indication, les images intermédiaires des deux séquences sont à dominante verte et les dernières à dominante rouge ;

- *Un changement uniforme de luminosité* en faisant varier l'ouverture du diaphragme de la caméra. On obtient pour l'image "Bureau" une séquence de 10 images naturelles d'intensités lumineuses différentes. Une partie d'entre elles est montrée à la figure 2.9. L'image de référence est la troisième parmi les cinq images représentées et a été

(référence) (40°) (80°) (120°) (180°)



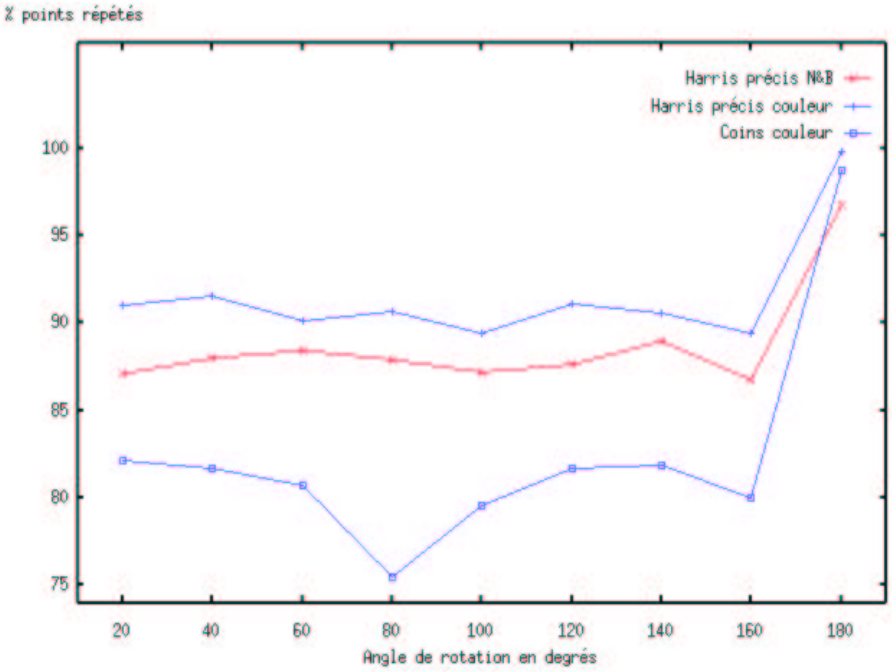
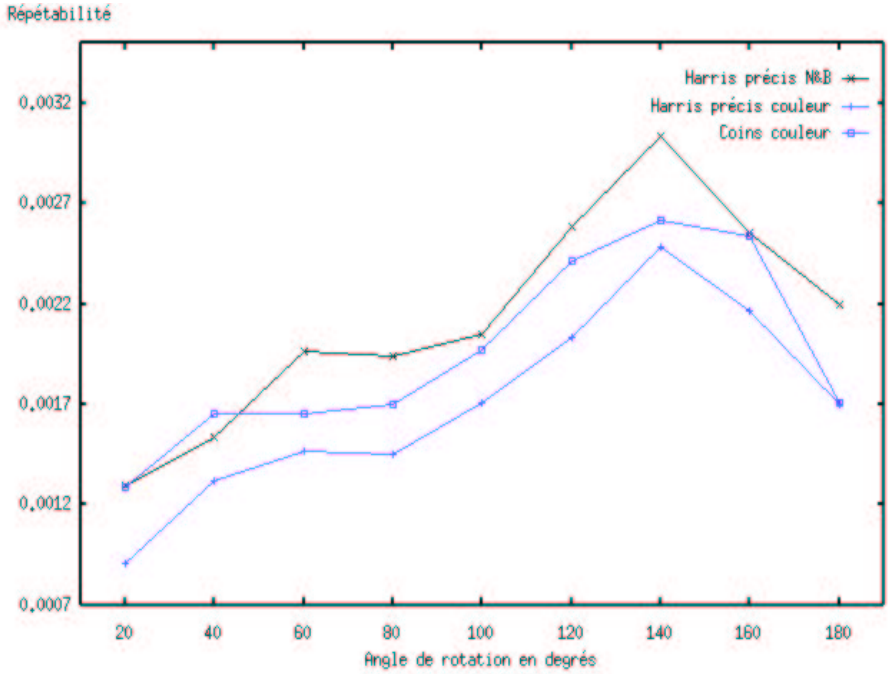


Figure 2.5 – *Séquence "Lézard" : Répétabilité R^1 suivant 9 rotations image.*

(référence) (40°) (80°) (120°) (180°)



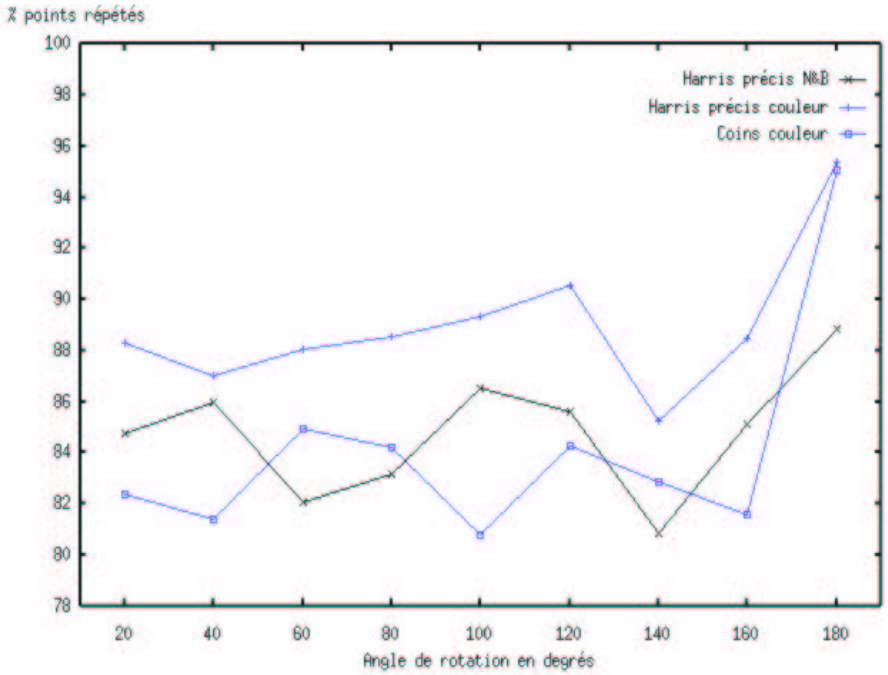


Figure 2.6 – *Séquence "Salle robotique" : Répétabilité R^1 suivant 9 rotations image.*

obtenue à partir d'une ouverture de diaphragme de 4, ce qui correspond à une image ni très sombre ni très claire.

Les mesures de répétabilité obtenues sont beaucoup plus faibles que celles obtenues pour les rotations image à la section précédente, nous pouvons d'ores et déjà noter que les 3 détecteurs ont une répétabilité excellente face aux changements affines de luminosité. Notons également que pour les détecteurs couleur, le traitement ne semble pas favoriser un plan couleur plus qu'un autre, en effet les mesures de répétabilité sont globalement constantes quelle que soit la transformation affine appliquée. Les deux premiers tests (cf. figures 2.7 et 2.8) montrent cependant clairement la supériorité du détecteur Harris en niveau de gris face aux changements affines de luminosité. Ce résultat s'explique par le nombre moins important de degrés de liberté du modèle d'illumination lorsqu'on travaille en niveau de gris.

Quant aux résultats concernant les changements uniformes de luminosité (cf. figure 2.9), nous constatons que les résultats se dégradent rapidement dès que l'image est sur/sous exposée, le pourcentage de points répétés chutant très rapidement jusqu'à 50%. Nous voyons ici que les détecteurs couleur, et en particulier le détecteur Harris couleur, donnent les meilleurs résultats.

2.2.2.3 Changement d'échelle

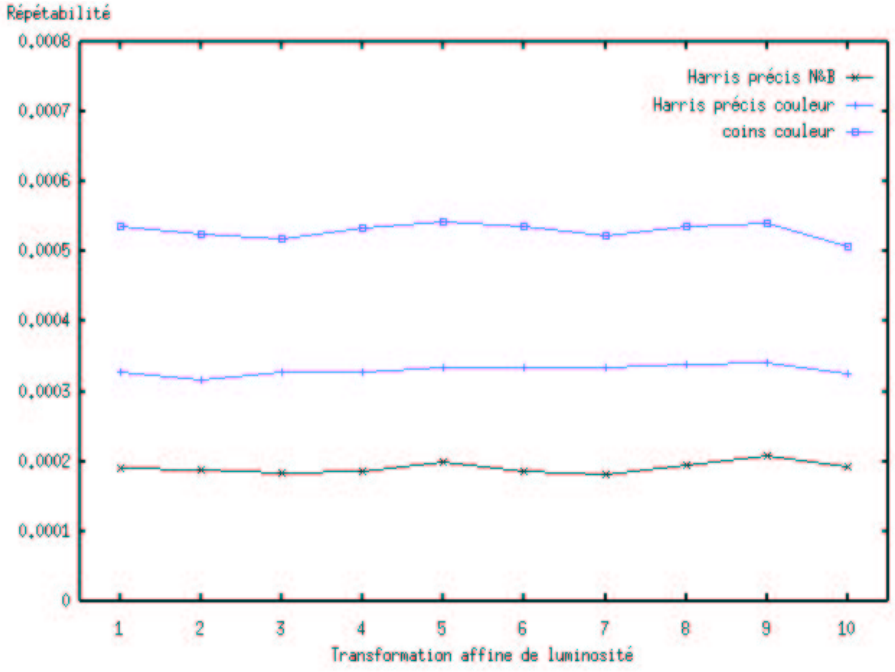
Pour obtenir un changement d'échelle, nous avons fait subir à la caméra virtuelle 9 translations selon son axe optique vers les images "Lézard" et "Salle robotique". Quelques exemples des images obtenues (les images de référence étant les dernières) et les mesures de répétabilité pour $\varepsilon = 1$ sont visibles aux figures 2.10 et 2.11. La détection a été réalisée dans tous les cas avec les mêmes paramètres de lissage gaussien.

Au vu des résultats obtenus, tous les détecteurs semblent très sensibles à un changement d'échelle. Les mesures de répétabilité sont en effet globalement supérieures à celles obtenues aux sections précédentes et les pourcentages de points répétés chutent très rapidement. On peut néanmoins remarquer que les détecteurs couleur donnent les meilleurs résultats, quelles que soient les séquences d'images. 85% des points sont répétés avec le détecteur de Harris Couleur pour le dernier changement d'échelle, qui correspond au facteur d'échelle le plus faible (d'une valeur de 2, se référer à la section 1.2.2.2 du chapitre 1).

2.2.2.4 Changement de point de vue

Pour mesurer la répétabilité des détecteurs face à un changement de point de vue, la caméra virtuelle a été déplacée 9 fois selon une rotation autour de la facette, en veillant à viser approximativement son centre de façon à en voir la plus grande partie. Le lecteur peut voir quelques exemples des images obtenues aux figures 2.12 et 2.13 sur les images "Lézard" et "Salle robotique".

Les résultats se dégradent pour les 3 détecteurs au fur et à mesure que le déplacement de la caméra augmente. Nous constatons que les deux détecteurs couleur obtiennent les



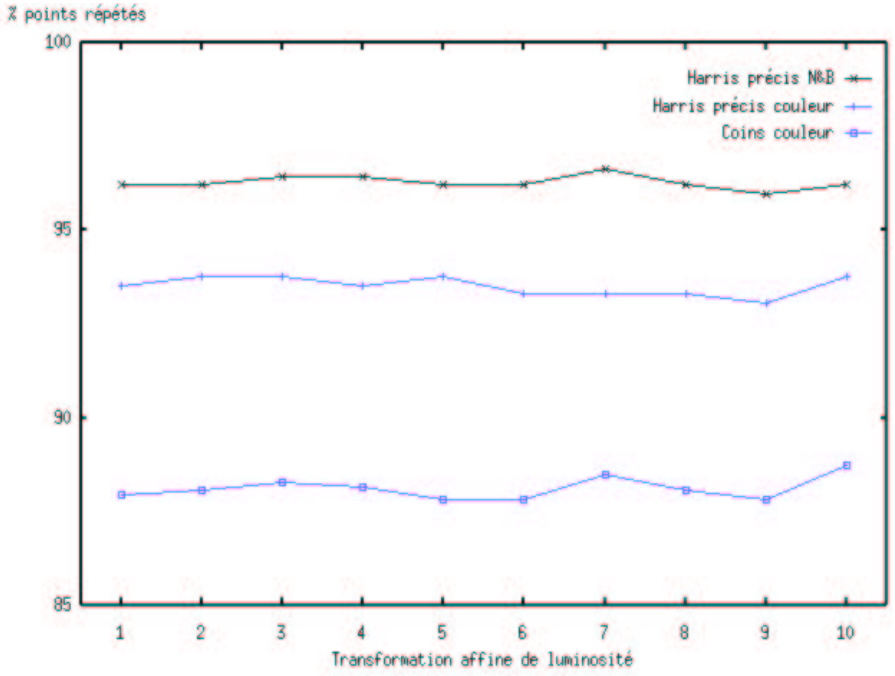
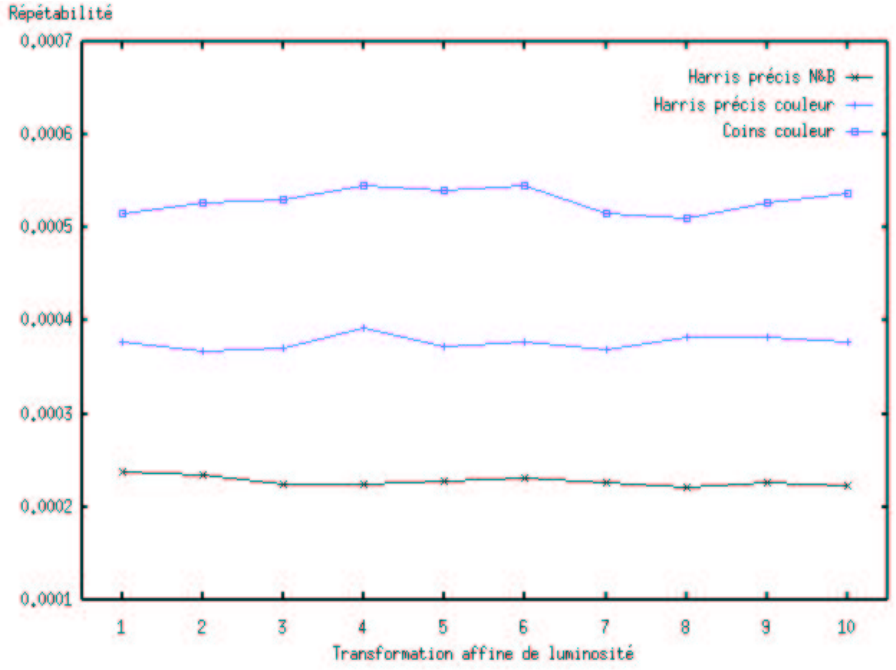


Figure 2.7 – *Séquence "Lézard" : Répétabilité R^1 suivant 10 changements affines de luminosité.*



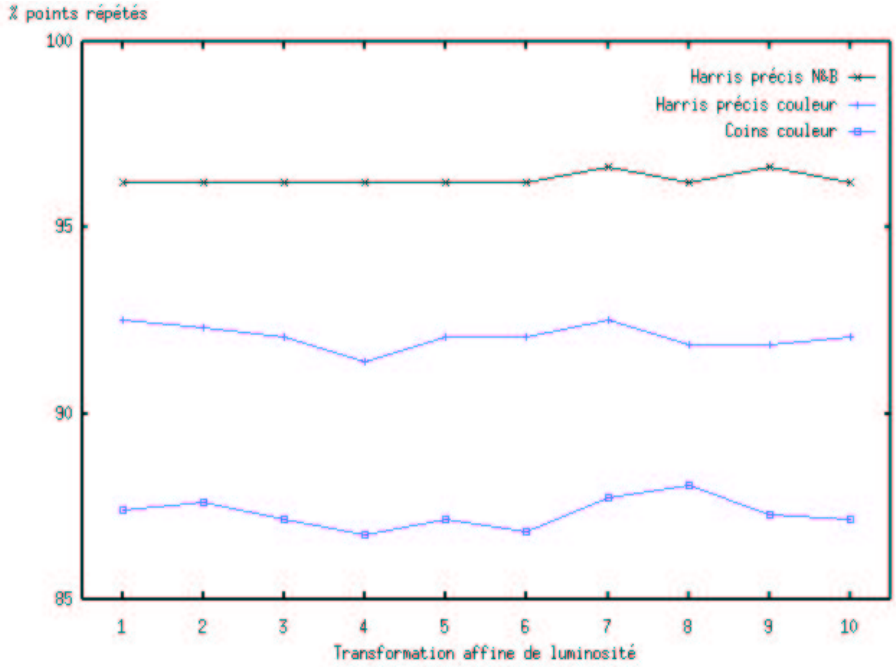
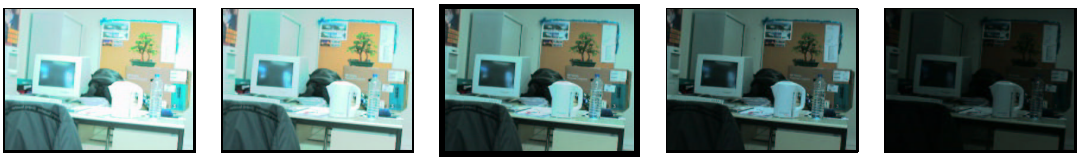
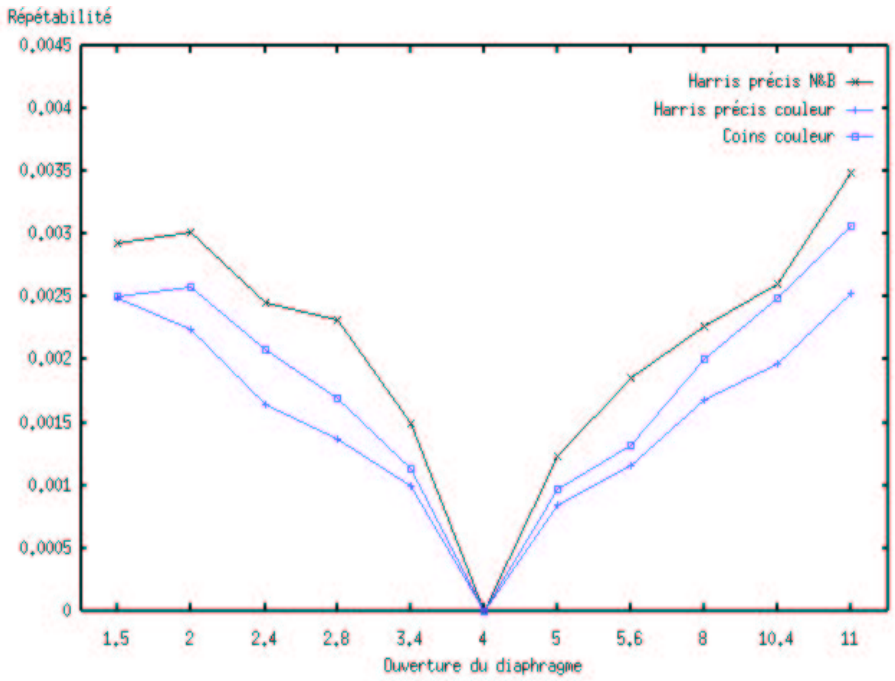


Figure 2.8 – *Séquence "Salle robotique" : Répétabilité R^1 suivant 10 changements affines de luminosité.*





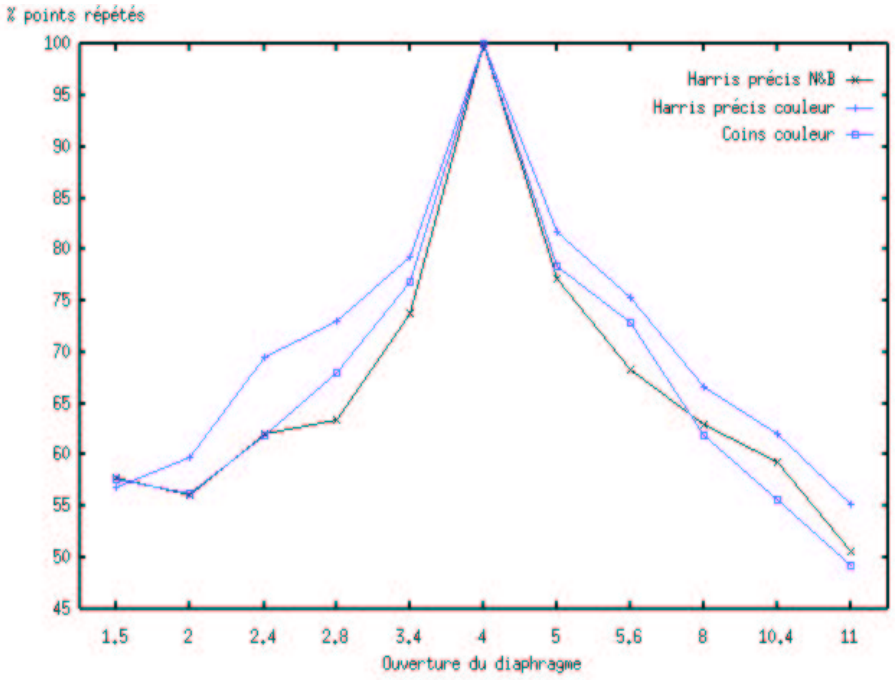
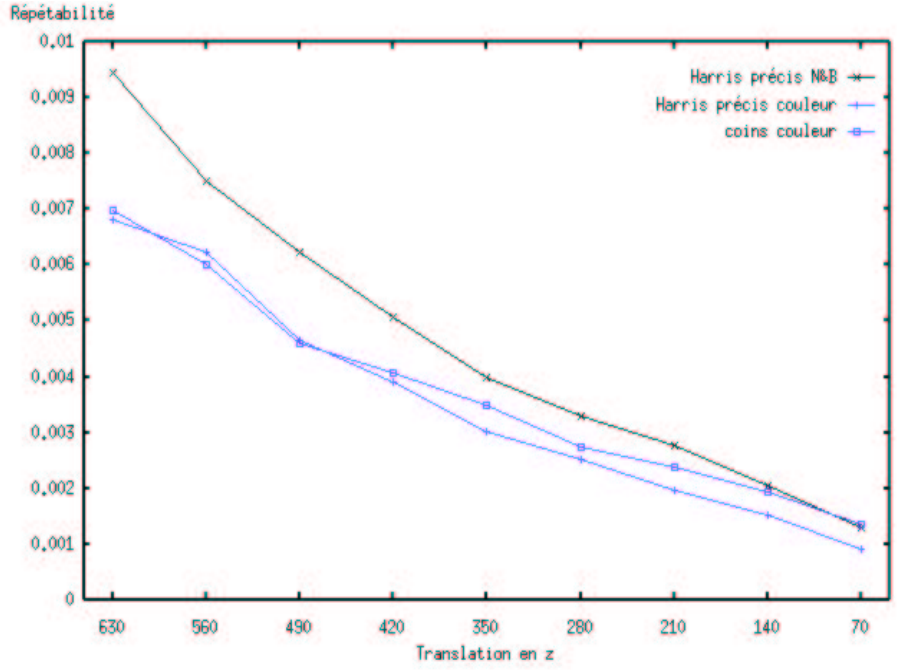


Figure 2.9 – *Séquence "Bureau" : Répétabilité R^1 suivant 10 changements uniformes de luminosité.*



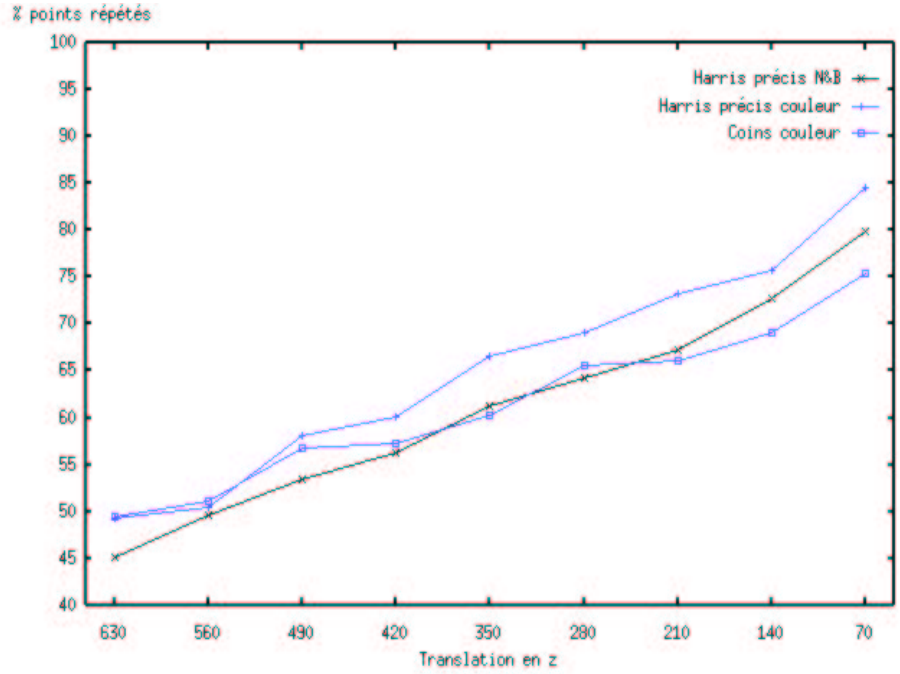
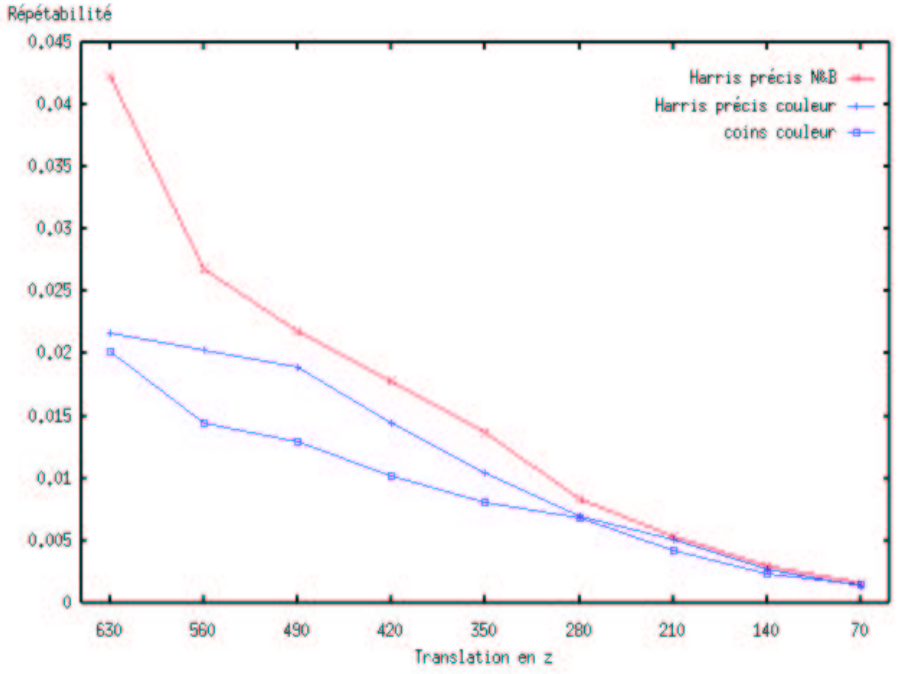


Figure 2.10 – *Séquence "Lézard" : Répétabilité R^1 suivant 10 changements d'échelle.*



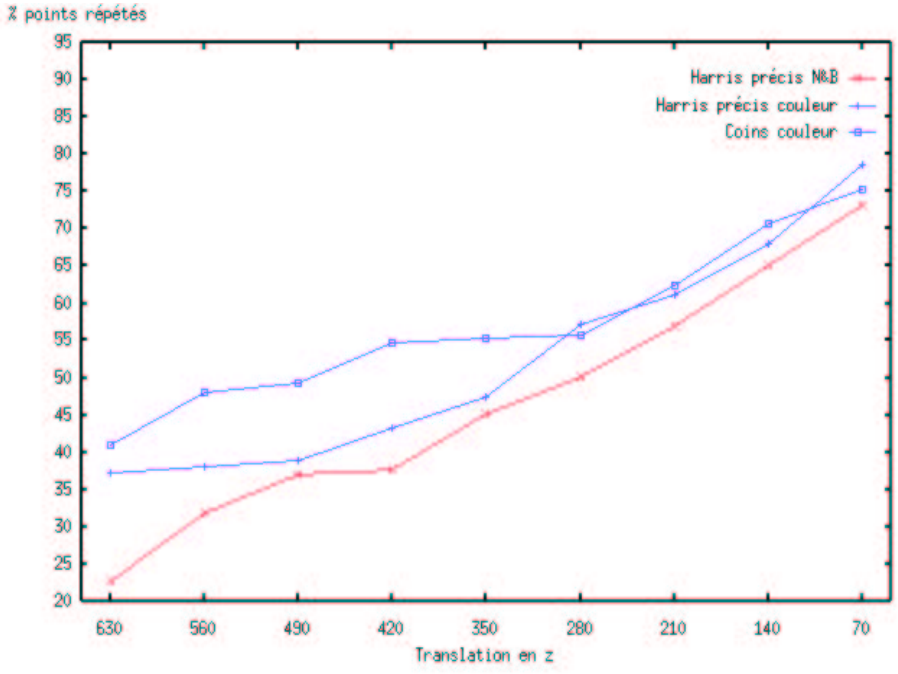
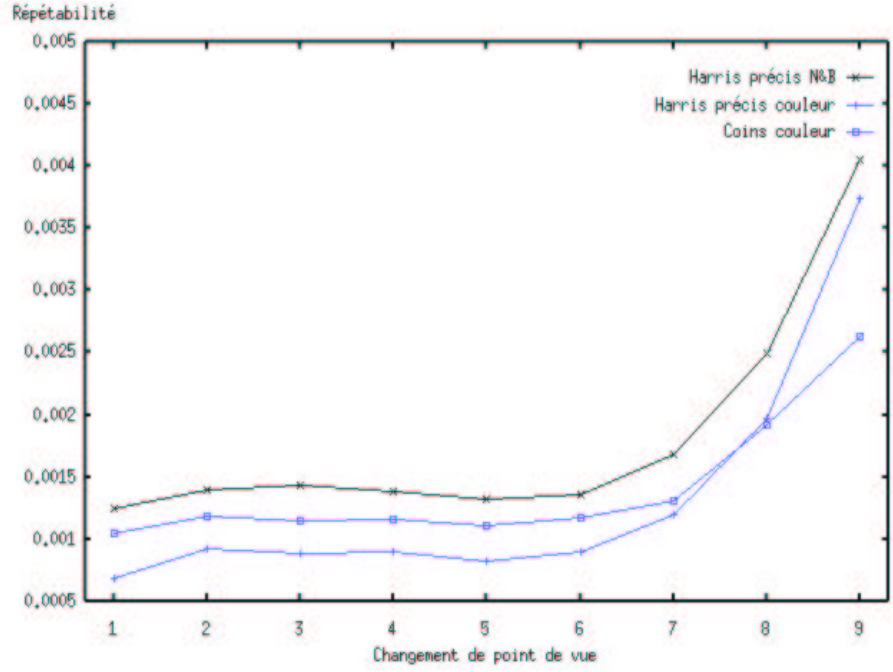


Figure 2.11 – *Séquence "Salle robotique" : Répétabilité R^1 suivant 10 changements d'échelle.*



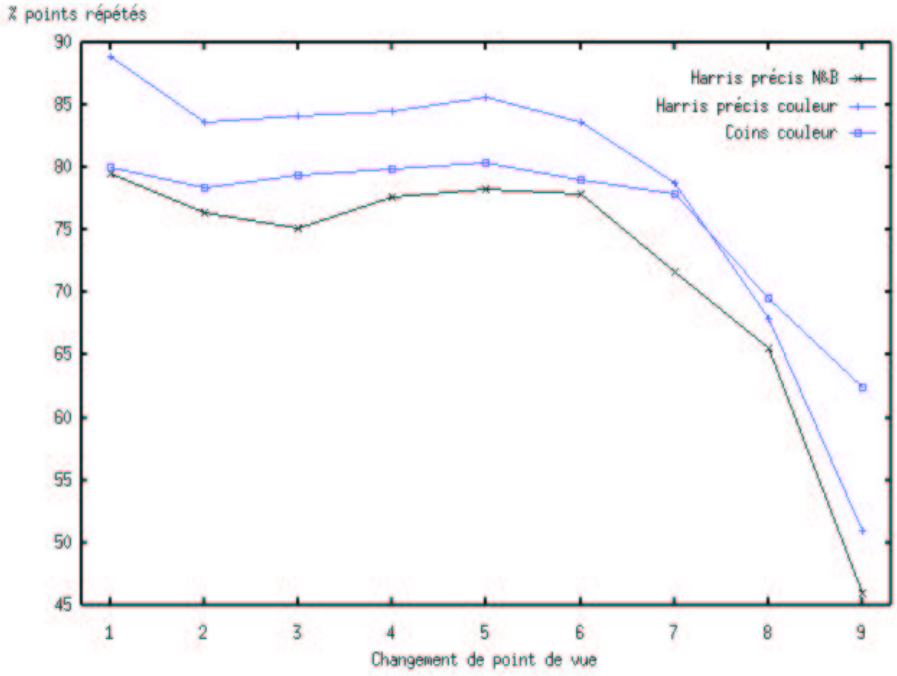
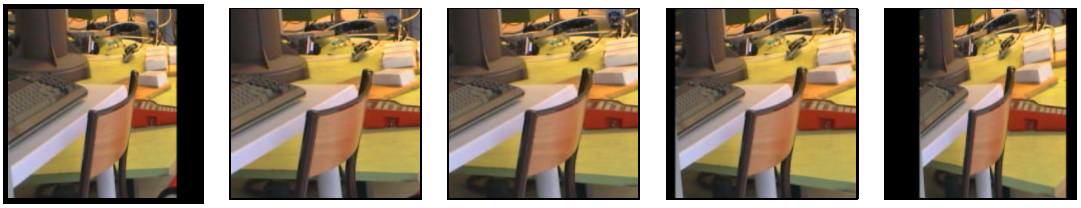
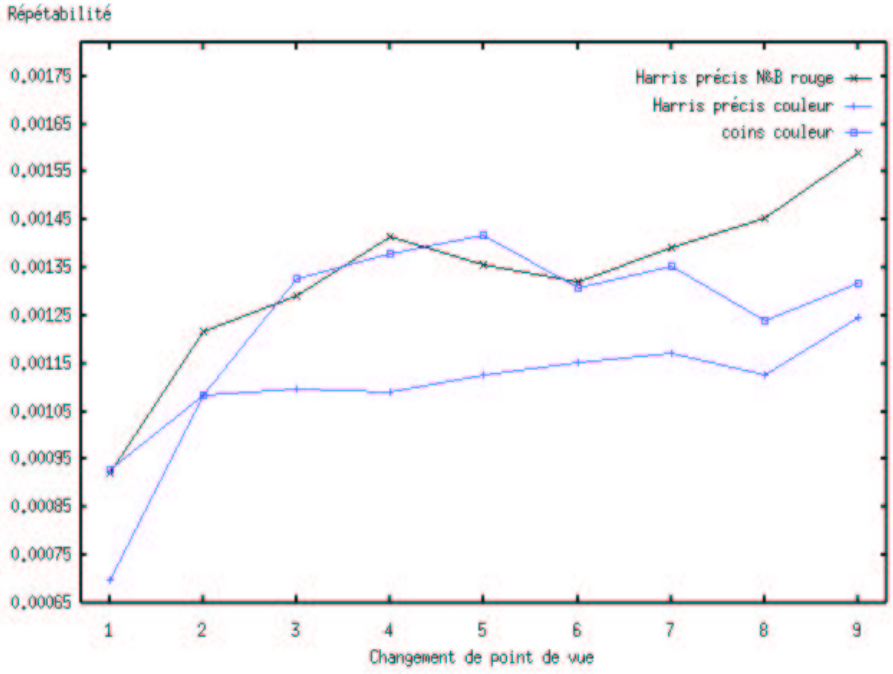


Figure 2.12 – *Séquence "Lézard" : Répétabilité R^1 suivant 9 changements de point de vue.*





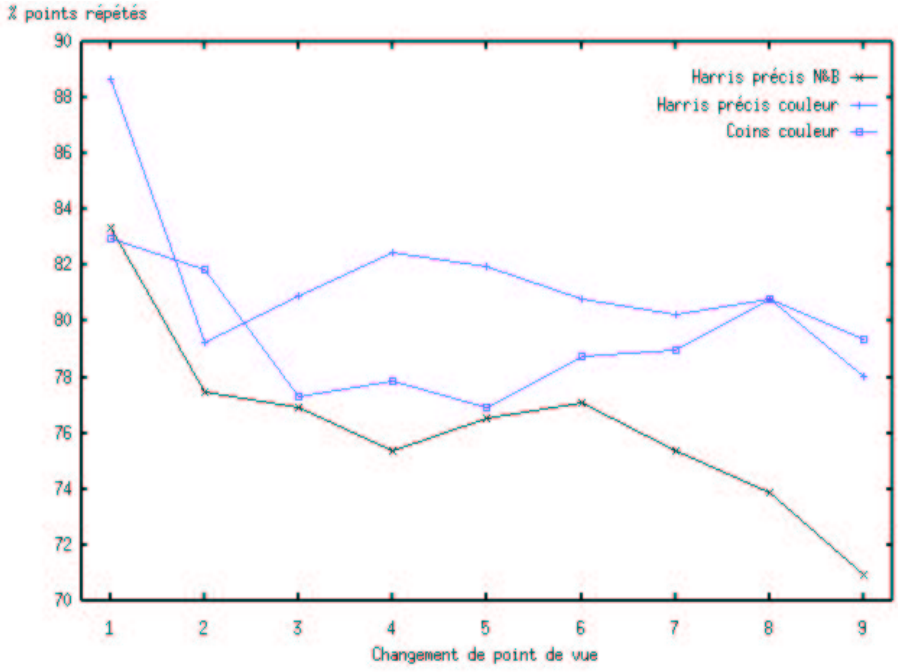


Figure 2.13 – *Séquence "Salle robotique" : Répétabilité R^1 suivant 9 changements de point de vue.*

meilleurs résultats.

2.2.2.5 Bruit

Les séquences bruitées ont été obtenues en appliquant un bruit blanc gaussien avec plusieurs valeurs de variance sur les images "Bureau" et "Lézard". En couleur, les trois plans ont été bruités simultanément. Pour l'image "Bureau" non directement disponible en niveau de gris, le bruit a été superposé à l'image obtenue par moyenne des trois plans couleur non bruités. Quelques vues et les résultats de répétabilité sont visibles aux figures 2.14 et 2.15. Les graphes montrent que l'information couleur améliore nettement la qualité de la détection, puisque c'est le détecteur Harris en niveau de gris qui offre les moins bons résultats. Ainsi, réaliser la détection à partir de plusieurs plans rend le calcul plus stable face au bruit.

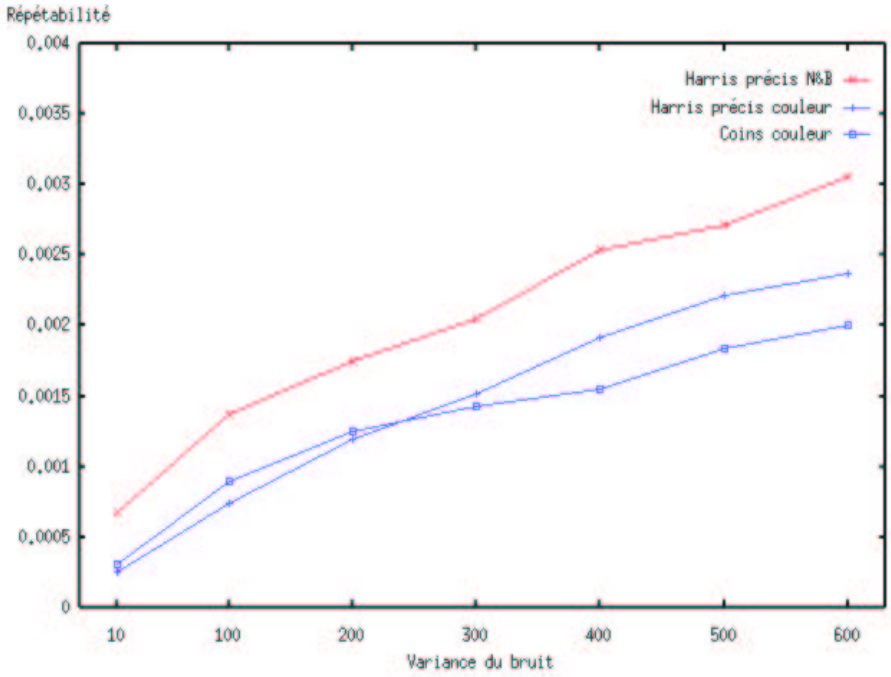
2.2.3 Précision de la détection

Il nous a semblé intéressant d'observer comment se comportent nos trois détecteurs lorsque l'on fait varier le rayon de voisinage ε. Les résultats de la figure 2.12 ont été obtenus pour un rayon de voisinage ε de 1. Nous avons repris la même image de référence ainsi que la vue associée au point de vue n°5. Les mesures de répétabilité ont alors été calculées en fonction de ε. En accord avec la formule de l'équation 2.7, nous l'avons fait varier entre 1 et 6 pixels, la détection ayant été réalisée en demi-pixel. Les mesures obtenues sont visibles à la figure 2.16. Il va sans dire que les résultats s'améliorent lorsque l'on tolère une plus grande erreur de localisation, le nombre de points répétés augmentant avec ε.

2.2.4 Mesure de la localisation projective

2.2.4.1 A partir d'images synthétiques

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 2.17 la grille utilisée sous quelques points de vue. Le graphe associé montre les mesures de $L_{\text{projectif}}$ obtenues. En premier lieu, notons que la mesure de localisation obtenue diminue avec le changement de point de vue ; elle devient même très faible pour la vue n°10 qui correspond à la dernière image de la figure. Cela ne traduit pas une augmentation de la précision de la détection avec la rotation, mais seulement une altération des distances entre points, due à la projection perspective. Il conviendra donc dans cette étude d'observer uniquement le comportement des détecteurs les uns relativement aux autres. Les deux détecteurs couleur donnent globalement les meilleurs résultats. Le détecteur de coins couleur est nettement supérieur aux deux autres, ce qui montre que c'est lui qui permet de détecter les points le plus précisément. Par conséquent, nous préconisons son utilisation pour des applications tel que le calibrage des caméras.



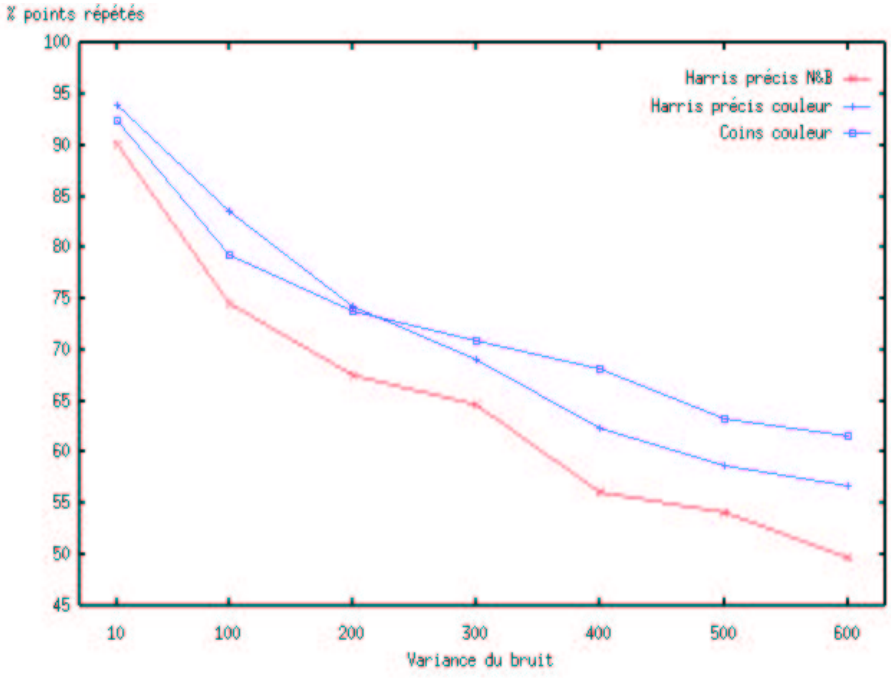
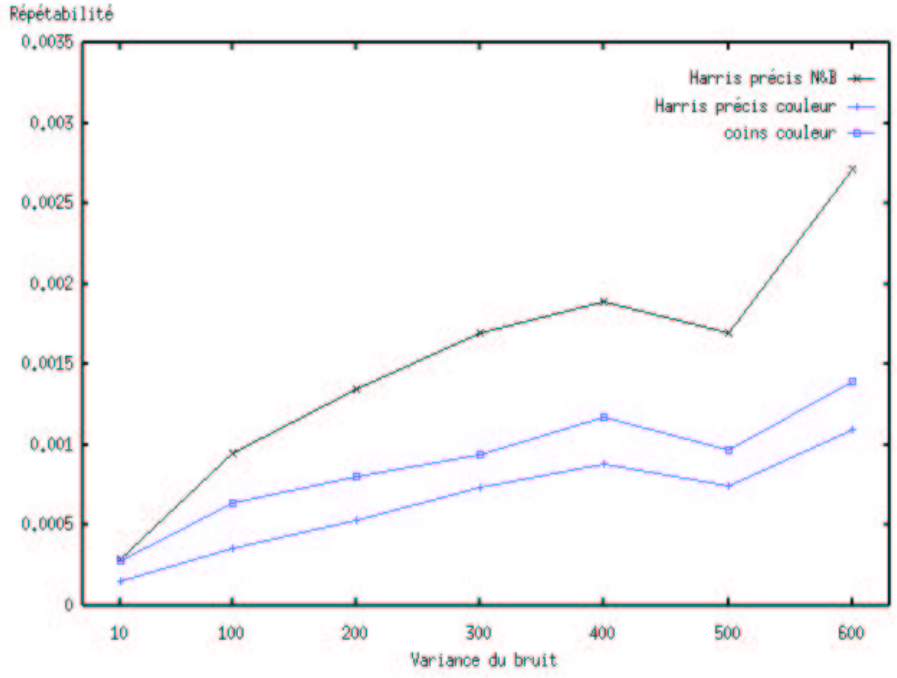


Figure 2.14 – *Séquence "Bureau" : Répétabilité R^1 suivant 7 images bruitées.*



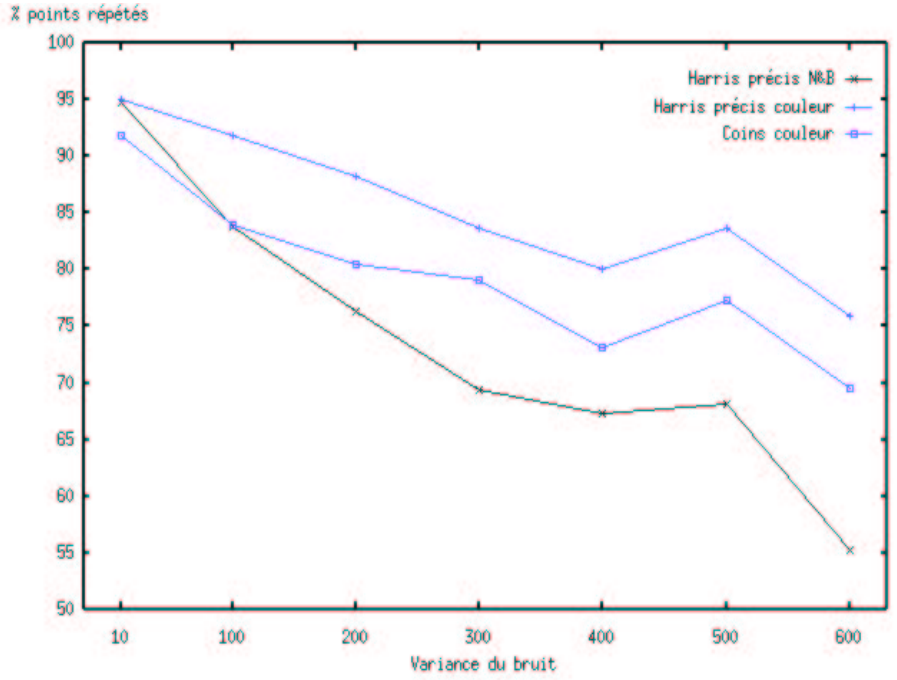
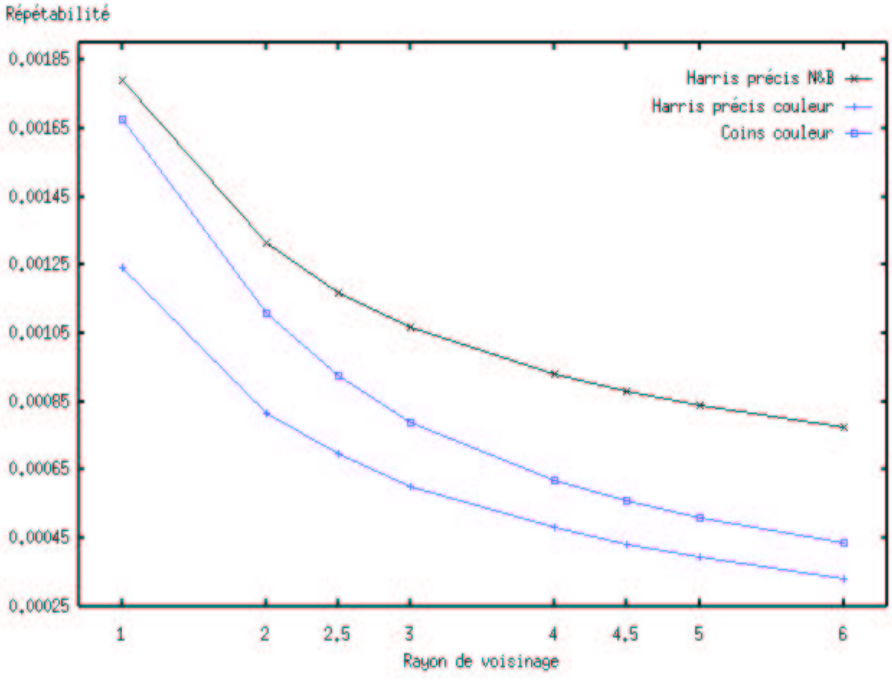


Figure 2.15 – *Séquence "Lézard" : Répétabilité R^1 suivant 7 images bruitées.*



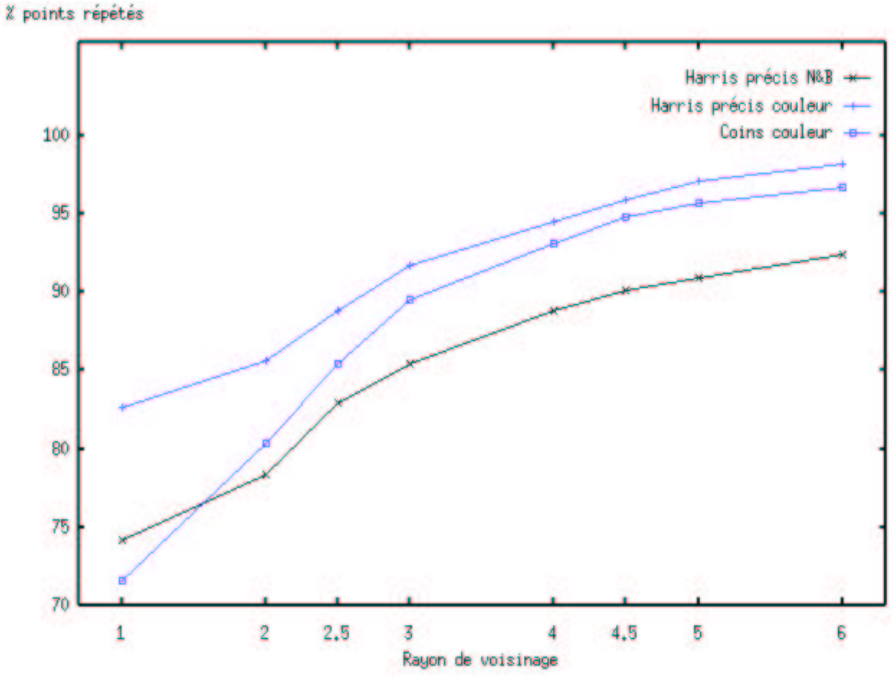


Figure 2.16 – *Paire "Lézard" : Répétabilité suivant 8 rayons de voisinage ε.*

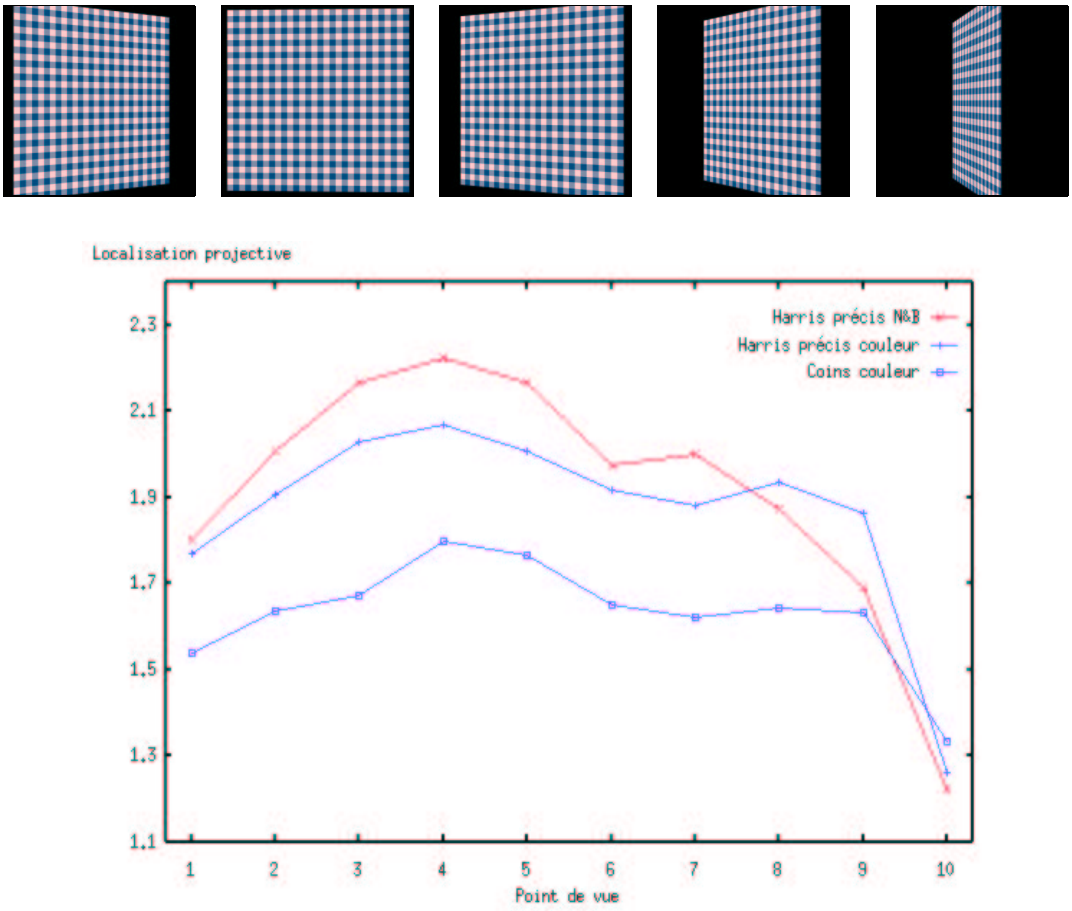
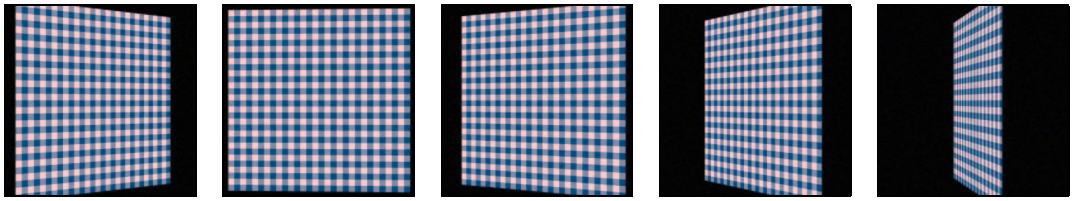


Figure 2.17 – *Localisation projective suivant 10 points de vue.*

2.2.4.2 Bruit

La figure 2.18 montre les mesures de localisation obtenues sur les images de la section précédente, auxquelles nous avons ajouté un bruit blanc gaussien de variance égale à 200. Les résultats sont globalement comparables à ceux obtenus à la section précédente à partir des images synthétiques.



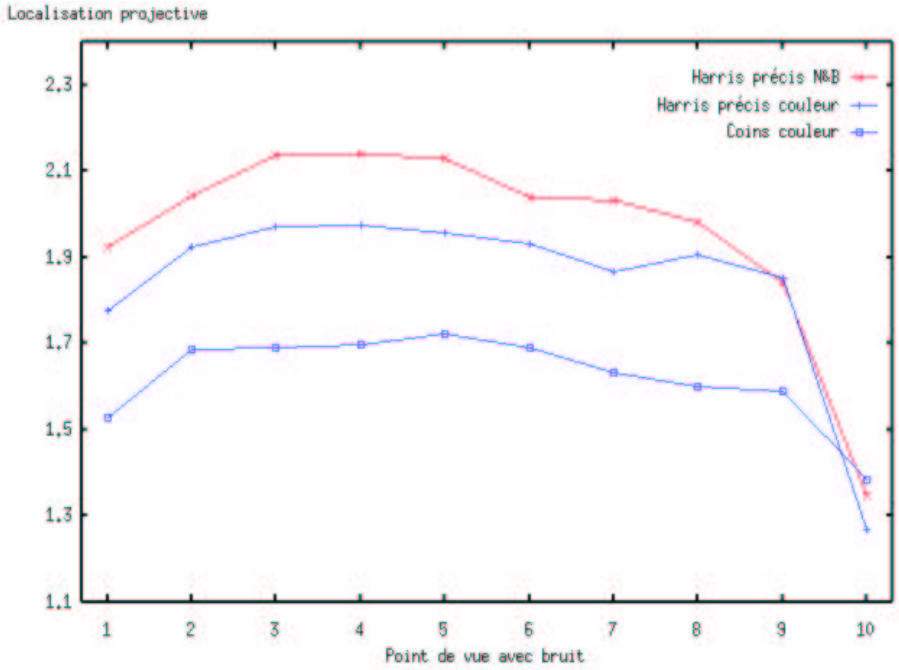


Figure 2.18 – *Localisation projective suivant 10 points de vue avec bruit.*

2.2.5 Mesure de la localisation euclidienne

2.2.5.1 A partir d'images synthétiques

Pour calculer la mesure de localisation euclidienne, nous avons utilisé les mêmes séquences d'images qui ont servi à la mesure de la localisation projective face aux différents points de vue avec et sans bruit. Pour chacune des deux séquences, 9 vues ont été mises en correspondance avec la vue de référence (la dernière image de chaque figure) en exploitant la mesure de répétabilité, à partir de la méthode d'appariement détaillée à la fin de la section

2.1.4. La reconstruction euclidienne a été réalisée suivant deux erreurs de répétabilité ε ($\varepsilon = 1$ et $\varepsilon = 2$). La figure 2.19 montre les résultats obtenus. D'un point de vue général, nous constatons que quels que soit le détecteur employé, les résultats s'améliorent lorsque le point de vue s'éloigne du point de vue de référence. Ce résultat est normal dans la mesure où la précision de la triangulation augmente avec l'écart de points de vue.

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 ce résultat à la relative mauvaise répétabilité de ce détecteur face aux changements de points de vue (se référer aux figures 2.12 et 2.13). Quant aux détecteurs Harris précis, les résultats semblent globalement meilleurs lorsque l'on tient compte de l'information couleur.

2.2.5.2 Bruit

La figure 2.20 montre les mesures de localisation euclidienne obtenues sur les images de la section précédente, auxquelles nous avons ajouté un bruit blanc gaussien de variance égale à 200. Les résultats pour les détecteurs couleur sont globalement comparables aux résultats précédents. En revanche, ces résultats se dégradent avec le bruit en ce qui concerne le détecteur en niveau de gris, pour devenir moins bons quel que soit le point de vue.

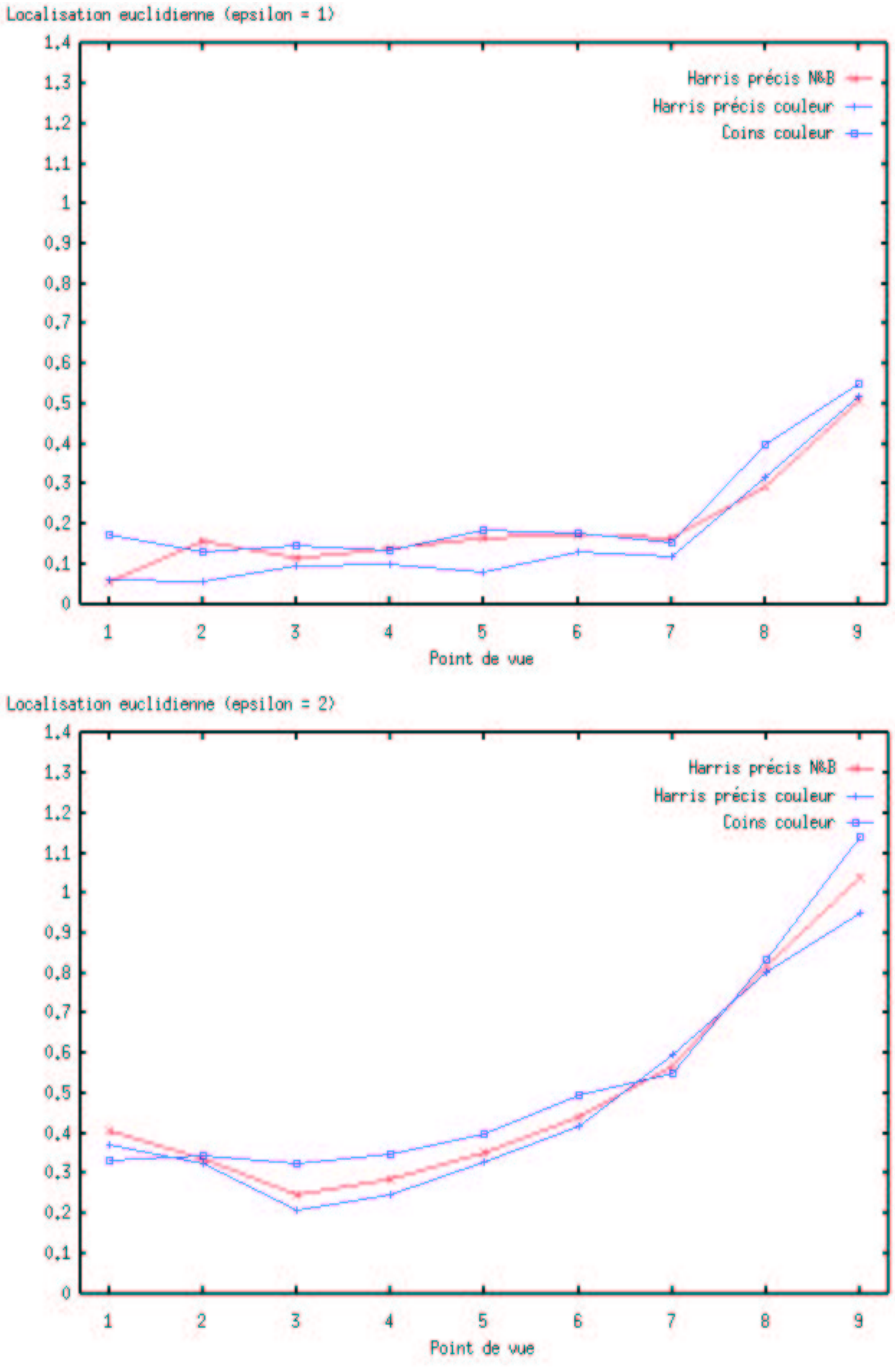


Figure 2.19 – *Mesures de localisation euclidienne suivant 9 points de vue, avec ε=1,2.*

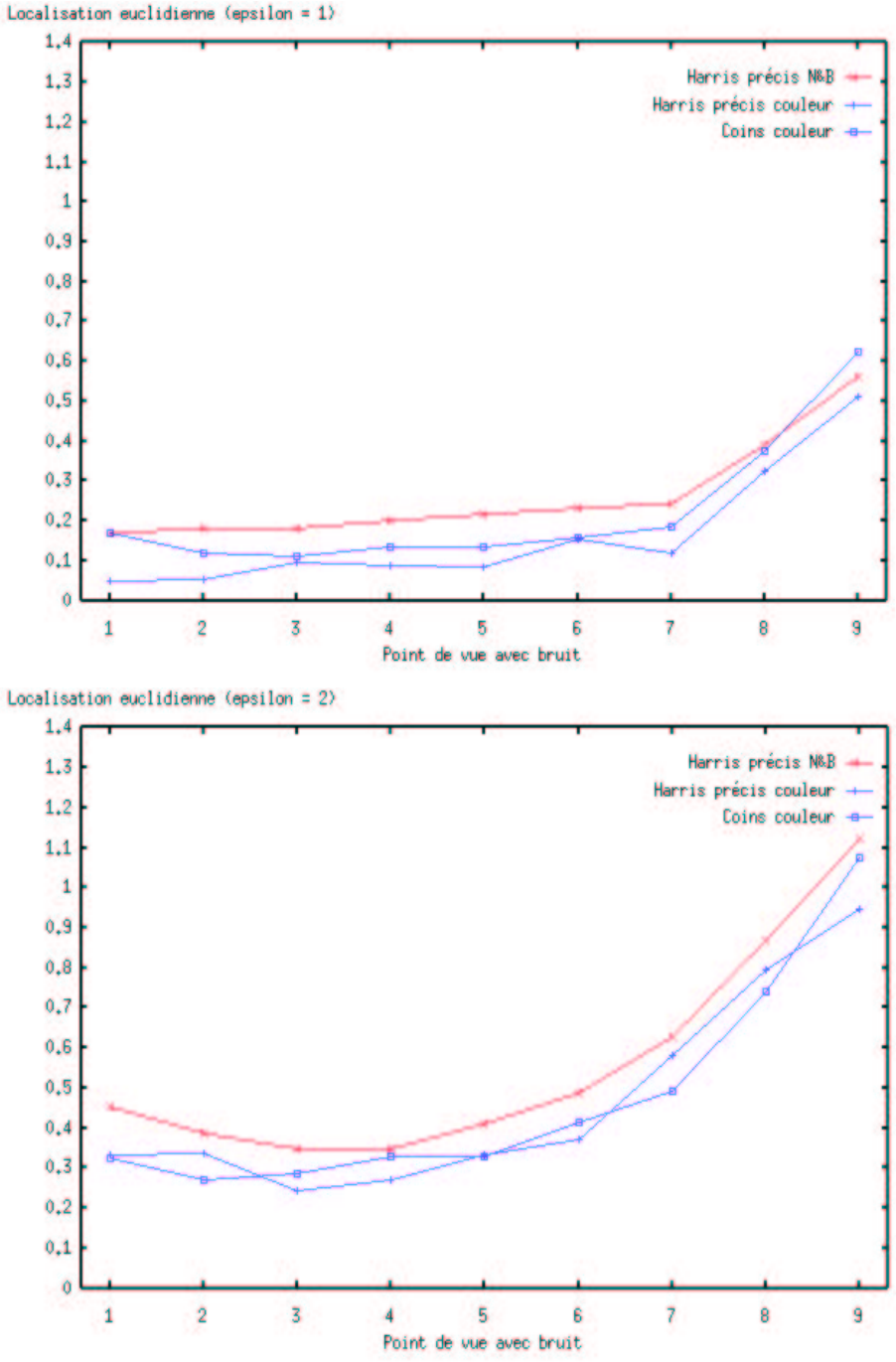


Figure 2.20 – *Mesures de localisation euclidienne suivant 9 points de vue avec bruit, avec* ε*=1,2.*

2.3 Conclusion

Dans cette partie, nous avons mené une évaluation comparative des deux détecteurs de points que nous avons mis en place pour la couleur. Les critères utilisés ici sont la répétabilité et la localisation (projective et euclidienne) des détecteurs. Les mesures ont été réalisées sur les détecteurs couleur ainsi que sur le détecteur Harris précis en niveau de gris, que nous avons considéré comme notre détecteur de référence.

Les résultats obtenus à travers les principales transformations de l'image sont globalement probants, puisqu'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, excepté pour les changements affines de luminosité où le détecteur en niveau de gris obtient les meilleurs résultats. L'information couleur est visiblement plus riche que les seuls niveaux de gris. Les résultats décevants obtenus pour les changements affines de luminosité s'expliquent par le fait que le cadre de l'évaluation est simplifié en passant aux niveaux de gris. Remarquons également que ces résultats ont été observés pour diverses images issues de capteurs différents (la liste des capteurs employés selon l'image est présentée à la table B.1 de l'annexe B).

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 le détecteur de coins apparaît comme équivalent au détecteur de Harris face aux changements d'échelle, mais moins répétable en ce qui concerne les rotations, les changements de luminosité, les changements de point de vue et même le bruit. Le détecteur de Harris couleur sera donc utilisé par la suite pour ce qui est de l'étape de mise en correspondance.

En ce qui concerne les mesures de localisation, 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 au calibrage de caméras par exemple. En revanche, le détecteur Harris précis couleur a 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.

Nous venons de prouver la supériorité des détecteurs couleur suivant les critères de répétabilité et de localisation mis en place. Nous les utilisons donc dans la suite de nos travaux et le choix entre le détecteur Harris couleur et le détecteur de coins couleur dépendra du but recherché.

Deuxième partie

Mise en correspondance d'images

Chapitre 3

Caractérisation locale des points d'intérêt - Etat de l'art

Nous passons en revue dans ce chapitre les différents travaux qui ont donné lieu à des méthodes permettant la caractérisation de points d'intérêt, aussi bien en niveau de gris qu'en couleur. Un certain nombre des approches décrites extrait une information locale autour du point d'intérêt, mais travaille uniquement sur des images en niveau de gris. En parallèle, les quelques méthodes rencontrées exploitant l'information couleur extraient une information beaucoup plus globale. Elles se révèlent cependant peu adaptées à la caractérisation et à la mise en correspondance de points d'intérêt.

Sommaire

Dans les chapitres précédents, nous avons vu comment détecter les points d'intérêt de l'image. Ces points ont été retenus parce qu'a priori ils sont à des endroits où le signal présente un contenu informatif important. Il s'agit maintenant de valider cette hypothèse en capturant cette information. Pour ce faire, il faut décrire localement la fonction signal au voisinage d'un point d'intérêt, de la manière la plus précise possible. Le descripteur obtenu doit posséder plusieurs caractéristiques pour que la mise en correspondance soit efficace : il doit être rapide à calculer, peu important en taille et pourvu d'une mesure de similarité adaptée.

Pour être le plus complet possible, la description recherchée doit également être invariante aux principales transformations de l'image, les vues considérées pouvant être quelconques. D'autre part, le support choisi est un support couleur. Nous avons démontré au chapitre 2 l'intérêt que cette information supplémentaire peut apporter dans la détection des points d'intérêt. Nous allons maintenant tenter de l'exploiter pour enrichir les caractérisations locales de points existantes.

Nous rappelons à la section 3.1 quelques notions utiles sur la théorie des invariants en Vision par Ordinateur. Puis la section 3.2 présente un état de l'art des différentes méthodes de caractérisation existant en niveau de gris, tant dans le domaine spatial que dans le domaine fréquentiel. A la section 3.3, nous passons en revue les rares méthodes de caractérisation couleur que nous avons rencontrées. Nous verrons que ces quelques méthodes sont pour la plupart trop globales pour être efficaces pour la caractérisation de points d'intérêt. Le fameux problème de la constance des couleurs sera également abordé. Il est en effet fondamental en couleur de mettre en place des modèles d'illumination permettant de caractériser efficacement une image indépendamment de la source lumineuse qui éclaire la scène. Enfin, les avantages et les inconvénients des approches qui auront été abordées sont discutés dans la conclusion.

3.1 Introduction aux invariants

Nous nous contentons ici d'une introduction sommaire à la théorie des invariants. Pour plus d'informations, le lecteur peut consulter [Mundy et Zisserman, 1992; Weiss, 1993; Gros et Quan, 1993].

En Vision par ordinateur, les deux types de transformations généralement étudiées sont les transformations de la scène tridimensionnelle vers l'image et les transformations de l'image. Dans [Burns et al., 1990], le théorème de Burns montre qu'il n'existe pas d'invariants pour ce premier type de transformation. Il en existe cependant pour certaines classes géométriques d'objets 3D, ce sont les invariants projectifs. De nombreux travaux ont été réalisés sur ce type d'invariants; on peut consulter entre autres [Morin, 1993; Zisserman et al., 1995; Quan et Mohr, 1995; Laveau, 1996].

Considérons maintenant les transformations de l'image. Soient deux vues I_1 et I_2 d'une même scène, prises dans des conditions différentes (différents paramètres intrinsèques et extrinsèques des caméras, changement d'illumination, etc). Nous nous intéressons alors aux deux types de transformation de l'image suivants :

1. Les transformations *photométriques* de l'image, soit :

$$I_2(x, y) = f(I_1(x, y))$$

Cette formulation traduit un changement d'illumination et opère donc uniquement sur le signal de l'image. f est traditionnellement une fonction de translation, une fonction affine des niveaux de gris ou plus généralement une transformation monotone ;

2. Les transformations *géométriques* de l'image, soit :

$$I_2(x, y) = I_1(g(x, y))$$

g peut tout simplement être une translation et plus généralement, dans le cas d'une scène plane, une homographie du plan projectif. Si l'on considère le changement de point de vue dans une scène quelconque, alors g n'existe pas.

Le problème qui se pose alors consiste à calculer des invariants en tenant compte de ces transformations. Différentes méthodes permettant de le faire existent, citons notamment les méthodes infinitésimales qui reposent sur les groupes de Lie [Van Gool et al., 1995] et les méthodes par généralisation et contrainte [Gros et Quan, 1992; Schulz-Mirbach, 1994]. Dans le cas où aucun invariant n'existe, il est possible de faire appel aux *quasi-invariants*. Ce type d'invariant est dû à Binford et remonte à la fin des années 60. On peut consulter [Binford et Lewitt, 1993] pour une définition plus précise. Retenons seulement une propriété qui nous sera très utile par la suite : les invariants au groupe des similitudes sont des quasi-invariants pour une transformation perspective, i.e. qu'ils sont localement invariants à une telle transformation.

Dans la suite de notre étude, nous considérons essentiellement les transformations de l'image suivantes : la translation, la rotation, le changement d'échelle, le changement d'illumination et le changement de point de vue.

3.2 Caractérisation en niveau de gris

Parmi toutes les approches iconiques de caractérisation de points d'intérêt, une méthode simple mais sûrement la plus répandue existe : un point est décrit par une zone d'intérêt de l'image ("template" en anglais), à savoir le niveau de gris qui lui est associé ainsi que ceux de ses pixels voisins. Ces valeurs sont stockées directement dans un vecteur. La méthode peut également être appliquée aux dérivées de l'image, comme par exemple au gradient [Crouzil et al., 1996] ou encore à l'image du laplacien [Nishihara, 1983]. Cette méthode de caractérisation est traditionnellement associée à la méthode de mise en correspondance par corrélation ("template matching"). Elle sera plus amplement décrite dans la section 5.1.2.1 du chapitre 5. Notons pour le moment que son principal désavantage est qu'elle n'est pas invariante aux transformations classiques de l'image énoncées plus haut.

Le point peut également être décrit par l'histogramme des niveaux de gris des voisins. Cette caractérisation est intéressante car elle est invariante à certaines des transformations de l'image, comme les transformations euclidiennes, le changement d'échelle (dans une

certaine mesure) et il est possible de l'adapter aux changements d'intensité. Elle est même très discriminante lorsqu'elle est appliquée à des images en couleur. C'est pourquoi nous la détaillons dans la section 3.3 réservée à la couleur.

De nombreuses autres méthodes de caractérisation locale existent. Nous présentons les principales dans les quatre prochaines sections. Il s'agit en premier lieu des caractérisations travaillant dans le domaine spatial, comme les invariants différentiels, les moments statistiques ou encore les invariants non paramétriques. L'image étant un signal, nous abordons également les descriptions fréquentielles, dans la dernière section.

3.2.1 Les invariants différentiels

3.2.1.1 Le jet local

Une fonction peut être approximée localement par ses dérivées. Si l'on sait les calculer en un point jusqu'à l'ordre n, alors la série de Taylor décrit cette fonction jusqu'à cet ordre, comme le montre l'équation 3.1 ci-dessous :

$$\begin{aligned} f(x_0+x, y_0+y) = f(x_0,y_0) + x\frac{\partial}{\partial x}f(x_0,y_0) + y\frac{\partial}{\partial x}f(x_0,y_0) + \ldots \\ \ldots + \sum_{p=1}^{n} x^p y^{n-p} \frac{\partial^n}{\partial x^p \partial y^{n-p}} f(x_0,y_0) + O(x^n, y^n) \end{aligned} \tag{3.1}$$

Il est donc possible de décrire une image en un point par l'ensemble des dérivées en ce point. Sous forme de vecteur, cet ensemble caractérise la géométrie locale du voisinage du point. Dans [Koenderink et Van Doorn, 1987], Koenderink a utilisé cette idée et a nommé le vecteur *jet local*. Les dérivées sont calculées numériquement de manière stable en utilisant un filtre passe-bas tel que la gaussienne et ses dérivées (cf. annexe C). Le jet local, noté $J_n(x,y,\sigma)$, est alors défini à l'ordre n, pour le point (x,y) et avec la taille σ de la gaussienne de la façon suivante :

$$J_n(x,y,\sigma) = \{I_{i_1 \ldots i_k}(x,y,\sigma) / k = 0, \ldots, n\} \tag{3.2}$$

où $I_{i_1 \ldots i_k}(x,y,\sigma)$ désigne la dérivée k^{ieme} de l'image par rapport aux variables $i_1, \ldots, i_k$ (x ou y dans notre cas). Si les dérivées sont calculées jusqu'à l'ordre 2, le jet local sera de dimension 5 et de dimension 9 jusqu'à l'ordre 3. On parle également de *jet local multi-échelle* lorsque l'on considère l'ensemble des dérivées calculées pour plusieurs tailles de gaussienne.

Notons enfin qu'il existe une autre approche assez répandue et notamment développée dans [Murase et Nayar, 1995] pour décrire le voisinage de points contenus dans des scènes naturelles. Murase et Nayar partent du principe que les pixels du voisinage sont plus ou moins corrélés. Ils procèdent alors à une analyse en composantes principales, afin d'éliminer l'information redondante. Cette caractérisation s'avère en fait équivalente à celle que nous venons de décrire, il a en effet été montré dans [Hancock et al., 1992] que c'est une très bonne approximation des dérivées du jet local.

Le principal inconvénient de la caractérisation présentée ici est que le vecteur n'est pas invariant aux diverses transformations de l'image. Nous présentons dans les paragraphes suivants plusieurs méthodes permettant de corriger ou au moins d'améliorer ce problème.

3.2.1.2 Les filtres directionnels

Le jet local présenté à la section précédente peut être interprété comme la projection du signal sur la base constituée de la gaussienne et de ses dérivées. Dans [Freeman et Adelson, 1991], Freeman et Adelson ont montré qu'à partir de cette base, il est possible de calculer ce qu'ils appellent des filtres directionnels, c'est à dire des filtres définis par des dérivées calculées dans n'importe quelle direction. Il est donc possible de recalculer le jet local en considérant des directions différentes, afin de tenir compte d'une éventuelle rotation appliquée à l'image. La direction du gradient, lorsque celui-ci n'est pas nul, est une solution possible puisqu'elle traduit l'orientation de l'image. Nous donnons à l'équation 3.3 les 3 premières dérivées I_n obtenues à l'ordre n dans la direction θ :

$$\begin{cases} I_1(\theta) = I_x \cos(\theta) + I_y \sin(\theta) \\ I_2(\theta) = I_{xx} \cos^2(\theta) + 2I_{xy} \sin(\theta)\cos(\theta) + I_{yy} \sin^2(\theta) \\ I_3(\theta) = I_{xxx} cos^3(\theta) + 3I_{xxy} cos^2(\theta) sin(\theta) + 3I_{xyy} sin^2(\theta) cos(\theta) + I_{yyy} sin^3(\theta) \end{cases} \tag{3.3}$$

Le principal inconvénient de cette méthode est que le résultat dépend fortement du calcul de l'angle θ, qui représente une importante source d'instabilité. Pour plus de précisions, consulter [Rao et Ballard, 1995] qui utilisent cette méthode et calculent ainsi un jet local ajustable invariant en rotation.

3.2.1.3 Les invariants de Hilbert

Hilbert [Hilbert, 1890] a montré que n'importe quel invariant au groupe des déplacements $SO(2)$ d'ordre fini pouvait s'exprimer comme un polynôme d'invariants irréductibles reposant sur la combinaison de dérivées du jet local. [Koenderink et Van Doorn, 1987; Salden et al., 1992; Florack et al., 1994; ter Haar Romeny, 1996] ont repris cette idée et ont proposé de caractériser un point par un vecteur contenant un ensemble complet de ces invariants. Si l'on considère une image scalaire, ces invariants représentent l'ensemble de base des primitives qui permettent de décrire toutes les propriétés locales intrinsèques de l'image. Cet ensemble est bien connu pour ses propriétés du premier et du second ordre et est rendu indépendant de la rotation image s'il est exprimé en coordonnées de *Gauge* comme suit :

$$I \quad I_\eta \quad I_{\eta\eta} \quad I_{\xi\eta} \quad I_{\xi\xi} \tag{3.4}$$

où η est le vecteur unitaire tel que $\eta = \frac{\nabla I}{|\nabla I|}$ et $\xi \perp \eta$. Notons que dans ce système de coordonnées, nous avons $I_\xi = 0$. Les ensembles d'ordre supérieur sont beaucoup plus compliqués.

Dans le cadre de la caractérisation de points d'intérêt, il est préférable de considérer, d'un point de vue géométrique et/ou numérique, une combinaison de ces cinq invariants, plutôt que ceux présentés en 3.4. Par exemple l'ensemble suivant s'avère plus performant

pour décrire un point en niveau de gris :

$$\left\{ \begin{array}{cc} I & I \\ I_{\eta\eta} + I_{\xi\xi} & I_{xx} + I_{yy} \\ I_\eta & \mid \nabla I \mid \\ \frac{I_{\xi\xi}}{I_\eta} & \frac{-I_x^2 I_{yy} + 2 I_x I_y I_{xy} - I_y^2 I_{xx}}{(I_x^2 + I_y^2)^{3/2}} \\ \frac{I_{\xi\eta}}{I_\eta} & \frac{I_x I_y (I_{yy} - I_{xx}) + I_{xy}(I_x^2 - I_y^2)}{(I_x^2 + I_y^2)^{3/2}} \end{array} \right. \tag{3.5}$$

Ces invariants sont obtenus à partir des éléments du jet local de l'équation 3.2 donc par la convolution du signal avec une gaussienne. Il est donc envisageable de les calculer pour différents supports de gaussienne.

Cette méthode de caractérisation est implémentée par Schmid dans [Schmid, 1996], pour faire de l'indexation d'images. Elle calcule les invariants jusqu'à l'ordre 3 pour obtenir une caractérisation suffisamment riche et les exprime en notation d'Einstein[1]. Cela donne le vecteur noté ici $\vec{v}_{rot}$ de neuf invariants suivant :

$$\vec{v}_{rot}[0..8] = \left(\begin{array}{c} I \\ I_i I_i \\ I_{ij} I_i I_j \\ I_{ii} \\ I_{ij} I_{ij} \\ \varepsilon_{ij} (I_{jkl} I_i I_k I_l - I_{jkk} I_i I_l I_l) \\ I_{iij} I_j I_k I_k - I_{ijk} I_i I_j I_k \\ -\varepsilon_{ij} I_{jkl} I_i I_k I_l \\ I_{ijk} I_i I_j I_k \end{array} \right) \tag{3.6}$$

où ε_{ij} représente le tenseur canonique anti-symétrique de Levy-Civita tel que :

$$\left\{ \begin{array}{l} \varepsilon_{12} = -\varepsilon_{21} = 1 \\ \varepsilon_{11} = \varepsilon_{22} = 0 \end{array} \right. \tag{3.7}$$

et où I représente la fonction de luminance convoluée avec une gaussienne et I_i la sommation des dérivées du jet local par rapport à toutes les composantes telle que :

$$\left\{ \begin{array}{l} I_i = \sum_i I_i = I_x + I_y \\ I_{ij} = \sum_i \sum_j I_{ij} = I_{xx} + I_{xy} + I_{yx} + I_{yy} \end{array} \right. \tag{3.8}$$

On a par exemple $\vec{v}_{rot}[1] = \sum_i (I_i I_i) = I_x^2 + I_y^2$ qui représente la magnitude du gradient.

Nous venons de présenter deux types de caractérisation locale invariantes à la rotation, l'une utilisant le jet local ajustable et l'autre les invariants de Hilbert. Voyons maintenant à la section suivante comment rendre ces caractérisations invariantes à d'autres transformations de l'image.

[1] Se référer à [ter Haar Romeny, 1996] pour une introduction à la théorie des tenseurs et à la notation contractée qui en découle.

3.2.1.4 Autres transformations de l'image

Invariance au changement d'échelle

Comme nous l'avons vu au chapitre 1, un changement d'échelle est dû soit au déplacement de la caméra soit à un changement de la distance focale de celle-ci, et il peut être décrit par un changement affine (se référer à l'équation 1.8).

L'appariement de deux images différant d'un changement d'échelle pose plusieurs difficultés majeures. Comme pour la détection des points d'intérêt, il est impératif de prendre en compte dans la caractérisation de ces points le changement de résolution entre les deux images et surtout son impact sur les données photométriques.

Schmid montre dans [Schmid, 1996] que le vecteur $\vec{v}_{rot}$ de l'équation 3.6 est robuste à un changement d'échelle allant jusqu'à 20% et Rao montre dans [Rao et Ballard, 1995] que le jet local ajustable tolère quant à lui des changements d'échelle jusqu'à 10%. Il est donc nécessaire de mettre en œuvre d'autres techniques.

D'après l'équation 1.9 (cf. chapitre 1), les dérivées n-ièmes de l'image sont théoriquement égales, à un facteur multiplicatif α^n près. Il est donc aussi possible de mettre en place des invariants au changement d'échelle faisant intervenir un rapport de dérivées. Malheureusement il s'avère que de tels invariants ne sont guère plus stables à un changement d'échelle que $\vec{v}_{rot}$ ou que le jet ajustable [Schmid, 1996]. En effet, supposons que les dérivées sont implémentées numériquement par convolution avec une gaussienne. Pour effectivement calculer un invariant, il est nécessaire d'adapter le support de la gaussienne au changement d'échelle. Le choix de ce paramètre permet alors d'effectuer un lissage plus ou moins important de l'image et donc de la décrire à un certain niveau d'échelle. Comme pour la détection des points d'intérêt, il est impératif d'envisager une approche multi-échelle (cf. l'équation 1.11 pour une définition de l'espace échelle). Il est ensuite nécessaire de discrétiser cet espace en échelle. Schmid propose de choisir un pas de discrétisation qui correspond à un changement d'échelle d'au plus 20%, puisque $\vec{v}_{rot}$ est stable au changement d'échelle jusqu'à ce pas. Les invariants définis précédemment (le jet local ajustable et le vecteur $\vec{v}_{rot}$) sont donc calculés pour plusieurs valeurs de σ, ce paramètre correspondant à une échelle particulière.

Cette méthode de caractérisation a jusqu'à présent été essentiellement utilisée pour faire de l'indexation d'images. Citons par exemple [Rao et Ballard, 1995] qui utilise un jet local ajustable défini jusqu'au troisième ordre et à 5 niveaux d'échelle, soit un vecteur invariant à la rotation et au changement d'échelle de $9 \times 5 = 45$ composantes ; ou encore [Schmid, 1996; Dufournaud et al., 2000] qui calculent aussi $\vec{v}_{rot}$ pour plusieurs supports de gaussienne.

Invariance au changement de luminosité

La première composante de $\vec{v}_{rot}$ étant le niveau de gris du point (éventuellement lissé), elle n'est bien sûr pas invariante au changement de luminosité. En revanche, il est facile de voir que la dérivation rend les autres composantes invariantes aux translations des niveaux de gris. Si l'on considère les transformations affines, il faut prendre en compte des rapports de dérivées. En effet, une transformation affine des niveaux de gris entre deux images I_1

et I_2 se modélise par :

$$I_2(x,y) = aI_1(x,y) + b \tag{3.9}$$

Les dérivées n-ièmes du signal sont donc de la forme :

$$I_2^{(n)}(x,y) = aI_1^{(n)}(x,y) \tag{3.10}$$

Par conséquent, n'importe quel quotient de deux dérivées est invariant à cette transformation affine. Pour rendre le vecteur $\vec{v}_{rot}$ invariant aux transformations affines de niveaux de gris, il suffit donc de diviser chacune de ses composantes $\vec{v}_{rot}[1..8]$ par la puissance adéquate de l'une d'entre elles, la magnitude du gradient par exemple. Le vecteur d'invariants obtenu proposé par [Schmid, 1996] est donc réduit à $9-2=7$ invariants. Il et noté ici $\vec{v}_{aff}$:

$$\vec{v}_{aff}[0..2] = \begin{pmatrix} \frac{\vec{v}_{rot}[2]}{(I_iI_i)^{\frac{3}{2}}} \\ \frac{\vec{v}_{rot}[3]}{(I_iI_i)^{\frac{1}{2}}} \\ \frac{\vec{v}_{rot}[4]}{I_iI_i} \end{pmatrix} \quad \text{et} \quad \vec{v}_{aff}[3..6] = \frac{1}{(I_iI_i)^2}\vec{v}_{rot}[5..8] \tag{3.11}$$

En ce qui concerne les autres transformations des niveaux de gris, Florack fait remarquer dans [Florack et al., 1994] que les isophotes ne sont pas modifiés sous l'action d'une transformation inversible de luminosité. Jusqu'au deuxième ordre, il existe deux invariants, la courbure des isophotes k et la courbure des lignes de plus grande pente μ, exprimées ici en notation d'Einstein :

$$k = \frac{\varepsilon_{ij}\varepsilon_{kl}I_iI_{jk}I_l}{(I_mI_m)^{\frac{3}{2}}} \quad \text{et} \quad \mu = \frac{\varepsilon_{ij}I_jI_kI_{ik}}{(I_mI_m)^{\frac{3}{2}}} \tag{3.12}$$

Invariance au changement de point de vue

Le jet local ajustable et le vecteur d'invariants $\vec{v}_{rot}$ présentés précédemment sont à la fois invariants à la rotation image et au changement d'échelle s'ils sont calculés dans un cadre multi-échelle. Ces deux types de caractérisation sont donc invariants au groupe des similitudes de l'image, et par conséquent quasi-invariants aux transformations perspectives [Binford et Lewitt, 1993]. Ils constituent donc une caractérisation robuste aux changements de point de vue. Nous aurons l'occasion de confirmer cette propriété dans le chapitre sur l'appariement.

3.2.2 Les moments

Les moments permettent également de caractériser une image. En effet, d'un point de vue stochastique, cette dernière peut être vue comme une distribution aléatoire, et par conséquent peut être caractérisée de façon unique par ses moments. Etant donné une image $I(x,y)$, le moment d'ordre $(p+q)$ est défini dans le cas discret de la façon suivante :

$$m_{pq} = \sum_x \sum_y x^p y^q I(x,y) \tag{3.13}$$

Les moments ont été utilisés pour la première fois en 1962 par Hu [Hu, 1962] pour des problèmes de reconnaissance en Vision par Ordinateur. Il a alors été démontré l'existence de combinaisons de moments qui sont invariantes à la translation, à la rotation et au changement d'échelle.

Teague a introduit en 1980 les moments de Zernike qui utilisent la théorie des polynômes orthogonaux du même nom pour décrire une image. Le lecteur peut consulter [Teague, 1980; Teh et Chin, 1988] pour un état de l'art sur les techniques basées sur les moments. Freeman et Saleh [Freeman et Saleh, 1988] ont également étendu l'utilisation des moments invariants en les considérant à la fois dans le domaine spatial et fréquentiel, permettant ainsi de capturer des informations à divers niveaux. Il existe aussi les moments de Legendre basés sur les polynômes du même nom. Une étude plus récente [Kim et Yuan, 1994] a montré que les moments de Zernike sont les moins sensibles au bruit et les moins redondants en information.

De nombreux travaux comme [Prokop et Reeves, 1992; Reiss, 1993; Flusser et al., 1994; Van Gool et al., 1996] ont ensuite été réalisés sur l'invariance des moments face aux transformations affines des coordonnées et aux changements d'intensité dans des images de niveaux de gris. Il faut cependant noter que les moments restent tout de même assez sensibles aux occultations. Mais l'inconvénient majeur de ces approches est qu'il est la plupart du temps nécessaire de calculer les moments à un ordre relativement grand si l'on souhaite obtenir une caractérisation suffisamment discriminante. Nous verrons dans la section 3.3.2.2 que l'introduction des moments couleur généralisés permet de remédier à ce problème.

3.2.3 Les invariants non paramétriques

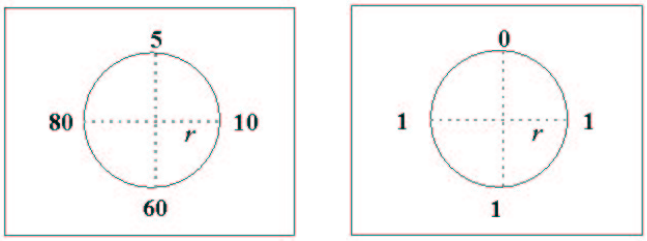
Dans cette section, nous présentons la méthode récente développée par Lan et Mohr et basée sur des invariants à la rotation non paramétriques [Lan et Mohr, 1997a]. Soient deux images I_1 et I_2 différant d'une rotation d'angle c, et deux pixels (r_1, θ_1) et (r_2, θ_2) exprimés en coordonnées polaires. Les auteurs partent du principe que l'ordre entre deux pixels reste inchangé sous une rotation de l'image. Ainsi si $I_1(r_1, \theta_1) > I_1(r_2, \theta_2)$ alors on a également $I_2(r_1, \theta_1 + c) > I_2(r_2, \theta_2 + c)$. Cette hypothèse leur permet de définir pour une image I la fonction de comparaison suivante :

$$C_I(r_1, \theta, r_2, \theta + \Delta_1) = \begin{cases} 1 & \text{si } I(r_1, \theta) > I(r_2, \theta + \Delta_1) \\ 0 & \text{sinon} \end{cases} \tag{3.14}$$

La figure 3.1 illustre cette définition pour $r_1 = r_2 = r$, $\theta = 0$ et $\Delta_1 = \frac{\pi}{2}$.

Pour obtenir un invariant à la rotation à partir de cette description, les auteurs appliquent ensuite les méthodes de généralisation et contrainte abordées à la section 3.1. En intégrant sur le groupe des rotations, ils obtiennent l'invariant N_I défini ci-après, pour n'importe quel autre couple de points $(r_3, \theta + \Delta_2)$ et $(r_4, \theta + \Delta_3)$:

$$N_I(r_1, r_2, r_3, r_4, \Delta_1, \Delta_2, \Delta_3) = \frac{1}{2\pi} \int_0^{2\pi} |C_I(r_1, \theta, r_2, \theta + \Delta_1) - C_I(r_3, \theta + \Delta_2, r_4, \theta + \Delta_3)| d\theta \tag{3.15}$$



Une image avec 4 niveaux de gris

La caractérisation non paramétrique correspondante

Figure 3.1 – *Un exemple de caractérisation non paramétrique*

Cette caractérisation a les avantages d'être invariante à la rotation image ainsi qu'aux transformations croissantes de la luminance. Elle est aussi moins sensible au bruit et aux occultations que les approches traditionnelles, puisque seul l'ordre entre les pixels est considéré.

Dans leurs expérimentations, Lan et Mohr réduisent N_I à quatre paramètres en posant : $r_1 = r_2 = s$, $r_3 = r_4 = t$ et $\Delta_3 = \Delta_1 + \Delta_2$. L'invariant devient :

$$N_I(s,s,t,t,\Delta_1,\Delta_2,\Delta_1+\Delta_2) = \frac{1}{2\pi}\int_0^{2\pi} |C_I(s,\theta,s,\theta+\Delta_1) - C_I(t,\theta+\Delta_2,t,\theta+\Delta_1+\Delta_2)| d\theta \tag{3.16}$$

15 valeurs sont utilisées pour les rayons et 6 pour les angles, ce qui conduit à un vecteur caractéristique contenant 8100 invariants. La grande dimension de ce vecteur impose alors de mettre en œuvre une technique d'appariement spécifique : trois tests sont d'abord réalisés pour éliminer la plus grande partie des faux candidats, avant d'employer pour les correspondants potentiels restants une métrique plus sophistiquée.

La méthode est testée puis comparée à celle de Schmid utilisant les invariants différentiels (présentée à la section 3.2.1.3). Les résultats obtenus montrent l'amélioration qu'apporte cette technique face aux transformations euclidiennes, aux changements d'intensité et aux changements de point de vue. Il semble en effet que les invariants non paramétriques soient moins sensibles au bruit que la méthode basée sur les invariants différentiels, qui se doit d'utiliser les dérivées de l'image jusqu'à l'ordre 3 pour être suffisamment discriminante.

3.2.4 Les descriptions fréquentielles

3.2.4.1 Les Transformées de Fourier et de Mellin

La caractérisation d'un signal f la plus connue dans le domaine des fréquences est sans aucun doute *la Transformée de Fourier* [Bracewell, 1978], dont nous rappelons la formulation continue dans le cas bi-dimensionnel à l'équation 3.17 ci-dessous :

$$F(u,v) = \int\int f(x,y) e^{-i(ux+vy)} dxdy \tag{3.17}$$

La transformée de Fourier d'une paire de fréquences $(u,v) \in \mathcal{R}^2$ est caractérisée par

une amplitude et une phase. Dans notre étude, l'intérêt de son utilisation (dans sa version continue comme discrète) réside dans le fait qu'au moins l'un de ces deux attributs reste constant à travers certaines transformations de l'image. Par exemple, sa phase est indépendante de la luminosité des images et son amplitude ne varie pas avec la translation. D'autre part, la phase est stable à un changement d'échelle jusqu'à 20% et peut donc être mise en oeuvre dans un contexte multi-échelle.

Des variantes existent sous certaines contraintes, comme notamment la transformée de Fourier *circulaire* [Grace et Spann, 1991], pour laquelle l'amplitude est rendue invariante à la rotation. Dans [Reddy et Chatterji, 1996], une méthode est également développée pour rendre cette caractérisation invariante à la translation, à la rotation et au changement d'échelle.

Un des principaux avantages de la caractérisation basée sur la Transformée de Fourier est qu'elle peut être calculée très efficacement, en comparaison avec d'autres méthodes de caractérisation, si l'on met en place son algorithme rapide de calcul.

Il existe également une autre transformée, *la Transformée de Mellin*, qui s'applique sur des nombres complexes dans sa forme la plus générale. Elle est traditionnellement utilisée uniquement sur des valeurs imaginaires pures, et sa formulation bi-dimensionnelle pour une paire de fréquences $(u, v) \in \mathcal{R}^2$ est alors la suivante :

$$z(u,v) = \int_0^\infty \int_0^\infty f(x,y) x^{iu-1} y^{iv-1} dxdy \tag{3.18}$$

La Transformée de Mellin peut être vue comme étant la Transformée de Fourier appliquée après un changement de variable exponentiel. Si elle est utilisée avec des valeurs imaginaines pures, alors son amplitude est invariante au changement d'échelle. Comme pour la Transformée de Fourier, il en existe une variante, appelée *la Transformée de Fourier-Mellin*, qui permet d'obtenir en plus l'invariance à la rotation [Sheng et Lejeune, 1991].

Des travaux ont été réalisés pour combiner les Transformées de Fourier et de Mellin et ainsi obtenir l'invariance à la fois à la translation, à la rotation et au changement d'échelle [Casasent et Psaltis, 1980]. Pour un état de l'art complet sur les Transformées de Fourier et de Mellin et plus généralement sur la théorie des invariants, le lecteur peut consulter [Wood, 1996].

Soulignons enfin l'inconvénient majeur de ces deux types de caractérisation : elles imposent une localisation en fréquence et non pas en espace, c'est-à-dire qu'elles ne permettent pas de dire quelles fréquences caractérisent un point donné de l'image, selon le principe d'incertitude. Elles s'avèrent donc difficilement exploitables telles quelles pour caractériser efficacement les points d'intérêt, et sont plutôt utilisées pour faire de la reconnaissance de formes ou de l'indexation d'images. Des travaux ont tout de même été réalisés pour les adapter à la localisation en espace ; nous présentons ainsi à la section suivante la Transformée de Gabor, qui peut être considérée comme une Transformée de Fourier locale en espace.

3.2.4.2 La Transformée de Gabor

Pour obtenir une bonne localisation en espace, il est possible d'adapter la Transformée de Fourier avec une fonction de fenêtre (on parle alors de Transformée de Fourier *à court terme* ou encore de Transformée de Fourier *à fenêtre glissante*). Cette fonction doit avoir la particularité d'être bien localisée en espace et en fréquence, comme typiquement la Gaussienne. Ainsi dans [Gabor, 1946], Gabor définit une transformée portant son nom qui utilise un fenêtrage gaussien permettant d'obtenir une bonne précision à la fois en fréquence et en espace. Le signal est convolué par un filtre dont l'expression est la suivante :

$$G_{\sigma w_x w_y}(x,y) = e^{i(w_x x + w_y y)} \frac{1}{\sigma 2\pi} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.19}$$

Dans le cadre du calcul d'appariements, une égalité de phase entre deux points signifie une grande probabilité qu'il s'agisse de points à mettre en correspondance, en sachant néanmoins qu'une même valeur de phase peut apparaître plusieurs fois pour des points différents. Pour être certain qu'une égalité de phase correspond à un appariement exact, une approche multi-échelle s'impose. Toutefois cette mesure est locale en fréquence mais directionnelle en espace. Elle résiste donc malheureusement peu à des rotations et des changements d'échelle faibles. Dans [Wu et Bhanu, 1995], Wu a alors proposé une implantation des filtres de Gabor dans plusieurs directions et à plusieurs échelles permettant de résoudre ce problème. Etant donné que ce filtre donne une information locale de la phase du signal, une utilisation classique est donc par exemple l'estimation de la disparité entre deux images [Sanger, 1988; Fleet et al., 1991].

Il faut tout de même noter que les résultats obtenus à partir de cette approche dépendent fortement de la taille de la fenêtre employée, c'est-à-dire du support de la Gaussienne. En effet, le choix de ce paramètre est subordonné à une connaissance a priori du signal que l'on désire analyser, information que l'on possède rarement lorsque le signal est une image.

3.2.4.3 Les ondelettes

La première définition de la Transformée en Ondelettes a été établie par Grossmann et Morlet [Grossmann et Morlet, 1984] en temps et échelle continus. La théorie des Ondelettes[2] repose sur l'idée que le signal peut être caractérisé par différentes échelles et différentes résolutions. En résumé, elle consiste à remplacer le paramètre de fréquence de la Transformée de Gabor par un paramètre d'échelle, ce qui conduit à une analyse temps-échelle et non plus temps-fréquence. L'analyse par ondelettes de la fonction f est définie par :

$$W_\psi(a,b) = \frac{1}{\sqrt{a}} \int f(x) \bar{\psi}(\frac{x-b}{a}) dx \tag{3.20}$$

où ψ est une fonction continue quelconque appelée fonction génératrice de l'ondelette, centrée en b et d'échelle a.

L'inconvénient majeur de la Transformée en Ondelettes telle qu'elle vient d'être présentée est que sa résolution fréquentielle est d'autant plus mauvaise que l'échelle est petite. Par

[2]Le lecteur peut consulter le livre de Meyer [Meyer, 1992] pour une présentation complète.

exemple, un signal ayant un spectre de fréquence étroit et essentiellement localisé aux hautes fréquences ne sera pas bien représenté. Une solution a ce problème a émergé au début des années 90 avec l'apparition de la théorie des décompositions atomiques (ou par paquets) d'ondelettes. Ces méthodes ont pour caractéristique commune l'analyse des signaux suivant trois paramètres physiques : le temps, la fréquence et l'échelle. Citons notamment la méthode de Mallat [Mallat, 1989; Mallat, 1999] qui a également étendu ses travaux dans le domaine de la Vision par Ordinateur au cas des signaux discrets. Comme pour la Transformée de Gabor, il reste néanmoins comme inconvénient le fait que les résultats obtenus dépendent fortement de la taille du voisinage sur lequel sont effectués les calculs. Il existe une autre transformée, la Transformée de Wigner qui permet un calcul de la fréquence en tout point. Nous ne développons pas cette approche ici car la représentation de cette transformée est difficile et lourde à calculer et par conséquent n'est pas adaptée à la caractérisation de points d'intérêt.

3.2.5 Bilan

Nous avons présenté dans cette section les principales approches permettant de caractériser les points d'intérêt issus d'une image monochrome. Une grande partie des techniques développées travaille directement dans le domaine de l'image : la corrélation, les invariants différentiels, les moments, les invariants non paramétriques.

La corrélation est la méthode de caractérisation la plus ancienne et probablement la plus répandue jusqu'à ces dernières années. Elle est petit à petit délaissée au profit de techniques moins coûteuses en temps de calcul et surtout plus robustes aux transformations de l'image telles que la rotation.

Nous avons également rencontré les approches statistiques qui voient l'image comme une distribution aléatoire des niveaux de gris, et qui exploitent notamment les moments ou encore les histogrammes pour caractériser celle-ci. Quelques travaux ont été réalisés suivant cette voie, mais nous n'avons pas jugé nécessaire de les développer outre mesure dans cette section, car en réalité ces approches s'avèrent bien plus intéressantes lorsqu'elles sont appliquées à des images en couleur. Nous ne manquerons donc pas d'y consacrer une part importante de la prochaine section.

Viennent ensuite les invariants non paramétriques, méthode récente et originale qui semble faire ses preuves, au vu des résultats présentés. Ils permettent en effet une caractérisation invariante à la rotation, aux transformations croissantes de la luminance, et semblent assez peu sensibles au bruit et aux occultations.

En ce qui concerne les approches définies dans le domaine spatial, nous conclurons sur les invariants différentiels. Depuis les travaux de Hilbert, de nombreuses études ont été réalisées sur ces ensembles d'invariants à la rotation, jusqu'à ce que Schmid valide la méthode en 1996, en les calculant jusqu'au troisième ordre pour faire de l'indexation dans des bases d'images. L'approche s'est avérée probante, les invariants pouvant être assez facilement rendus invariants à certains changements de luminosité et s'avérant plutôt robustes aux changements de point de vue. Il semble que cette méthode iconique soit en voie de remplacer la classique corrélation. Elle paraît prometteuse pour l'appariement, l'indexation ou encore la reconnaissance d'images.

Une image étant après tout un signal, il s'est avéré nécessaire pour être complet de passer en revue les méthodes de description travaillant dans le domaine fréquentiel. Les transformées de Fourier et de Mellin, outils classiques en traitement du signal, se révèlent fort peu appropriées pour caractériser des points de l'image, puisqu'elles n'admettent qu'une localisation en fréquence. Des adaptations pour améliorer leur localisation en espace ont bien été tentées, citons la transformée de Gabor ou encore les ondelettes, mais celles-ci admettent trop d'inconvénients pour être efficaces lors de la caractérisation de points.

Plus généralement, les nombreuses approches que nous venons de passer en revue nous montrent que les niveaux de gris de l'image possèdent à eux seuls un contenu informatif suffisamment riche pour caractériser efficacement les points d'intérêt d'une image. La puissance toujours plus grande des ordinateurs disponibles sur le marché et l'engouement croissant pour les techniques du multimédia nous entraînent naturellement vers l'utilisation de données plus complexes comme les images en couleur. Qu'advient-il alors de toutes les méthodes de caractérisation qui viennent d'être présentées ? En existe-t-il de nouvelles, plus performantes et spécifiques à l'information couleur ? Ou peut-on en rester aux simples niveaux de gris ? Nous allons explorer ces questions à la section suivante, consacrée à la caractérisation en couleur.

3.3 Caractérisation en couleur

Le signal couleur peut être décomposé de diverses manières en trois composantes [Luong, 1991]. Il est alors naturellement possible de considérer une image en couleur comme la simple superposition d'images en niveau de gris, et par conséquent d'exploiter pour chacune de ces composantes les approches en niveau de gris venant d'être décrites. A partir des années 90, on a vu néanmoins l'émergence de nouvelles méthodes dédiées à la couleur. Nous allons les passer en revue dans cette section. Nous verrons alors que l'information couleur ouvre de nouvelles possibilités.

A ce jour, la plus grande partie des techniques de caractérisation spécifiques à la couleur provient de l'indexation d'images. Ces techniques sont basées sur une approche *globale* de l'image. Les études réalisées consistent à utiliser la couleur pour caractériser l'image entière le plus efficacement possible, et ceci en tenant compte des principales transformations de l'image. Étant donné le contexte dans lequel sont réalisés ces travaux - l'indexation - une des préoccupations principales est l'obtention de l'invariance face aux changements d'illumination. On parle alors du problème de la *constance des couleurs*, qui consiste à trouver une description couleur de la surface des objets qui soit indépendante de l'illumination. La constance des couleurs est un thème récurrent dès lors qu'une nouvelle caractérisation utilisant l'information couleur est mise en place. Il nécessite en premier lieu de disposer d'un modèle d'illumination. C'est la raison pour laquelle nous présentons dans une première section les différents modèles d'illumination susceptibles d'être utilisés dans la suite de cette étude. Les méthodes de caractérisation globales sont ensuite détaillées à la section 3.3.2. Il y est montré qu'elles fournissent des résultats intéressants. Étant donné qu'elles sont globales, elles ne permettent cependant pas de caractériser directement les points d'intérêt. Il est donc nécessaire de les adapter à des techniques plus *locales*, ce qui semble relativement délicat à réaliser. En réalité, nous avons rencontré très peu de

techniques utilisant l'information couleur pour caractériser l'image localement. Il en existe quelques unes que nous décrivons à la section 3.3.3, avant de présenter celle que nous avons mise en place.

3.3.1 Les modèles d'illumination

Une scène peut être soumise à deux types de changement d'illumination :

- Un changement d'intensité d'une ou plusieurs teintes de la source émettrice. On parle alors de changement *interne* de la source ;
- Un déplacement de la source émettrice. On parle dans ce cas d'un changement *externe*.

Nous présentons dans les deux sections suivantes les différentes approximations rencontrées qui permettent de modéliser ces deux types de changement d'illumination. Puis, avant de passer en revue les méthodes de caractérisation de points d'intérêt intégrant pour la plupart ces modèles, nous détaillons dans la section 3.3.1.3 quelques méthodes de normalisation de l'image. Un pré-traitement de l'image est en effet une solution pour rendre celle-ci indépendante du modèle d'illumination considéré.

3.3.1.1 Changement interne

Dès lors qu'ils se veulent réalistes, les modèles d'illumination sont extrêmement complexes. Pour des raisons évidentes de simplification, on a l'habitude de considérer la formation des images couleur dans les mondes de type Mondrian[3]. Dans ce contexte, la lumière réfléchie sur une surface dépend des propriétés spectrales de la réflectance de la surface et de la lumière incidente. Si l'on considère un système de vision, elle est capturée par un ensemble de n capteurs de la caméra. Les surfaces étant supposées lambertiennes, la valeur $\rho_k(x)$ mémorisée par chaque capteur à la position x est donnée par l'équation 3.21 :

$$\rho_k(x) = \int_w S_\lambda(x_{obj}) E_\lambda(x_{obj}) R_{\lambda,k} \, d\lambda \qquad 1 \leq k \leq n \tag{3.21}$$

où λ représente la longueur d'onde appartenant au spectre visible w, $R_{(\lambda,k)}$ la réponse du k^{eme} capteur pour la longueur d'onde λ, $E_\lambda(x_{obj})$ la lumière incidente et $S_\lambda(x_{obj})$ la réflectance spectrale au point x_{obj} de la surface projeté en x sur l'image. Ce modèle décrit précisément la formation des images pour l'œil humain et les caméras couleur.

Les trois grandeurs $R_{(\lambda,k)}$, $E_\lambda(x_{obj})$ et $S_\lambda(x_{obj})$ venant d'être définies peuvent être décomposées sur des bases de dimension réduite selon des modèles linéaires, afin d'obtenir une dimensionalité faible. Étudions le cas de l'approximation de la réflectance spectrale :

Approximation de la réflectance spectrale

[3]Un Mondrian est un monde simplifié composé de surfaces planes lambertiennes soumises à une source d'illumination ponctuelle qui émet uniformément dans toutes les directions.

La réflectance spectrale de la surface peut être approchée par un modèle linéaire fini, dont la robustesse a été prouvée [Maloney, 1986] et tel que :

$$S_\lambda(x) = \sum_{1 \leq i \leq m} \sigma_i(x)\ s_{i,\lambda} \tag{3.22}$$

où les $s_{i,\lambda}$ sont un ensemble de fonctions de base supposées constantes sur tout l'objet, et les $\sigma_i(x)$ les m éléments du vecteur $\sigma(x) = (\sigma_1(x), ..., \sigma_m(x))^T$ représentant les poids de la fonction de réflectance et uniquement dépendants du point. Il décrit la distribution de la réflectance spectrale pour l'objet. Notons $\rho(x) = (\rho_1(x), ..., \rho_n(x))^T$ le vecteur des n mesures de capteurs. Étant donnée A une matrice $n \times n$, il est possible d'écrire pour $m = n$:

$$\rho(x) = A\sigma(x) \tag{3.23}$$

Remarquons que A ne dépend que de l'illumination. En considérant cette approximation, il est alors possible de modéliser les changements d'illumination selon le modèle linéaire ci-dessous.

Modèle linéaire de changement d'illumination

Soient deux illuminations caractérisées par les matrices respectives A et $\tilde{A}$ non singulières. Les réponses des capteurs pour chaque type d'illumination sont alors liées par la transformation linéaire suivante :

$$\tilde{\rho}(x) \approx M\rho(x) \quad \text{avec} \quad M = \tilde{A}A^{-1} \tag{3.24}$$

Compte tenu des hypothèses faites, le modèle d'illumination linéaire est, avec neuf paramètres, le plus complet des modèles connus. On rencontre cependant d'autres approximations plus simples. Citons rapidement le modèle *scalaire* à un seul paramètre, très spécialisé [Brainard et al., 1989], pour lequel les plans couleur sont multipliés par le même scalaire. Vient ensuite le modèle *diagonal* à trois paramètres, associé principalement à Finlayson et présenté ci-dessous.

Modèle diagonal de changement d'illumination

Le modèle diagonal, étudié notamment par [Land, 1977; Forsyth, 1990; Finlayson, 1995] est une version simplifiée du modèle linéaire. Il s'avère cependant suffisant lorsque les capteurs possèdent une bande spectrale étroite, par exemple lorsqu'ils sont sensibles à une seule longueur d'onde. Sa pertinence dépend donc des réponses fournies par les capteurs. Cette idée a donné lieu à quelques travaux, mais c'est Finlayson dans [Finlayson et al., 1994] qui l'a finalisée. Il propose d'utiliser certaines combinaisons linéaires des réponses, plus enclines à minimiser l'erreur d'approximation. Le modèle diagonal devient donc dans sa version *généralisée* :

$$\mathcal{T}\tilde{\rho}(x) = D\mathcal{T}\rho(x) \tag{3.25}$$

où $\mathcal{T}$ représente la transformation d'affinage ("sharpening" en anglais) des réponses données par les capteurs, D étant la matrice diagonale. Finlayson introduit également plusieurs méthodes pour calculer $\mathcal{T}$.

En résumé, le modèle linéaire est le plus complet des modèles d'illumination, sous l'hypothèse du monde Mondrian, c'est à dire sans tenir compte par exemple des inter-réflexions ou des reflets spéculaires. Le modèle diagonal, plus simple, semble toutefois suffisant dans sa version généralisée, bien qu'en réalité aucune comparaison avec d'autres modèles n'ait été réalisée à ce jour.

D'autres systèmes plus complexes peuvent être envisagés. Il est par exemple possible d'intégrer aux modèles existants une translation des composantes de la couleur. Ainsi dans [Gros et al., 1997b], 11 modèles d'illumination internes sont évalués, incluant les modèles diagonal et linéaire. La principale préoccupation des auteurs est de déterminer quel modèle d'illumination possède le meilleur ratio qualité/complexité dans un contexte de caractérisation *locale* de points d'intérêt. Il est montré que le calcul d'un modèle linéaire n'est pertinent que pour de grandes images. Lorsqu'on ne s'intéresse qu'à des sous-parties de l'image, les paramètres non diagonaux ne s'avèrent pas significatifs, ce qui rend les modèles diagonaux amplement suffisants. En ce qui concerne les paramètres de translation, l'étude montre qu'ils sont significatifs, même pour des petites zones de l'image. Ils le sont d'autant plus lorsque les deux images sont très différentes. Ainsi ils apportent une nette amélioration en cas de reflets spéculaires [Wolff, 1994] par exemple, aspect jusqu'alors non traité avec les modèles classiques, du fait de l'hypothèse simplificatrice du monde Mondrian. Une des conclusions est que le modèle diagonal de Finlayson ne se justifie que pour des images très petites, ou très peu différentes. Le modèle alors préconisé est le modèle diagonal enrichi d'un vecteur de translation tel que :

$$\tilde{\rho}(x) = D\rho(x) + T \tag{3.26}$$

Il possède six degrés de liberté pour un espace des couleurs à trois composantes. Signalons tout de même que le modèle diagonal au sens de Finlayson est malgré tout le modèle d'illumination que l'on va être amené à retrouver pour la plupart des approches, qu'il s'agisse des méthodes de caractérisation globales, locales ou encore des méthodes de normalisation.

3.3.1.2 Changement externe

Dans le cas d'un déplacement de la source de lumière, Finlayson suppose que chaque pixel est multiplié par un facteur. Ainsi un pixel (r, v, b) est transformé en (r', v', b') tel que :

$$(r', v', b') = \delta(r, v, b) \tag{3.27}$$

Le facteur multiplicatif δ dépend de la position et de l'orientation de la source par rapport aux deux positions successives de cette source et à celles de la caméra, et varie donc de pixel en pixel. Si l'image contient n pixels, on obtient le modèle d'illumination de l'équation 3.28 ci-après :

$$\begin{pmatrix} r'_1 & v'_1 & b'_1 \\ r'_2 & v'_2 & b'_2 \\ \vdots & \vdots & \vdots \\ r'_n & v'_n & b'_n \end{pmatrix} = \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_n \end{pmatrix} \begin{pmatrix} r_1 & v_1 & b_1 \\ r_2 & v_2 & b_2 \\ \vdots & \vdots & \vdots \\ r_n & v_n & b_n \end{pmatrix} \tag{3.28}$$

Le modèle de transformation entre deux images est composé d'un coefficient multiplicatif pour chaque pixel de l'image. Cette solution constitue bien sûr une approximation de la vérité.

3.3.1.3 Normalisation de l'image

Normaliser l'image est une solution pour rendre celle-ci indépendante d'un modèle d'illumination. C'est un pré-traitement qui s'avère essentiel si les descripteurs mis en place ensuite ne permettent pas eux-mêmes d'intégrer le modèle.

Considérons toutes les images qui peuvent se déduire d'une image donnée I par un modèle de transformation tel que ceux venant d'être étudiés. Les images ainsi dérivées sont dites similaires les unes par rapport aux autres à travers le modèle d'illumination considéré. Cette notion de similarité définit une relation d'équivalence entre toutes ces images ainsi qu'une classe d'équivalence que nous appellerons $\mathcal{C}(I)$. Normaliser l'image I consiste donc à caractériser la classe $\mathcal{C}(I)$ par une image représentative notée $\overline{I}$. Cette image normalisée doit être indépendante du modèle de transformation, c'est à dire qu'elle doit pouvoir être obtenue à partir de n'importe quelle image de $\mathcal{C}(I)$. Considérons deux images I_1 et I_2 différant d'un changement d'illumination et leurs classes d'équivalence respectives $\mathcal{C}(I_1)$ et $\mathcal{C}(I_2)$ caractérisées par $\overline{I}_1$ et $\overline{I}_2$. Si les deux images $\overline{I}_1$ et $\overline{I}_2$ sont identiques, cela signifie que les deux images non normalisées I_1 et I_2 proviennent de la même classe $\mathcal{C}(I_1)$=$\mathcal{C}(I_2)$. Elles seront par conséquent considérées comme similaires à travers le modèle d'illumination.

Le choix de la méthode de normalisation est bien sûr fonction du modèle d'illumination considéré. Ainsi, pour un changement interne de la source lumineuse modélisé par le modèle diagonal de Finlayson, chaque canal doit être normalisé indépendamment. Pour éliminer les trois paramètres du modèle, il est possible de diviser chaque niveau de gris de l'image par la moyenne des niveaux de gris du plan correspondant, ou encore de ramener à 1 la norme du vecteur constitué des niveaux de gris de chaque plan. En ce qui concerne les déplacements de la source, il faut en revanche considérer les trois plans simultanément. On peut par exemple multiplier la couleur de chaque pixel par un facteur de telle sorte qu'on obtienne $r + v + b = 1$. Dans ces conditions, le problème majeur consiste à réaliser une normalisation conjointe pour les deux modèles à la fois. Finlayson y apporte une solution en développant une méthode itérative présentée ci-après :

Méthode itérative de Finlayson

Finlayson a mis en place une méthode itérative pour normaliser l'image face à des changements d'illumination internes et externes. Chaque itération comporte deux étapes. Dans la première, l'image est normalisée pour le changement externe, et dans la deuxième, le résultat de la première étape est normalisé pour les changements internes. Le processus est assez rapide et converge en quelques itérations. Soient $(r_i^{(p)}, v_i^{(p)}, b_i^{(p)})$ un pixel de l'image

à l'étape p et n le nombre de pixels à traiter. La valeur du pixel à l'itération $p+1$ est :

$$\begin{cases} r_i^{(p+1)} = \frac{n r_i^{(p)}}{r_i^{(p)}+v_i^{(p)}+b_i^{(p)}} \left(\sum_{j=1}^{n} \frac{r_j^{(p)}}{r_j^{(p)}+v_j^{(p)}+b_j^{(p)}} \right)^{-1} \\ v_i^{(p+1)} = \frac{n v_i^{(p)}}{r_i^{(p)}+v_i^{(p)}+b_i^{(p)}} \left(\sum_{j=1}^{n} \frac{v_j^{(p)}}{r_j^{(p)}+v_j^{(p)}+b_j^{(p)}} \right)^{-1} \\ b_i^{(p+1)} = \frac{n b_i^{(p)}}{r_i^{(p)}+v_i^{(p)}+b_i^{(p)}} \left(\sum_{j=1}^{n} \frac{b_j^{(p)}}{r_j^{(p)}+v_j^{(p)}+b_j^{(p)}} \right)^{-1} \end{cases} \quad (3.29)$$

Normaliser à l'itération p les niveaux de gris de chaque plan par l'expression $r_i^{(p)} + v_i^{(p)} + b_i^{(p)}$ permet de s'affranchir des paramètres δ_i du modèle de changements internes de l'équation 3.28. Le résultat obtenu pour chaque pixel est ensuite divisé par la moyenne des niveaux de gris ainsi calculés sur toute l'image. Le processus itératif permet de converger vers une solution indépendante des trois paramètres du modèle diagonal préconisé par Finlayson. Une méthode directe, démontrée dans [Gros et al., 1997b], permet d'arriver au même résultat, mais au prix de calculs assez complexes et plutôt instables.

Nous avons présenté dans cette partie les approximations des divers changements d'illumination rencontrées en Vision par Ordinateur. Nous avons également vu comment ces modèles pouvaient être intégrés dans un processus de normalisation de l'image. Nous allons maintenant passer en revue les différentes méthodes de caractérisation, globales puis locales. La plupart d'entre elles intègrent plus ou moins bien ces modèles d'illumination.

3.3.2 Les approches globales

La majeure partie des descriptions couleur existantes sont basées sur une approche globale de l'image, qui est alors décrite au moyen d'un unique vecteur de caractéristiques. Les travaux réalisés jusqu'à présent dans ce contexte voient l'image comme une distribution aléatoire des couleurs et la décrivent donc par des critères stochastiques. Les histogrammes et les moments sont les critères les plus fréquemment utilisés. Chacun d'eux fait l'objet d'une des sections suivantes. Nous présentons également dans une troisième section les quelques autres approches que nous avons rencontrées. Enfin toutes ces méthodes sont discutées dans la dernière partie.

3.3.2.1 Les histogrammes

Les méthodes d'indexation à partir de données couleur ont été introduites par [Swain et Ballard, 1991]. Il est possible de caractériser une image en calculant l'histogramme couleur qui lui est associé. La méthode consiste à comptabiliser le nombre de fois qu'une couleur apparaît dans l'image. En pratique, le nombre de couleurs étant très grand, il faut discrétiser l'espace des couleurs en les regroupant en "paquets". Plus formellement, si l'espace couleur discret utilisé est composé de n couleurs, alors l'histogramme couleur $\mathcal{H}$ d'une image I contenant N pixels est un vecteur $(h_{c_1}, h_{c_2}, ..., h_{c_n})$ où chaque élément h_{c_i} représente le nombre de pixels de couleur c_i dans I. Les éléments h_{c_i} doivent respecter la contrainte $\sum_{i=1}^{n} h_{c_i} = N$. D'autres processus d'indexation portent sur les histogrammes

cumulatifs [Stricker et Orengo, 1995]. Il faut ensuite mettre en place une mesure de similarité pour comparer les histogrammes construits. Les méthodes de comparaison existantes sont décrites dans la section 5.1.3 du chapitre sur la mise en correspondance.

Le principe est intéressant car les histogrammes sont invariants à la translation et à la rotation image, varient très peu au changement de point de vue et au changement d'échelle, sont également très peu sensibles aux occultations et enfin n'imposent pas nécessairement de travailler avec des images d'une scène rigide (comme par exemple un vêtement disposé de différentes manières sur plusieurs photos) [Swain et Ballard, 1991].

Clustering

La principale difficulté rencontrée lors de la construction des histogrammes couleur est bien sûr la discrétisation des couleurs. Le regroupement ("clustering" en anglais) des couleurs en "paquets" constitue une forme grossière de regroupement des distributions de couleurs en un ensemble plus restreint. Une réduction trop importante du nombre de couleurs utilisées mène à des performances d'indexation très instables [Zhang et al., 1995a]. Dans ce contexte, on rencontre différentes méthodes de regroupement : certains auteurs utilisent une table de référence des couleurs [Mehtre et al., 1995]. Elle contient le meilleur ensemble de couleurs pour toutes les images de la base, les couleurs de chaque image étant ensuite réduites à cet espace.

D'autres comme [Kankanhalli et al., 1996] préfèrent effectuer le regroupement indépendamment pour chaque image, de telle sorte que chaque image est représentée par sa propre table des couleurs. Le travail d'appariement de deux images nécessite alors en premier lieu de mettre en correspondance les tables de couleurs associées à chaque image.

D'autres encore, comme [Drew et al., 1998], transforment au préalable l'image en coordonnées chromatiques {r,v}[4] pour ramener l'espace 3D des couleurs à une description 2D et ainsi réduire la taille de l'histogramme, qui peut même être alors compressé efficacement [Li et al., 1999].

Sensibilité aux changements d'illumination

Depuis Swain, des travaux ont été réalisés pour pallier à certains inconvénients des histogrammes couleur, comme par exemple leur sensibilité aux changements de luminosité. Dans [Funt et Finlayson, 1995], Funt et Finlayson y apportent une solution appelée méthode CCCI pour "Color Constant Color Indexing", sous l'hypothèse du modèle d'illumination diagonal. Ils calculent des histogrammes de rapports de couleur : l'indexation n'est plus réalisée sur la couleur elle-même mais sur les dérivées du logarithme des couleurs, qui traduit en fait un rapport de couleurs uniquement fonction des albédos et donc indépendant de l'illumination. Les auteurs notent que ceci revient à effectuer de manière grossière une détection de contour sur le logarithme de l'image. La méthode s'avère meilleure que celle de Swain sur des images ayant subi un changement d'illumination, mais cependant moins bonne lorsque l'éclairage reste constant. En outre, les rapports de couleur utilisés sont très sensibles au bruit pour les intensités faibles.

[4]Coordonnées chromatiques : $r = \frac{R}{R+V+B}$, $v = \frac{V}{R+V+B}$ et $b = 1 - r - v$.

Comment tenir compte de l'information spatiale ?

Plus généralement, une critique attribuée aux histogrammes est la perte de la distribution spatiale des niveaux de gris. Mieux vaut leur associer d'autres méthodes d'indexation de plus haut niveau, comme l'indexation de formes par exemple. Ils semblent néanmoins être une bonne approche comme premier filtre (simple et rapide) dans la base, avant d'envisager des méthodes plus coûteuses. Quelques travaux ont tout de même été réalisés pour pallier ce problème, citons notamment [Zhang et al., 1995a; Smith et Chang, 1996; Pass et Zabih, 1996]. Ils ont tous en commun le découpage de l'image en régions pour lesquelles un histogramme couleur est calculé.

Un autre type d'approche plus récent consiste à enrichir l'histogramme d'une certaine information structurelle. Cette approche a été choisie dans [Huang et al., 1999] où la notion de *corrélogramme couleur* est mise en place. Les auteurs définissent un corrélogramme comme étant un histogramme étendu, qui exprime la façon dont la corrélation spatiale des couleurs évolue avec la distance. Sa formulation est donnée à l'équation 3.30 :

$$\gamma_{c_i,c_j}^{(k)}(I) = \Pr_{p_1 \in I \ / \ I(p_1)=c_i} [p_2 \in I \ / \ I(p_2) = c_j \text{ et } |p_1 - p_2| = k] \tag{3.30}$$

Pour chaque pixel de l'image de couleur c_i, le corrélogramme donne la probabilité que le pixel situé à la distance k du pixel considéré soit de couleur c_j. L'efficacité de cette structure dépend fortement du choix de la distance, qui doit être suffisamment grande pour que la caractérisation soit pertinente, mais cependant raisonnable pour garder des temps de calcul et des coûts de stockage exploitables. Les auteurs développent à ce propos une méthode rendant le calcul du corrélogramme plus efficace.

Une autre méthode développée par Siggelkow et Schael dans [Siggelkow et Schael, 1999] rend également l'histogramme couleur plus discriminant en y introduisant une information de voisinage. L'approche est basée sur la méthode de calcul des invariants par généralisation et contrainte de [Schulz-Mirbach, 1994]. Mirbach avait proposé de calculer la moyenne des valeurs de niveau de gris d'une image sur une orbite associée au groupe des transformations euclidiennes de l'image. Pour une translation (t_0, t_1) et une rotation d'angle φ, il formule la solution sous forme d'intégrale, calculée sur toute l'image $I_{(M \times N)}$ et de la forme :

$$A_f(I) = \frac{1}{2\pi NM} \int_{t_0=0}^{N} \int_{t_1=0}^{M} \int_{\varphi=0}^{2\pi} f(g(I)) d\varphi dt_1 dt_0 \tag{3.31}$$

où f représente une fonction locale qui dépend uniquement des niveaux de gris de I et g une transformation euclidienne de l'image fonction des paramètres (t_0, t_1) et φ. L'information de voisinage que l'on souhaite voir apparaître dans l'invariant dépend du choix de f. Par exemple, prendre $f(I) = I(0,0)$ revient simplement à calculer la moyenne des niveaux de gris de l'image, alors que $f(I) = I(0,0).I(1,0)$ permet en plus de faire entrer dans la moyenne les niveaux de gris des voisins du point (t_0, t_1) situés sur un cercle de rayon 1 autour du point. L'invariant obtenu est représenté ci-après à l'équation 3.32.

$$A_f(I) = \frac{1}{2\pi NM} \int_{t_0=0}^{N} \int_{t_1=0}^{M} I(t_0, t_1) \int_{\varphi=0}^{2\pi} I(\cos\varphi + t_0, \sin\varphi + t_1) d\varphi dt_1 dt_0 \tag{3.32}$$

Dans le cas discret, t_0 et t_1 sont des entiers et l'angle φ est choisi selon un pas de

discrétisation, les niveaux de gris des points inter-pixels ainsi calculés étant estimés par interpolation bi-linéaire.

Cette approche permet de tenir compte d'une certaine information locale, mais le choix de caractériser l'image par une moyenne ne permet pas de traiter le problème des occultations ou des changements d'arrière-plan par exemple. C'est pourquoi Siggelkow et Schael proposent de remplacer cette simple moyenne par un histogramme couleur au sens de Swain, afin de mieux préserver l'information structurelle locale. Cet histogramme est calculé sur toutes les valeurs $f(g(I))$ de l'équation 3.31. Si $f(I) = I(0,0)$, alors il correspond à l'histogramme couleur de Swain. Pour comprendre les avantages de cette variante, considérons les textures de la figure 3.2, qui sont composées de 2 couleurs : blanc (B) et gris (G).

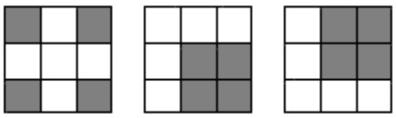


Figure 3.2 – *Textures binaires caractérisées par le même histogramme couleur* $\mathcal{H} = (5_B, 4_G)$.

Ces trois textures possèdent le même histogramme couleur $\mathcal{H}$ et ne sont donc pas caractérisables avec la méthode classique de Swain. En revanche, si l'on considère la méthode de Siggelkow et Schael avec $f(I) = I(0,0).I(1,0)$ (voir l'équation 3.32), alors on obtient le même histogramme pour T_2 et T_2' et un histogramme différent pour T_1. Par exemple, si le pas de discrétisation de l'angle φ a été choisi de façon à ne considérer que les voisins directs de chaque pixel en connexité-8, alors on obtient :

$$\left| \begin{array}{lcl} \mathcal{H}^*(T_1) & = & (0_{GG}, 16_{BB}, 24_{BG}) \\ \mathcal{H}^*(T_2) = \mathcal{H}^*(T_2') & = & (12_{GG}, 10_{BB}, 18_{BG}) \end{array} \right. \tag{3.33}$$

La méthode permet ainsi de conserver les caractéristiques locales des points, tout en étant invariante à la translation et à la rotation. Elle peut encore être affinée en combinant plusieurs histogrammes calculés pour différents noyaux f. Remarquons enfin que cette approche s'avère plus lourde à mettre en œuvre que les histogrammes couleur classiques. Une méthode est également développée dans [Siggelkow et Schael, 1999] pour en réduire la complexité.

Il faut tout de même bien avoir à l'esprit que l'indexation à partir des histogrammes en général est très coûteuse en temps de calcul, les histogrammes devant avoir une certaine taille pour être suffisamment discriminants et les bases d'images devenant de plus en plus volumineuses. Cet inconvénient est une des raisons pour lesquelles certains auteurs ont orienté leurs travaux vers des critères stochastiques moins "gourmands", comme les moments. Cette autre approche fait l'objet de la prochaine section.

3.3.2.2 Les moments

Si l'on considère la distribution des couleurs d'une image en terme de mélange probabiliste, alors il vient naturellement à l'idée, après les histogrammes, d'utiliser les moments statistiques pour indexer des images. C'est en premier lieu un bon moyen d'échapper à la discrétisation de l'espace des couleurs, propre aux histogrammes.

Dans la section précédente, nous avons présenté les travaux de Stricker et Orengo [Stricker et Orengo, 1995] à propos des histogrammes couleur. Ceux-ci proposent également une méthode de caractérisation basée sur les moments de la distribution couleur. Ils utilisent jusqu'au troisième ordre les moments centrés de chaque plan couleur, ceux-ci étant décrits dans l'espace TSI. Le premier moment correspond à la moyenne, le second à la variance et le troisième à la distorsion des niveaux de gris. Ils montrent alors que la combinaison pondérée de ces mesures, associée à une fonction de similarité, donnent de meilleurs résultats d'indexation qu'avec les histogrammes cumulés, tout en étant plus efficace, puisque les index ne contiennent que les éléments essentiels de la distribution couleur.

Healey et Slater dans [Healey et Slater, 1994] caractérisent eux aussi l'image à partir de moments issus de la distribution des couleurs. Ils approchent la réflectance spectrale de la surface par un modèle linéaire fini (cf. la section 3.3.1.1), et en déduisent que les histogrammes couleur d'un même objet exposé à deux sources lumineuses différentes sont reliés par une transformation affine de leurs coordonnées. L'espace ainsi défini est alors décrit par les valeurs propres de la matrice des moments centrés calculés sur ces histogrammes et que l'on doit à [Taubin et Cooper, 1992]. La caractérisation obtenue est invariante aux transformations linéaires des coordonnées et aux changements d'illumination modélisés par un modèle linéaire (cf. la section 3.3.1.1). Elle apparaît plus robuste que la méthode CCCI basée sur les histogrammes (décrite à la section précédente), surtout face aux niveaux de gris de faible intensité. Elle rend aussi le processus d'indexation plus efficace puisque seulement six valeurs de moments doivent être passées en revue pour chaque image. On lui reprochera quand même d'être plus "globale" que les méthodes à base d'histogrammes, et donc de moins bien se comporter en présence de faibles occultations par exemple.

Une méthode de reconnaissance de texture couleur est également développée dans [Wang et Healey, 1998] à partir des moments de Zernike [Teague, 1980]. La méthode de caractérisation obtenue est invariante à la rotation, au changement d'échelle et aux variations d'illumination.

Dans [Mindru et al., 1999], Mindru, Moons et Van Gool définissent la notion de moments couleur généralisés. Étant donnés une image I telle que $I(x,y) = (R(x,y), V(x,y), B(x,y))$ représente le vecteur des niveaux de gris du pixel (x,y) de I et Ω une région appartenant à I, le moment couleur généralisé de Ω d'ordre $p+q$ et de degré $a+b+c$ est tel que :

$$M_{pq}^{abc} = \int\int_{\Omega} x^p y^q R^a(x,y) V^b(x,y) B^c(x,y) dxdy \tag{3.34}$$

Remarquons par exemple que les moments couleur généralisés M_{pq}^{000} correspondent aux moments d'ordre $p+q$ de la région Ω, et que les moments M_{00}^{abc} représentent les moments non centrés d'ordre $a+b+c$ de la distribution couleur des valeurs RVB de Ω.

En considérant des combinaisons des moments ainsi définis, jusqu'au premier ordre et

au deuxième degré seulement, Mindru, Moons et Van Gool ont mis en place des invariants aux transformations affines de l'image et aux changements d'illumination selon le modèle diagonal avec translation. Ces invariants sont répartis en trois classes, selon le nombre de plans couleur impliqués. On obtient par exemple pour chacun des plans couleur les deux invariants suivants :

$$\begin{cases} B_{02} = \frac{M^2_{00}\ M^0_{00}}{(M^1_{00})^2} \\ B_{12} = \frac{M^2_{10}\ M^0_{01}\ M^1_{00} + M^1_{10}\ M^2_{01}\ M^0_{00} + M^0_{10}\ M^1_{01}\ M^2_{00} - M^2_{10}\ M^1_{01}\ M^0_{00} - M^1_{10}\ M^0_{01}\ M^2_{00} - M^0_{10}\ M^2_{01}\ M^1_{00}}{M^2_{00}\ M^1_{00}\ M^0_{00}} \end{cases} \tag{3.35}$$

où M^i_{pq} représente M^{i00}_{pq}, M^{0i0}_{pq} ou M^{00i}_{pq} selon le plan considéré. Les invariants impliquant deux plans couleur à la fois sont au nombre de 10 et il en existe 21 pour les trois plans [Mindru et al., 1999].

Les auteurs comparent ensuite leur approche avec celle de Reiss [Reiss, 1993] qui utilise des moments invariants aux mêmes transformations de l'image mais calculés jusqu'au quatrième ordre. Ils montrent que l'approche couleur permet d'améliorer nettement le processus de reconnaissance et que les invariants sont plus robustes aux occultations, ce qui faisait jusqu'alors défaut aux caractérisations basées sur les moments classiques par rapport à celles basées sur les histogrammes.

Comme toutes les descriptions basées sur la distribution des couleurs, les méthodes utilisant les moments ont le principal inconvénient de ne pas tenir compte de l'information structurelle contenue dans la zone de l'image étudiée. De même que pour les histogrammes, elles peuvent néanmoins servir de premier filtre pour réduire le nombre d'appariements potentiels durant le processus d'appariement. Citons tout de même les travaux récents de Healey et Slater [Healey et Slater, 1997] qui améliorent leur méthode de caractérisation [Healey et Slater, 1994] en introduisant des filtres spatiaux dans le processus d'indexation. Les six invariants sont calculés sur l'image de référence, mais aussi sur l'image obtenue par exemple par différence de filtres gaussiens. L'ensemble des $6n$ invariants devient plus discriminant et permet ainsi de différencier plusieurs textures possédant la même distribution de couleurs, l'information capturée par chacun des sextuples d'invariants étant différente. Le choix des filtres à employer dépend des applications envisagées. Pour obtenir l'invariance à la rotation, il convient d'utiliser un filtre qui soit invariant à la rotation, la magnitude du gradient de l'image peut être utilisée par exemple.

D'autres méthodes de description basées sur les moments ont été développées encore plus récemment, comme par exemple dans [Thai et Healey, 1998] où la description est faite à partir de moments complexes calculés sur des images couleur traitées au préalable par des filtres de Gabor. Nous ne détaillerons pas ces approches ici car elles s'appliquent à des cas particuliers d'images, des textures périodiques plus précisément, et ne peuvent donc guère être exploitées pour l'appariement de points d'intérêt extraits d'images quelconques.

3.3.2.3 Autres approches

Nous présentons dans cette section trois autres approches permettant de caractériser une image en couleur. Contrairement aux histogrammes et aux moments qui viennent d'être passés en revue, ces méthodes ne font pas ici l'objet d'une section à part entière,

car ce sont des techniques moins classiques dans le domaine de l'indexation ou de la reconnaissance. Elles nous ont cependant semblé suffisamment pertinentes pour être référencées dans cette étude.

La première approche applique à la couleur la classique mesure de corrélation[5] : Kondepudy et Healey proposent dans [Kondepudy et Healey, 1994] de caractériser une texture couleur 3D par six mesures de corrélation : les trois fonctions d'autocorrélation C_{RR}, C_{VV} et C_{BB} ainsi que les corrélations croisées C_{RV}, C_{RB} et C_{VB} entre les trois plans couleur. Cela suppose bien sûr que les bandes couleur ne varient pas les unes par rapport aux autres. La caractérisation obtenue est invariante aux transformations euclidiennes de l'image et peut facilement être rendue invariantes aux changements affines de luminosité.

Dans [Finlayson et al., 1996], Finlayson, Chatterjee et Funt proposent quant à eux une méthode basée sur des mesures d'angles. Cela donne une caractérisation invariante aux transformations euclidiennes de l'image et aux changements d'illumination selon le modèle diagonal généralisé de Finlayson (défini à l'équation 3.25). Ils indexent les objets par les angles de la distribution couleur, c'est à dire les angles formés par chacun des plans couleur représentés sous forme de vecteurs et pris deux à deux. En effet, seules les normes de ces vecteurs varient avec l'illumination, leur direction restant constante. Ils ajoutent également à cette caractérisation les angles de couleurs entre les distributions couleur des contours. La méthode est comparée à l'approche CCCI utilisant les histogrammes et décrite à la section 3.3.2.1 ainsi qu'à celle de Healey et Slater fondée sur les moments et présentée à la section 3.3.2.2. Les résultats obtenus semblent privilégier l'approche basée sur les angles, la méthode de Healey et Slater étant la moins satisfaisante. Les auteurs attribuent ce résultat au fait que les moments extraient une information beaucoup plus globale que les histogrammes ou les angles.

Comme nous l'avons vu au long de ce chapitre, de nombreuses approches modélisent la distribution des couleurs en terme de mélange probabiliste, la plupart du temps sous forme d'histogrammes et de moments. Une autre alternative basée sur le même principe a été développée très récemment dans [Biernacki et Mohr, 1999], où la distribution est envisagée sous forme de mélange gaussien des couleurs résistant aux changements de luminosité. La qualité des appariements obtenue est du même ordre qu'avec les histogrammes. Mais le principal avantage de la méthode est qu'elle permet un gain important dans la taille de l'index. En effet, le nombre de classes partitionnant l'espace des couleurs est déterminé automatiquement pour chaque image, contrairement aux méthodes basées sur les histogrammes où il est imposé a priori. L'inconvénient majeur réside dans le temps d'indexation qui est plutôt long.

3.3.2.4 Discussion

Toutes les techniques que nous venons de décrire considèrent l'image globalement. La majeure partie d'entre elles se proclament invariantes à la translation, à la rotation et au changement d'illumination. Bien que l'on puisse trouver des tests expérimentaux montrant les bons résultats de chacune de ces méthodes par rapport aux autres, on ne trouve pas de résultats synthétiques les comparant toutes sur un jeu unique de données. Il est donc

[5] Consulter la table 5.1 pour une revue des principales mesures de corrélation existantes.

difficile de tirer des conclusions quant à leur pertinence. En dépit de cela, quelques points importants ressortent :

Les descriptions basées sur les moments permettent de s'affranchir de la discrétisation de l'espace des couleurs. Elles permettent également une représentation dans un espace de moindre dimension, contrairement aux histogrammes qui nécessitent l'emploi de techniques d'appariement de forte complexité, comme nous le verrons au prochain chapitre. En revanche, les approches basées sur les histogrammes ou les angles semblent moins sensibles aux occultations que celles utilisant les moments, bien que des efforts aient été réalisés récemment pour pallier ce problème [Mindru et al., 1999].

Plus généralement, le principal avantage des descriptions globales est qu'elles incluent la notion d'invariance aux changements d'illumination. En revanche, un point faible pour la plupart est de ne pas assez tenir compte de la distribution spatiale de l'information couleur. Elles peuvent néanmoins s'avérer être une bonne approche comme premier filtre dans la base d'images, avant d'envisager des méthodes mettant en jeu des descriptions plus locales ou des primitives de plus haut niveau.

La question que nous nous posons alors dans le contexte de notre étude est la suivante : peut-on utiliser les méthodes globales pour caractériser les points d'intérêt ? Cela revient à les adapter de manière à les rendre *locales*. Cela semble difficile, puisque toutes les méthodes reposent sur la comparaison de distributions de couleurs et sont donc basées sur des mesures statistiques. Or lorsqu'elles sont envisagées localement, ces distributions ne contiennent qu'un faible nombre de pixels. Il n'est pas assuré que les méthodes globales puissent fonctionner correctement avec de si petites populations. D'autre part, les techniques développées sont la plupart du temps peu robustes aux occultations importantes. Ceci constitue un inconvénient non négligeable dans le cas de la caractérisation de points d'intérêt où le fond peut compter pour une part importante dans la distribution. En tout cas, il serait au moins utile de mener une étude sur la dimension minimale que doit avoir la distribution locale pour que ces méthodes restent efficaces.

Une autre solution consiste à mettre en place des techniques de caractérisation spécifiques à des populations couleur locales. Jusqu'à présent, très peu de travaux ont suivi cette voie. Les quelques méthodes rencontrées sont passées en revue dans la prochaine section.

3.3.3 Les approches locales

Très peu de travaux ont été réalisés sur la caractérisation locale de points d'intérêt à partir de l'information couleur. Nous avons seulement trouvé quelques implémentations locales des histogrammes couleur. Ainsi dans [Ennesser et Medioni, 1995], des histogrammes locaux sont construits sur plusieurs zones de l'image de tailles variables puis intersectés avec l'histogramme de l'objet recherché. Les auteurs montrent que cette approche donne de meilleurs résultats qu'avec la méthode classique d'indexation de Swain et Ballard. Il existe aussi l'approche de Zhang [Zhang et al., 1995a] qui segmente l'image en régions et calcule pour chacune d'elles un histogramme.

Cependant ces deux approches ne sont guère satisfaisantes puisque, une fois encore, elles utilisent des méthodes statistiques qui demeurent suspectes pour des petites populations. Restent alors deux directions qui nous semblent intéressantes. La première exploite le

paradigme du hachage géométrique dans le cadre de l'invariance aux changements d'illumination ; elle est détaillée dans la prochaine section. En ce qui concerne la seconde direction, nous proposons une approche qui étend à la couleur la technique basée sur les invariants différentiels de Hilbert, présentée à la section 3.2.1.3. Cette étude est l'objet du chapitre 4.

3.3.3.1 Hachage des couleurs

Schwartz, Wolfson et Lamdan ont développé une technique d'indexation utilisant le paradigme du hachage géométrique de Wolfson[6]. Cette technique s'applique sur des ensembles de points, sous différentes transformations géométriques. L'idée maîtresse consiste à stocker dans la table des informations géométriques indépendantes de la position, de l'orientation et de l'échelle des objets.

Si l'on se place dans le contexte de l'invariance aux changements d'illumination, il est alors possible d'étendre le principe du hachage à la couleur. En effet, considérons le modèle diagonal comme modèle d'illumination. Ce modèle traduit que le rapport des couleurs en deux points est invariant aux changements d'illumination. Étant donné un point de référence, les couleurs de tous les autres points peuvent être définies en terme de coordonnées de couleur par rapport à la couleur de ce point de référence. Les valeurs du rapport des couleurs peuvent donc être calculées pour tout point j, dès qu'un point de base i a été fixé. Une table de hachage peut alors être construite en prenant pour index ces valeurs. Chaque entrée de la table pointe sur toutes les paires de points qui ont les rapports de couleur indexés. Il est donc parfaitement possible de caractériser un point d'intérêt et son voisinage local par une table de hachage des couleurs. Il suffira ensuite, lors de l'appariement, de calculer les rapports de couleur du voisinage associé au point que l'on veut apparier, puis de passer en revue la table et d'ajouter un vote dès qu'un rapport de couleurs similaire est rencontré. Le correspondant potentiel du point sera alors celui dont le voisinage possède le plus grand nombre de votes.

L'approche est intéressante sur plusieurs points :

- Elle n'utilise aucun outil statistique, donc est bien adaptée aux petites populations ;
- Elle est invariante aux changements d'illumination dans la mesure où le modèle diagonal est valide ;
- Elle est particulièrement robuste au bruit et aux occultations, de part la grande quantité de votes impliqués ;
- L'algorithme est facilement parallélisable [Wolfson et Rigoutsos, 1997] ;
- Elle n'est basée sur aucune information géométrique, donc doit être robuste aux translations et aux rotations.

Son inconvénient majeur réside dans le coût important qu'impose la construction de la table ; il est en effet nécessaire de considérer tous les couples de points du voisinage des points d'intérêt considérés. Elle semble donc difficilement exploitable pour mettre en correspondance un grand nombre de points entre images.

[6] Consulter [Wolfson et Rigoutsos, 1997] pour un état de l'art.

3.4 Conclusion

Ce chapitre a présenté un état de l'art des méthodes de caractérisation locale de points d'intérêt, en noir et blanc comme en couleur. La plus grande partie des approches rencontrées sont dédiées aux images monochromes et travaillent tant dans le domaine spatial que fréquentiel. De notre étude ressort l'approche basée sur les invariants différentiels de Hilbert. Cette méthode de caractérisation en niveau de gris est invariante aux transformations euclidiennes de l'image, robuste au changement d'échelle si elle est envisagée dans un contexte multi-échelle, peut être rendue invariante à certaines transformations des niveaux de gris et enfin se comporte relativement bien dans le cas de changement de point de vue. Elle a été validée récemment par Schmid pour l'appariement de points entre images, avec l'indexation d'images comme objectif.

Plusieurs méthodes dédiées à la couleur ont également été passées en revue. Elles sont malheureusement pour la plupart peu adaptées à la caractérisation locale de points, car essentiellement développées pour décrire l'image dans son intégralité, l'application visée le plus souvent étant l'indexation d'images. Dans ce contexte, l'essentiel des travaux menés sur la couleur depuis ces dix dernières années a été concentré sur la constance des couleurs, c'est-à-dire la capacité à rendre la caractérisation invariante aux changements d'illumination. Les méthodes d'ailleurs développées sont pour la plupart pertinentes. Certaines d'entre elles sont même invariantes aux transformations euclidiennes de l'image. Malheureusement, elles semblent inadaptées à des approches plus "locales" comme la caractérisation de points d'intérêt. En effet, la plupart d'entre elles considère l'image comme une distribution aléatoire des couleurs et par conséquent la caractérise à l'aide d'outils statistiques, les histogrammes couleur par exemple. Il s'avérerait donc risqué de les employer pour des approches locales où la taille de la distribution aurait toutes les chances d'être trop faible.

N'ayant trouvé dans la littérature aucune méthode de caractérisation locale dédiée à la couleur réellement concluante, nous avons choisi d'orienter nos travaux dans cette voie. Utiliser sur chaque canal couleur les méthodes développées en niveau de gris est la première idée qui vient à l'esprit. Dans ce contexte, le choix de la caractérisation basée sur les invariants différentiels de Hilbert nous semble judicieux, d'une part parce que celle-ci a fait ses preuves en noir et blanc, mais surtout parce qu'elle offre la possibilité d'exploiter l'information inter-canal, alors que les autres approches se contentent de multiplier la caractérisation par le nombre de plans. Cette étude fait l'objet du prochain chapitre. Nous montrons comment il est possible de mettre en évidence des informations spécifiques à la couleur, en plus bien sûr des invariants existants pour chaque plan couleur.

Chapitre 4

Les invariants différentiels couleur

Les méthodes de caractérisation qui viennent d'être présentées ne s'avèrent pas concluantes pour la caractérisation de points d'intérêt. Nous présentons donc dans ce chapitre une nouvelle approche de caractérisation locale, qui exploite l'information couleur. La méthode est basée sur les invariants différentiels de Hilbert. Elle permet de mettre en place une caractérisation invariante à la rotation image. Nous montrons que l'information couleur permet de simplifier considérablement la mise en place de cette caractérisation, tout en la rendant plus robuste au bruit. Une technique issue du problème de la constance des couleurs est également présentée pour rendre cette approche invariante aux changements d'illumination.

Sommaire

Dans ce chapitre, nous étendons à la couleur les invariants différentiels de Hilbert jusqu'alors utilisés en niveau de gris. Cette technique a été présentée à la section 3.2.1.3 du chapitre précédent. Elle permet de caractériser localement le point d'intérêt en exploitant l'information contenue dans les niveaux de gris de son voisinage et est invariante aux transformations orthogonales de l'image.

Les invariants ainsi définis sont bien sûr appliqués à chaque plan couleur. Nous montrons à la section 4.1 qu'il existe également deux invariants inter-canaux spécifiques à la couleur. Ces invariants supplémentaires permettent alors de mettre en place une caractérisation couleur robuste, qui est présentée à la section 4.2. Contrairement aux méthodes globales, les méthodes locales tiennent d'avantage compte de la distribution spatiale mais ont l'inconvénient de ne pas être invariantes aux changements d'illumination. Il faut donc mettre en œuvre une technique pour remédier à ce problème. Ceci fait l'objet de la section 4.3, dans laquelle plusieurs types de changements d'illumination sont passés en revue et plusieurs méthodes de normalisation sont proposées. La mise en place des invariants différentiels couleur ainsi que l'étude des changements d'illumination ont fait l'objet de trois publications [Montesinos et al., 1998; Gouet et al., 1998b; Montesinos et al., 2000].

4.1 Deux invariants spécifiques à la couleur

Considérons une image couleur {R,V,B} ainsi que le groupe des rotations spécifié par un paramètre (l'angle de la rotation). Il y a par plan couleur 5 dérivées jusqu'au deuxième ordre, sans compter l'intensité lumineuse. Les invariants sont donc dans ce cas au nombre de $(5+1)\times 3-1 = 17$. Nous avons comptabilisé 5 invariants différentiels en niveau de gris, soit $5 \times 3 = 15$ invariants pour la couleur. Le tableau 4.1 résume la dimension du vecteur invariant et le nombre de degrés de liberté associés en fonction des trois premiers ordres.

	Ordre 1	Ordre 2	Ordre 3
Dérivées	**2** *(x y)*	**3** *(xx xy yy)*	**4** *(xxx xxy xyy yyy)*
Degrés de liberté *(ndg + dérivées)*	**3** *(ndg + 2 dérivées)*	**6** *(+ 3 dérivées)*	**10** *(+ 4 dérivées)*
Invariants N&B *(les d.d.l. - 1)*	**2** *(3 d.d.l. - 1)*	**5** *(6 d.d.l. - 1)*	**9** *(10 d.d.l. - 1)*
Invariants couleur	**2**	-	-
Invariants N&B + couleur *(les d.d.l. par plan - 1)*	2×3+2=**8** *(3×3-1)*	5×3+2=**17** *(6×3-1)*	9×3+2=**29** *(10×3-1)*

Table 4.1 – *Cardinalité de l'ensemble des invariants différentiels à la rotation.*

Il existe donc deux invariants supplémentaires inter-canaux, à prendre dans l'ensemble suivant [Montesinos et al., 1998] :

$$\nabla R.\nabla V \quad \nabla R.\nabla B \quad \nabla V.\nabla B \tag{4.1}$$

Remarquons qu'ils n'impliquent que les dérivées du premier ordre.

4.2 Une caractérisation couleur robuste

Notre solution consiste à n'utiliser que les invariants de Hilbert à l'ordre 1. La caractérisation qui serait alors obtenue pour des images en niveau de gris 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 $\vec{v}_{col}$ et tel que :

$$\vec{v}_{col}(\vec{x}, \sigma) = \begin{pmatrix} R \\ \|\nabla R\|^2 \\ V \\ \|\nabla V\|^2 \\ B \\ \|\nabla B\|^2 \\ \nabla R.\nabla V \\ \nabla R.\nabla B \end{pmatrix} \tag{4.2}$$

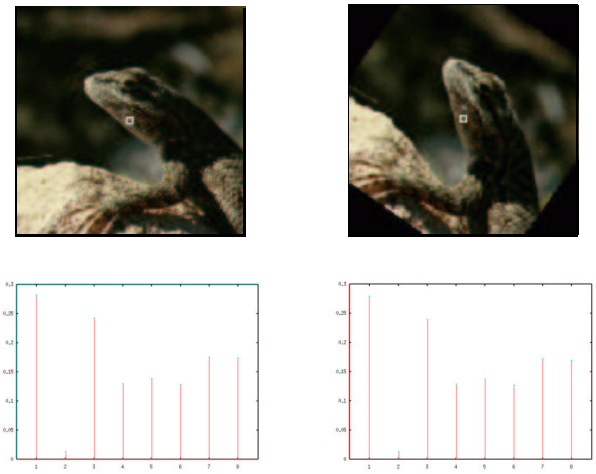
La simplification que nous permet l'information couleur présente deux avantages certains :

- La caractérisation obtenue est robuste au bruit, puisqu'elle n'implique que les dérivées de l'image d'ordre 1,
- La complexité de la méthode est faible, puisque seulement quatorze images sont nécessaires au calcul (R, R_x, R_y, $\mid \nabla R \mid$, V, V_x, V_y, $\mid \nabla R \mid$, B, B_x, B_y, $\mid \nabla R \mid$, $\nabla R.\nabla V$, $\nabla R.\nabla B$). L'utilisation des ordres supérieurs rendrait la méthode difficilement exploitable en terme d'occupation mémoire/disque, surtout si un filtrage sub-pixel est envisagé (cf. annexe C).

Il serait nécessaire d'aller jusqu'à l'ordre 3 en niveau de gris [Koenderink et Van Doorn, 1987; Salden et al., 1992; Schmid, 1996] pour obtenir un vecteur d'invariants de dimension similaire (avec neuf invariants). L'invariant "perdu" est compensé par la robustesse de la méthode. La figure 4.1 montre les vecteurs d'invariants de l'équation 4.2 calculés sur deux images différant d'une rotation image de 35° et obtenus pour le point central de l'image qui est aussi le centre de la rotation. Elle montre clairement l'invariance de la caractérisation à la rotation.

4.3 Changement d'illumination

La méthode de caractérisation mise en place ici tient compte de la distribution spatiale de l'information couleur mais n'est en revanche pas invariante aux changements d'illumination. Nous présentons dans les deux sections suivantes les méthodes mises en place pour pallier ces inconvénients, qu'il s'agisse de changements internes ou externes. La pertinence des approches présentées est démontrée sur plusieurs séries d'images en couleur différant de ces changements d'illumination.



Les 8 invariants couleur

Les 8 invariants couleur soumis à une rotation de 35°

Figure 4.1 – *Les deux vecteurs d'invariants couleurs obtenus pour le centre de la rotation.*

4.3.1 Changement interne

Le modèle d'illumination choisi pour les changements interne est le modèle diagonal enrichi d'une translation, qui a été présenté à l'équation 3.26. Il a été en effet jugé comme possédant le meilleur rapport qualité/complexité lorsque les images sont petites [Gros et al., 1997b], ce qui est notre cas puisque la caractérisation est locale.

A partir du modèle établi, deux directions sont envisageables pour rendre la caractérisation invariante. La première solution consiste à modifier le vecteur $\vec{v}_{rot}$ pour le rendre invariant à ce type de transformation, tel que l'a fait Schmid dans [Schmid, 1996] pour les invariants différentiels en niveau de gris. Le modèle d'illumination ayant six degrés de liberté, le vecteur tel que nous l'avons défini serait réduit à $8 - 6 = 2$ invariants (les angles inter-plans). Cette solution ne convient évidemment pas, un vecteur à deux composantes étant bien sûr trop pauvre. Il faudrait alors au moins considérer les invariants jusqu'à l'ordre 2. On obtiendrait dans ce cas $17 - 6 = 11$ invariants aux changements d'illumination (se reporter au tableau 4.1 pour le décompte des degrés de liberté). Cette solution n'apparaît pas non plus comme la plus optimale. En effet, l'utilisation des dérivées jusqu'au deuxième ordre affaiblit la caractérisation face au bruit, d'autant plus que la normalisation des invariants face aux changements d'illumination conduirait à calculer des rapports de dérivées (cf. section 3.2.1.4 du chapitre 3), rendant la caractérisation encore plus sensible au bruit.

La seconde direction consiste en une normalisation de l'image, tel que cela a été présenté à la section 3.3.1.3. Cette solution a l'avantage de prendre en compte les six paramètres du modèle sans appauvrir la caractérisation basée sur les invariants différentiels du premier ordre. De plus, les calculs sont dans ce cas réalisés uniquement à l'ordre 0 en tenant compte d'une information de voisinage, et sont ainsi supposés stables face au bruit.

4.3.1.1 Notre méthode de normalisation

Pour éliminer les six paramètres du modèle diagonal avec translation, la solution que nous avons adoptée [Gouet et al., 1998b; Montesinos et al., 2000] consiste à normaliser les trois canaux indépendamment les uns des autres. Pour chaque pixel (x, y) d'une image I, le niveau de gris normalisé associé au plan I_p est obtenu de la façon suivante, pour $\beta \neq \gamma$:

$$\overline{I_p(x,y)} = \frac{I_p(x,y) - \mathcal{F}^{\alpha}_{(a,b)\in I_p} I_p(a,b)}{\mathcal{F}^{\beta}_{(a,b)\in I_p} I_p(a,b) - \mathcal{F}^{\gamma}_{(a,b)\in I_p} I_p(a,b)} \tag{4.3}$$

où $\mathcal{F}^{i}_{x\in X} g(x)$ désigne la i^{eme} valeur de $g(x)$ dans l'ensemble X, pour des valeurs de i rationnelles et quantifiées entre 0 et 1. Par exemple, $i = 1$ implique le maximum de $g(x)$, $i = \frac{1}{2}$ sa valeur médiane, etc. Nous avons choisi dans nos expérimentations $\alpha = \frac{1}{2}$, $\beta = \frac{3}{4}$ et $\gamma = \frac{1}{4}$. L'équation 4.3 peut donc se réécrire :

$$\overline{I_p(x,y)} = \frac{I_p(x,y) - \mathrm{med}_{12}}{\mathrm{med}_{34} - \mathrm{med}_{14}} \tag{4.4}$$

Le choix $\alpha = \frac{1}{2}$ conduit à considérer la médiane de la distribution des niveaux de gris du plan I_p (notée med_{12}), et les choix $\beta = \frac{1}{4}$ et $\gamma = \frac{3}{4}$ impliquent les médianes respectives med_{34} et med_{14} des deux moitiés de distribution issues de I_p. Cette notion est illustrée par le schéma 4.2 ci-après. Réalisé pour chaque pixel et sur chaque canal, ce traitement permet d'obtenir une image indépendante des paramètres D et T du modèle d'illumination.

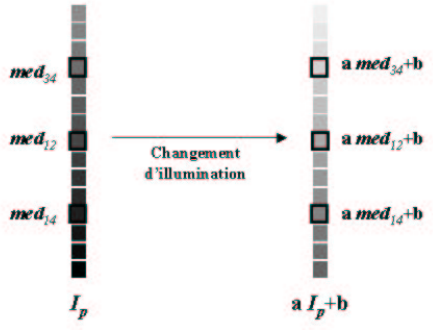


Figure 4.2 – *Comportement des médianes des niveaux de gris soumis à un changement d'illumination affine.*

Normalisation locale

Le processus de normalisation que nous venons de présenter a l'inconvénient d'être sensible aux changements de composition de l'image puisqu'il est appliqué globalement sur chaque plan. C'est par exemple le cas lorsque les deux images sont prises sous des points de vue différents. Nous proposons alors une variante plus locale pour pallier ce problème [Montesinos et al., 2000]. Pour chaque pixel du plan couleur considéré, les médianes sont calculées localement dans une fenêtre circulaire centrée sur le pixel étudié. Ainsi les propriétés locales des pixels sont préservées. Cette approche requiert la donnée d'un paramètre, le diamètre de la fenêtre à considérer : plus les images diffèrent, plus petit le paramètre doit être.

4.3.1.2 Exemples de normalisation

Les images utilisées normalisées tout au long de ce chapitre sont issues des séquences "Lézard", "Bureau" et "Bonzaï". Le lecteur peut consulter l'annexe B pour avoir des précisions sur la façon dont les images de référence ont été acquises.

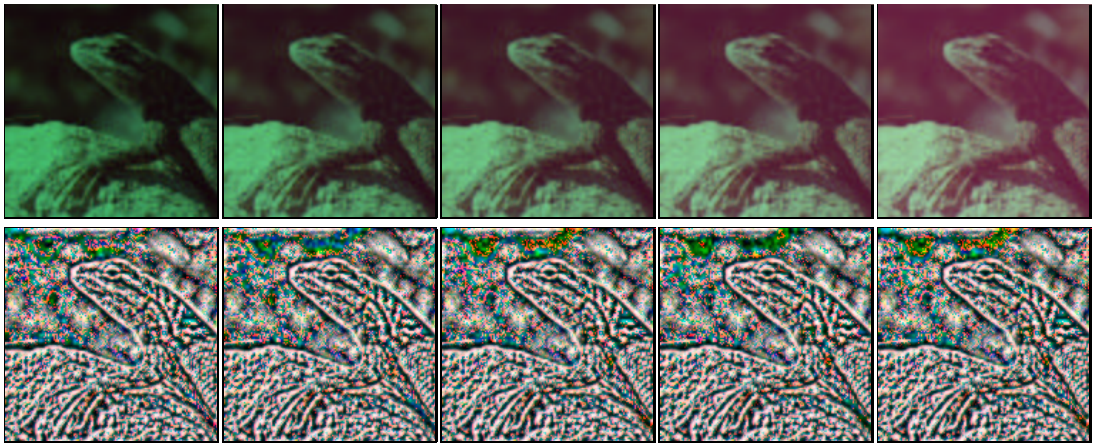
Les figures 4.3 et 4.4 montrent des exemples de normalisation locale sur cinq images différant d'un changement d'illumination interne. Pour la série "Lézard", les changements sont réalisés en modifiant synthétiquement chacun des canaux couleur selon une transformation linéaire des niveaux de gris. Un bruit blanc gaussien est également ajouté pour plus de réalisme. Les images normalisées obtenues sont présentées sur la seconde ligne. Le diamètre de la fenêtre utilisée est égal à 9. Pour la série de la figure 4.4, nous avons utilisé des images prises avec des ouvertures de diaphragme différentes. Ici aussi le diamètre de la fenêtre de normalisation est de 9. Dans les deux séries, les images sont classées par ordre croissant de la transformation opérée par rapport à la première image.

Les graphes associés à ces deux séries traduisent la comparaison deux à deux des images de chaque série avant et après normalisation. A chaque couple d'images (I_i, I_j), on associe les distances entre les couleurs de tous les couples de pixels $(I_i(x,y), I_j(x,y))$ et c'est la médiane de ces valeurs qui est représentée sur le graphe. La distance entre deux couleurs est tout simplement la moyenne des trois différences de niveaux de gris en valeur absolue ; elle varie entre 0 et 255.

Les résultats obtenus sur les deux séries sont similaires. Les graphes "en dent de scie" obtenus avant normalisation traduisent la dégradation croissante des distances plus le changement d'illumination est important. On constate sans surprise que les distances sont considérablement réduites après la normalisation. Les résultats les meilleurs sont obtenus avec la série de la figure 4.3. La médiane des erreurs y est pratiquement nulle pour tous les couples d'images normalisées passés en revue. Ce n'est pas surprenant dans la mesure où les changements d'illumination ont été opérés synthétiquement directement à partir du modèle diagonal avec translation. Pour l'autre figure, les médianes des erreurs après normalisation sont toutes inférieures à 5, ce qui constitue une amélioration conséquente. Ces résultats traduisent la pertinence du modèle d'illumination choisi et de la technique de normalisation. Pour la seconde série, les images normalisées conservent tout de même quelques différences, le modèle choisi demeurant une approximation de la réalité.

Avantage par rapport aux méthodes de normalisation classiques

D'autres combinaisons permettent également de s'affranchir des paramètres du modèle d'illumination, comme la normalisation dans l'intervalle [0..1] utilisant les extrema ($\alpha = \gamma = 0$ et $\beta = 1$ en référence à l'équation 4.3), ou encore la normalisation centrée-réduite. L'avantage du choix de la médiane comme outil de normalisation est qu'elle s'avère moins sensible au bruit que toute autre donnée statistique. Ce choix nous a en effet été confirmé en testant les trois méthodes sur plusieurs images différant de changements d'illumination et en comparant les images obtenues. Les résultats obtenus à partir des images de la figure 4.4 sont présentés à la figure 4.5. C'est la méthode travaillant avec les médianes qui fournit les plus petites différences entre images normalisées.



(1) (2) (3) (4) (5)

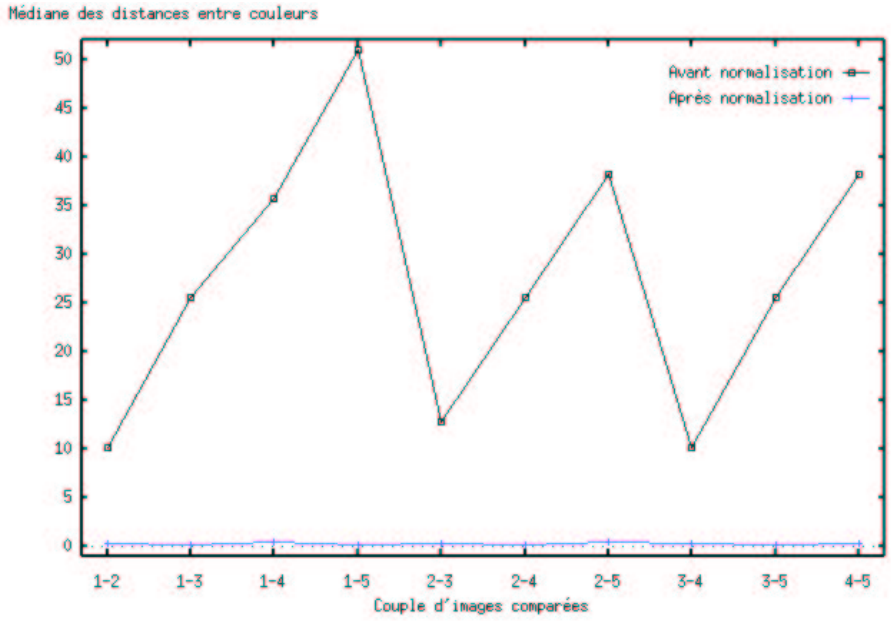


Figure 4.3 – *Série "Lézard" avec changement d'illumination interne (synthétique) : normalisation locale.*

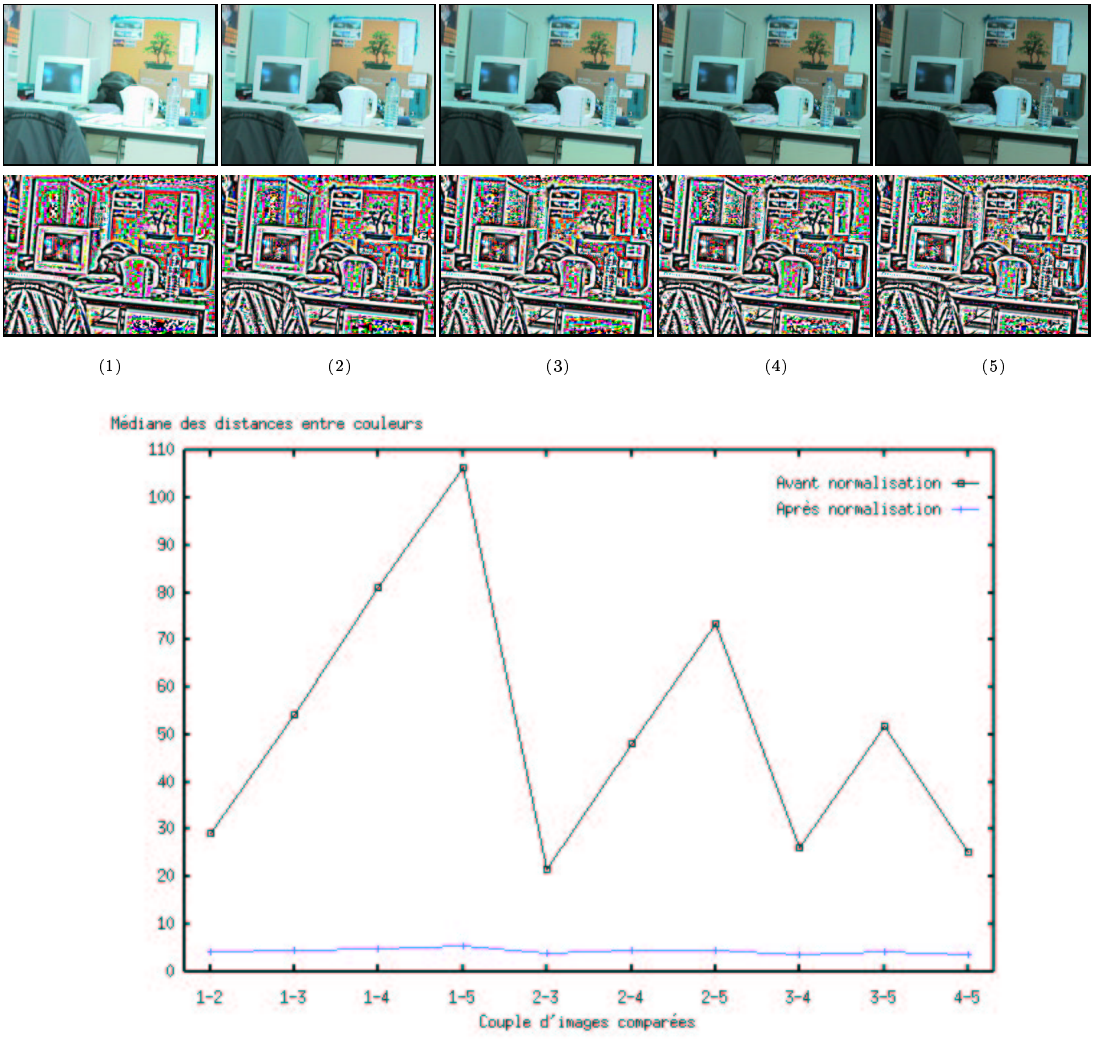


Figure 4.4 – *Série "Bureau" avec changement d'illumination interne (ouverture du diaphragme) : normalisation locale.*

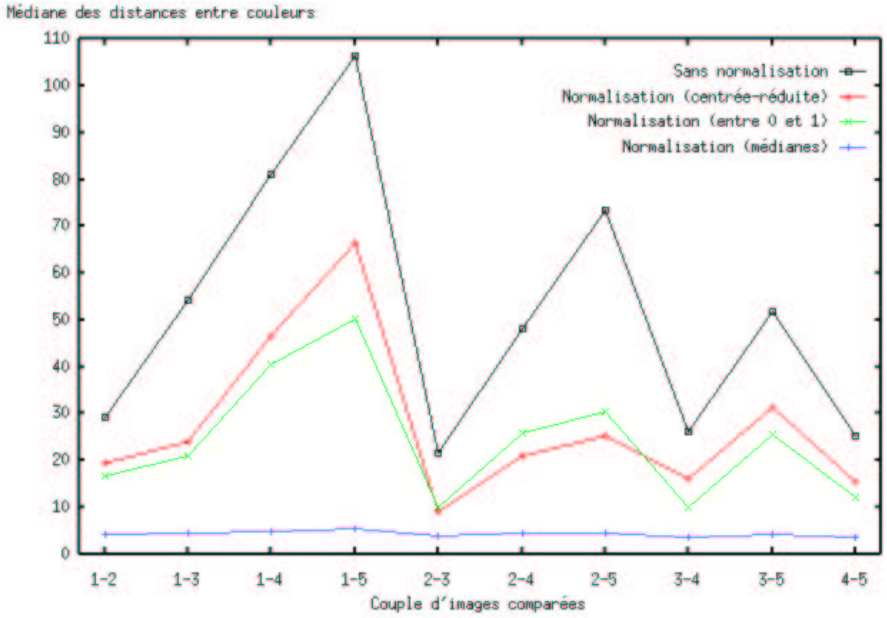


Figure 4.5 – *Comparaison de la méthode de normalisation basée sur les médianes avec les méthodes classiques.*

Un exemple de vecteur caractéristique avant et après normalisation

Un exemple de vecteur caractéristique $\vec{v}_{col}$ face aux changements internes d'illumination est présenté à la figure 4.6. Les vecteurs d'invariants sont calculés sur deux points d'images différant d'un changement d'illumination synthétique[1] et bruité. Les deux diagrammes de gauche représentent le vecteur caractéristique avant normalisation et ceux de droite le représentent après une normalisation réalisée localement. Nous voyons clairement que le traitement qui a été opéré sur les images rend la caractérisation invariante tout en préservant son pouvoir discriminant.

4.3.2 Changement externe

L'approximation utilisée pour les changements d'illumination externes a été présentée à la section 3.3.1.2 du chapitre 3. C'est un modèle à n paramètres si l'image contient n pixels. Il est donc bien sûr impossible de rendre le vecteur des huit invariants indépendant de ce modèle. La solution consiste alors à normaliser l'image.

4.3.2.1 La méthode de normalisation

Une façon de supprimer chacun des paramètres δ_i consiste à multiplier chaque pixel (r, v, b) par un facteur de telle sorte que l'on ait par exemple :

$$r + v + b = 1 \tag{4.5}$$

[1] avec comme paramètres : D_{11}=0.5, D_{22}=0.4, D_{33}=0.3, T_1=0.3,T_2=0.2 et T_3=0.1.

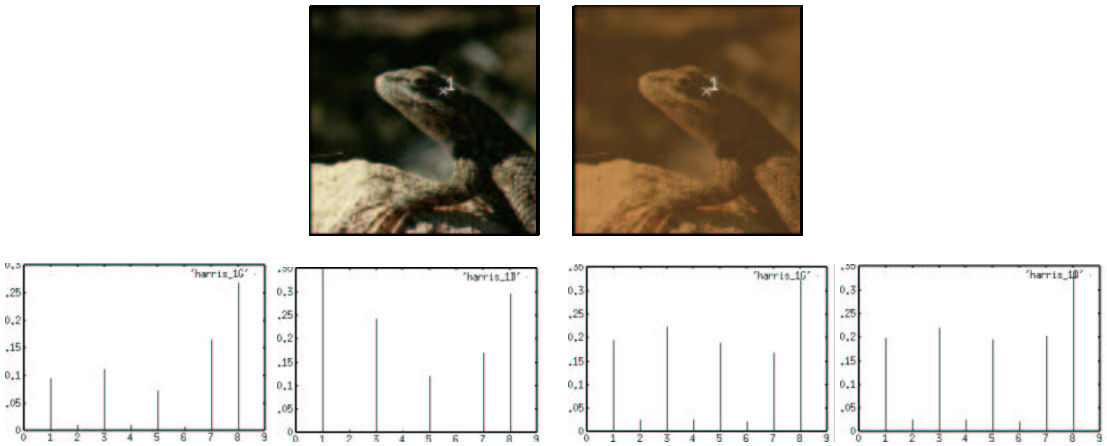


Figure 4.6 – *Les deux vecteurs d'invariants couleurs obtenus avant et après normalisation.*

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, à un éventuel changement de position de la source près, et ceci bien sûr dans la mesure où le modèle utilisé est correct. Contrairement à la technique mise en place à la section précédente pour des changements internes, cette normalisation est purement locale puisqu'elle est met en jeu un paramètre différent pour chaque pixel ; en outre chacun de ces paramètres dépend des trois canaux.

4.3.2.2 Exemples de normalisation

Le lecteur peut voir à la figure 4.7 une série d'images différant de changements d'illumination externes. La scène est relativement compliquée. Elle est en effet soumise à un éclairage naturel à peu près constant provenant d'une fenêtre située sur la droite, ainsi qu'à une lampe non visible située sur la gauche et déplacée à chaque prise. Elle contient également un certain nombre d'objets situés sur plusieurs plans, qui occasionnent des ombres diverses selon la position de la lampe. Notons enfin qu'il y a une forte couleur dominante : le bleu. Les images normalisées sont présentées sur la deuxième ligne de la figure ; elles ont été traitées en tenant compte de la contrainte de l'équation 4.5.

On constate que la normalisation permet de s'affranchir de la position des sources de lumière et donc des ombres. Sur les images normalisées, seules les formes ressortent. Comme le montre le graphe associé à la figure, les images sont dans l'ensemble très différentes avant la normalisation. Citons par exemple les couples d'images 1-5 et 2-5 pour lesquels la médiane des erreurs atteint près de 40, sur une échelle de niveaux de gris de 0 à 255. En revanche, la normalisation permet d'unifier considérablement ces mêmes images, puisque les médianes des erreurs chutent toutes en dessous de 5.

Ce traitement particulièrement local a permis de rendre les images indépendantes des changements d'illumination externes. A titre de comparaison, nous avons également souhaité mettre en œuvre sur ces mêmes images la technique de normalisation dédiée

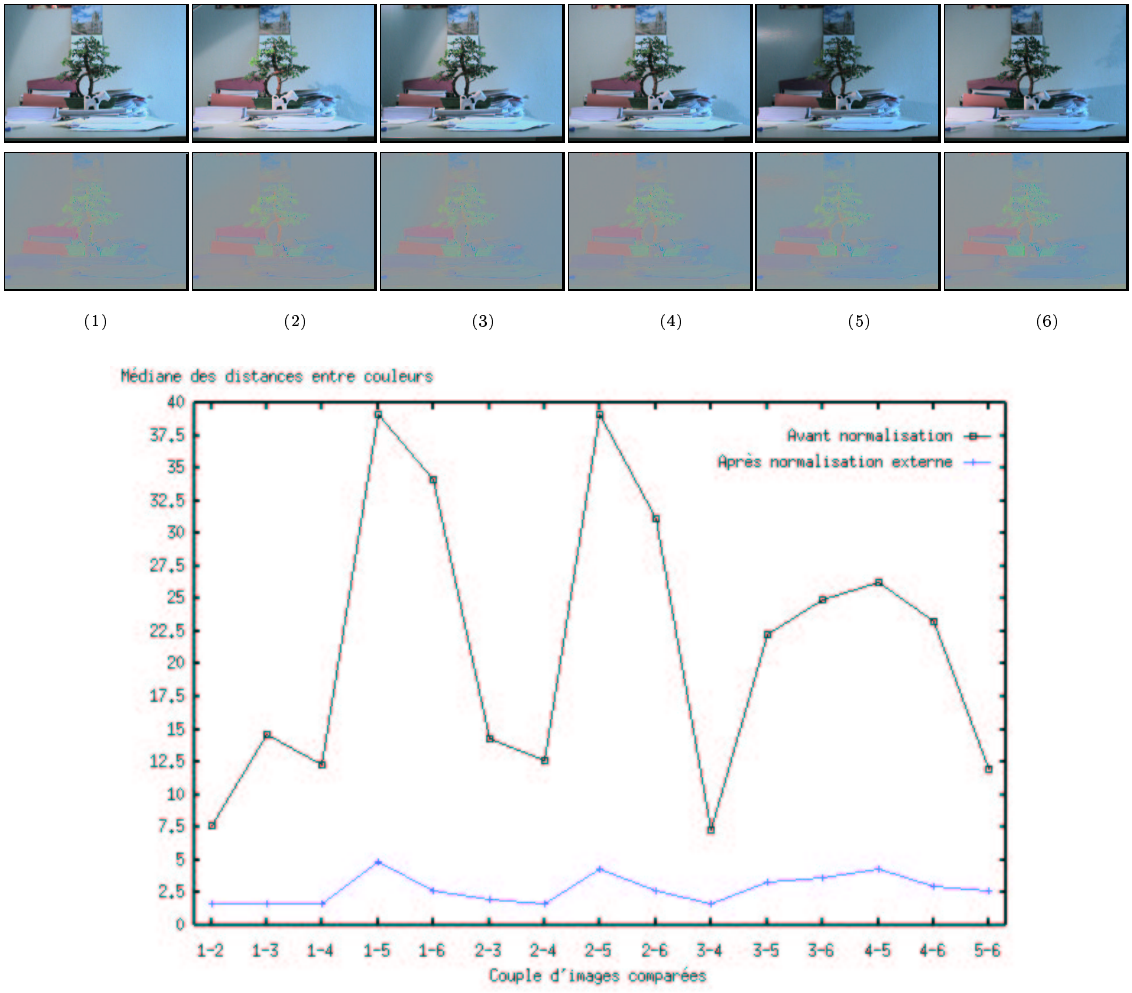


Figure 4.7 – *Série "Bonzaï" avec changement d'illumination externe (déplacement de la lampe) : normalisation.*

jusqu'alors aux changements internes, et bien évidemment dans sa version locale. Les calculs ont été réalisés pour deux tailles de voisinage et les résultats sont présentés à la figure 4.8.

Le graphe montre que la normalisation locale donne des résultats intéressants, bien qu'elle ne soit pas a priori dédiée aux changements d'illumination externes. On constate que la normalisation est d'autant plus efficace que la taille du voisinage est réduite. Les résultats sont même globalement meilleurs qu'avec la méthode traditionnelle (comparer avec le graphe de la figure 4.7) pour une taille de voisinage de 3.

4.3.3 Changement complexe

Par changement complexe d'illumination, on entend la combinaison des modèles internes et externes présentés aux deux sections précédentes.

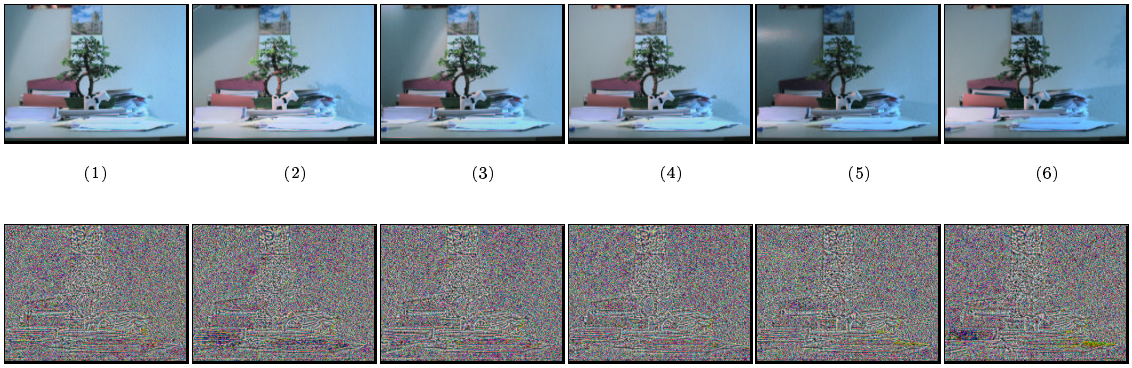
Considérons maintenant la série d'images de la figure 4.9. Les images de la première ligne diffèrent à la fois de changements externes (toujours le déplacement de la lampe) et de changements internes réalisés synthétiquement par des transformations affines sur chacun des plans couleur, avec un bruit additionnel. La méthode de normalisation face aux changements externes a été appliquée selon la contrainte de l'équation 4.5 et les images résultats sont présentées sur la deuxième ligne de la figure. Une normalisation locale, à partir d'un fenêtrage de taille 21, a également été réalisée sur les images initiales pour parer aux changements internes. Les images normalisées sont présentées sur la troisième ligne.

En examinant le graphe associé à toutes ces images, on constate que la normalisation externe n'apporte aucune amélioration face à ce type de changement complexe. Les médianes des erreurs sont en effet à peine meilleures, voire moins bonnes qu'avant la normalisation. En revanche, les résultats obtenus par la normalisation locale sont une fois de plus particulièrement probants, puisque les médianes calculées sont toutes inférieures à 3.

Les tests réalisés pour cette série d'images ne tiennent pas compte des deux modèles d'illumination conjointement. Pour ce faire, il faudrait envisager une méthode de normalisation permettant d'éliminer les $6 + n$ paramètres mis en jeu. Finlayson l'a réalisé de manière itérative, mais pour le modèle diagonal simple (se reporter à la section 3.3.1.3 pour le détail de la méthode). Notre approche intégrant un paramètre de translation additionnel, la méthode s'avère plus complexe. Nous n'avons d'ailleurs pas cherché à la mettre en place. En effet, à travers les diverses séries d'images testées, nous avons constaté que les résultats obtenus sur ce type de changement complexe avec une simple normalisation locale sont très satisfaisants. Mettre en œuvre une technique intégrant également les changements externes n'améliorerait donc pas les résultats de façon significative, d'autant plus que la méthode de normalisation externe possède un inconvénient majeur, qui est abordé au paragraphe suivant.

Influence de la normalisation sur les invariants

Pour être à la fois invariant à la rotation et aux changements d'illumination, le vecteur $\vec{v}_{col}$ défini à l'équation 4.2 doit être calculé à partir de l'image normalisée. La normalisation



(1) (2) (3) (4) (5) (6)

Normalisation locale (taille du voisinage = 3)

Normalisation locale (taille de voisinage = 9)

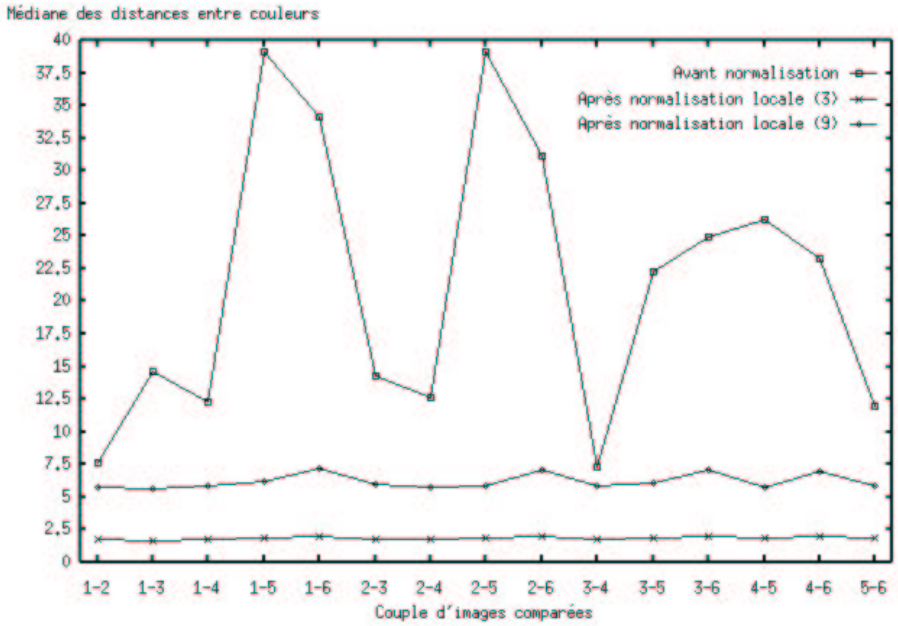
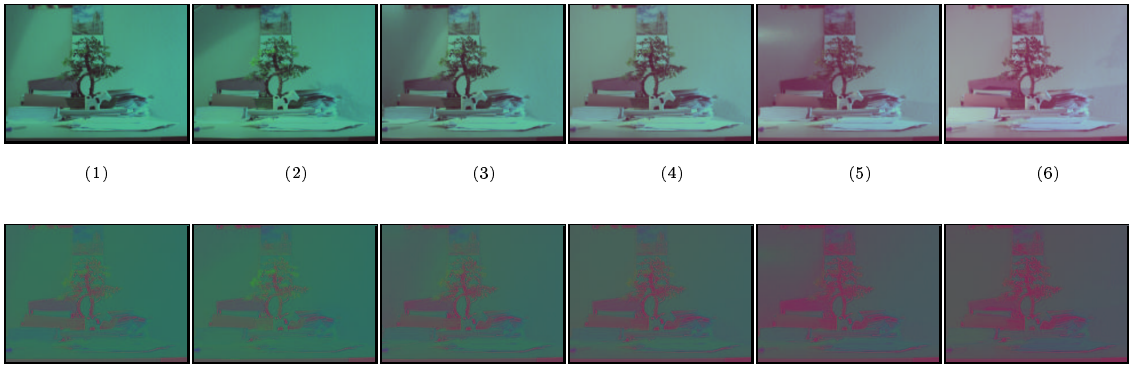
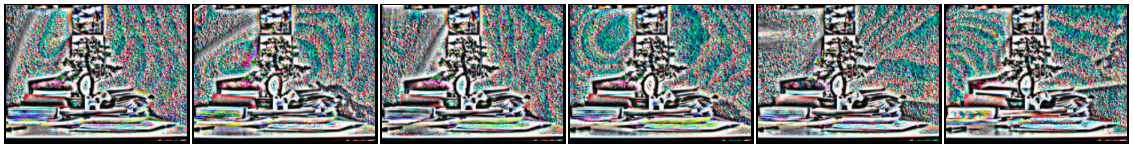


Figure 4.8 – *Même série "Bonzaï" mais avec deux normalisations locales.*



Normalisation externe



Normalisation interne (locale avec fenêtre de voisinage = 21)

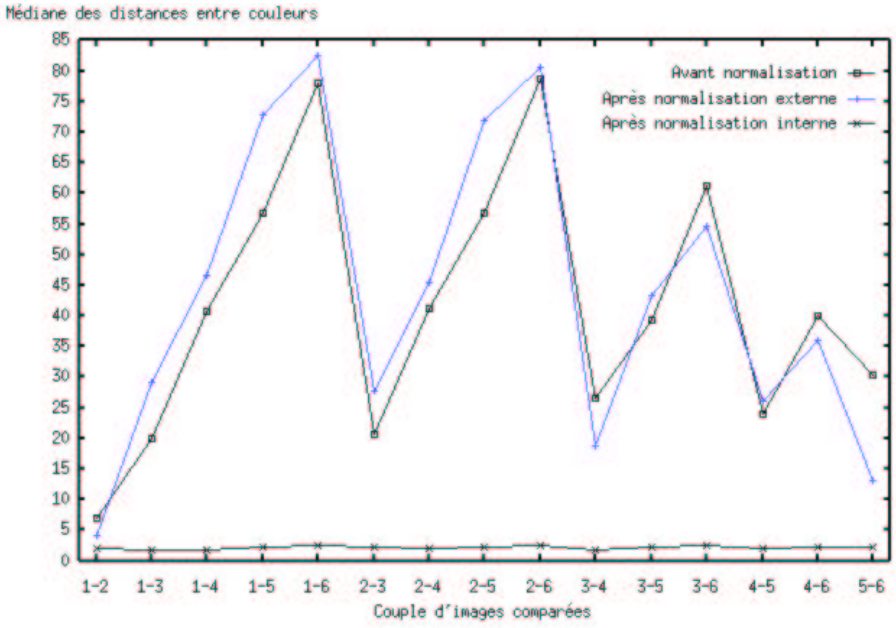


Figure 4.9 – *Série "Bonzai" avec changement d'illumination complexe : normalisation.*

face aux changements internes, qu'elle soit locale ou non, ne modifie en rien le pouvoir discriminant de $\vec{v}_{col}$. En effet, le pré-traitement de l'image est effectué sur chacun des plans couleur, les uns indépendamment des autres et ne corrèle donc en aucun cas l'information couleur. Pour s'en assurer, l'expérience suivante a été réalisée : trois scores de corrélation ont été calculés entre les trois plans couleur d'une image pris deux à deux sur un ensemble de points caractéristiques, ceci avant et après normalisation. Il s'avère que les scores obtenus ne sont pas modifiés par la normalisation de l'image. Une image ayant ses trois plans couleur fortement (resp. faiblement) corrélés le sera tout autant après normalisation.

En revanche, il n'en est pas de même pour la méthode de normalisation face aux changements externes présentée à la section 4.3.2. En effet, la contrainte de normalisation de l'équation 4.5 rend les trois canaux couleur inter-dépendants. L'image après normalisation correspond à une description en coordonnées chromatiques $\left\{r = \frac{R}{R+V+B}, v = \frac{V}{R+V+B}\right\}$. Le vecteur d'invariants $\vec{v}_{col}$ est donc malheureusement réduit à $3 \times 2 - 1 = 5$ invariants indépendants.

Quelle solution adopter face à un changement complexe d'illumination ?

Suite à ces observations, nous avons pris le parti de ne pas normaliser l'image face aux changements d'illumination externes, afin de conserver la pertinence de chacun des invariants. Les tests qui ont été présentés aux figures 4.8 et 4.9 montrent que, dès l'instant où la normalisation est envisagée localement, le modèle diagonal avec translation, utilisé seul, permet de modéliser les changements internes comme externes de manière tout à fait satisfaisante.

Le lecteur peut également observer aux figures 4.10 et 4.11 le résultat de la normalisation locale sur les invariants couleurs, calculés sur plusieurs images différant de changements complexes d'illumination. Quatre images de la séquence "Bonzaï" ont été sélectionnées pour ce test. Elles diffèrent de changements d'illumination externes (toujours le déplacement de la lampe) et nous leur avons fait subir un changement interne en appliquant une transformation linéaire des niveaux de gris, et ajouté un bruit blanc gaussien. Ces images sont représentées sur la deuxième ligne de la figure 4.10. Les mêmes images après normalisation locale (taille du voisinage = 21) sont représentées sur la deuxième ligne de la figure 4.11. Plusieurs points Harris couleur ont été extraits d'une des images originales. Lors de la détection, le seuil a été placé suffisamment haut de façon à isoler des points à l'endroit où le signal possède un contenu informatif le plus riche possible. Quatre points ont été obtenus, ils sont présentés sur la première ligne des deux figures. Les quatre dernières lignes de chaque figure montrent les huit invariants couleur obtenus pour chacun de ces points, avant et après normalisation selon la figure considérée.

Nous constatons sans surprise à la figure 4.10 que pour tous les points, la caractérisation varie avec les images. Les différences atteignent même 50% pour certains invariants, comme en atteste par exemple le niveau de gris du plan bleu (l'invariant n°5 sur les graphes) calculé au point P_3, qui varie du simple au double entre la première et la dernière image.

En revanche, les invariants de la figure 4.11, obtenus quant à eux après normalisation, varient beaucoup moins d'une image à l'autre et ce, quelque soit le point considéré. On constate des écarts de l'ordre de 10% dans le pire des cas. Il convient d'attribuer ces différences à l'approximation du modèle d'illumination utilisée.

Les 4 points

Image 1 Image 2 Image 3 Image 4

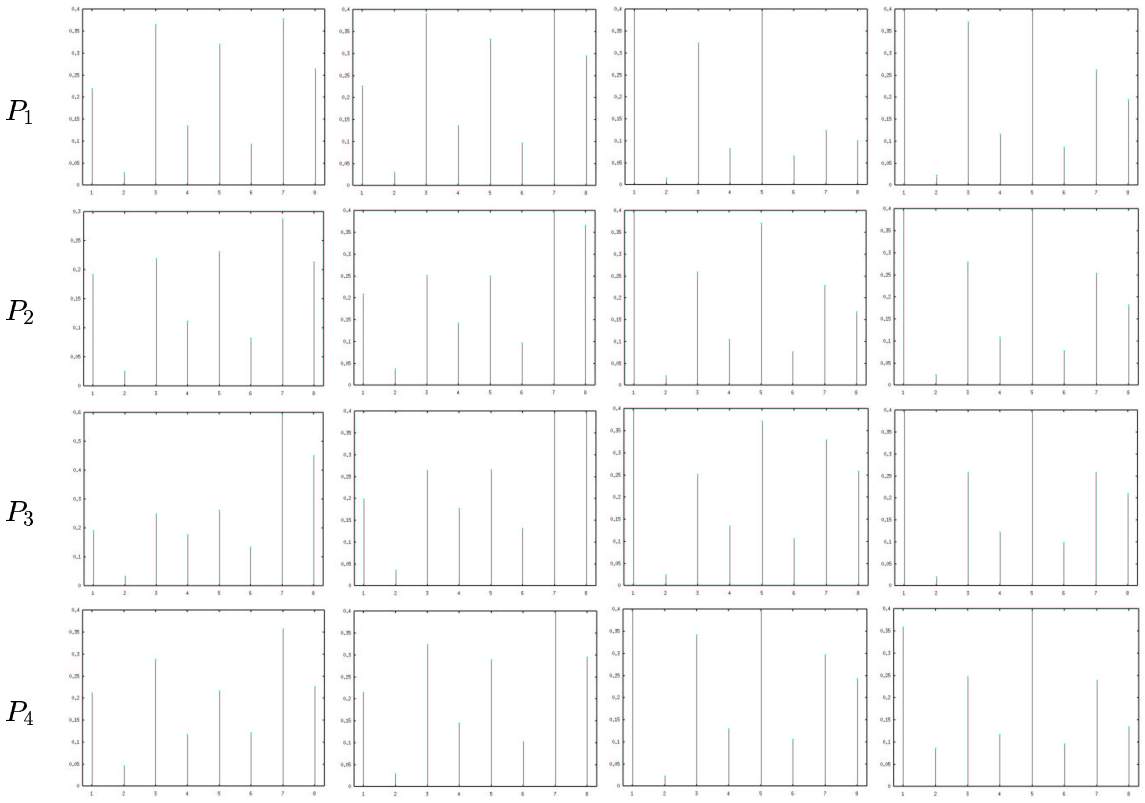
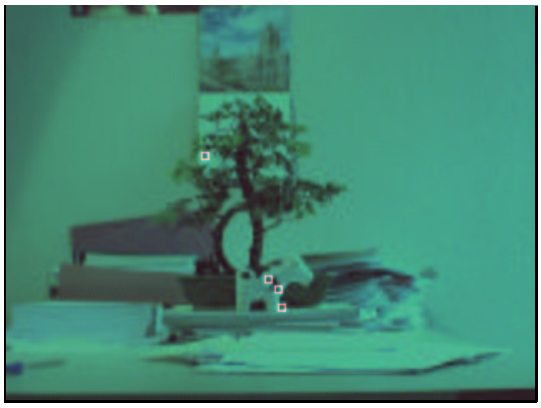


Figure 4.10 – *Série "Bonzaï" avec changement d'illumination complexe : invariants obtenus pour 4 points avant normalisation.*



Les 4 points

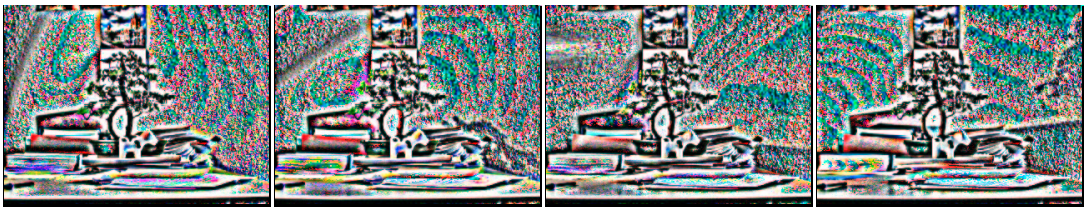


Image 1 Image 2 Image 3 Image 4

Figure 4.11 – *Série "Bonzaï" avec changement d'illumination complexe : invariants obtenus pour les 4 mêmes points après normalisation.*

En conclusion de cette étude sur le changement d'illumination, nous dirons que la normalisation d'une image n'est pas une opération simple à réaliser. En effet, l'estimation d'un modèle d'illumination réaliste est délicate. Heureusement dans notre cas, une approximation telle que celle que nous venons de présenter et valider (le modèle diagonal avec translation) suffit, étant donné qu'elle sera utilisée dans un contexte de caractérisation *locale*. L'étape de normalisation même est également complexe à mettre en place. Celle présentée ici s'avère satisfaisante. Notons tout de même qu'il faut être capable de fixer efficacement la taille de la fenêtre lorsqu'elle est employée localement.

4.4 Conclusion

La méthode de caractérisation que nous avons présentée dans ce chapitre a le principal avantage d'éliminer un des inconvénients qui faisait jusqu'alors défaut aux invariants différentiels en niveau de gris, à savoir leur sensibilité au bruit. Elle réduit en effet à l'ordre un le calcul des invariants différentiels, tout en restant aussi riche. La caractérisation couleur obtenue gagne ainsi en simplicité mais surtout en robustesse.

La deuxième étape de nos travaux a ensuite consisté à adapter le problème de la constance des couleurs à cette nouvelle méthode de caractérisation. Le modèle d'illumination adopté pour les changements internes est le modèle diagonal de Finlayson mais enrichi d'une translation. Il s'avère posséder le meilleur rapport qualité/complexité dans le cadre de la caractérisation locale d'images. Le cas des changements externes d'illumination a également été étudié. Pour ne pas appauvrir la caractérisation invariante à la rotation basée sur huit invariants différentiels, nous avons choisi de pré-traiter l'image afin de la rendre invariante à ces modèles d'illumination. La méthode de normalisation mise en place a l'avantage d'être robuste au bruit puisqu'elle travaille à l'ordre zéro. Elle a été également implémentée localement pour ne pas perdre les propriétés locales des pixels. Des tests ont été réalisés sur plusieurs séries d'images différant de changements d'illumination internes comme externes. Ils ont permis d'évaluer cette approche en comparant pixel par pixel les images normalisées. Les résultats obtenus montrent que les images traitées par cette méthode deviennent bien moins dépendantes des changements d'illumination, les quelques différences subsistant étant inhérentes au modèle d'illumination utilisé, qui n'est bien sûr qu'une approximation de la réalité. Les images obtenues ont toutefois été jugées satisfaisantes, puisque les invariants différentiels, calculés à partir de ces images pour plusieurs points, ne varient plus que très sensiblement, malgré les changements d'illumination complexes mis en jeu. La répercussion de ces normalisations sur les invariants couleur n'a toutefois été évaluée que "visuellement". Il faudra mettre en place des méthodes automatiques permettant de vérifier que $\vec{v}_{col}$ est bien rendu invariant aux changements d'illumination. Ce travail sera réalisé dans les prochaines chapitres traitant de la mise en correspondance.

Une fois de plus, nous avons montré l'intérêt de l'utilisation de la couleur. Cette information supplémentaire permet de mettre en place des caractérisations de points d'intérêt plus robustes. Notons également que la méthode de normalisation face aux changements externes, même si nous ne l'exploiterons pas par la suite, permet de s'affranchir des ombres. Ceci constitue une révolution par rapport aux images en niveau de gris, pour lesquelles le déplacement de la source lumineuse représente un bruit pratiquement incontournable. De

nouvelles possibilités de mise en correspondance s'offrent alors, que ce soit pour l'indexation d'images ou les applications liées à la stéréoscopie. Cette voie semble intéressante à explorer.

Dans la suite de cette thèse, nous utiliserons comme descripteur de points d'intérêt le vecteur $\vec{v}_{col}$ défini à l'équation 4.2, calculé dans un contexte multi-échelle avec plusieurs largeurs de gaussienne. La caractérisation sera ainsi invariante aux transformations euclidiennes de l'image, aux changements d'échelle et donc robuste aux changements de point de vue. Les images seront au préalable normalisées localement pour s'affranchir des changements d'illumination.

Une fois les points d'intérêt détectés et caractérisés, l'étape suivante consiste à les mettre en correspondance en comparant les caractérisations obtenues, et ceci au travers des différentes transformations de l'image allant jusqu'au changement de point de vue dans le cas le plus général. Les méthodes de mise en correspondance existantes sont passées en revue dans le prochain chapitre, avant de présenter notre approche au chapitre 6.

Chapitre 5

Méthodes d'appariement - Etat de l'art

Dans ce chapitre, nous passons en revue les principales méthodes de mise en correspondance qui exploitent les diverses caractérisations présentées au chapitre 3. L'objectif de ces approches est de mettre en place une mesure de similarité permettant de les comparer. Nous voyons également que certaines de ces méthodes exploitent des informations géométriques plus globales pour améliorer l'appariement. Les principales techniques de mise en correspondance dense, consacrées à l'appariement de l'intégralité des images, sont aussi présentées.

Sommaire

La mise en correspondance de points d'intérêt entre deux images est un problème très coûteux. Afin de réduire au maximum l'espace de recherche du correspondant d'un point et de limiter le risque de faux appariements, il est nécessaire de mettre en place des contraintes.

Une première contrainte de nature photométrique est la similarité qui peut exister entre les points provenant d'un même point 3D, sous l'hypothèse de l'invariance ou quasi-invariance par projection. Les ensembles de caractérisation de points qui ont été présentées tout au long des chapitres 3 et 4 représentent des critères pour juger si deux points se ressemblent, à condition d'être capable de les comparer. Définir une mesure de similarité pour comparer ces vecteurs de caractéristiques représente un travail complexe qui a suscité, et suscite encore de nombreux travaux. Nous passons en revue dans la section 5.1 les principales méthodes de comparaison que nous avons rencontrées.

D'autres contraintes existent et sont pour la plupart de nature géométrique. Il s'agit en premier lieu de la géométrie épipolaire, qui est la seule contrainte physique reliant les deux images. Il est également possible de définir des contraintes semi-locales qui tiennent comptent de l'information contenue dans le voisinage du point. L'étude de ces contraintes fait l'objet de la section 5.2. Nous présentons ensuite à la section 5.3 les principales méthodes de mise en correspondance dense existantes.

5.1 Appariement épars selon l'approche iconique

5.1.1 Principe : le calcul des scores d'appariement

La mise en correspondance de deux ensembles de points issus de deux images commence par le calcul d'un score d'appariement pour chaque paire de points. Les scores obtenus indiquent quels couples ont de bonnes chances de constituer un appariement correct ou non. Ce score est calculé en comparant la caractérisation associée à chaque point. Dès l'instant où le descripteur du point ne se ramène pas à une simple valeur scalaire, toute la difficulté de la méthode repose sur le choix d'une bonne mesure de comparaison. La définition du critère d'appariement est importante car elle conditionne la qualité des résultats.

Dans les sections 5.1.2, 5.1.3 et 5.1.4, nous passons en revue les différents types de mesures de similarité rencontrés selon les principales méthodes de caractérisation présentées précédemment. Une fois la comparaison effectuée pour tous les couples de points possibles, il s'agit de retenir les couples associés aux meilleurs scores, qui ont toutes les chances de correspondre à de bons appariements. Ce travail fait l'objet de la section 5.1.5.

5.1.2 Appariement basé sur une partie du signal

Plusieurs techniques réalisent la comparaison d'images ou de sous-images directement à partir du signal, sans passer par une caractérisation spécifique. Nous présentons à la section 5.1.2.1 la classique méthode de corrélation qui compare deux à deux les pixels des deux zones à apparier. Il existe également une méthode moins locale, qui s'applique dans le domaine spectral et qui exploite le spectre de puissance croisé des deux signaux à

apparier. Cette approche porte le nom de corrélation de phase et est présentée à la section 5.1.2.2. Comme la corrélation, la distance de Hausdorff définit une mesure de similarité entre deux ensembles de points. Elle sera présentée à la section 5.1.2.3 .

5.1.2.1 La corrélation du signal

La corrélation du signal, encore appelée "template matching" dans la littérature, est la méthode de mise en correspondance la plus ancienne. Elle peut être interprétée comme une caractérisation élémentaire du signal autour d'un point : tout point est caractérisé par l'ensemble des valeurs du signal sur son voisinage.

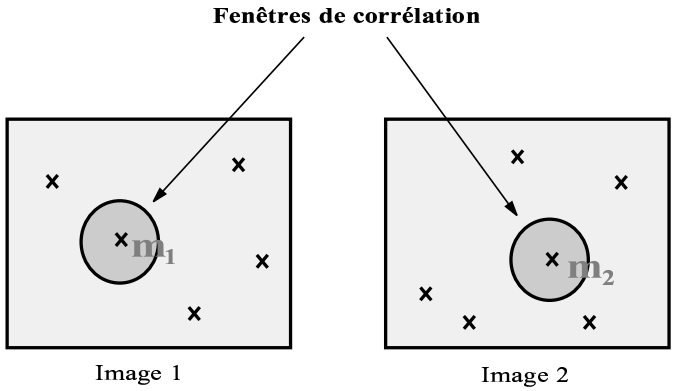


Figure 5.1 – *Calcul du score de corrélation entre deux points m_1 et m_2.*

Comme le montre en deux dimensions la figure 5.1, on calcule un score de corrélation entre les fenêtres de corrélation entourant les deux points à traiter. Nous avons résumé dans le tableau 5.1 les formules de corrélation les plus classiques (SSD[1], NSSD[2], NCC[3], ZNSSD[4], ZNCC[5]) , avec leurs principales propriétés. F_i est le vecteur qui contient les valeurs du signal de l'image I_i dans la fenêtre de corrélation du point considéré et $\overline{F_i}$ la moyenne de ces valeurs. Tous les scores excepté SSD sont compris dans l'intervalle [-1..1].

SSD	$scr(F_1, F_2) = \|F_1 - F_2\|^2$	
NSSD	$scr(F_1, F_2) = \frac{\|F_1 - F_2\|^2}{\|F_1\| . \|F_2\|}$	$scr(F_1, F_2) = scr(aF_1, aF_2)$
NCC	$scr(F_1, F_2) = \frac{F_1 . F_2}{\|F_1\| . \|F_2\|}$	$scr(F_1, F_2) = scr(a_1F_1, a_2F_2)$
ZNSSD	$scr(F_1, F_2) = \frac{\|(F_1 - \overline{F_1}) - (F_2 - \overline{F_2})\|^2}{\|F_1 - \overline{F_1}\| . \|F_2 - \overline{F_2}\|}$	$scr(F_1, F_2) = scr(aF_1 + b_1, aF_2 + b_2)$
ZNCC	$scr(F_1, F_2) = \frac{(F_1 - \overline{F_1}) . (F_2 - \overline{F_2})}{\|F_1 - \overline{F_1}\| . \|F_2 - \overline{F_2}\|}$	$scr(F_1, F_2) = scr(a_1F_1 + b_1, a_2F_2 + b_2)$

Table 5.1 – *Tableau des principales mesures de corrélation.*

Remarquons que la formule ZNCC est invariante aux transformations affines. Elle est

[1]Sum of Squared Differences
[2]Normalized Sum of Squared Differences
[3]Normalized Cross Correlation
[4]Zero mean Normalized Sum of Squared Differences
[5]Zero mean Normalized Cross Correlation

donc bien appropriée au calcul des scores de corrélation entre images dont les niveaux de gris diffèrent d'une transformation affine.

L'algorithme de corrélation a bien sûr pour support les images de niveaux de gris. Mais il est également possible de l'appliquer à des images calculées en précision sub-pixel ainsi qu'à certaines des images dérivées comme par exemple l'image du laplacien. Dans [Nishihara, 1983], Nishihara montre d'ailleurs que les résultats sont meilleurs en utilisant le laplacien car les pics de corrélation obtenus sont plus marqués. Dans [Crouzil et al., 1996], Crouzil travaille sur les images des dérivées du premier ordre : il utilise un score de similarité entre les gradients des pixels des fenêtres de corrélation pour calculer ce qu'il nomme la mesure de corrélation du gradient des deux fenêtres. Dans le cas d'images en couleur, il est possible de calculer un score de corrélation pour chacun des plans couleur [Lemestre, 1997; Devernay, 1997]. Il s'agit ensuite de choisir entre plusieurs opérateurs (produit, addition, minimum, maximum, etc) pour obtenir un score unique.

Un des inconvénients de la corrélation est que lorsque les deux images possèdent un grand nombre de points et que la fenêtre de corrélation est grande, le temps de calcul des scores de corrélation est relativement long. Il est néanmoins possible d'améliorer ce temps si la disparité entre les deux images est faible, en n'effectuant les calculs que sur une fenêtre de points, comme dans [Zhang et al., 1995b]. Bien sûr ceci introduit une contrainte sur le mouvement entre les caméras. Une autre solution est de travailler sur des fenêtres plus petites, les erreurs d'appariement ainsi générées pouvant être ensuite éliminées à l'aide par exemple de contraintes semi-locales comme celles de la relaxation, décrite plus loin dans ce chapitre. On trouve dans la littérature d'autres approches qui tentent de rendre la corrélation plus efficace. Celle de Chou et Chen [Chou et Chen, 1990] transforme les images en niveaux de gris en images binaires tout en stockant les moments. La fonction de corrélation est ensuite mise en place en utilisant un ET logique et une addition sur chaque paire de pixels. Zabih utilise aussi cette méthode avec la distance de Hamming ; elle est détaillée dans [Lan et Mohr, 1997b]. Il faut néanmoins remarquer que ces solutions sont moins performantes que la corrélation traditionnelle puisqu'il y a perte d'information.

L'inconvénient majeur de la corrélation est qu'elle n'est pas invariante à n'importe quel point de vue : l'image doit avoir été prise dans des conditions voisines afin que les fenêtres de corrélation se superposent correctement par simple translation. Une solution envisageable dans le cas d'une rotation est de considérer des fenêtres de corrélation circulaires et de leur faire subir cette rotation afin que les signaux soient superposés correctement (le centre de rotation sera le point considéré et l'angle pourra être déterminé en comparant par exemple les directions des gradients des deux points, si la calibration du système est inconnue). Dans [Brand, 1995] une solution impliquant des fenêtres de corrélation déformables est proposée. L'auteur suppose que les masques sont suffisamment petits pour qu'on puisse considérer que deux masques en correspondance sont déformés par une transformation affine. Celle-ci n'est pas quelconque, les coins des fenêtres respectant nécessairement la contrainte épipolaire.

Enfin la méthode peut aussi échouer à la frontière d'objets occultants. Dans [Lotti et Giraudon, 1993], un algorithme de corrélation pour réaliser la mise en correspondance d'images aériennes est présenté. Les auteurs calculent pour chaque pixel une fenêtre adaptative contrainte par une carte de contours extraite des images. Dans [Lan et Mohr, 1997b], une solution est également apportée à ce problème en mettant en place une méthode de

corrélation partielle qui permet de calculer les scores de corrélation uniquement sur la bonne partie de la fenêtre à corréler. Elle permet d'obtenir de meilleurs résultats aux frontières d'objets occultants mais suppose que le signal obéit à une transformation affine d'une image à l'autre et donne de moins bons résultats que les méthodes traditionnelles dans les autres régions de l'image.

Dans le cas de changements d'intensité plus complexes que les changements affines, il existe la classe des corrélations non paramétriques. Elles exploitent des modèles non paramétriques et sont donc moins sensibles au bruit. En addition, elles gèrent mieux les effets non linéaires puisqu'aucune hypothèse de linéarité n'est utilisée. Citons notamment [Zabih et Woodfill, 1994] qui utilise une mesure non paramétrique et locale basée uniquement sur les niveaux de gris de l'image ainsi qu'une mesure, non paramétrique et locale également, qui tient d'avantage compte de l'information spatiale. La corrélation selon Bhat [Bhat et Nayar, 1996] utilise quant à elle une mesure non paramétrique qui gère très bien les effets non linéaires, comme les reflets spéculaires par exemple, ainsi que le bruit. Malheureusement, Le points faible de ces approches est qu'elles sont très sensibles aux transformations géométriques de l'image, et qu'elles gèrent mal les occultations.

5.1.2.2 La corrélation de phase

Nos avons vu à la section 3.2.4.1 du chapitre précédent que le signal pouvait également être caractérisé dans le domaine spectral par la Transformée de Fourier notamment. Nous allons voir dans cette section qu'il est possible de comparer deux images à partir de cette Transformée, en exploitant le théorème du décalage qui lui est associé. La méthode porte le nom de *corrélation de phase*.

Soient deux signaux bi-dimensionnels $f_1(x, y)$ et $f_2(x, y)$ différant d'une translation (x_0, y_0) et $F_1(\omega_x, \omega_y)$ et $F_2(\omega_x, \omega_y)$ leurs Transformées de Fourier respectives[6]. Le théorème du décalage[7] nous permet de déduire :

$$\frac{F_2(\omega_x, \omega_y)F_1^*(\omega_x, \omega_y)}{|F_2(\omega_x, \omega_y)F_1^*(\omega_x, \omega_y)|} = e^{j(\omega_x x_0 + \omega_y y_0)} \tag{5.1}$$

où $*$ dénote le complexe conjugué de la transformée. La partie gauche de l'équation correspond au spectre de puissance croisé des deux signaux. Il est alors très facile de déterminer les paramètres x_0 et y_0, puisque la Transformée de Fourier inverse de la partie droite de l'équation 5.1 est une impulsion de Dirac centrée en (x_0, y_0) :

$$F^{-1}\left(\frac{F_2(\omega_x, \omega_y)F_1^*(\omega_x, \omega_y)}{|F_2(\omega_x, \omega_y)F_1^*(\omega_x, \omega_y)|}\right) = F^{-1}\left(e^{j(\omega_x x_0 + \omega_y y_0)}\right) = \delta(x_0, y_0) \tag{5.2}$$

Le calcul du spectre de puissance croisé permet non seulement de déterminer si deux signaux sont identiques à une translation près, mais aussi de calculer les paramètres de cette translation. En Vision par Ordinateur, f_1 et f_2 sont des images. Cela implique donc de considérer la Transformée de Fourier dans sa version discrète. Malgré les approximations engendrées, les résultats de l'équation 5.2 restent valables [Kuglin et Hines, 1975].

[6] Se reporter à l'équation 3.17 du chapitre précédent pour la définition de la Transformée de Fourier.

[7] Conservation de l'amplitude : $F_2(\omega_x, \omega_y) = F_1(\omega_x, \omega_y)e^{j(\omega_x x_0 + \omega_y y_0)}$.

Dans la pratique, le signal déphasé obtenu contient un pic significatif au point de translation, mais aussi quelques pics additionnels supposés suivre une distribution normale centrée en zéro [Kuglin et Hines, 1975]. L'amplitude de tous ces pics dépend bien sûr du bruit contenu dans les deux images mais aussi de la taille de leur partie commune, une partie commune importante induisant un pic principal très prononcé par rapport aux autres pics.

Cette méthode de comparaison est particulièrement robuste au bruit et aux variations d'illumination non uniformes, ce qui la rend bien plus appropriée que les méthodes de corrélation classiques. Notons également qu'elle est parfaitement applicable sur des images ayant subi un lissage. En effet, le théorème de convolution conduit dans le domaine spectral à un simple facteur multiplicatif qui disparaît avec le quotient de l'équation 5.2.

On peut néanmoins reprocher à la méthode, telle qu'elle vient d'être présentée, de ne fonctionner que pour des valeurs entières de (x_0, y_0). Il existe des méthodes d'interpolation permettant d'estimer le décalage sous-pixellique. Dans [Shekarforoush et al., 1996], une méthode plus robuste, basée sur l'estimation de la décomposition polyphase du spectre de puissance croisé, est proposée.

La corrélation de phase classique traite de l'appariement d'images différant d'une simple translation. Des travaux plus récents [Reddy et Chatterji, 1996] ont été réalisés en vue d'étendre la méthode au cas de la rotation et du changement d'échelle. Les auteurs proposent ainsi une méthode permettant de comparer deux images à travers ces transformations, mais aussi de déterminer les paramètres de translation, de rotation et les facteurs d'échelle (un facteur dans chaque direction) impliqués. Les paramètres estimés lors des diverses expérimentations sont très précis, un peu moins cependant pour les changements d'échelle qui restent corrects jusqu'à un changement de 1.8. Ajoutons également que la méthode est tout de même beaucoup plus complexe que la version classique, puisqu'elle nécessite le calcul de trois Transformées de Fourier et de trois Transformées de Fourier inverses.

5.1.2.3 La distance de Hausdorff

Comme la corrélation, la distance de Hausdorff est une mesure de similarité permettant de comparer deux ensembles finis de points. Classiquement, sa formulation pour deux ensembles A et B est la suivante :

$$H(A, B) = \max(h(A, B), h(B, A)) \tag{5.3}$$

avec

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{5.4}$$

Ainsi $h(A, B)$ (resp. $h(B, A)$) est faible lorsque chaque point de A (resp. B) est proche d'un quelconque point de B (resp. A), et par conséquent, $H(A, B)$ traduit la similarité de A et de B lorsque les deux distances directes sont toutes les deux faibles à la fois.

Utilisée telle quelle, cette distance est trop fragile. Elle devient par exemple très grande lorsque un seul point de A est très éloigné de tous les points de B, et n'est donc pas

acceptable pour comparer deux images ou sous-images, bruitées la plupart du temps. Une solution avantageuse consiste à remplacer l'équation 5.4 par :

$$h^f(A,B) = f^{eme}_{a\in A} \min_{b\in B} \|a-b\| \tag{5.5}$$

où $f^{eme}_{x\in X} g(x)$ désigne la f^{eme} valeur de $g(x)$ dans l'ensemble X, pour des valeurs de f quantifiées entre 0 et 1. Par exemple, pour $f = 1$, on retrouve la distance de Hausdorff classique de l'équation 5.4 et la médiane des minima est donnée en prenant $f = \frac{1}{2}$. La distance ainsi définie est appelée *distance de Hausdorff partielle* [Huttenlocher et al., 1993].

Ses principaux avantages sont d'être robuste au bruit et aux occultations, surtout si l'on considère la médiane dans le calcul ($f = \frac{1}{2}$). Elle a jusqu'à présent essentiellement été utilisée pour mettre en correspondance deux images dans un cadre d'indexation d'images. Les deux ensembles de points considérés A et B sont respectivement le modèle recherché et une image de la base. Plusieurs travaux ont été réalisés pour l'améliorer en la rendant indépendante des différentes transformations de l'image. On les doit essentiellement à [Huttenlocher et al., 1993] pour l'invariance à la translation et à [Huttenlocher et Rucklidge, 1993] pour l'invariance au changement d'échelle. Plus récemment, dans [Rucklidge, 1997], c'est l'invariance aux transformations affines des coordonnées qui est étudiée, afin de traiter le cas d'objets plans projetés par projection perspective.

5.1.3 Comparaison d'histogrammes

Dans la section 3.3.2.1 du chapitre précédent, nous avons présenté une approche de caractérisation couleur très répandue et basée sur les histogrammes. Diverses variantes plutôt satisfaisantes de ces histogrammes existent, intégrant pour certaines l'invariance aux changements d'illumination. Pour procéder à l'appariement, il reste maintenant à comparer ces histogrammes entre eux. La tâche se révèle délicate, d'une part parce que la mise en place d'une distance entre histogrammes est loin d'être triviale et d'autre part parce que l'algorithme de mise en correspondance doit être capable de gérer efficacement la masse d'information souvent imposante que contiennent les histogrammes. Nous avons recensé plusieurs méthodes de comparaison, elles sont présentées dans la suite de cette section.

5.1.3.1 Intersections d'histogrammes

La première méthode, sans doute la plus référencée, est celle proposée par Swain et Ballard [Swain et Ballard, 1991]. Elle est basée sur l'intersection d'histogrammes. L'histogramme du modèle est comparé aux histogrammes des images de la base selon la formule ci-après, n étant le nombre de couleurs de l'espace discret des couleurs :

$$Inter(H,M) = \frac{\sum_{i=1}^{n} min(h_{c_i}, m_{c_i})}{\sum_{i=1}^{n} m_{c_i}} \tag{5.6}$$

La valeur obtenue est entre 0 et 1, 1 indiquant une bonne ressemblance. Swain met également en place dans [Swain, 1993] une méthode empirique pour rendre l'appariement

moins coûteux en temps de calcul. Seules les couleurs les plus représentées dans les images et le modèle sont comparées. On obtient ainsi une intersection d'histogrammes partielle.

Dans [Funt et Finlayson, 1995], Funt et Finlayson utilisent également l'intersection d'histogrammes sur des histogrammes qui ont été rendus insensibles aux changements d'illumination par la méthode CCCI.

Un avantage de cette fonction d'intersection est qu'elle néglige explicitement les pixels du fond qui peuvent apparaître dans l'histogramme du modèle mais pas dans l'histogramme de la base. Elle n'implique pas non plus un lourd niveau de calcul puisque seulement deux opérations (le minimum et l'addition) sont nécessaires par cellule d'histogramme. La complexité est linéaire par rapport à la dimension n de l'histogramme : $O(n)$.

Un inconvénient de la fonction d'intersection est que toutes les cellules sont traitées de manière égale, impliquant l'équiprobabilité des couleurs. Ennesser et Medioni ont proposé dans [Ennesser et Medioni, 1993] une pondération de la fonction d'intersection par des poids w_i pour chaque cellule. Ce poids w_i doit être choisi proportionnel à l'importance de la couleur c_i associée. L'application des ces fonctions pondérées permet d'améliorer la discrimination d'objets.

Dans ce contexte de caractérisation par histogramme couleur, une image peut également être vue comme étant un point dans l'espace de l'histogramme de dimension n. Diverses méthodes de calcul de distance dans cet espace ont alors été proposées.

5.1.3.2 Mesures de distance

Si l'on considère le cadre général des espaces métriques, l'intersection d'histogrammes telle qu'elle a été présentée à l'équation 5.6 ne peut pas être considérée comme une distance, le calcul n'étant pas symétrique à cause de la normalisation au dénominateur. Pour mesurer la distance entre deux histogrammes, la distance L_1 est utilisée par certains auteurs comme notamment [Mehtre et al., 1995; Stricker et Orengo, 1995] ou encore [Huang et al., 1999] pour la comparaison de corrélogrammes[8]. Sa formulation est la suivante :

$$d_{L_1}(H, M) = \sum_{i=1}^{n} |h_{c_i} - m_{c_i}| \tag{5.7}$$

En réalité, comme Swain le remarque dans [Swain, 1993], il existe sous certaines conditions une similitude entre l'intersection des histogrammes et la norme L_1. Ainsi si l'on suppose que toutes les images contiennent le même nombre de pixels N, on obtient la relation suivante :

$$1 - Inter(H, M) = \frac{d_{L_1}(H, M)}{2N} \tag{5.8}$$

Cependant la norme L_1 ne s'avère pas la plus adaptée pour la comparaison d'histogrammes, car elle ne tient pas suffisamment compte de la ressemblance entre les couleurs.

[8] Se référer à l'équation 3.30 du chapitre précédent pour la définition des corrélogrammes couleur.

D'autres types de normes plus sophistiquées ont alors été mises en place, comme la distance euclidienne dans sa version généralisée définie par [Niblack et al., 1993] et présentée à l'équation 5.9 :

$$d_{L_2,A}(H,M) = \sqrt{(H-M)A(H-M)^T} \tag{5.9}$$

A représente une matrice de pondération permettant d'inclure dans le calcul la ressemblance entre les couleurs. On rencontre parfois dans la littérature le carré de cette mesure, on parle dans ce cas de distance quadratique [Schiele, 1997]. Notons tout de même que dans la pratique, la matrice employée est bien souvent la matrice identité. La mesure effectuée est alors une simple distance euclidienne, et souffre des mêmes inconvénients que la norme L_1. La comparaison porte le nom de Test de Cramer - von Mises.

Dans [Stricker et Orengo, 1995] où la caractérisation est réalisée à partir d'histogrammes cumulatifs, les auteurs réalisent leurs expérimentations avec les normes L_1 et L_2. Ils introduisent également la norme L_∞ (appelée Test de Kolmogorov - Smirnov) définie ci-après :

$$d_{L_\infty}(H,M) = \max_{1 \leq i \leq n} |h_{c_i} - m_{c_i}| \tag{5.10}$$

Stricker montre que le test de Kolmogorov - Smirnov fournit de meilleurs résultats que les deux autres tests et plus généralement que les comparaisons à partir d'histogrammes cumulatifs sont plus robustes que celles basées sur les histogrammes classiques au sens de Swain. Ce test est également utilisé et évalué par Sethi dans [Sethi et Patel, 1995], qui le juge comme étant le meilleur après celui du χ^2 présenté à la section suivante.

5.1.3.3 Tests d'hypothèses

Comme nous l'avons vu dans la partie dédiée à la caractérisation, un histogramme couleur peut également être considéré comme la réalisation d'une variable aléatoire donnant les couleurs dans une image. Dans ce contexte, la comparaison d'histogrammes peut se ramener à un problème de tests d'hypothèses, dans lequel il s'agit de déterminer si les deux histogrammes peuvent provenir de la même distribution. Cette représentation a été introduite simultanément par les équipes de Sethi [Sethi et Patel, 1995] et de Stricker [Stricker et Orengo, 1995]. Plusieurs mesures sont mises en place. Il s'agit en premier lieu du rapport de ressemblance de Yakimovsky, présenté à l'équation 5.11 ci-après :

$$y(H,M) = \frac{(\sigma^2)^{n_H+n_M}}{(\sigma_H^2)^{n_H}(\sigma_M^2)^{n_M}} \tag{5.11}$$

où σ^2 est la variance de la population formant les deux histogrammes, σ_I^2 la variance de l'histogramme I et n_H et n_M les tailles de chacune des populations.

Le test de dissimilarité du χ^2 est également employé pour déterminer la similitude entre deux distributions discrètes. Il est basé sur l'hypothèse que les distributions considérées sont gaussiennes. La mesure la plus communément rencontrée est donnée par :

$$\chi^2(H,M) = \begin{cases} \sum_j \frac{(h_{c_j}-m_{c_j})^2}{(h_{c_j}+m_{c_j})^2} & \text{si} \quad h_{c_j}+m_{c_j} \neq 0 \\ 0 & \text{sinon} \end{cases} \tag{5.12}$$

Les expériences menées par les deux équipes sur un ensemble de données expérimentales ont montré que cette dernière mesure conduit à de meilleurs résultats que l'intersection d'histogrammes, la distance euclidienne généralisée ou encore le test de Kolmogorov - Smirnov. Elle a tout de même l'inconvénient de ne pas être métrique, l'inégalité triangulaire n'étant pas satisfaite. Schiele propose alors dans [Schiele, 1997] une fonction modifiée qui elle, est métrique. L'auteur précise qu'elle donne des taux de reconnaissance élevés par rapport aux autres fonctions, mais qu'elle n'a pas la même qualité de résultats que les statistiques χ^2 originales.

Toujours dans [Schiele, 1997], une étude est menée sur la stabilité de la plupart des fonctions de comparaison que nous venons de présenter. Leur robustesse est évaluée en présence de bruit gaussien, de flou (simulé par un filtrage répété utilisant la moyenne), mais aussi face aux rotations de l'image, aux changements d'illumination. Les tests révèlent un comportement instable de la fonction d'intersection en présence de bruit, de flou et de rotations image. Les résultats sont même les plus mauvais pour le changement d'éclairage. Encore une fois, ce sont les tests basés sur le χ^2 qui s'avèrent les plus stables quelles que soient les transformations considérées.

Citons pour finir un exemple d'application récent utilisant cette mesure de comparaison : nous avions entre autre détaillé dans la partie caractérisation (cf. section 3.3.2.1) les travaux de Siggelkow et Schael sur les histogrammes couleur, qui visaient à améliorer les histogrammes au sens de Swain pour mieux tenir compte de la distribution spatiale des couleurs tout en gardant l'invariance à la rotation. Dans [Siggelkow et Schael, 1999], ils utilisent cette méthode de caractérisation, accompagnée du test du χ^2 de l'équation 5.12 pour faire de la classification de textures.

5.1.4 Appariement basé sur les invariants différentiels

Dans le chapitre précédent sur la caractérisation, nous avons passé en revue un certain nombre d'invariants différentiels qui permettent de caractériser les points d'intérêt. Que l'on considère le jet local (cf. section 3.2.1.1), les filtres directionnels (cf. section 3.2.1.2) ou encore les invariants de Hilbert (cf. section 3.2.1.3), il s'agit à présent d'être capable de comparer les ensembles d'invariants engendrés. C'est ainsi que l'on pourra procéder à l'appariement des points.

Le problème est donc de décider si deux ensembles d'invariants sont similaires. La première idée qui vient à l'esprit est d'utiliser la distance euclidienne L_2, mais elle s'avère inappropriée ici car les différentes composantes du vecteur d'invariants n'ont pas nécessairement le même ordre de grandeur ni les mêmes tolérances au bruit. Il faut donc envisager d'autres méthodes plus sophistiquées. Les plus utilisées sont la corrélation et la distance de Mahalanobis, elles sont présentées ci-après. Le problème de la mise en correspondance dans un contexte multi-échelle est également traité à la section 5.1.4.3.

5.1.4.1 La corrélation

Il est possible de calculer un score de corrélation entre les deux vecteurs à comparer. Dans [Rao et Ballard, 1995], Rao utilise cette solution avec la formule de corrélation NCC

(cf. table 5.1). Il obtient de bons résultats, ceci dû en grande partie au fait qu'il travaille sur des vecteurs de grande dimension contenant le jet local directionnel jusqu'à l'ordre 3 et à 5 niveaux d'échelle, soit 45 invariants. Il montre en effet que les espaces de grande dimension tendent à être orthogonaux. On entend par vecteurs orthogonaux deux vecteurs pour lesquels on est incapable de dire s'ils se ressemblent beaucoup ou pas du tout. Donc si la taille du vecteur d'invariants est importante, seulement une minorité des scores calculés sera très proche de 1 alors que la majorité sera localisée autour de 0, ce qui les rend plus discriminants dans la recherche des bons couples. On comprend donc aisément l'intérêt de cumuler les invariants différentiels à plusieurs niveaux d'échelle lorsque cet outil de comparaison est utilisé.

5.1.4.2 La distance de Mahalanobis

Une autre solution, plus rigoureuse mais plus difficile à mettre en œuvre, est d'utiliser la distance de Mahalanobis, en modélisant les composantes du vecteur par des variables aléatoires gaussiennes. C'est la solution adoptée par Schmid [Schmid, 1996] pour comparer les invariants différentiels calculés jusqu'au troisième ordre en niveau de gris.

La distance de Mahalanobis est une distance statistique qui utilise la matrice de covariance Λ des composantes du vecteur. Elle est donnée par l'équation ci-après, pour deux vecteurs $\vec{v}_1$ et $\vec{v}_2$ à comparer :

$$\delta(\vec{v}_1, \vec{v}_2) = (\vec{v}_1 - \vec{v}_2)^T \Lambda^{-1} (\vec{v}_1 - \vec{v}_2) \tag{5.13}$$

Cette distance se révèle être particulièrement bien appropriée pour comparer les vecteurs d'invariants car la matrice Λ permet de tenir compte dans le calcul de la variation de chacune des composantes ainsi que de leur éventuelle corrélation. La qualité des résultats dépend donc très fortement de la représentativité de cette matrice. Mais son estimation est très complexe à réaliser de façon théorique puisque la forme du signal autour du point d'intérêt est quelconque. Une solution consiste à l'estimer de manière empirique. Comme c'est une grandeur statistique, son estimation nécessite la donnée de n observations de ces composantes. La méthode employée est la suivante : étant donné un point d'intérêt apparaissant sur n images, les n vecteurs d'invariants qui lui sont associés sont calculés, ce qui permet d'estimer la matrice de covariance associée à ce point. Ce calcul est très délicat à réaliser car une erreur de seulement un pixel dans l'appariement des n observations engendrera des erreurs dans toute la matrice. Il doit aussi tenir compte du bruit des images utilisées. Afin donc de minimiser les erreurs mais aussi d'obtenir une matrice représentative de la variété des points envisageables, le calcul est effectué pour k séquences de points sur des scènes les plus variées possibles. La matrice de covariance globale est enfin calculée en prenant la moyenne des k matrices calculées. La qualité des résultats obtenus dépend bien sûr de la qualité de l'appariement des observations, du nombre de séquences de points utilisées et de la variété des images choisies.

La distance de Mahalanobis est également intéressante car c'est une variable aléatoire qui suit la distribution du χ^2 à p degrés de liberté, pour des vecteurs d'invariants de dimension p. Il est donc possible d'utiliser une table de cette distribution pour seuiller la distance et ainsi rejeter les appariements qui ont la plus grande probabilité d'être faux, c'est-à-dire ceux qui correspondent aux plus grandes valeurs de la distance.

Changement de base

Une alternative existe pour simplifier l'utilisation de la distance de Mahalanobis lors de la comparaison des vecteurs d'invariants. En réalité, si n est la dimension des vecteurs traités, alors l'ensemble des vecteurs qui peuvent correspondre à un vecteur donné se situe dans un ellipsoïde à n dimensions, centré autour de ce vecteur. Il est alors possible de réaliser un changement de base permettant l'utilisation de la distance euclidienne pour comparer les deux vecteurs.

La matrice de covariance Λ étant réelle symétrique et semi-définie positive, il est possible de la décomposer comme suit :

$$\Lambda^{-1} = P^T D P = P^T \sqrt{D}\sqrt{D} P \tag{5.14}$$

où P est orthogonale et D positive diagonale. La distance de Mahalanobis δ peut donc être décomposée de la façon suivante :

$$\begin{aligned} \delta(\vec{v}_1, \vec{v}_2)^2 &= (\vec{v}_2 - \vec{v}_1)^T P^T \sqrt{D}\sqrt{D}(\vec{v}_2 - \vec{v}_1) \\ &= \left[\sqrt{D}P(\vec{v}_2 - \vec{v}_1)\right]^T \left[\sqrt{D}P(\vec{v}_2 - \vec{v}_1)\right] \\ &= \left[\sqrt{D}P\vec{v}_2 - \sqrt{D}P\vec{v}_1\right]^T \left[\sqrt{D}P\vec{v}_2 - \sqrt{D}P\vec{v}_1\right] \\ &= \left\|\sqrt{D}P\vec{v}_2 - \sqrt{D}P\vec{v}_1\right\|^2 \end{aligned} \tag{5.15}$$

Calculer la distance de Mahalanobis entre deux vecteurs est donc équivalent à transformer ces deux vecteurs en les multipliant par la matrice $\sqrt{D}P$ puis à calculer la distance euclidienne entre les deux vecteurs obtenus.

5.1.4.3 Contexte multi-échelle

Nous avons vu dans la section 3.2.1.4 du chapitre précédent qu'il est conseillé de calculer les vecteurs d'invariants à plusieurs niveaux échelles, afin de rendre la caractérisation encore plus discriminante, mais également pour être capable de gérer les changements d'échelle entre deux images. Dans ce dernier cas, le facteur d'échelle étant en général inconnu, on a en effet recours à une approche multi-échelle dans laquelle les dérivées discrètes sont calculées sur plusieurs supports. Comme elle a intégré la détection des points d'intérêt, l'approche multi-échelle doit alors également gérer la comparaison des vecteurs caractéristiques. Il s'agit d'être capable de sélectionner pour chaque couple de points comparés les vecteurs qui correspondent au bon facteur d'échelle entre les deux images.

Dans ces conditions, une première méthode de comparaison a été développée par Rao dans [Rao et Ballard, 1995]. La méthode de caractérisation qu'il utilise est celle basée sur les filtres directionnels ; elle a été présentée à la section 3.2.1.2. Les invariants sont calculés à cinq niveaux d'échelle. L'appariement est alors réalisé en comparant par corrélation les vecteurs d'invariants de chaque paire de points en leur ayant fait au préalable subir un ou plusieurs «décalages», permettant d'envisager plusieurs changements d'échelle. On retient ensuite les appariements correspondant au décalage pour lequel la mise en correspondance

a obtenu les meilleurs scores. Bien sûr cette méthode suppose une connaissance au moins approximative du changement d'échelle existant entre les deux images, pour appliquer le décalage adéquat. Les expérimentations menées montrent cependant que cette technique ne résiste pas à des changements d'échelle supérieurs à 10%.

Dans [Schmid et Mohr, 1996], Schmid développe une méthode qui ne nécessite pas la connaissance du changement d'échelle. Les supports σ_k de gaussienne sont calculés par rapport à un support de référence σ_0 que l'on fait varier d'un facteur 1.2^n, n étant un entier. Par exemple, pour $\sigma_0 = 1$ et $n = 4$, on obtient les supports 0.48, 0.58, 0.69, 0.83, 1, 1.2, 1.44, 1.73, 2.07. Les invariants sont alors calculés par rapport à toutes ces échelles. Puis, pour chaque point, on compare le vecteur d'invariants associé au support σ_0 aux vecteurs d'invariants des autres points. Le couple de points associé aux supports (σ_0, σ_j) ayant le meilleur score d'appariement est conservé. Le processus est ensuite réalisé dans l'autre sens, en retenant les meilleurs couples ayant (σ_i, σ_0) pour supports. Parmi les deux listes d'appariements obtenues, les couples qui seront définitivement conservés sont ceux qui se sont sélectionnés mutuellement et pour lesquels les supports σ_i et σ_j sont complémentaires, i.e. $\sigma_j = \sigma_i^{-1}$. Les auteurs montrent que les résultats sont satisfaisants (avec des taux d'appariements corrects autour de 90%) pour des changements d'échelle allant jusqu'à 1.5 et se dégradent nettement au delà.

La mise en correspondance au travers des changements d'échelle se révèle en réalité plus délicate à mettre en œuvre que prévu. Outre l'impact important du changement de résolution sur la détection des points ainsi que sur les données photométriques, il s'avère que la partie commune entre les deux images à apparier peut se retrouver très réduite. Le risque de faux appariements en est donc augmenté, particulièrement lorsque l'on utilise une description locale du signal pour réaliser cet appariement.

Dans [Dufournaud et al., 2000], une approche plus complète est proposée pour tenter de pallier ces problèmes. Un espace d'échelle est construit sur l'image possédant la plus haute résolution (la plus proche de la scène). Les auteurs prennent 20 niveaux d'échelle pour leurs expérimentations. Pour trouver parmi tous les facteurs s_i celui qui s'approche le plus du facteur réel, ils définissent un critère $C(s_i)$ qui doit être maximum lorsque l'appariement des points entre les deux images est correct pour l'échelle s_i considérée. Ce critère correspond au nombre de points appariés correctement par l'algorithme de mise en correspondance et est déterminé automatiquement en utilisant la transformation affine qui lie les deux images (cf. l'équation 1.8). Celle-ci est estimée de façon robuste par la méthode RANSAC sur les points qui ont été appariés par des méthodes locales [Schmid, 1996]. Cette vérification par application d'une contrainte globale permet d'accepter ou de rejeter les appariements à partir d'une contrainte physique réelle : le mouvement de la caméra. A l'issu de la mise en correspondance, le facteur du changement d'échelle le plus proche de la réalité est également déterminé. Il faut noter qu'à partir du moment où la méthode robuste estime correctement la transformation globale, les points sont nécessairement appariés correctement. Les auteurs précisent que cette estimation est toujours fiable si le taux de mauvais appariements initiaux est inférieur à 50%. Les expérimentations menées se révèlent probantes pour des facteurs d'échelle allant jusqu'à 6. Remarquons tout de même que la qualité de l'appariement se dégrade lorsque la scène a un relief trop prononcé, le changement d'échelle entre les deux images n'étant plus dans ce cas exactement le même partout.

5.1.5 La procédure d'appariement

Une fois les scores d'appariement calculés pour tous les couples de points, il s'agit de déterminer ceux qui ont la plus grande probabilité d'être bien appariés, en étudiant les scores obtenus. Le but de la procédure d'appariement consiste même à ne garder que des paires dans lesquelles chaque point n'est impliqué qu'une seule fois, pour respecter la contrainte d'unicité[9] selon laquelle un point doit avoir un correspondant unique dans l'autre image [Marr et Poggio, 1976]. Plusieurs approches sont possibles :

5.1.5.1 Algorithme "Winner Takes All"

C'est la méthode la plus simple. Elle consiste à fournir un correspondant à chaque point de la première image en ne retenant que le correspondant associé au meilleur score d'appariement. Elle respecte la contrainte d'unicité mais a les principaux inconvénients de ne pas être symétrique (elle ne fournirait en effet pas les mêmes appariements en partant de l'autre image) et d'imposer un correspondant à un point qui pourrait être occulté dans l'autre image. Une première amélioration consiste donc à réaliser une vérification croisée. La méthode est présentée dans la prochaine section.

5.1.5.2 Mise en correspondance par appariement croisé

La méthode de mise en correspondance par appariement croisé fournit directement un ensemble de couples de points qui a l'avantage d'être symétrique. Chaque point apparié n'est impliqué que dans un seul appariement mais tous les points ne trouvent pas nécessairement de correspondant.

Pour chaque point des deux images, on sélectionne le couple qui a le score le plus élevé. Les paires de points retenues sont celles qui ont été mutuellement sélectionnées, comme l'illustre la figure 5.2.

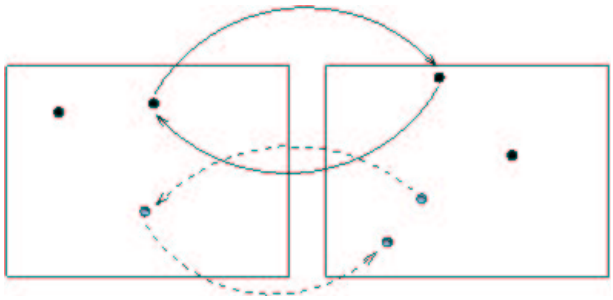


Figure 5.2 – *Illustration de la mise en correspondance par appariement croisé.*

Cette méthode est plus satisfaisante que la précédente. Mais dans la pratique plusieurs points d'intérêt d'une même image peuvent posséder des caractéristiques similaires. Cela conduit à calculer pour un même point des scores d'appariement très proches, dits *ambigus*,

[9]Rappelons d'ailleurs que cette contrainte peut être violée pour des objets transparents ou lorsque plusieurs objets sont alignés.

pour ces correspondants potentiels. En présence de bruit, il est donc possible d'être amené à sélectionner à tort un couple de points, sous prétexte qu'il possède le score le plus élevé. Dans [Rao et Ballard, 1995], Rao suggère l'utilisation de vecteurs de caractéristiques de grande dimension pour diminuer cette probabilité d'erreur. Ceci n'est pas praticable pour tous les types de caractérisation, comme notamment avec les invariants différentiels qu'il faudrait calculer à des ordres de dérivation plus importants. Dans ce contexte, la solution par appariement croisé apparaît trop directe pour donner des résultats réellement satisfaisants. L'utilisation de méthodes moins "sévères", comme le seuillage accompagné de contraintes semi-locales, nous semble plus approprié.

5.1.5.3 Utilisation d'une fonction de seuillage

Le seuillage des scores d'appariement obtenus consiste à éliminer de l'ensemble des appariements les couples possédant un score inférieur à une certaine valeur donnée. Dans le cas de la corrélation (cf. section 5.1.2.1) ou de la distance de Hausdorff (cf. section 5.1.2.3), ce travail est délicat puisqu'on ne connaît pas a priori la valeur du seuil qu'il faut considérer. Il est la plupart du temps choisi arbitrairement par rapport à la moyenne ou la médiane de l'ensemble des scores. En revanche, la distance de Mahalanobis (cf. section 5.1.4.2) étant une variable aléatoire qui suit la distribution du χ^2, il est possible d'utiliser une table de cette distribution pour calculer le seuil. En ce qui concerne les normes L_1 et L_2 entre histogrammes (cf. section 5.1.3.2), une étude a été menée dans [Stricker et Swain, 1994] portant entre autre sur l'estimation du seuil à considérer.

Une fois les appariements improbables éliminés, il reste un ensemble de couples de points qui constituent les appariements potentiels. Ces couples ne respectent en général pas la contrainte d'unicité. Afin d'éliminer les éventuelles ambiguïtés de cet ensemble, il est d'usage d'utiliser des contraintes de cohérence basées sur le voisinage du point ainsi que des contraintes géométriques, comme nous allons le voir dans la prochaine section.

5.2 Contraintes géométriques

La liste de contraintes géométriques que nous énumérons dans cette section n'est pas exhaustive. Elle donne les principales contraintes utilisées par les algorithmes de stéréoscopie. Il s'agit tout d'abord à la section 5.2.1 d'une contrainte basée sur la géométrie épipolaire du système stéréoscopique. Puis à la section 5.2.2, nous développons la technique dite de relaxation, qui utilise quant à elle des contraintes semi-locales basées sur le voisinage.

D'autres contraintes existent, comme la contrainte d'ordre, de continuité ou encore la limite du gradient de disparité[Pollard et al., 1985]. Mais elles sont peu utilisées dans la pratique car elles ne s'appliquent pas dans de nombreux cas. Selon les applications envisagées, d'autres contraintes plus spécifiques sont parfois utilisées, comme par exemple lorsqu'on a une connaissance a priori du contenu de la scène à traiter ou des limites maximales et minimales de sa profondeur.

5.2.1 La géométrie épipolaire

La contrainte épipolaire est la seule relation géométrique qui lie deux images stéréoscopiques dans un contexte non calibré. Pour d'avantage de détails, le lecteur peut consulter la section 7.3.1.1 du chapitre 7 qui lui est dédiée. Cette contrainte est très utile ici puisqu'elle permet de ramener le problème bidimensionnel de la mise en correspondance à un problème unidimensionnel de recherche le long des lignes épipolaires conjuguées. C'est la contrainte géométrique la plus puissante et elle est toujours vérifiée.

5.2.2 Contraintes semi-locales : la relaxation

Il existe une classe d'algorithmes itératifs très populaires en Vision par Ordinateur, utilisés entre autres par [Faugeras et Berthod, 1981; Horaud et Monga, 1993; Zhang et al., 1995b; Laveau, 1996; Schmid, 1996], dits algorithmes de relaxation. Les problèmes qu'ils permettent de résoudre s'expriment sous la forme de la minimisation d'une fonction d'énergie globale, appelée alors critère de relaxation. Cette fonction décrit les interactions locales entre les primitives que l'on souhaite extraire et un ensemble d'observations. Cette minimisation, non linéaire dans la plupart des cas et portant sur un nombre très élevé de variables, est usuellement menée par les algorithmes de relaxation. Dans le cadre de recherche qui est le nôtre, ces approches semblent bien appropriées pour pallier le problème des correspondances ambiguës, en permettant de générer un ensemble de correspondances cohérent. Il faut donc définir un critère de relaxation qui prenne en compte ces considérations, l'étape de la relaxation consistant à le minimiser.

Le critère de relaxation traditionnellement utilisé est basé sur des contraintes de voisinage du point. Dans ce contexte, la relaxation consiste à propager, inhiber ou renforcer des hypothèses d'appariements par rapport à leur voisinage. Une probabilité initiale mesurant la qualité de l'hypothèse locale est affectée à chaque appariement. Cette probabilité est révisée suivant la compatibilité du voisinage. De cette manière, les points d'intérêt se réorganisent en propageant ces contraintes. Après un certain nombre d'itérations, le système converge vers un ensemble de correspondances non ambiguës suffisamment grand. Notons que cette technique impose des coûts de calcul relativement importants.

Une contrainte de voisinage

La contrainte de voisinage qui est classiquement utilisée prend en compte la configuration relative des points d'intérêt. Ainsi une paire de points (m_1, m_2) est considérée comme étant un bon appariement si dans le voisinage de m_1 on trouve un assez grand nombre de points bien appariés avec des points du voisinage de m_2. Cette idée est illustrée par la figure 5.3.

Cette contrainte est valable quelles que soient les transformations géométriques ou photométriques existant entre les deux images à apparier. Elle est notamment utilisée par [Zhang et al., 1995b; Schmid, 1996].

La contrainte de voisinage est généralement accompagnée de contraintes géométriques qui permettent une vérification supplémentaire des appariements trouvés. Pour [Schmid, 1996] par exemple, la contrainte est basée sur la conservation des angles, qui est toujours

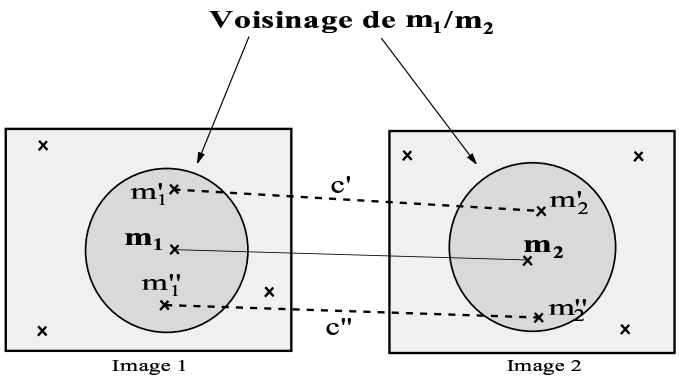


Figure 5.3 – *Contraintes de voisinage autour d'un couple de points* (m_1, m_2) *en correspondance.*

vérifiée dans le cas des similitudes entre images. Ainsi l'angle défini par deux voisins d'un point doit être constant pour toutes les vues de ce point. D'autres contraintes sont également envisageables, comme la distance entre le point étudié et ses voisins [Zhang et al., 1995b]. Nous verrons dans le prochain chapitre qu'il est possible de mettre en œuvre un système de contraintes géométriques sophistiqué permettant de réduire considérablement le nombre de faux appariements.

5.2.3 Autres contraintes géométriques

La géométrie épipolaire est la contrainte géométrique bi-linéaire qui existe entre deux images. Si plus de deux caméras sont utilisées pour la mise en correspondance, il est possible d'utiliser des contraintes multi-linéaires d'ordre supérieur, afin de rendre les appariements encore moins ambigus. Dans le cas de trois caméras, les trilinéarités sont la plus connue des contraintes géométriques. Elle ont été introduites par Shashua [Shashua, 1994] dans un contexte de reconnaissance d'objets à partir de vues différentes, et sont équivalentes au tenseur trifocal de Hartley [Hartley, 1994]. Ces contraintes sont exprimées par quatre relations qui existent entre les trois projections d'un point de la scène tridimensionnelle. Comme pour la géométrie épipolaire, il est possible de les estimer à partir de correspondances éparses, lorsque la calibration n'est pas disponible. Pour une plus ample description des trilinéarités, le lecteur peut se reporter au chapitre 7 de ce mémoire, et plus généralement à la thèse de Laveau [Laveau, 1996] pour la géométrie des systèmes de N caméras.

5.3 Mise en correspondance dense

L'appariement dense consiste à mettre en correspondance autant de pixels que possible dans les deux images. Vouloir apparier les pixels uniquement à l'aide des méthodes présentées dans la section 5.1 n'est pas applicable, l'espace de recherche étant bien trop important. Il est fondamental d'utiliser des contraintes géométriques supplémentaires telles que la contrainte épipolaire (cf. section 5.2.1), la limite de disparité, les contraintes de

cohérence globale (les appariements ne sont cherchés que dans des régions voisines des appariements des voisins). Parmi toutes les contraintes qui peuvent être envisagées, la géométrie épipolaire est toujours vérifiée dans le cas de scènes rigides et permet de réduire l'espace de recherche des correspondances. Lorsque les caméras sont calibrées, elle est directement disponible. Si ce n'est pas le cas, des méthodes robustes permettant son estimation existent (elles seront développées au chapitre 7) ; mais elles nécessitent la donnée d'un certain nombre de mises en correspondance éparses. Il est donc classique de procéder au préalable à la mise en correspondance de points d'intérêt avant d'envisager une mise en correspondance dense.

Des méthodes d'optimisation ont également été mises en place en vue d'améliorer les résultats de mise en correspondance et d'en réduire encore la complexité. Les principales méthodes rencontrées sont passées en revue :

5.3.1 Rectification des images

Lorsque la contrainte épipolaire est utilisée pour procéder à l'appariement, le balayage du faisceau s'avère fastidieux. C'est pourquoi on effectue couramment une rectification des images avant de les apparier. Cette opération consiste à reprojeter les images originales sur un plan parallèle à la droite (C_1, C_2) passant par les centres optiques des deux caméras, appelé alors plan de rectification.

Par cette transformation, les épipôles des images rectifiées sont repoussés à l'infini le long de (C_1, C_2) qui correspond aussi à la direction des abscisses des images rectifiées. Les lignes épipolaires se retrouvent ainsi alignées avec les lignes de l'image, ce qui rend le parcours de la droite plus simple à implémenter. A l'issue de la mise en correspondance, les images sont ensuite "dérectifiées".

La rectification a cependant l'inconvénient d'introduire dans le traitement deux étapes d'interpolation qui sont sources d'imprécision. Dans le cas où l'angle de convergence entre les deux caméras est faible (les lignes épipolaires sont quasi horizontales), l'étape de rectification entraîne une perte de précision sans vraiment apporter d'information complémentaire. Pour d'avantage de détails, le lecteur peut consulter [Devernay, 1997] qui consacre un chapitre de sa thèse à l'examen des méthodes de rectification, selon des points de vue géométriques puis algébriques et présente également une méthode de rectification locale.

5.3.2 Extension des méthodes d'appariement épars

Il s'agit simplement ici d'appliquer la procédure d'appariement épars présentée à la section 5.1.5, mais sur un plus grand nombre de points, avec au moins la contrainte épipolaire et le plus souvent une contrainte de disparité.

La méthode est la suivante : pour chaque pixel p_1 de la première image, on calcule un score d'appariement (selon une des méthodes présentées à la section 5.1) pour tous les pixels appartenant à la droite épipolaire associée à p_1. Il est ensuite possible de déterminer le correspondant potentiel de p_1 en utilisant la méthode WTA (cf. section 5.1.5.1) ou plus

efficacement, la méthode par appariement croisé (cf. section 5.1.5.2). Si c'est la seconde méthode qui est employée, la carte de disparité obtenue comporte des zones non renseignées, correspondant le plus souvent à des occultations. Cette approche a été utilisée pour la première fois dans [Fua, 1991]. Il est également possible de lui adjoindre une fonction de seuillage.

Une variante de cette technique, que l'on doit à [Pollard et al., 1985], consiste à calculer le score d'appariement entre deux pixels en prenant la somme des scores d'appariement entre les pixels de leurs voisinages respectifs. Cette idée n'est d'ailleurs pas sans rappeler la contrainte de voisinage de la relaxation mais appliquée ici sur des pixels (cf. section 5.2.2). La technique d'appariement est également sensiblement différente, puisque que dès qu'un couple est retenu après avoir fait la vérification croisée, les pixels impliqués dans l'appariement sont éliminés de l'espace de recherche. Cela conduit à construire un ensemble de mises en correspondance plus important qu'avec la méthode classique par appariement croisé, et certainement moins fiable puisque des points qui n'auraient pas dû avoir de correspondant avec la méthode classique, à cause de scores d'appariements trop faibles, peuvent se retrouver appariés. Cependant la technique mérite d'être citée, de par la nouveauté du calcul des scores d'appariement. En effet, elle permet d'intégrer une contrainte de gradient de disparité, en plus de la contrainte épipolaire et de la contrainte d'unicité.

Jusqu'à présent, c'est traditionnellement la corrélation qui est utilisée pour calculer les scores d'appariements, avec toutes les contraintes que cette technique impose (cf. section 5.1.2.1). Pour diminuer les temps de calcul trop coûteux, les adeptes de la corrélation, citons bien sûr [Hannah, 1989], travaillent sur l'image à plusieurs niveaux de résolution. Ils utilisent l'approche hiérarchique[10] développée initialement par Moravec [Moravec, 1977]. Des travaux très récents [Lew et Huang, 1999] permettent même d'optimiser la recherche dans le graphe de la hiérarchie en utilisant l'algorithme A*, issu du domaine de l'intelligence artificielle.

5.3.3 Résolution par programmation dynamique

Les algorithmes de programmation dynamique sont apparus il y a une quinzaine d'années. Ils appliquent cette technique de recherche opérationnelle à l'appariement de deux listes de points ordonnées. Le principe est une recherche de chemin optimal dans un graphe. Celui-ci est construit en plaçant en abscisse les points de la première image et en ordonnées les points de la seconde. Chaque nœud du graphe représente un appariement potentiel. Une fonction de coût élémentaire est définie pour chaque portion du chemin reliant deux nœuds successifs. Pour apparier les deux listes de points, on examine les scores d'appariement de chaque nœud (i, j). Le score obtenu détermine si le nœud étudié constitue un appariement valide, puis en fonction du résultat, les nœuds $(i+1, j)$, $(i, j+1)$ et $(i+1, j+1)$ sont examinés à leur tour, en tenant compte du coût élémentaire de chacune de ces transitions. L'algorithme de programmation dynamique permet de trouver le chemin optimal menant du premier au dernier nœud, le coût total du chemin étant égal à la somme des coûts élémentaires des transitions le composant. La contrainte d'unicité est

[10] Consulter [Dyer, 1987] pour un excellent état de l'art sur les approches hiérarchiques et leurs applications.

implicitement respectée, ainsi que la contrainte d'ordre. Dans le cadre de la mise en correspondance d'images, il est nécessaire d'intégrer au calcul la géométrie épipolaire, la notion d'ordre n'ayant un sens que sur les lignes épipolaires conjuguées. La bonne marche du système dépend en fait du réglage des coûts élémentaires entre chaque nœud. Ils s'avèrent très délicats à régler, car ils peuvent entièrement changer le comportement de l'algorithme. De même, la contrainte d'ordre s'avère être un inconvénient pour de nombreuses scènes, où les "croisements" entre appariements ne sont pas rares[11].

5.3.4 Approches énergétiques

Il s'agit ici d'optimiser globalement la mise en correspondance dense entre les deux images. Chaque configuration de mise en correspondance représente une certaine énergie, que l'on essaie de minimiser.

Une implémentation utilisant un recuit simulé est proposée dans [Ouali et al., 1996]. L'ensemble des appariements constitue la configuration du système. L'énergie est alors calculée à partir de la valeur de corrélation des points, d'un terme de lissage (sauf sur les contours pour ne pas lisser les ruptures de disparité) ainsi que d'une contrainte sur le nombre de points occultés. Cette dernière contrainte sert à éviter de converger vers une solution pour laquelle aucun point ne serait apparié et où l'énergie engendrée serait nulle. Les contraintes d'unicité et épipolaire sont assurées par la forme des configurations autorisées. Cependant, l'implémentation n'est réalisée que sur des images de très petites dimensions et les pondérations des différents termes d'énergie ne sont pas connues, rendant ainsi l'évaluation de la méthode difficile.

Dans [Robert et Deriche, 1996], le problème est posé de la même façon mais ici on cherche à minimiser l'énergie sur la profondeur Z de chaque point. La fonction d'énergie est composée d'un terme d'énergie $M_{12}(Z)$ ainsi que d'un terme de régularisation $S(Z)$. $M_{12}(Z)$ définit une mesure de similarité entre les points en correspondance dans les deux images. Dans la pratique, il intègre le niveau de gris du point mais également des données multi-dimensionnelles comme le gradient ou la couleur. En outre, les systèmes comportant plus de deux oculaires peuvent facilement être gérés par une simple sommation sur $M_{1j}(Z)$. Le terme régularisateur $S(Z)$ permet de contraindre la forme de la fonction de profondeur Z. Il doit préserver les discontinuités. Les auteurs choisissent un terme de filtrage anisotropique qui n'opère pas dans les directions perpendiculaires aux contours. Les résultats obtenus sur un couple d'images synthétiques sont très bons. Cependant, la méthode nécessite un étalonnage des caméras pour le calcul des profondeurs Z.

5.3.5 Régularisation des appariements

Les algorithmes de mise en correspondance dense ne calculent généralement pas un appariement rigoureusement dense. De plus, ils ne renvoient une valeur de disparité que pour les pixels où le calcul est possible, c'est-à-dire les pixels non occultés. Dans ce contexte, la carte de disparité finale comportera des zones non renseignées. L'étape de régularisation consiste à remplir les zones manquantes par une information de disparité.

[11] Citons l'exception des images aériennes, pour lesquelles la contrainte d'ordre est toujours respectée.

Un filtrage simple ne peut pas convenir, car il aurait pour effet de lisser la carte et les frontières d'occultation deviendraient mal définies. Un filtre médian convient mieux ; il préserve les contours, mais a l'inconvénient d'étaler les appariements isolés en des amas de faux appariements. Il est aussi possible d'appliquer un filtrage anisotropique. Lorsque la calibration des caméras est connue, on peut également envisager d'approcher les points reconstruits connus par un modèle de surface 3D continu. Le modèle mathématique donne alors une valeur de disparité interpolée pour les points manquants.

5.4 Conclusion

Dans ce chapitre, nous avons fait un état de l'art des différentes méthodes permettant de mettre en correspondance deux images. Les principales méthodes permettant de comparer les caractérisations présentées dans la première partie de cette thèse ont tout d'abord été passées en revue. Nous nous sommes d'avantage intéressés à la comparaison des vecteurs d'invariants différentiels, puisque c'est cette approche, dans sa version couleur, que nous avons exploitée pour caractériser les points d'intérêt. Dans ce cadre, la méthode de comparaison la plus robuste est celle qui utilise la distance de Mahalanobis. Rappelons cependant qu'elle fait intervenir la matrice de covariance des composantes du vecteur, dont l'estimation est empirique et relativement fastidieuse à mettre en œuvre.

Toujours dans le contexte des invariants différentiels de Hilbert, si l'on veut considérer les changements d'échelle entre images, il est nécessaire de mettre en place une approche multi-échelle pour les comparer. Si le facteur d'échelle est connu, alors un lissage gaussien adapté permettra de comparer les vecteurs à des échelles différentes. S'il est inconnu, alors il faudra envisager de calculer les invariants à plusieurs niveaux d'échelle, ou alors, comme l'a récemment proposé [Dufournaud et al., 2000], estimer ce facteur d'échelle à l'aide d'une méthode RANSAC. Dans tous les cas, il est très important de considérer également la détection des points d'intérêt dans un contexte multi-échelle.

Vient ensuite l'étape délicate du seuillage des scores d'appariements obtenus (scores de corrélations, intersections d'histogrammes ou encore distances). Nous avons expliqué qu'il n'est pas possible de se contenter simplement de sélectionner les couples associés aux meilleurs scores. Il est plus judicieux de mettre en œuvre une fonction de seuillage, éliminant ainsi les couples de points qui ont toutes les chances d'être des faux appariements. Les couples de points restants pouvant être ambigus, c'est-à-dire pouvant être impliqués dans plusieurs appariements, il est nécessaire d'utiliser d'autres types de contraintes permettant d'exhiber les bonnes mises en correspondance. Ces contraintes sont de nature géométrique. Il y a bien sûr la contrainte épipolaire, mais surtout les contraintes semi-locales exploitées par l'algorithme de relaxation. Cet algorithme itératif permet de "désambiguiser" les appariements en mettant en jeu des contraintes de voisinage.

Une fois l'ensemble de mises en correspondance éparses obtenu, il est possible d'envisager la mise en correspondance dense des deux images. Celle-ci requiert en effet la connaissance de la géométrie épipolaire pour être réalisable. Si elle n'est pas disponible, alors il s'agira de l'estimer à partir des appariements des points d'intérêt.

Les techniques d'appariement qui viennent d'être présentées donnent pour la plupart

des résultats de mise en correspondance intéressants, dans la mesure où le nombre de points d'intérêt mis en jeu n'est pas très important. Pour des grands nombres de points, la mise en correspondance devient beaucoup trop lourde à mettre en œuvre. Or il est pourtant souvent nécessaire de travailler avec des grands ensembles de points ; c'est le cas par exemple pour la reconstruction tridimensionnelle de la scène à partir d'images stéréoscopiques. C'est pourquoi nous proposons dans le prochain chapitre une méthode de mise en correspondance robuste pouvant s'appliquer sur de grands nombres de points. nous montrerons que cette approche permet d'obtenir des résultats d'appariements supérieurs dans des temps de calcul beaucoup plus raisonnables.

Chapitre 6

Une méthode de mise en correspondance robuste

Nous avons présenté au chapitre 4 une nouvelle méthode de caractérisation de points d'intérêt. Ce chapitre se consacre en premier lieu à la mise en place d'une métrique permettant de l'exploiter pour l'étape de la mise en correspondance. Des contraintes géométriques invariantes aux principales transformations de l'image sont également introduites. Puis, nous présentons une méthode de mise en correspondance robuste rendant efficace l'appariement d'un très grand nombre de points. Le processus complet d'appariement est ensuite testé et comparé aux approches traditionnelles.

Sommaire

Dans le précédent chapitre, nous avons passé en revue les méthodes de mise en correspondance existantes. Notre méthode de caractérisation consiste à utiliser les invariants différentiels couleur présentés au chapitre 4. Nous avons vu que le vecteur d'invariants obtenu est traditionnellement comparé avec la distance de Mahalanobis. Cette grandeur est la plus rigoureuse, mais est généralement difficile à estimer. C'est pourquoi nous présentons à la section 6.1 la méthode d'appariement que nous avons mise en place pour comparer deux vecteurs d'invariants. A cette méthode, s'ajoutent l'étape de la relaxation. Nous présentons également les contraintes semi-locales géométriques que nous utilisons.

Malheureusement, comme pour les méthodes classiques de mise ne correspondance, notre approche s'avère posséder une trop grande complexité pour gérer efficacement les grands nombres de points. C'est pourquoi nous proposons à la section 6.2 un algorithme incrémental de mise en correspondance. Il consiste à réaliser l'appariement par itérations, en ajoutant au fur et à mesure des itérations des points à apparier. Ces travaux ont fait l'objet d'une publication [Gouet et al., 1998a].

La méthode complète de mise en correspondance est ensuite évaluée à la section 6.3. L'algorithme de base, présenté dans la première section, est d'abord évalué face aux principales transformations de l'image, permettant ainsi de tester la robustesse de notre caractérisation basée sur les invariants différentiels couleur. Enfin le processus incrémental est étudié puis validé sur de grands ensembles de points. Il est comparé à la version non incrémentale correspondante. Nous verrons que les résultats obtenus en l'utilisant sont nettement supérieurs, qu'il s'agisse des taux d'appariements corrects obtenus ou des temps de calcul.

6.1 La méthode d'appariement

6.1.1 Comparaison des vecteurs

Nous avons vu à la section 5.1.4 du chapitre 5 comment il était possible de comparer deux vecteurs d'invariants. La méthode la plus robuste est sans aucun doute la distance de Mahalanobis (cf. 5.1.4.2). Elle est cependant très difficile à estimer. Nous avons donc utilisé une méthode intermédiaire, beaucoup plus simple et qui fournit pourtant d'excellents résultats, comme nous le verrons plus loin. Comme chaque composante du vecteur d'invariants est défini dans un intervalle spécifique, il n'est pas possible d'appliquer directement la norme euclidienne pour réaliser la comparaison. C'est pourquoi nous normalisons chaque composante du vecteur dans un intervalle fixé, à partir des extrema de la distribution de la composante issue des vecteurs des deux images. Ce changement de base simplifié est envisageable dans la mesure où le vecteur d'invariants n'implique que des dérivées gaussiennes d'ordre un, et est par conséquent robuste au bruit. On calcule ensuite une norme euclidienne entre les deux vecteurs normalisés. Les distances obtenues sont seuillées pour ne garder que les couples de points susceptibles de correspondre à des appariements corrects. Enfin, nous employons un algorithme de relaxation (cf. section 5.2.2) pour éliminer les ambiguïtés dans les appariements restants. Les contraintes semi-locales de voisinage et géométriques utilisées sont détaillées à la prochaine section.

6.1.2 Contraintes semi-locales de relaxation

Idéalement si les invariants utilisés pour décrire les points étaient complètement discriminants, seul le critère mis en place ci-dessus serait suffisant et permettrait l'appariement des points. Mais ce n'est pas le cas et une image peut comporter de nombreux points ayant une caractérisation similaire. De plus, l'inconvénient de l'algorithme d'appariement employé est qu'il peut générer un ensemble de couples non cohérent dans le sens où un point peut être impliqué dans plusieurs correspondances de points. Nous utilisons un algorithme de relaxation pour éliminer les correspondances ambiguës de cet ensemble de correspondances potentielles. Il s'agit donc de définir en premier lieu un critère de relaxation, l'étape de la relaxation consistant à le minimiser.

6.1.2.1 Le critère de relaxation

Le critère de relaxation est défini pour un ensemble $\mathcal{E}$ de correspondances de points. A chacun des couples potentiels (m_1, m_2) de $\mathcal{E}$ où m_1 appartient à la première image et m_2 à la seconde, on associe un score de relaxation $\mathcal{SM}(m_1, m_2)$ qui est toujours positif et qui mesure la force de l'agencement (m_1, m_2). Le critère de relaxation noté C est donc défini comme la somme des scores de relaxation de toutes les correspondances de points de $\mathcal{E}$:

$$C(\mathcal{E}) = \sum_{(m_{1i}, m_{2j}) \in \mathcal{E}} \mathcal{SM}(m_{1i}, m_{2j}) \tag{6.1}$$

Ainsi minimiser ce critère revient à choisir l'ensemble des bonnes correspondances de points.

6.1.2.2 Le score de relaxation

Un moyen de diminuer le risque d'avoir des ambiguïtés est de filtrer les mises en correspondance en ajoutant des contraintes semi-locales de cohérence basées sur le voisinage d'un point, ainsi que des contraintes géométriques.

La contrainte de voisinage

Comme dans [Zhang et al., 1995b; Schmid, 1996], nous partons du principe qu'une paire de points (m_1, m_2) est un bon appariement si dans le voisinage $\mathcal{V}(m_1)$ de m_1 on trouve un assez grand nombre de points bien appariés avec des points du voisinage $\mathcal{V}(m_2)$ de m_2. Cette idée est illustrée par la figure 6.1.

Les contraintes géométriques

Nous imposons également à ces points voisins des contraintes de nature géométrique. Ces contraintes vont dépendre de la nature des transformations reliant les deux images. La méthode de description basée sur les invariants différentiels couleur que nous avons présentée est invariante au groupe des similitudes, c'est-à-dire à la rotation et aux changements d'échelle. Il faut donc mettre en place des contraintes géométriques qui tiennent compte de ces degrés de liberté.

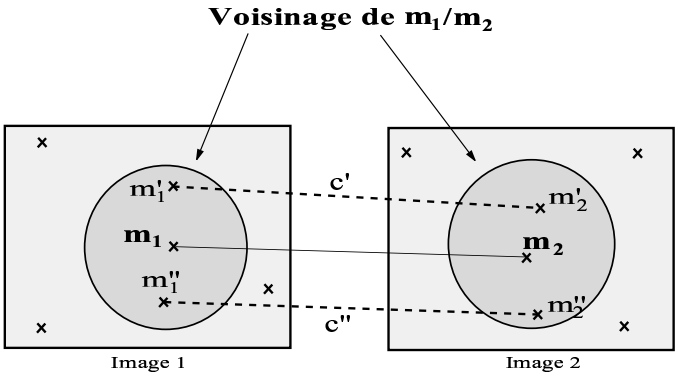


Figure 6.1 – *Contrainte de voisinage relative au couple de points* (m_1, m_2).

Notre méthode consiste à utiliser, en plus du nombre de points bien appariés entre les deux voisinages $\mathcal{V}(m_1)$ et $\mathcal{V}(m_2)$, la conservation des angles entre les points voisins. Dans [Schmid, 1996], Schmid propose de tenir compte de l'angle défini par les deux voisins du point considéré. Cet angle doit être globalement constant pour toutes les vues de ce point. Nous avons, quant à nous, mis en place une contrainte angulaire basée sur le gradient multi-spectral des points considérés. Cette contrainte a une complexité moins importante que celle utilisée par Schmid, car elle ne considère ici qu'un seul voisin à la fois. Comme l'illustre la figure 6.2, l'angle entre le gradient du point étudié et celui des voisins doit être constant d'une vue à l'autre. Une telle conservation est vérifiée dans le cas des similitudes entre images.

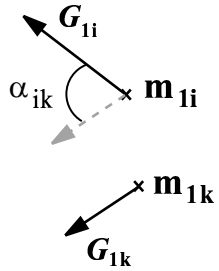


Figure 6.2 – *Contrainte géométrique angulaire entre le gradient du point considéré* m_{1i} *et celui de son voisin* m_{1k}.

Cette contrainte est alors exprimée pour le couple (m_{1i}, m_{2j}) par le score de relaxation noté $\mathcal{SM}$ suivant :

$$\mathcal{SM}(m_{1i}, m_{2j}) = c_{ij} \sum_{m_{1k} \in \mathcal{V}(m_{1i})} \left[\max_{m_{2l} \in \mathcal{V}(m_{2j})} P(m_{1i}, m_{2j}; m_{1k}, m_{2l}) \right] \tag{6.2}$$

avec

- $P(m_{1i}, m_{2j}; m_{1k}, m_{2l}) = c_{kl}.\alpha(m_{1i}, m_{2j}; m_{1k}, m_{2l})$;
- c_{ij} et c_{kl} : scores d'appariement de respectivement (m_{1i}, m_{2j}) et (m_{1k}, m_{2l}) calculés à l'étape précédente (par comparaison des vecteurs d'invariants) et positifs ;

- $\alpha(m_{1i}, m_{2j}; m_{1k}, m_{2l}) = f(|angle(\vec{G}_{m_{1i}}, \vec{G}_{m_{1k}}) - angle(\vec{G}_{m_{2j}}, \vec{G}_{m_{2l}})|)$, $\vec{G}_m$ représentant le gradient au point m ;
- $f(x) = \frac{1-x}{\xi_a}$ si $x < \xi_a$ et 0 sinon ;

La force ainsi calculée n'est pas symétrique : plusieurs couples de points (m_{1k_1}, m_{1k_2}) de $\mathcal{V}(m_{1i})$ peuvent avoir un poids P maximum pour le même couple de points (m_{2l_1}, m_{2l_2}) de $\mathcal{V}(m_{2j})$. Dans ce cas, le calcul symétrique ne donnerait pas les mêmes résultats. Notre algorithme rend le calcul symétrique en ne prenant en compte que le couple (m_{1k_1}, m_{1k_2}) associé au poids maximum calculé le plus grand. Le gradient multi-spectral est calculé selon la méthode développée à la section 1.3.1.1 du chapitre 1.

6.1.2.3 L'algorithme de relaxation

La minimisation du critère de relaxation C est faite de façon itérative. On calcule d'abord le critère de relaxation $C(\mathcal{E})$ à partir de l'ensemble des couples de points produit à l'étape de la mise en correspondance. Puis à chaque itération, on sélectionne les couples de points pour lesquels le score de relaxation $\mathcal{SM}$ est supérieur à celui des couples dont un des deux points est impliqué dans le couple étudié. Ces derniers couples sont alors éliminés de $\mathcal{E}$. On recalcule ensuite $C(\mathcal{E})$ et on réitère jusqu'à ce que $C(\mathcal{E})$ ne décroît plus. L'algorithme peut être résumé en 6.1 :

Répéter
  Pour *chaque couple* $(m_{1\alpha}, m_{2\beta}) \in \mathcal{E}$ **Faire**
    Sélectionner $(m_{1\alpha}, m_{2\beta})$;
    Pour *chaque couple* $(m_{1\alpha}, m_{2j}) \in \mathcal{E}$ *tel que* $j \neq \beta$ **Faire**
      Si $\mathcal{SM}(m_{1\alpha}, m_{2\beta}) < \mathcal{SM}(m_{1\alpha}, m_{2j})$ **Alors** Désélectionner $(m_{1\alpha}, m_{2\beta})$;
    Si $(m_{1\alpha}, m_{2\beta})$ *est sélectionné* **Alors**
      Pour *chaque couple* $(m_{1i}, m_{2\beta}) \in \mathcal{E}$ *tel que* $i \neq \alpha$ **Faire**
        Si $\mathcal{SM}(m_{1\alpha}, m_{2\beta}) < \mathcal{SM}(m_{1i}, m_{2\beta})$ **Alors** Désélectionner $(m_{1\alpha}, m_{2\beta})$;
  Eliminer de $\mathcal{E}$ les couples non sélectionnés ;
  Calculer $C(\mathcal{E})$;
Tant que $C(\mathcal{E})$ *décroît*;

Algorithme 6.1 – *Le processus de relaxation.*

En sortie, l'ensemble $\mathcal{E}$ contient des correspondances de points cohérentes. Le processus itératif est arrêté dès qu'on ne peut plus sélectionner de couples c'est-à-dire lorsque $C(\mathcal{E})$ ne décroît plus. Il converge obligatoirement puisqu'à chaque itération, au moins un couple de points est ôté de $\mathcal{E}$ dont la dimension est finie.

6.1.2.4 Le degré d'ambiguïté

Le calcul du score de relaxation $\mathcal{SM}$ n'élimine pas entièrement la notion d'ambiguïté d'une correspondance de points. En effet, un couple est sélectionné dès que sa force d'agencement est la plus importante parmi celles des autres couples engendrés par les points impliqués ; cependant cette sélection peut s'avérer ambiguë si cette force est tout juste

supérieure aux autres. On a donc choisi de définir pour chaque couple de points sélectionné un degré d'ambiguïté noté $\mathcal{UA}$ et compris entre 0 et 1 tel que :

$$\mathcal{UA}(m_{1i}, m_{2j}) = 1 - \frac{\max[\mathcal{SM}(m_{1i}, m_{2k}), \mathcal{SM}(m_{1l}, m_{2j})]}{\mathcal{SM}(m_{1i}, m_{2j})} \tag{6.3}$$

$\mathcal{SM}(m_{1i}, m_{2k})$ et $\mathcal{SM}(m_{1l}, m_{2j})$ représentent les forces d'agencement des couples impliquant m_{1i} et m_{2j} et immédiatement inférieures à $\mathcal{SM}(m_{1i}, m_{2j})$. Plus formellement :

$$\begin{cases} \exists k \neq j \ / \ \mathcal{SM}(m_{1i}, m_{2k}) > \mathcal{SM}(m_{1i}, m_{2k'}) \ \forall k' \neq k, j. \\ \exists l \neq i \ / \ \mathcal{SM}(m_{1l}, m_{2j}) > \mathcal{SM}(m_{1l'}, m_{2j}) \ \forall l' \neq l, i. \end{cases} \tag{6.4}$$

Le couple (m_{1i}, m_{2j}) sélectionné est non ambigu si sa force $\mathcal{SM}(m_{1i}, m_{2j})$ est très supérieure à celles des couples impliquant m_{1i} et m_{2j}. Dans ce cas, $\mathcal{UA}(m_{1i}, m_{2j})$ est proche de 1. Le processus de relaxation est donc modifié afin de prendre en compte le critère d'ambiguïté : pour être retenu, un agencement de points sélectionné doit avoir un degré d'ambiguïté suffisamment fort. Dans la pratique, à chaque itération on trie les forces d'agencement et les degrés d'ambiguïté de chaque couple par ordre décroissant et on ne sélectionne que les couples qui possèdent les meilleurs scores $\mathcal{SM}$ et $\mathcal{UA}$. Pour plus de détails, consulter [Zhang et al., 1995b].

6.1.2.5 Elimination des faux appariements

A l'issue de la relaxation, un ensemble de mises en correspondances non ambiguës est obtenu. Nous estimons alors la géométrie épipolaire à partir de ces appariements, en utilisant une méthode robuste de type LMedS (présentée à la section 7.6.1.3 du chapitre 7) qui met en œuvre des critères adéquats visant à minimiser les distances des points aux droites épipolaires. Cette méthode permet de mettre en place une ultime phase d'élimination des faux appariements pouvant subsister, c'est-à-dire ceux qui ne respectent pas la contrainte épipolaire.

6.1.3 Limites de la méthode

Le principal inconvénient de la méthode de mise en correspondance qui vient d'être présentée est en premier lieu sa complexité. Si aucune information de disparité n'est disponible entre les deux images (c'est le cas lorsque les caméras ne sont pas calibrées), alors pour m points dans la première image et n dans la seconde, la méthode de comparaison seule a une complexité en $O(m \times n)$; pour la relaxation, elle est en $O(m^2 \times n^2)$ dans la pire des configurations. En conséquence, un autre inconvénient est le nombre de couples ambigus résultants, qui augmente avec le nombre de points à apparier. Cela rend l'algorithme de relaxation plus coûteux en temps de calcul et finalement peut générer un plus grand nombre de mauvais appariements. En résumé, le processus d'appariement décrit précédemment est efficace jusqu'à 200 ou 300 points mais devient rapidement inutilisable au delà. Il est pourtant souvent nécessaire de mettre en correspondance un très grand nombre de points pour réaliser des cartes de profondeur denses entre deux images. Nous

proposons dans la section suivante une solution pour améliorer les résultats de mise en correspondance sur de grands nombres de points.

6.2 Un algorithme de mise en correspondance incrémental

Notre approche [Gouet et al., 1998a] consiste à utiliser des contraintes géométriques qui permettront de localiser les points à apparier beaucoup plus précisément. Si la disparité entre les images n'est pas connue, nous devons trouver cette zone nous-mêmes. Supposons que nous ayons à notre disposition un ensemble $\mathcal{M}$ d'appariements corrects entre les deux images. Nous montrons alors dans la section suivante comment cette donnée nous fournit des informations sur la zone recherchée.

6.2.1 L'information géométrique disponible

Nous présentons dans cette section deux types de contraintes géométriques. Elles vont nous permettre de réduire considérablement la zone de recherche du correspondant d'un point dans l'autre image. Les contraintes mises en place ici sont invariantes aux principales transformations de l'image.

6.2.1.1 La géométrie épipolaire

Si $\mathcal{M}$ contient au moins sept appariements, alors il est possible d'estimer la géométrie épipolaire du système des deux caméras (on obtient trois solutions avec sept appariements et une solution unique pour un nombre supérieur, cf. section 7.6.1 du chapitre 7). Celle-ci est caractérisée par une matrice fondamentale $F_{(3\times3)}$ qui vérifie $m_2^T F m_1 = 0$ pour deux points appariés m_1 et m_2. Cette équation traduit le fait que le point m_2 dans la seconde image est situé sur la droite épipolaire $F m_1$ et réciproquement que le point m_1 de la première image est situé sur la droite $F^T m_2$. Dès l'instant où F est connue, il est facile de voir que la complexité de la méthode d'appariement est réduite, puisque la zone de recherche du point devient une droite. Dans la suite, la matrice fondamentale estimée à partir de l'ensemble de mises en correspondances $\mathcal{M}$ sera noté $F_{\mathcal{M}}$.

6.2.1.2 La triangulation de Delaunay

Dans cette section, nous définissons une contrainte de semi-planarité locale, basée sur la triangulation de Delaunay.

Considérons un point tridimensionnel P appartenant à un triangle T, (p_1, p_2) ses projections sur deux images et (t_1, t_2) les projections du triangle. Il est facile de montrer qu'un triangle est transformé en un triangle par transformation projective, ainsi t_1 et t_2 sont également des triangles. Le point p_1 est nécessairement situé dans le triangle t_1 et a nécessairement son correspondant p_2 dans t_2. Si P n'appartient pas à T, la position de p_2 relativement à t_2 est fonction de la disparité. L'expérience a montré qu'il suffit de considérer t_2 et ses plus proches triangles voisins. Ainsi la disparité est inversement

proportionnelle au nombre de points qui peuvent être appariés et donc proportionnelle à l'aire des triangles traités. C'est la raison pour laquelle la combinaison de t_2 avec ses plus proches voisins représente une zone qui toutes les chances de contenir p_2.

Nous venons donc de définir une zone de l'image qui contient le correspondant p_2 d'un point p_1. La triangulation est calculée sur la première image, puis "appliquée" sur la deuxième à partir des mises en correspondance : un triangle de la seconde image doit avoir ses trois sommets appariés avec les sommets d'un même triangle dans la première image. Nous avons opté pour une triangulation de Delaunay en ce qui concerne la première image, parce qu'elle produit des triangles les plus équi-angulaires possibles [Boissonnat et Teillaud, 1986]. La deuxième triangulation engendrée dans l'autre image n'est alors pas obligatoirement une triangulation de Delaunay. Nous noterons $\mathcal{T}_{\mathcal{M}}$ les deux triangulations construites à partir des appariements $\mathcal{M}$.

En combinant les deux contraintes géométriques présentées ci-dessus, la recherche d'un correspondant dans une image est réduite à un segment, si le point étudié appartient à un triangle de la triangulation de Delaunay. Si ce n'est pas le cas, alors la recherche est réduite à un ensemble de segments, voire au pire la droite épipolaire entière. La figure 6.3 illustre cette idée : (a, a'), (b, b') et (c, c') représentent des appariements. Un point m de l'image 1 situé dans le triangle (a, b, c) a son correspondant dans l'image 2 sur la droite épipolaire $F.m$ et dans le triangle (a', b', c'). Réciproquement, le point m' a son correspondant sur la droite $F^T.m'$ et dans le triangle (a, b, c).

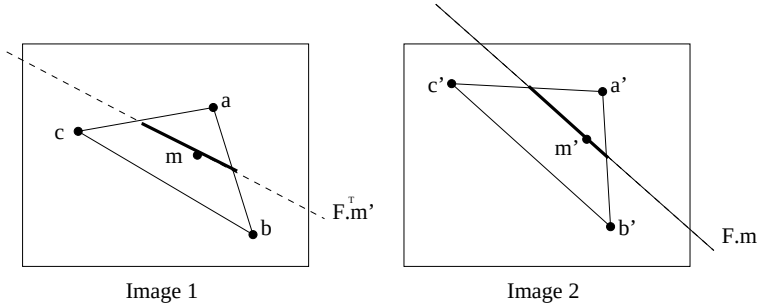


Figure 6.3 – *Mise en correspondance à l'aide des contraintes géométriques.*

Ces nouvelles contraintes géométriques permettent de réduire l'espace de recherche du point à apparier et ce, quelles que soient les images. Elles vont alors être intégrées dans notre processus incrémental de mise en correspondance. Ce processus est présenté à la section suivante.

6.2.2 L'algorithme incrémental de mise en correspondance

Les contraintes présentées ci-dessus supposent que nous ayons à notre disposition un ensemble de points appariés pour initialiser le nouveau processus d'appariement. En effet, notre approche consiste à mettre en place une méthode d'appariement incrémentale qui calcule à l'itération i un ensemble d'appariements $\mathcal{M}^i$ à partir des contraintes géométriques associées aux appariements $\mathcal{M}^{i-1}$ de l'itération $i-1$. L'algorithme peut alors être décomposé en 6 étapes résumées en 6.2 [Gouet et al., 1998a].

Répéter

· Extraire (ou ajouter) les points d'intérêt dans les deux images. On obtient les ensembles de points $\mathcal{P}_1^i$ et $\mathcal{P}_2^i$;

· Caractériser chacun des points à l'aide du vecteur caractéristique $\vec{v}_{col}$ (cf. equation 4.2) ;

· Pour chaque point $p_{1k} \in \mathcal{P}_1^i$ et $p_{2l} \in \mathcal{P}_2^i$, estimer la zone de recherche dans l'autre image, à partir des contraintes géométriques $F_{\mathcal{M}_{i-1}}$ et $\mathcal{T}_{\mathcal{M}_{i-1}}$ si elles existent. On obtient $\mathcal{A}_{1k}$ et $\mathcal{A}_{2l}$;

· Comparer les vecteurs caractéristiques de p_{1k} et p_{2l} qui vérifient : $p_{1k} \in \mathcal{A}_{2l}$ et $p_{2l} \in \mathcal{A}_{1k}$. L'ensemble d'appariements $\mathcal{M}_\star^i$ avec de possibles ambiguïtés est obtenu ;

· Eliminer les appariements ambigus de $\mathcal{M}_\star^i$ par la relaxation (cf. section 6.1.2). On obtient l'ensemble d'appariements $\mathcal{M}^i$ sans ambiguïtés ;

· Calculer les contraintes géométriques : la triangulation $\mathcal{T}_{\mathcal{M}^i}$ et la matrice fondamentale $F_{\mathcal{M}^i}$ associée aux appariements $\mathcal{M}^i$;

Tant que $\mathcal{M}^i$ *n'est pas assez grand*;

Algorithme 6.2 – *Mise en correspondance incrémentale.*

Etant donné que les appariements $\mathcal{M}^i$ sont calculés à partir de ceux obtenus à l'itération $i-1$, il est très important d'obtenir le plus grand nombre de bons appariements à chaque itération. Cette condition est vérifiée la plupart du temps parce que les contraintes géométriques permettent d'éliminer la majeure partie des appariements incorrects. En effet, les points qui ne vérifient pas la contrainte épipolaire sont éliminés et l'expérience a montré que la plupart des points restants incorrects sont eux aussi isolés grâce à la contrainte de triangulation, ce que ne font généralement pas les approches classiques de mise en correspondance. Il est intéressant de remarquer que la taille des zones de recherche (des segments) diminue au fur et à mesure des itérations. Ainsi le processus d'appariement gagne en temps de calcul et voit les ambiguïtés diminuer à chaque itération.

La principale difficulté de notre approche réside dans l'estimation du premier ensemble $\mathcal{M}^0$ pour lequel aucune contrainte géométrique n'est disponible. Ces appariements doivent pourtant être de très bonne qualité puisqu'ils permettent d'initialiser tout le processus incrémental. Notre solution consiste à estimer une matrice fondamentale $F_{\mathcal{M}^0}$ de manière robuste à la fin de la relaxation et dans un premier temps d'éliminer tous les appariements qui ne sont pas consistants avec $F_{\mathcal{M}^0}$. Puis les faux appariements restants sont éliminés en ne gardant qu'un faible pourcentage des meilleurs appariements (ceux qui possèdent le meilleur score de relaxation. L'expérience a montré que retenir uniquement 40% des appariements permettait d'obtenir une excellente base pour les itérations suivantes. Notons que la contrainte $F_{\mathcal{M}^0}$ peut être ré-estimée à partir de ces nouveaux appariements avant d'envisager la première itération.

6.3 Evaluation de la méthode

Il est bien sûr possible d'évaluer visuellement les résultats de notre méthode de mise en correspondance. Cependant, comme pour les détecteurs de points d'intérêt, il nous a semblé plus judicieux de mettre en œuvre une méthode d'évaluation quantitative et automatique. Nous allons voir dans la section suivante que plusieurs possibilités s'offrent à nous, selon le type de scène tridimensionnelle considérée.

6.3.1 Les méthodes d'évaluation automatique

Lorsque le changement de point de vue d'une image à l'autre est quelconque, il n'existe pas de relation directe entre deux projections d'un point de la scène tridimensionnelle. Il est cependant possible d'établir une relation particulière, une transformation homographique, lorsque la scène est plane. La technique employée est développée dans la prochaine section. Nous présenterons également à la section 6.3.1.2 une méthode plus générale basée sur la géométrie épipolaire.

6.3.1.1 Cas de scènes planes

Si les deux images sont issues d'une scène plane, alors tous les appariements $(m_1, m2)$ doivent être en correspondance homographique (cf. section 7.3.3 du chapitre 7). Ils doivent donc vérifier :

$$m_2 = H_{12}.m_1 \tag{6.5}$$

Une méthode pour évaluer la qualité de la mise en correspondance consiste donc à calculer les distances entre les points et leurs correspondants idéaux selon H_{12}. Soit d_{ij} l'erreur calculée, en terme de distance, pour le couple de points appariés (m_{1i}, m_{2j}) et telle que :

$$d_{ij} = \frac{dist(H_{12}.m_{1i}, m_{2j}) + dist(m_{1i}, H_{21}.m_{2j})}{2} \tag{6.6}$$

où $dist(m_1, m_2)$ représente la distance euclidienne entre les deux points m_1 et m_2. La mesure de répétabilité mise en place pour évaluer les détecteurs de points d'intérêt intégrait la distribution des distances afin de tenir compte de la précision de la détection. La notion de précision n'étant pas nécessaire ici, nous définissons un critère d'évaluation des mises en correspondance qui se contente de comptabiliser le nombre de points correctement appariés selon la contrainte homographique. On obtient le critère $\mathcal{N}_\varepsilon$ suivant :

$$\mathcal{N}_\varepsilon = |\{(m_{1i}, m_{2j}) \ / \ d_{ij} < \varepsilon\}| \tag{6.7}$$

où ε représente la taille du voisinage dans lequel on s'autorise à chercher le correspondant du point étudié. Il est en effet bien sûr impossible de ne considérer que les couples de points associés à des distances nulles, les points détectés étant une approximation de la réalité. Comme pour la mesure de répétabilité, le choix d'une valeur pour ce paramètre peut être guidé par l'équation 2.7, selon la précision sub-pixellique envisagée.

Toute la difficulté de cette méthode d'évaluation réside dans le calcul de H_{12}. Il est possible de l'estimer directement à partir des appariements trouvés, à l'aide d'une technique de moindres carrés médians. La méthode peut cependant s'avérer biaisée si les appariements ne sont pas précis dans une proportion suffisamment significative, puisqu'elle se sert des appariements pour établir un critère permettant justement de les évaluer. Certains auteurs, comme [Brand, 1995; Schmid, 1996], préfèrent donc s'appuyer sur des cibles circulaires très précisément appariées dans les images et vérifiées manuellement.

6.3.1.2 Cas de scènes quelconques

Si la scène est quelconque, alors le seul lien géométrique existant entre deux points appariés est la géométrie épipolaire, décrite par la matrice fondamentale F_{12}. Un appariement (m_1, m_2) peut être considéré comme correct si le point m_2 appartient à la droite épipolaire correspondant au point m_1. Comme pour l'homographie, on obtient la distance suivante :

$$d'_{ij} = \frac{dist(F_{12}.m_{1i} = 0, m_{2j}) + dist(F_{21}.m_{2j} = 0, m_{1i})}{2} \tag{6.8}$$

où $dist(d, m)$ représente cette fois la distance euclidienne entre la droite d et le point m. Comme pour l'homographie, on obtient la mesure $\mathcal{N'}_\varepsilon$ suivante :

$$\mathcal{N'}_\varepsilon = |\{(m_{1i}, m_{2j}) \ / \ d_{ij} < \varepsilon\}| \tag{6.9}$$

Dans ce cas, ε permet de traduire l'appartenance ou la non appartenance dans l'autre image du correspond du point étudié à la droite épipolaire associée.

Une telle évaluation n'est pas exempte d'erreur puisque deux points d'un appariement faux peuvent vérifier la contrainte épipolaire. Cependant la probabilité d'un tel événement est suffisamment faible pour que cet estimateur fournisse une bonne évaluation du nombre d'appariements corrects. Comme avec l'homographie, il est possible d'utiliser une méthode robuste pour estimer la matrice fondamentale F_{12}. Mais l'estimation s'avère encore plus biaisée, car le calcul de F_{12} est encore plus sensible à la précision des appariements.

6.3.2 Cadre de l'évaluation

Dans la section suivante, nous allons évaluer la méthode de mise en correspondance qui vient d'être présentée. Les points d'intérêt utilisés sont les points de Harris Précis Couleur calculés en précision demi-pixel [Montesinos et Dattenny, 1997] avec comme paramètres du lissage gaussien $\sigma = 1$ et $\widetilde{\sigma} = 2$. Pour le seuillage par maxima locaux, le diamètre de la fenêtre est de 15 et la valeur du seuil varie en fonction de l'image, de façon à obtenir approximativement le même nombre de points dans les deux images. Les points extraits sont ensuite caractérisés par le vecteur d'invariants différentiels couleur $\vec{v}_{col}$, calculé à l'aide des dérivées d'ordre un avec un paramètre de lissage gaussien plus important ($\sigma = 3$) pour minimiser le bruit. Les images sont normalisées face aux changements d'illumination selon la méthode introduite à la section 4.3 du chapitre 4. La méthode d'appariement est celle qui a été détaillée à la section 6.1. Elle est intégrée ou non, selon le cas, dans le processus de mise en correspondance incrémental présenté à la section 6.2.

La robustesse de la mise en correspondance est évaluée par rapport aux principales transformations de l'image, c'est-à-dire par rapport aux translations, aux rotations 2D, aux changements d'illumination, aux changements d'échelle et aux changements de point de vue. Nous avons tout d'abord utilisé la méthode d'évaluation présentée à la section 6.3.1.1 et basée sur l'estimation d'une homographie. Les scènes utilisées sont donc des scènes planes. Nous avons travaillé à partir des mêmes scènes synthétiques bruitées que celles utilisées pour évaluer la répétabilité des détecteurs de points au chapitre 2. La méthode permettant l'estimation de l'homographie est similaire à celle développée à la section 2.1.4 du même chapitre. On obtient ainsi une homographie que l'on peut qualifier d'exacte. Puis nous avons considéré le cas de scènes quelconques pour évaluer le processus complet de mise en correspondance sur de grands ensembles de points. La technique d'évaluation est donc celle développée à la section 6.3.1.2.

6.3.3 Résultats

Nous présentons ici deux types de résultats :

- En premier lieu, les résultats de l'évaluation du processus d'appariement des points d'intérêt face aux quatre principales transformations de l'image (rotation, changement d'illumination, changement d'échelle et changement de point de vue). Seul l'algorithme de base (cf. 6.1) est utilisé sur un nombre de points raisonnable. Les résultats sont présentés sous forme de graphe, mettant en évidence le nombre d'appariements qui auraient dus être trouvés (les appariements effectifs), le nombre d'appariements trouvés et le nombre d'appariements corrects $\mathcal{N}_\varepsilon$ selon la contrainte homographique (cf. équation 6.7). Dans tous les cas, la valeur du seuil de répétabilité ε est fixé à 2, sachant que la détection des points est réalisée en précision demi-pixel. Cette étude fait l'objet des quatre prochaines sections.
- C'est ensuite le processus de mise en correspondance complet, incluant donc l'algorithme incrémental, qui est testé. Les images proviennent d'une scène non plane et différent à la fois en point de vue et en illumination (ce dernier étant naturel et donc de nature inconnue). Les résultats sont comparés avec ceux obtenus avec la méthode d'appariement de base et sont présentés sous forme de tableaux. On constate une nette amélioration, tant dans la qualité des mises en correspondance obtenues que dans les temps de calcul. Cette étude fait l'objet de la section 6.3.3.5.

Dans tous les cas, les pourcentages d'appariements corrects que nous serons amenés à présenter sont calculés par rapport au nombre d'appariements effectifs.

6.3.3.1 Rotation image

La figure 6.4 présente les résultats de mise en correspondance sur l'image "Lézard" qui a été soumise à 6 rotations dans le plan image. Un bruit gaussien a été ajouté pour plus de réalisme. Quelques unes des images obtenues sont visibles sur la première ligne de la figure. Les résultats de mise en correspondance sont excellents pour ce type de transformation. Le taux d'appariements le plus faible est en effet de 92% et a été obtenu pour la rotation de 150°. Au contraire, c'est avec la rotation de 90° que les résultats sont les meilleurs, avec une

seule erreur d'appariements (278 points étaient à appariés et 277 l'ont été correctement). Ces résultats montrent la robustesse du vecteur d'invariants $\vec{v}_{col}$ face à une rotation image quelconque.

(référence) (60°) (90°) (150°) (180°)

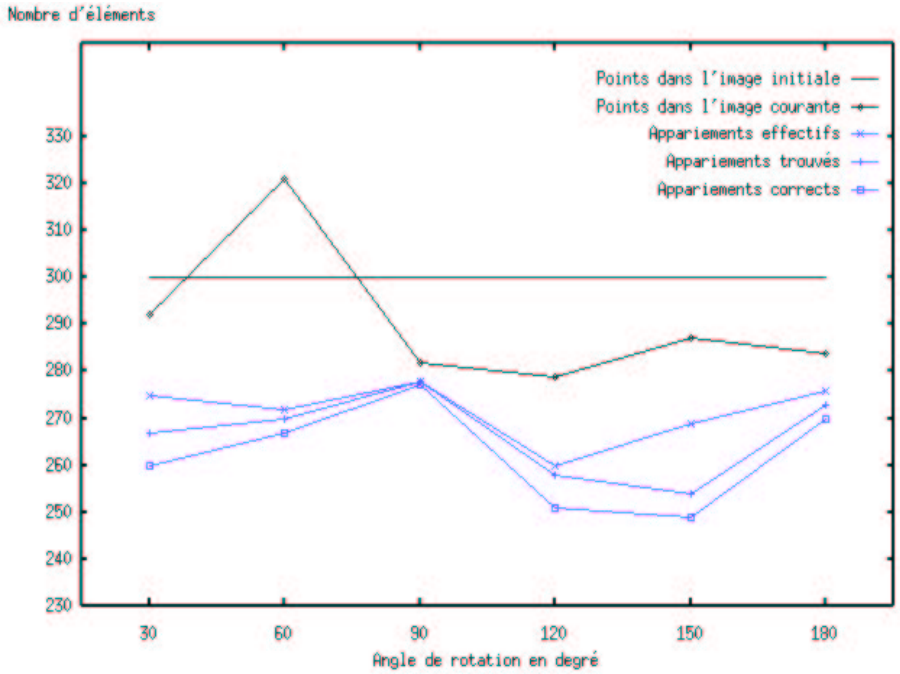
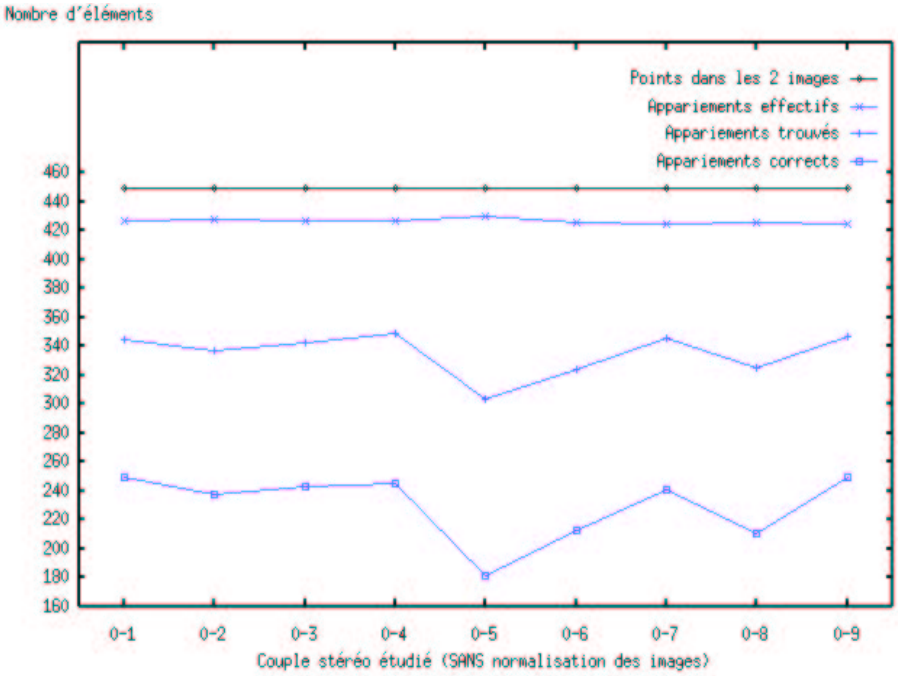


Figure 6.4 – *Séquence "Lézard" : Résultats d'appariement face à 6 rotations image.*

6.3.3.2 Changement d'illumination

Dans cette section, la figure 6.5 montre les résultats de mise en correspondance sur l'image "Lézard" qui a été soumise à 9 changements d'illumination interne. Ces transformations ont été réalisées synthétiquement selon le modèle diagonal avec translation qui a été présenté à la section 4.3.1 du chapitre 3. Un bruit gaussien a été ajouté. Quelques unes des images obtenues sont présentées sur la première ligne de la figure. Les 9 mises en correspondance ont été réalisées en prenant l'image n°0 (encadrée en gras sur la figure) comme première image du couple. Elles ont été réalisées avant d'avoir normalisé les images, puis après normalisation. Celle-ci a été réalisée localement, à l'aide d'une fenêtre de normalisation de diamètre 21 pixels (les images utilisées ont pour dimension (400×400)).

(0) (3) (6) (8) (9)



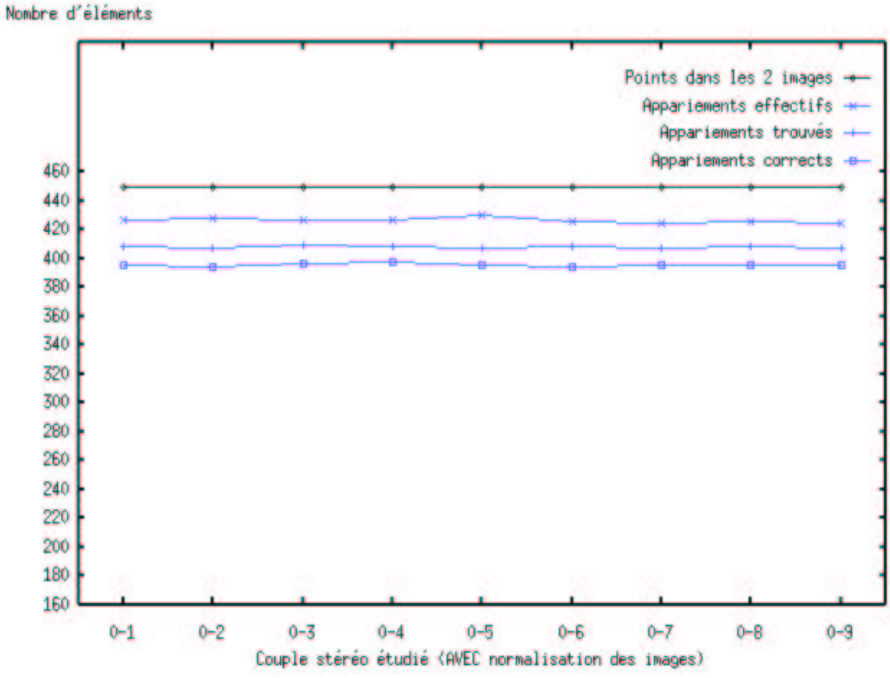
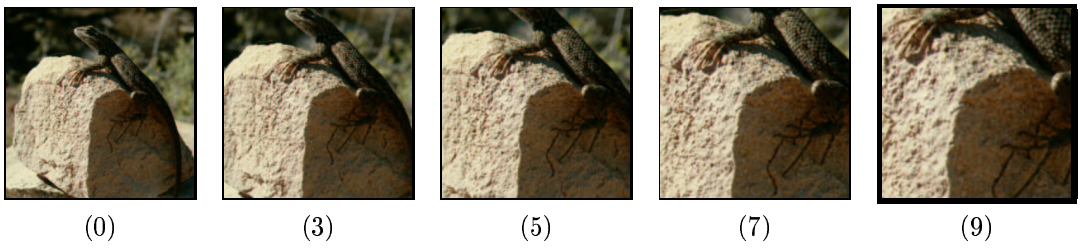


Figure 6.5 – *Séquence "Lézard" : Résultats d'appariement face à 9 changements d'illumination interne (avant et après normalisation des images).*

Les résultats obtenus montrent clairement l'intérêt de normaliser les images avant de les mettre en correspondance. Sans pré-traitement, la moyenne des taux d'appariements est autour de 50% de points correctement appariés. En revanche, après normalisation, ce taux avoisine les 92% pour tous les changements traités.

6.3.3.3 Changement d'échelle

Nous avons utilisé ici les images ainsi que les points Harris Couleur qui ont servi à la mesure de la répétabilité des détecteurs de points d'intérêt face au changement d'échelle (cf. section 2.2.2.3 du chapitre 2). La caméra virtuelle a subi 9 translations selon son axe optique, en s'éloignant de la facette photographiée. Quelques unes des vues sont présentées sur la première ligne de la figure 6.6. C'est la dernière vue (la n°9) qui a servi de référence pour toutes les mises en correspondance.



(0) (3) (5) (7) (9)

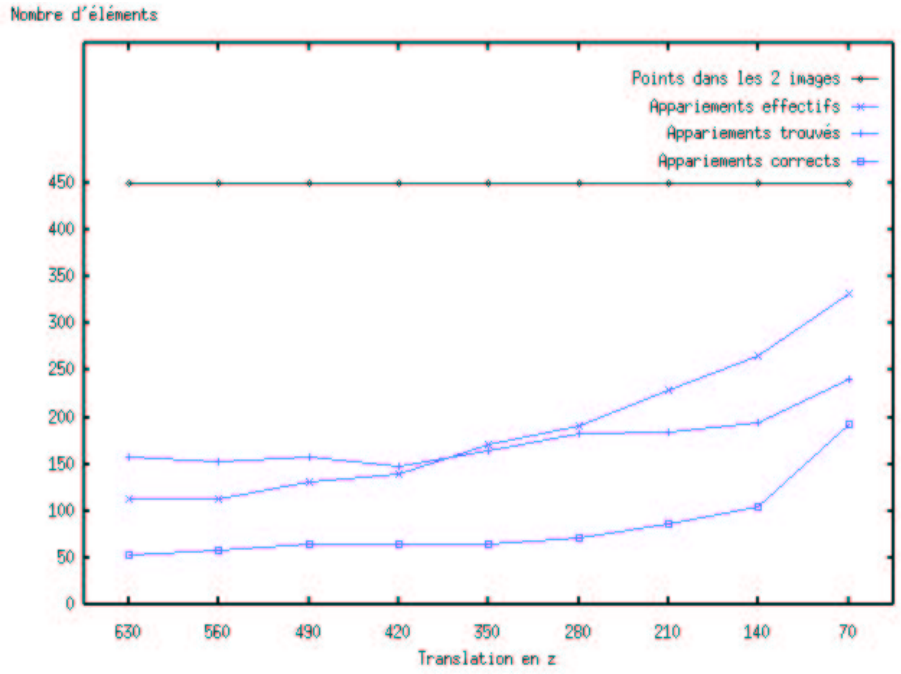


Figure 6.6 – *Séquence "Lézard" : Résultats d'appariement face à 9 changements d'échelle.*

Comme on peut le voir sur le graphe de la figure, le processus de mise en correspondance est très sensible aux changements d'échelle. On constate en premier lieu que le

nombre d'appariements qui doivent être trouvés est faible par rapport au nombre de points détectés, ce qui s'expliquent par le fait que la partie commune entre les couples d'images à apparier se retrouve relativement réduite. En ce qui concerne les appariements trouvés, on obtient un taux d'à peine 60% d'appariements corrects pour le dernier couple d'images, qui est associé au changement d'échelle le plus faible (le facteur d'échelle est de 2 dans ce cas). Ces mauvais résultats sont à attribuer au fait que le détecteur et la caractérisation à base d'invariants sont peu robustes aux changements d'échelle (consulter respectivement les mesures de répétabilité de la figure 2.11 au chapitre 2 et la section 3.2.1.4 du chapitre 3). Il est nécessaire d'intégrer leur calcul dans un contexte multi-échelle, comme le propose [Dufournaud et al., 2000]. La méthode a été présentée à la section 5.1.4.3 du chapitre 5. Remarquons pour finir qu'en ce qui concerne les changements d'échelle plus importants, le taux d'appariements corrects est d'environ 40% ; quelques appariements ont tout de même pu être formés, essentiellement grâce à l'étape de la relaxation pour laquelle les contraintes de voisinage et d'angles sont robustes à ce type de transformation (cf. section 6.1.2).

6.3.3.4 Changement de point de vue

La séquence d'images utilisée est la même qu'à la section 2.2.2.4 du chapitre 2. La caméra virtuelle a été déplacée 9 fois selon une rotation autour de la facette, en veillant à viser approximativement son centre de façon à en voir la plus grande partie. Le lecteur peut voir quelques exemples des images obtenues sur la première ligne de la figure 6.7. C'est la première image (image 0 encadrée en gras) représentée qui a été appariée aux 9 autres vues de la séquence. Exactement 450 points ont été détectés dans toutes les images.

On constate sur le graphe que les résultats se dégradent plus l'angle entre les deux caméras augmente. Les meilleurs scores sont obtenus avec l'image n°5, avec approximativement 93% d'appariements corrects. C'est pour l'image la plus éloignée de l'image de référence (la n°9), que l'on obtient le moins bon pourcentage : 86%.

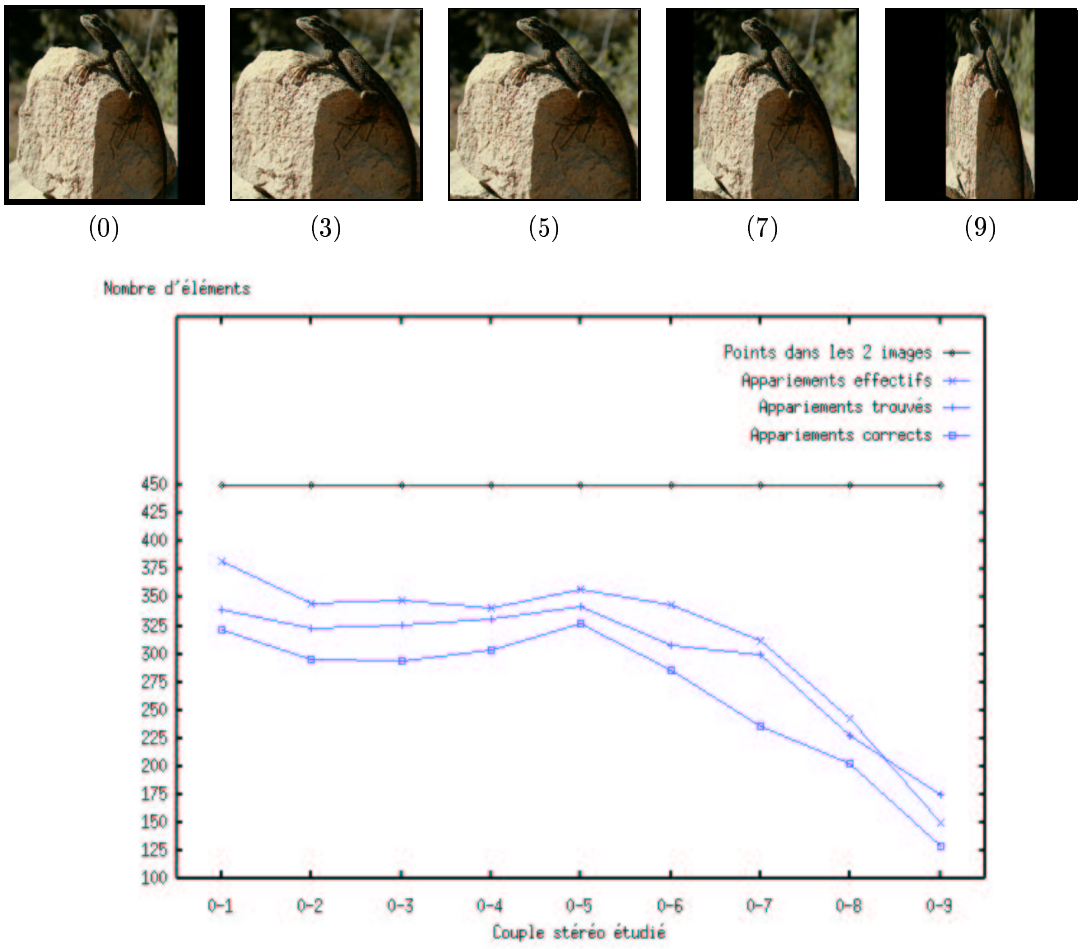


Figure 6.7 – *Séquence "Lézard" : Résultats d'appariement face à 9 changements de point de vue.*

6.3.3.5 Evaluation du processus incrémental

Dans cette section, les images testées proviennent d'une scène tridimensionnelle non plane. Le critère d'évaluation utilisé est donc celui basé sur la contrainte épipolaire (cf. équation 6.9). Nous étudions les d'appariements corrects obtenus à l'issue de la mise en correspondance. Il s'agit donc des appariements qui sont en correspondance épipolaire, c'est-à-dire les points dont le correspondant est dans le pire des cas à la distance ε de la droite épipolaire associée. La valeur pour ε est fixée à 3 dans toutes les expériences, sachant que la détection des points, la mise en correspondance et l'estimation de la matrice fondamentale sont réalisées en précision demi-pixel.

L'appariement est d'abord réalisé sans puis avec l'algorithme incrémental, et les deux ensembles de résultats obtenus sont comparés. Lorsque celui-ci est utilisé, la matrice fondamentale est estimée à l'aide d'une méthode linéaire robuste (cf. section 7.6.1 du chapitre 7). La géométrie épipolaire ainsi estimée n'a pas besoin d'être exacte, puisqu'elle n'est utilisée que pour éliminer les couples de points qui ne sont pas en correspondance épipolaire. Comme la méthode d'estimation est linéaire, elle peut-être implémentée efficacement tout en fournissant une géométrie très précise loin des épipôles. La triangulation de la première image est une triangulation de Delaunay semi-dynamique [Boissonnat et Teillaud, 1986]. Elle est implémentée de manière incrémentale, par insertion de points, pour être le plus efficace possible avec l'algorithme de mise en correspondance : à l'itération i, la nouvelle triangulation est calculée en insérant les points $\mathcal{P}_1^i$ issus de $\mathcal{M}^i$ dans la triangulation $\mathcal{T}_{\mathcal{M}^i}$ calculée à l'itération précédente.

Les figures 6.8 and 6.9 montrent les résultats de mise en correspondance de (220×231) points d'intérêt sur deux images aux points de vue et illumination différents. La géométrie épipolaire superposée a été estimée à partir de tous les appariements obtenus. La première figure présente les appariements (appelés $\mathcal{M}$ par la suite) obtenus à partir de l'algorithme de base, alors que la seconde montre les appariements ($\mathcal{M}_{inc}$) à l'issu du processus de mise en correspondance incrémental.

Comparons les résultats obtenus :

- Temps de calcul : 90' ont été nécessaires pour calculer l'ensemble $\mathcal{M}$. Seulement trois itération et quelques minutes ont permis de calculer $\mathcal{M}_{inc}$. Ce net gain de temps provient essentiellement de la complexité devenue réduite lors de la comparaison des vecteurs d'invariants, puisque l'aire de recherche est plus limitée. Une autre raison est que le nombre de couples ambigus est beaucoup moins important au début de la relaxation, qui par conséquent est beaucoup plus rapide ;
- Qualité des résultats : nous obtenons de meilleurs résultats avec la méthode incrémentale (170 appariements tous corrects) qu'avec la méthode de base (155 appariements dont un très petit nombre faux). Examinons les plus précisément : à la figure 6.8, la deuxième droite épipolaire en partant du haut sur chaque image est associée à l'appariement n°142 (sur le socle au second plan dans la scène). Cet appariement est faux, mais peut être éliminé en utilisant la contrainte épipolaire. Cependant, considérons les mêmes points à la figure 6.9. Les appariements obtenus (n°29 et n°60) avec la méthode incrémentale sont corrects. Cet exemple démontre l'emploi de contraintes géométriques permet d'apparier plus de points. Considérons

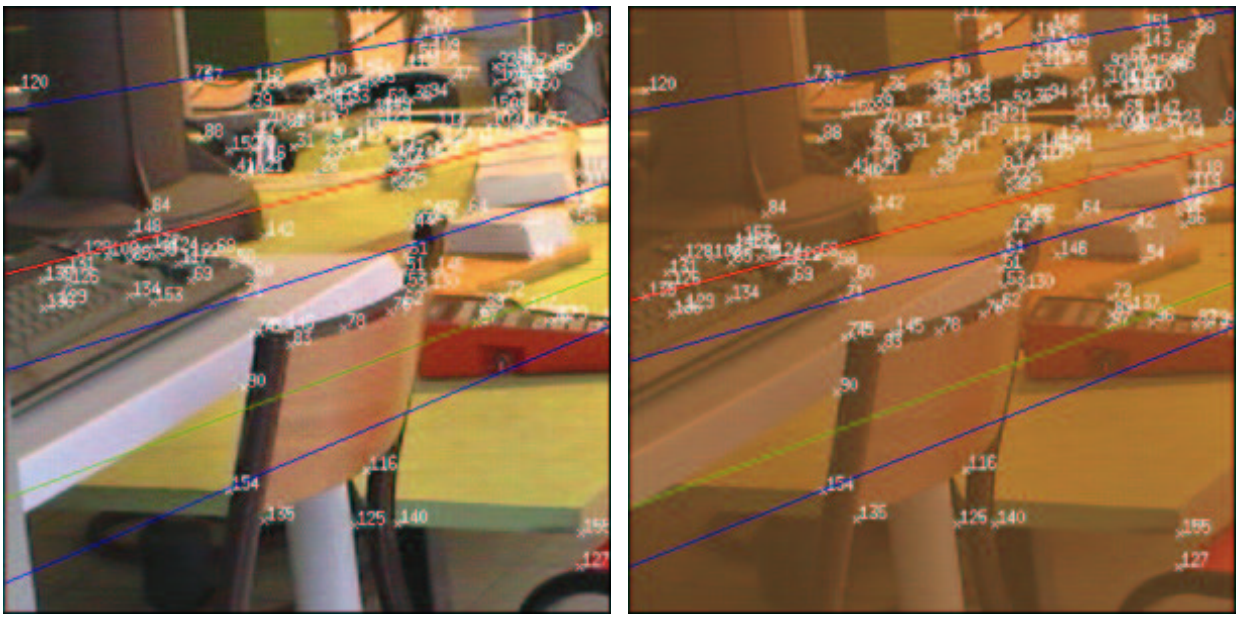


Figure 6.8 – *Images "Salle robotique" avec changements de point de vue et d'illumination. 155 appariements ont été trouvés avec la méthode de base. Les droites épipolaires superposées correspondent aux appariements* $\{73,142,71,97,154\}$.

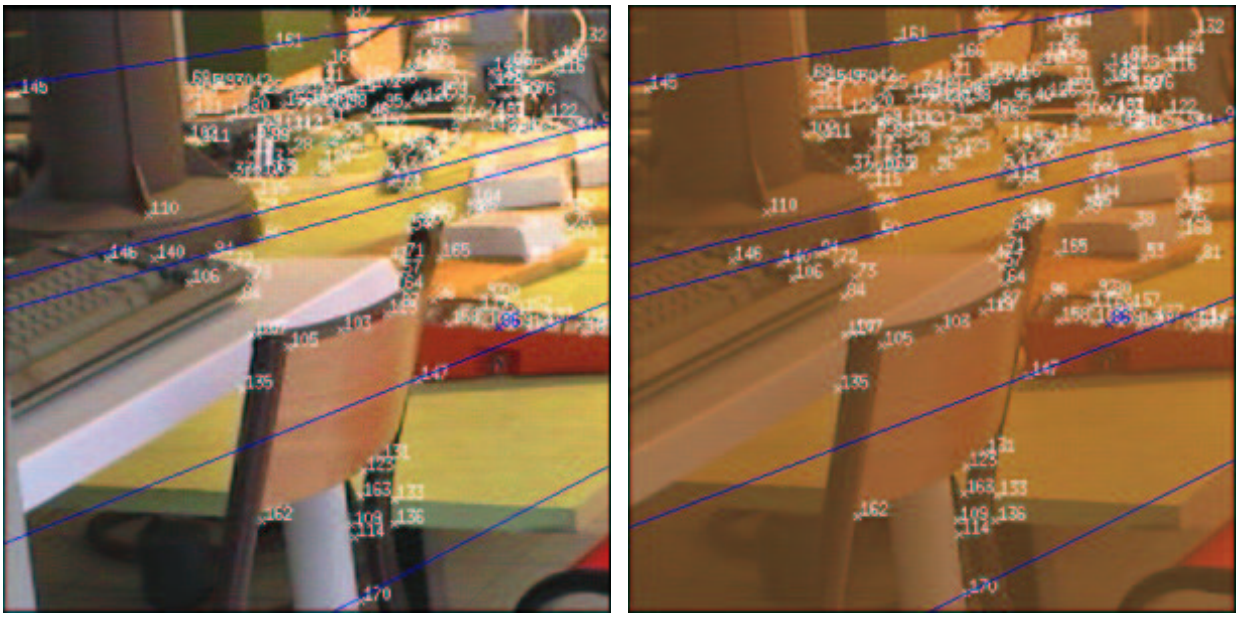


Figure 6.9 – *Images "Salle robotique" avec changements de point de vue et d'illumination. 170 appariements ont été trouvés à l'aide de l'algorithme incrémental. La géométrie épipolaire finale* $F_{\mathcal{M}^2}$ *superposée correspond aux appariements* $\{161,29,60,147,170\}$.

maintenant les deux deuxièmes droites épipolaires en partant du bas à la figure 6.8. Elles sont associées à l'appariement n°97 (sur le boîtier de commande à droite de la chaise) qui vérifie la contrainte épipolaire mais est incorrect. Sur l'autre figure, les mêmes points ont été correctement appariés (appariement n°86 représenté d'une autre couleur). Cet autre exemple nous montre que la méthode incrémentale permet d'augmenter le taux d'appariements corrects.

Détail des itérations

La figure 6.10 montre les contraintes géométriques appliquées à la dernière itération de l'appariement incrémental sur deux images aux points de vue et illuminations différents. 367 et 269 points ont été détectés dans les deux images et 164 points ont été correctement appariés en trois itérations. Seulement 175 de ces points auraient pu être appariés, les points de vue étant relativement différents. Le taux d'appariements corrects est de 94%.

Les informations géométriques superposées sur la première ligne de la figure sont la géométrie épipolaire $F_{\mathcal{M}^1}$ et sur la seconde, les triangulations $\mathcal{T}_{\mathcal{M}^1}$. Ces deux résultats ont été calculés à l'issue de l'itération 1 (deuxième itération) et ont été utilisés pour réaliser l'appariement de la dernière itération. Les calculs ont été réalisés sur une station Sun Ultra 5, 333Mhz, avec 256Mo de mémoire. Les points d'intérêt ont été appariés en 100" approximativement. En comparaison, l'algorithme de base a mis 48' pour donner une solution avec un taux d'appariements incorrects supérieur.

Le lecteur peut consulter les tableaux 6.1 et 6.2 qui résument respectivement les pourcentages d'appariements corrects et les temps de calcul obtenus pour chaque itération sur les trois premières lignes du tableau, et les données correspondantes mais pour la méthode de base non incrémentale sur la dernière ligne. Signalons que le taux d'appariements corrects de la première itération est volontairement faible (62%). Nous n'avons en effet conservé qu'un petit pourcentage (40%) des appariements formés vérifiant la contrainte épipolaire, de manière à avoir une très forte probabilité que l'ensemble $\mathcal{M}^0$ obtenu comprenne des appariements tous corrects. On peut aussi remarquer que la durée de la relaxation diminue avec les itérations. Les aires de recherche sont en effet de plus en plus réduites lors de la phase d'appariement (il y a de plus en plus de triangles), réduisant ainsi à chaque fois le nombre d'appariements ambigus. Au contraire, dans la version non incrémentale, la relaxation est beaucoup plus gourmande en temps de calcul (47'), le nombre d'appariements ambigus étant nettement plus important au début du processus.

Itération	Nombre de points	Nombre d'appariements			%	Nombre de triangles
		effectifs	trouvés	corrects		
0	(125,109)	58	36	36	62%	-
1	(211,151)	83	74	72	87%	60
2	(367,269)	175	166	164	94%	130
-	(367,269)	175	170	156	89%	-

Table 6.1 – *Mise en correspondance de (367,269) points Harris couleur : résultats de mise en correspondance. Les trois premières lignes correspondent aux trois itérations de l'algorithme incrémental. La dernière ligne correspond à la méthode de base non incrémentale.*

Figure 6.10 – *Résultats d'appariements $\mathcal{M}^2$ sur deux images "Toys" avec changement de point de vue et d'illumination. Les droites épipolaires superposées correspondent aux appariements $\{70,107,144,117,140\}$.*

Itération	Temps de calcul			
	Points	Appariement	Relaxation	Total
0	15”	5”	23”	43”
1	10”	3”	21”	34”
2	7”	2”	14”	23”
-	7”	14”	47’	47’21”

Table 6.2 – *Mise en correspondance de (367,269) points Harris couleur : temps de calcul. Les trois premières lignes correspondent aux trois itérations de l'algorithme incrémental. La dernière ligne correspond à la méthode de base non incrémentale.*

Exemple sur un très grand nombre de points

Les figures 6.11, 6.12 et 6.13 présentent des résultats de mise en correspondance sur deux images ayant des points de vue et illumination différentes. 1170 et 1035 points Harris couleur ont été extraits. La mise en correspondance a été réalisée de manière incrémentale et a permis d'obtenir 403 appariements. Malgré le grand nombre de points mis en jeu, seulement quelques minutes ont été nécessaire pour réaliser l'appariement complet et le taux d'appariements corrects obtenu est de 95%. La dernière figure montre la géométrie épipolaire finale obtenue à l'issue des itérations. Cette fois, son estimation a été réalisée à l'aide d'une méthode non linéaire robuste de type LMedS (cf. section 7.6.1.3 du chapitre 7) pour obtenir la plus géométrie exacte possible.

Figure 6.11 – *(1170,1035) points d'intérêt Harris Couleur extraits de deux images, avec changement de point de vue et d'illumination.*

Figure 6.12 – *Résultats de la mise en correspondance des points présentés à la figure 6.11. L'algorithme incrémental a permis d'apparier 403 points.*

Figure 6.13 – *Géométrie épipolaire obtenue à partir des appariements présentés à la figure 6.12. Les droites épipolaires tracées dans une image correspondent dans l'autre image aux quatre coins du toit de la grande maison, au sommet de la tente dans le haut de la scène et à un coin du panneau au premier plan.*

6.4 Conclusion

Dans ce chapitre, nous avons présenté et évalué une méthode de mise en correspondance de points d'intérêt robuste. Nous avons en premier lieu développé une méthode permettant de comparer les vecteurs d'invariants couleur. Cette méthode est en théorie sous-optimale par rapport à la distance de Mahalanobis traditionnellement utilisée, car elle n'intègre pas un modèle de bruit. Cependant elle a l'avantage d'être beaucoup plus facile à mettre en œuvre, et de bien se comporter avec notre caractérisation dans la mesure où celle-ci n'implique que des dérivées de l'image d'ordre un et est donc peu bruitée.

Ensuite un algorithme de relaxation basé sur des contraintes semi-locales est utilisé pour éliminer les ambiguïtés dans les appariements formés. Les contraintes que nous avons mises en place sont des contraintes de voisinage ainsi que des contraintes géométriques angulaires invariantes au groupe des similitudes. Elles permettent d'obtenir un ensemble de mises en correspondance consistant au travers des principales transformations de l'image. Les tests réalisés à la section des résultats montrent que la méthode se comporte très bien face aux rotations, aux changements d'illumination et aux changements de point de vue, d'après les taux d'appariements corrects obtenus (dont les extrema sont récapitulés dans le tableau 6.3 ci-après). Ceux-ci sont en effet pour la plupart supérieurs à 90% pour ces transformations. Le taux de 86% obtenu pour les changements de point de vue n'est pas considéré comme représentatif, car il résulte d'un changement extrême (voir la vue n°9 de la figure 6.7). Ces résultats satisfaisants permettent également de valider la pertinence de la méthode de caractérisation basée sur les invariants différentiels couleur que nous avons présentée au chapitre 4.

Type de transformation	Taux d'appariements corrects
Rotation image	92-99 %
Changement d'illumination	92 %
Changement d'échelle	40-60 %
Changement de point de vue	86-93 %

Table 6.3 – *Récapitulatif des taux d'appariements corrects obtenus au travers des principales transformations de l'image.*

En ce qui concerne les changements d'échelle, les résultats obtenus ne sont pas très bons (60% pour la transformation la plus faible, de facteur d'échelle 2). Ceci est dû au fait que la détection et la caractérisation des points Harris Couleur n'ont pas été réalisées dans un contexte multi-échelle. Il faudra envisager de le faire, en développant par exemple une méthode similaire à celle proposée par [Dufournaud et al., 2000] et présentée à la section 5.1.4.3 du chapitre 5. Elle semble être une solution efficace pour traiter ces changements jusqu'à un facteur d'échelle de 6.

La deuxième étape de notre travail a ensuite consisté à mettre en place une méthode de mise en correspondance efficace face à de grands nombres de points d'intérêt. La complexité des approches proposées (incluant notre propre méthode d'appariement) engendre d'une part des temps de calcul très importants et d'autre part un plus grand nombre d'ambiguïtés difficilement éliminées à l'étape de la relaxation. Nous avons donc proposé un algorithme incrémental de mise en correspondance, basé sur deux contraintes géométriques invariantes

aux transformations de l'image : la contrainte épipolaire et une triangulation construite sur les points appariés.

Dans la section des résultats, plusieurs mises en correspondance utilisant ce principe ont été réalisées et étudiées. Les résultats obtenus sont encore améliorés par rapport à ceux calculés sans le processus incrémental, en termes de qualité des appariements formés et de temps de calcul. Les trois exemples traités (figures 6.9, 6.10 et 6.12) ont permis d'obtenir respectivement 100%, 94% et 95% d'appariements corrects. Les temps de calcul sont également considérablement réduits par rapport aux approches traditionnelles.

A la fin du processus complet de mise en correspondance qui vient d'être présenté, nous avons obtenu un ensemble d'appariements $\mathcal{M}$. En addition, nous avons également calculé la géométrie épipolaire du système des deux caméras caractérisée par $F_{\mathcal{M}}$, ainsi qu'une triangulation $\mathcal{T}_{\mathcal{M}}$ des points appariés. Ces deux contraintes nous ont été très utiles lors du processus de mise en correspondance incrémental, mais leur utilité ne s'arrête pas là. Elles peuvent en effet être maintenant exploitées pour réaliser par exemple la reconstruction tridimensionnelle de la scène, ou encore la synthèse de nouvelles vues de cette scène. Ce travail fait l'objet de la prochaine partie de ce mémoire. Dans le prochain chapitre, nous commençons par rappeler les différentes notions permettant de définir les systèmes de vision admettant deux et trois oculaires, dans un contexte non calibré.

Troisième partie

Application à la synthèse de vues intermédiaires

Chapitre 7

Géométrie des systèmes binoculaires et trinoculaires non calibrés

L'objectif de ce chapitre est de rappeler au lecteur les principales notions requises pour manipuler les systèmes de vision constitués de deux ou de trois caméras, dans un contexte non calibré. Dans un premier temps, les différents modèles de caméra sont passés en revue. Puis nous présentons les entités caractérisant les systèmes binoculaires, notamment la géométrie épipolaire. Ceux-ci représentent la configuration minimale nécessaire pour envisager une reconstruction tridimensionnelle. Nous nous intéressons également aux systèmes trinoculaires, qui manipulent des entités indispensables au transfert d'images. Enfin, nous passons en revue les différentes méthodes permettant l'estimation des diverses entités mises en jeu. Les caméras n'étant pas calibrées, les seules informations exploitables pour ces estimations sont les mises en correspondance de points.

Sommaire

Le but de ce chapitre est de rappeler le formalisme et les relations qui caractérisent un système de vision constitué d'au plus trois caméras. Nous décrivons d'abord à la section 7.1 le formalisme qui permet de modéliser une caméra. Puis, après avoir rappelé les degrés de liberté d'un système de N caméras à la section 7.2, nous présentons les différentes entités existant dans les systèmes binoculaires à la section 7.3 et dans les systèmes trinoculaires à la section 7.4. Nous nous arrêtons ici à l'étude de la géométrie de trois caméras, les applications envisagées au prochain chapitre ne nécessitant pas en théorie la donnée de caméras supplémentaires. En effet, la donnée de deux vues d'une scène correspond à la configuration minimale permettant sa reconstruction tridimensionnelle, et la donnée de trois caméras correspond à la configuration minimale permettant de mettre en place des méthodes de transfert d'images.

Les diverses entités que nous manipulons ici sont liées entre elles et reliées à la matrice de projection caractérisant la caméra par des relations qui sont ensuite passées en revue à la section 7.5.

Cette étude est réalisée dans un contexte de caméras non calibrées et de scènes rigides. Nous ne nous donnons donc aucune information sur la géométrie des caméras. Les seules données disponibles sont les appariements de points établis à l'aide du processus de mise en correspondance décrit depuis le début de cette thèse. Nous présentons donc également à la section 7.6 les différentes méthodes d'estimation de tous les paramètres mis en jeu dans les systèmes binoculaires et trinoculaires à partir de ces seules données. Le lecteur pourra constater que nous insistons particulièrement sur le calcul de la géométrie épipolaire. L'estimation de la matrice fondamentale s'avère en effet relativement délicate à réaliser, alors que cette entité représente une donnée essentielle dans l'étape de la mise en correspondance des points, comme dans le processus de reconstruction même.

Pour une excellente introduction à la géométrie des systèmes multi-oculaires, le lecteur peut se référer à l'ouvrage de Hartley et Zisserman paru récemment [Hartley et Zisserman, 2000].

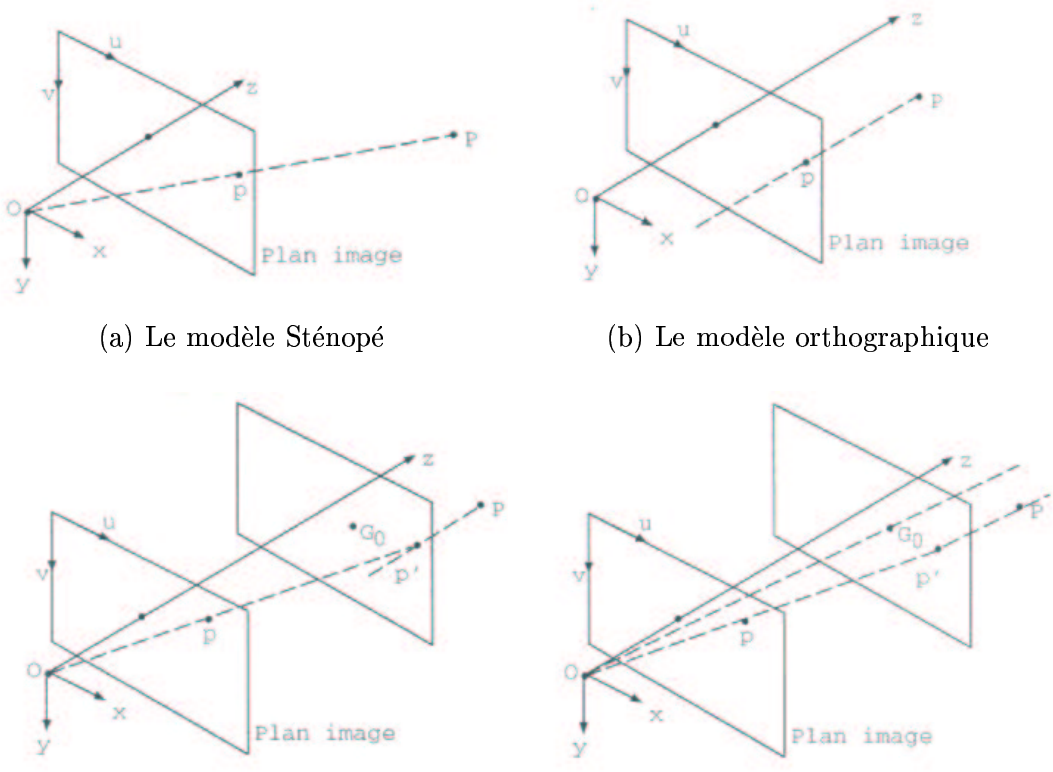
7.1 Le modèle de caméra

Le modèle le plus couramment utilisé est le modèle *Sténopé*. D'autres modèles existent, certains ont l'avantage de la simplicité alors que d'autres prennent mieux en compte l'optique de la caméra, comme par exemple la distorsion, mais au prix de paramètres additionnels. Ces modèles seront présentés à la section 7.1.2.

7.1.1 Le modèle Sténopé

Le modèle Sténopé considère la transformation comme une projection perspective pure (cf. figure 7.1.a). Il constitue une bonne approximation pour les projections effectuées par les caméras CCD.

Une caméra est représentée par un plan rétinien P_r et un centre optique C qui n'appartient pas à P_r. L'image d'un point M de l'espace est la trace du rayon (CM) sur le plan P_r. Cc est la distance focale où c est la projection orthogonale de C sur P_r est est



(a) Le modèle Sténopé

(b) Le modèle orthographique

(c) Le modèle perspectif faible

(d) Le modèle paraperspectif

Figure 7.1 – *Les principaux modèles de projection.*

appelé point principal.

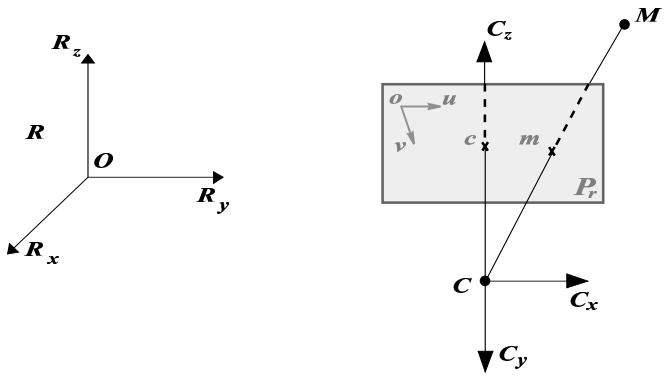


Figure 7.2 – *Le modèle Sténopé.*

Soit $\mathcal{R}(O, R_x, R_y, R_z)$ un repère orthonormé de l'espace, $\mathcal{R}_c(C, C_x, C_y, C_z)$ un autre repère orthonormé d'origine C tel que l'axe des z soit orthogonal à P_r et $\mathcal{R}_a(o, u, v)$ un repère affine du plan P_r. L'expression analytique de la projection de M en m se décompose en trois parties distinctes :

- un changement de repère de l'espace (visant à exprimer $\mathcal{R}$ dans $\mathcal{R}_c$) ;
- une projection (visant à exprimer m dans $\mathcal{R}_c$) ;
- un changement de repère dans le plan (visant à exprimer m dans $\mathcal{R}_a$).

Cette opération de projection dépend de plusieurs paramètres : la rotation R et la translation T décrivent le changement de repère entre $\mathcal{R}$ et $\mathcal{R}_c$. θ est l'angle entre les vecteurs u et v, α_u et α_v sont les facteurs d'échelle sur les axes. u_0 et v_0 sont les coordonnées de c dans le repère $\mathcal{R}_a$. En pratique, θ est toujours proche de $\frac{\pi}{2}$ et le rapport $\frac{\alpha_u}{\alpha_v}$ varie peu. La caméra est donc représentée par R, T, α_u, α_v, θ, u_0 et v_0. R et T sont appelés les paramètres extrinsèques puisqu'ils ne sont fonction que de la position et de l'orientation de la caméra dans l'espace et ils dépendent chacun de trois paramètres. Les autres paramètres sont appelés paramètres intrinsèques puisqu'ils sont fonction uniquement de la caméra ; ils dépendent en tout de cinq paramètres. L'opération de projection est donc fonction de 11 paramètres au total.

Formulation projective

Pour un point M donné de l'espace représenté par le vecteur $M_(X_r, Y_r, Z_r)$ dans le repère $\mathcal{R}$, ses coordonnées (u, v) s'expriment dans le repère $\mathcal{R}_a$ du plan P_r en formulation

projective de la façon suivante :

$$\begin{pmatrix} s.u \\ s.v \\ s \end{pmatrix} = P \begin{pmatrix} X_r \\ Y_r \\ Z_r \\ 1 \end{pmatrix} \quad \text{avec} \quad P = AP_0D$$

$$D = \begin{pmatrix} R & T \\ 0_3^T & 1 \end{pmatrix} \qquad P_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} \alpha_u & -\alpha_u \cos\theta & u_0 \\ 0 & \alpha_v \sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7.1}$$

D exprime le changement de repère de l'espace, P_0 exprime la projection et A le changement de coordonnées dans le plan.

L'équation 7.1 est bien sûr définie à une constante multiplicative près s comme toutes les équations projectives. P est une matrice (3×4) définie à un facteur d'échelle près et de rang 3. On retrouve la dépendance en 11 paramètres.

7.1.2 Autres modèles

Nous présentons dans cette section les principaux modèles de caméra existants en dehors du modèle perspectif pur. Ces modèles sont spécifiques à certains types de scènes ou certaines applications. Ils sont décrits ici du plus simple au plus complet. Une comparaison approfondie peut être trouvée dans [Horaud et al., 1995]. Nous verrons également dans la dernière section comment la distorsion peut être prise en compte.

7.1.2.1 Le modèle orthographique

Le modèle de projection orthographique (cf. figure 7.1.b) est un modèle perspectif dont le centre optique est à l'infini. Les rayons optiques arrivent donc tous parallèlement sur le plan rétinien. C'est un modèle très simplifié et généralement utilisé si la distance focale est très grande ou si la scène est très éloignée de la caméra. Son manque de réalisme ont conduit à la mise en place de modèles plus complets comme le modèle perspectif ou ceux décrits ci-après.

7.1.2.2 Le modèle perspectif faible

Dans le cas où la taille de la scène est relativement petite par rapport à la distance entre la caméra et la scène, on suppose que les points subissent d'abord une projection orthogonale sur un plan perpendiculaire à l'axe optique et passant par le centre de gravité G_0 des points de la scène observée (cf. figure 7.1.c). Ensuite, l'image de chacun de ces points subit une projection perspective sur le plan rétinien. Ici la projection perspective correspond tout simplement à une homographie, puisque tous les points appartiennent au même plan de l'espace.

7.1.2.3 Le modèle paraperspectif

Ce modèle est une variante du modèle précédent. Les points subissent dans un premier temps une projection parallèle sur le plan perpendiculaire à l'axe optique et passant par G_0 le centre de gravité de la scène, selon une direction définie par $\vec{OG_0}$. Les points obtenus subissent ensuite une projection perspective sur le plan image (cf. figure 7.1.d). Ce modèle a l'avantage, par rapport au précédent, de tenir compte d'une éventuelle translation latérale importante de la scène par rapport à l'axe optique. Il tient compte du facteur d'échelle, de l'apparence d'objets à la périphérie de la caméra. Il est toutefois moins complet que le modèle Sténopé, et les calculs ne sont pas plus simples. Il est tout de même utilisé dans certaines applications [Poelman et Kanade, 1994].

7.1.2.4 Modélisation des distorsions

Trois hypothèses implicites sont contenues dans les modèles de caméra précédents : d'une part, l'image se forme sur une surface qui est effectivement plane. D'autre part, les pixels sont régulièrement espacés dans le plan image. Enfin la caméra ne présente pas de distorsion optique.

Les deux première hypothèses sont en général bien vérifiées pour les caméras CCD. En revanche, la distorsion optique n'est pas négligeable. Elle transforme les coordonnées image idéales en coordonnées x' et y' :

$$\begin{cases} x' = x + \delta_x(x, y) \\ y' = y + \delta_y(x, y) \end{cases} \tag{7.2}$$

Il existe deux types de distorsion. Les distorsions radiales sont dues au fait que les lentilles ne sont pas parfaitement minces et qu'elles présentent des défaut de courbure. Leur effet est de déplacer les points image symétriquement au centre de l'image. Les distorsions tangentielles apparaissent lorsque l'assemblage des lentilles sur l'objectif n'est pas parfait (lentilles non alignées ou non positionnées perpendiculairement à l'axe principal, ...). Elles sont souvent accompagnées de distorsions radiales.

Très vite, le modèle Sténopé a été complété afin de mieux représenter le phénomène physique de la formation des images. Les photogrammètres se sont penchés sur ce problème en ajoutant aux paramètres internes de la caméra des paramètres prenant en compte les distorsions radiales et tangentielles [Brown, 1966; Beyer, 1992]. Dans la communauté scientifique de la Vision par Ordinateur, des travaux ont également été réalisés pour prendre en compte les distorsions radiales dans le calibrage des caméras ; citons la méthode de Tsai [Tsai, 1986]. La correction apportée à un point (x, y) exprimé dans un repère dont l'origine est confondue avec le centre des distorsions radiales, est réalisée en translatant ce point par un vecteur $(\Delta x_r, \Delta y_r)$ défini comme suit :

$$\begin{cases} \Delta x_r = x(k_1 r^2 + k_2 r^4 + ...) \\ \Delta y_r = y(k_1 r^2 + k_2 r^4 + ...) \end{cases} \tag{7.3}$$

avec $r^2 = x^2 + y^2$. Le nombre de paramètres k_i estimés dépend de la précision que l'on souhaite obtenir sur le calcul de la distorsion et de la qualité de l'objectif utilisé.

Ce modèle a ensuite été complété par l'ajout de paramètres modélisant les distorsions tangentielles [Li et Schenk, 1990]. La correction des distorsions tangentielles pour un point (x, y) se modélise sous forme polynômiale par une translation $(\Delta x_t, \Delta y_t)$ définie par :

$$\begin{cases} \Delta x_t = [p_1(r^2 + 2x^2) + 2p_2xy].[1 + p_3r^2 + p_4r^4 + ...] \\ \Delta y_t = [p_2(r^2 + 2y^2) + 2p_1xy].[1 + p_3r^2 + p_4r^4 + ...] \end{cases} \tag{7.4}$$

Les expériences ont montré que le premier ordre était généralement suffisant pour que la différence entre le point exact et le point projeté dans l'image soit inférieure à 0.1 pixel.

D'autres méthode plus récentes existent. Il est ainsi possible d'annuler la distorsion en extrayant des contours qui sont censés représenter des segments, puis en cherchant la meilleure transformation de ces contours vers des segments de droite [Brand, 1995; Devernay et Faugeras, 1995]. L'avantage de ces méthodes est qu'elles opèrent de manière indépendante sur chaque image et ne nécessitent aucune estimation préalable de paramètres intrinsèques et/ou extrinsèques. La seule contrainte requise est la présence de segments dans l'image.

7.2 Degrés de liberté d'un système de N caméras

Nous considérons ici N caméras quelconques, sachant que dans les sections suivantes, N vaudra 2 et 3. Les degrés de liberté du système ainsi formé correspondent au nombre de paramètres nécessaires pour le décrire. Ce nombre dépend du type des transformations qui doivent laisser les paramètres du système invariants. Un degré de liberté correspond donc à un invariant scalaire indépendant[1].

Depuis le début de ce chapitre, nous nous sommes placés dans le formalisme des matrices de projection. Nous avons vu que chacune des matrices de projection possède 11 paramètres. Nous considérons le groupe des transformations le plus général, c'est-à-dire celui des transformations projectives (cf. annexe E), caractérisé par une homographie (4×4), soit par 15 coefficients. Alors le nombre de degrés de liberté α du système des N caméras est :

$$\alpha = 11N - 15 \tag{7.5}$$

Dans le cas d'une seule caméra, ce nombre n'est bien sûr pas négatif mais est simplement nul. Pour plus de détails concernant le comptage des invariants selon différents groupes de transformation, le lecteur peut consulter [Mundy et Zisserman, 1992].

Nous allons maintenant passer en revue les entités permettant de décrire la géométrie de systèmes comprenant deux puis trois caméras. Il existe une théorie très complète basée sur les déterminants qui permet de définir des relations à partir desquelles on peut dériver les bilinéarités (la matrice fondamentale), les trilinéarités et plus généralement les N-linéarités. On la doit à [Triggs, 1994; Faugeras et Mourrain, 1995]. Pour une introduction à cette méthodologie le lecteur peut également consulter [Laveau, 1996].

[1] Prenons l'exemple des invariants différentiels couleur mis en place au chapitre 4. Ceux-ci sont invariants à la rotation, qui possède un seul paramètre, son angle. Au premier ordre, il existe $3 \times 3 = 9$ grandeurs pour les trois plans couleur. On obtient donc $9 - 1 = 8$ invariants.

7.3 La stéréovision binoculaire

Après avoir défini le formalisme de la projection sur une caméra et rappelé quelques notions sur les invariants, nous nous intéressons maintenant aux relations géométriques qui existent entre deux vues de la même scène. Les deux caméras peuvent être deux caméras physiquement distinctes, ou une seule caméra observant la scène à partir de deux positions différentes obtenues par déplacement. La scène est considérée comme un ensemble d'objets statiques, au sens où ces objets ne bougent pas d'une vue à l'autre.

La géométrie qui décrit alors un couple d'images d'une même scène s'appelle *géométrie épipolaire*. Elle est caractérisée par une matrice (3×3) singulière. Cette matrice est connue sous le nom de matrice *essentielle* lorsqu'elle intègre les paramètres intrinsèques des caméras et de matrice *fondamentale* lorsqu'elle tient compte des paramètres extrinsèques uniquement.

7.3.1 La matrice essentielle

La matrice essentielle a été introduite pour la première fois en 1981 par Longuet-Higgins [Longuet-Higgins, 1981] dans le contexte de l'étude du mouvement. Les deux caméras sont supposées avoir des paramètres intrinsèques connus. Dans ce contexte, la matrice essentielle est définie comme étant une matrice (3×3) notée E_{12} et qui vérifie pour tout point M de l'espace, associé à ses projections m_1 et m_2 exprimées en coordonnées projectives[2] de $\mathcal{P}^2$ dans les deux images :

$$m_2^T E_{12} m_1 = 0 \tag{7.6}$$

Cette matrice est déterminée à une constante multiplicative près et est de rang 2. D'autres propriétés la caractérisent :

- Elle possède deux valeurs singulières non nulles et égales,
- Elle vérifie les deux relations 7.7 et 7.8 suivantes :

$$(l_1^2 + l_2^2 + l_3^2)^2 = 4(\|l_1 \wedge l_2\|^2 + \|l_1 \wedge l_3\|^2 + \|l_2 \wedge l_3\|^2) \tag{7.7}$$

où l_i représentant les vecteurs lignes de E.

$$trace^2(EE^T) = 2.trace((EE^T)^2) \tag{7.8}$$

Soit (R, t) le déplacement entre les deux caméras, illustré par la figure 7.3 et tel que, pour tout point M de la scène :

$$M_{final} = RM_{initial} + t \tag{7.9}$$

La translation t est exprimée dans le repère final et vaut $t = C'C$. La matrice R exprime la rotation autour d'un axe passant par le centre optique C'. On rencontre parfois la convention inverse dans la littérature.

[2] Soit $m_i = (m_i^1, m_i^2, s)^T$.

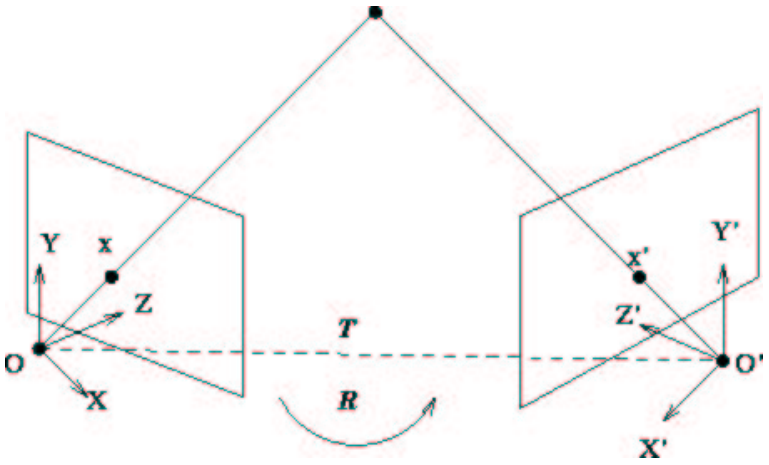


Figure 7.3 – *Contrainte de Longuet-Higgins.*

Alors il est possible de prendre pour E_{12} la solution ci-dessous, $\widehat{t}$ représentant la matrice antisymétrique relative à t :

$$E_{12}^T = \widehat{t}R \tag{7.10}$$

Les conditions sur le rang et la trace de E sont nécessaires et suffisantes pour assurer sa décomposition sous cette forme. Il existe des techniques permettant de retrouver R et t à partir de E. Elles sont présentées à la section 7.6.1.5.

7.3.1.1 Interprétation géométrique : la géométrie épipolaire

La relation 7.6 signifie l'appartenance de m_2, le correspondant de m_1, à une droite donnée, appelée *droite épipolaire* associée à m_1 et notée $\Lambda(m_1)$ à la figure 7.4. De même, le point m_1 de la première image est situé sur la droite épipolaire conjuguée de $D(m_1)$. Cette droite correspond à la projection dans la deuxième image de la ligne de vue (m_1, C_1), C_1 étant le centre optique de la première caméra. On peut également voir les deux droites épipolaires conjuguées comme les intersections respectives des deux plans rétiniens I_1 et I_2 avec le plan épipolaire défini par M et les centres optiques des deux caméras. Notons que les deux faisceaux de droites épipolaires engendrés sont en correspondance homographique. Dans sa thèse [Luong, 1992], Luong présente une méthode permettant de déterminer cette homographie.

7.3.2 La matrice fondamentale

La matrice fondamentale est un concept clé lorsque les images ne sont pas calibrées. Luong a montré dans [Luong, 1992] qu'il existe une entité plus appropriée que la matrice essentielle pour l'étude des caméras qui est une matrice (3×3) appelée matrice fondamentale et notée F_{12}. Elle est reliée à la matrice essentielle par une simple relation linéaire :

$$F_{12} = A_2^{-1T} E_{12} A_1^{-1} \tag{7.11}$$

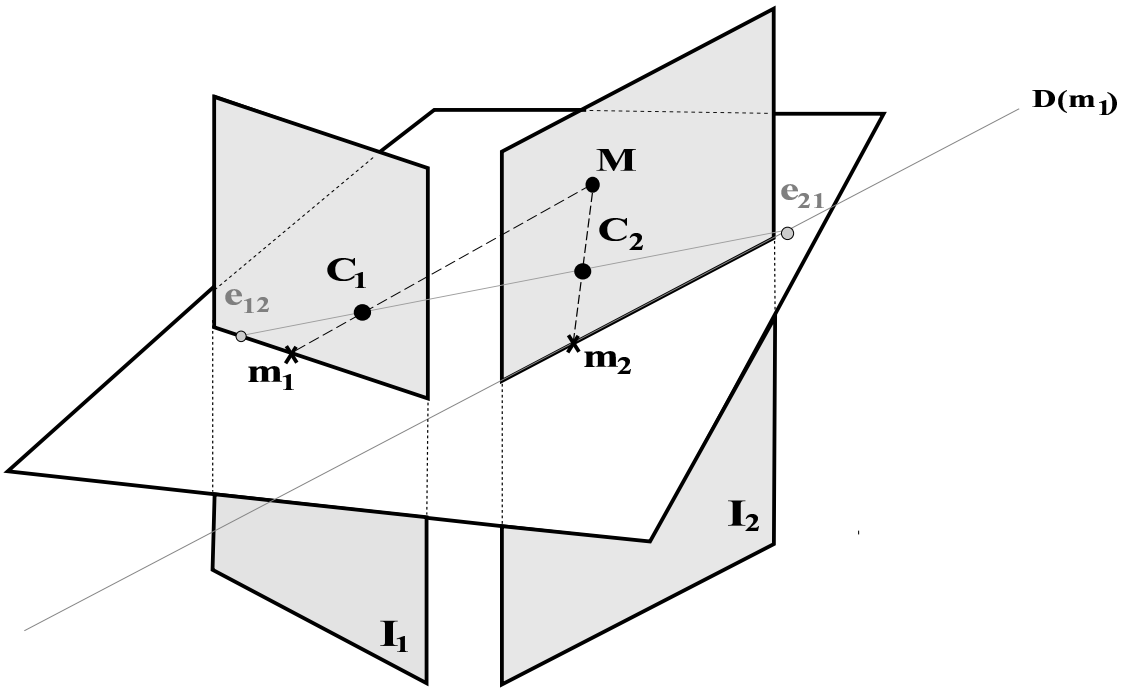


Figure 7.4 – *Géométrie épipolaire d'un système de deux caméras.*

où A_1 et A_2 sont les matrices des paramètres intrinsèques définies à l'équation 7.1.

L'équation exprimant la relation entre les pixels m_1 et m_2 se transforme donc en :

$$m_2^T F_{12} m_1 = 0 \tag{7.12}$$

Comme pour la matrice essentielle, $F_{12}m_1 = 0$ représente la droite épipolaire sur laquelle se trouvent les correspondants possibles de m_1 dans la seconde image. Réciproquement, $F_{12}^T m_2 = 0$ traduit la droite épipolaire associée à m_2 dans la première image. Ces deux droites sont en correspondance homographique. Le noyau de F_{12} (resp. F_{12}^T) est l'épipôle e_{12} (resp. e_{21}) qui correspond à l'image du centre optique de la seconde (resp. première) caméra vu de la première (resp. seconde).

Comme nous l'avons vu à l'équation 7.5, le système ainsi décrit possède $11 \times 2 - 15 = 7$ paramètres. En effet, la matrice fondamentale est une matrice (3×3) définie à un facteur d'échelle près et caractérisée par un déterminant nul. Elle ne dépend donc que de 7 paramètres indépendants.

En résumé, la matrice fondamentale contient toute l'information d'un système binoculaire projectif et c'est la seule contrainte géométrique qui existe entre deux images lorsque les paramètres intrinsèques et extrinsèques des caméras sont inconnus. Lorsque cette matrice est disponible, on dit que le système binoculaire est *faiblement calibré.*

7.3.3 Une configuration particulière : les scènes planes

Considérons des points de l'espace M_i se trouvant en configuration générale dans un même plan Π, et leur image m_i. Les points tridimensionnels et leurs projections respectives sont reliés par une relation projective linéaire[3] de $\mathcal{P}^3$ vers $\mathcal{P}^2$. Celle-ci se restreint dans ce cas à une relation entre le plan Π et le plan de l'image obtenue, et est donc une homographie h du plan $\mathcal{P}^2$. Si nous avons à notre disposition deux images m_i et m'_i des mêmes points de la scène, alors la relation entre ces deux projections est $h' \circ h^{-1}$, comme l'illustre la figure 7.5. C'est également une relation homographique caractérisée par une matrice inversible H et telle que :

$$m'_i = H m_i \tag{7.13}$$

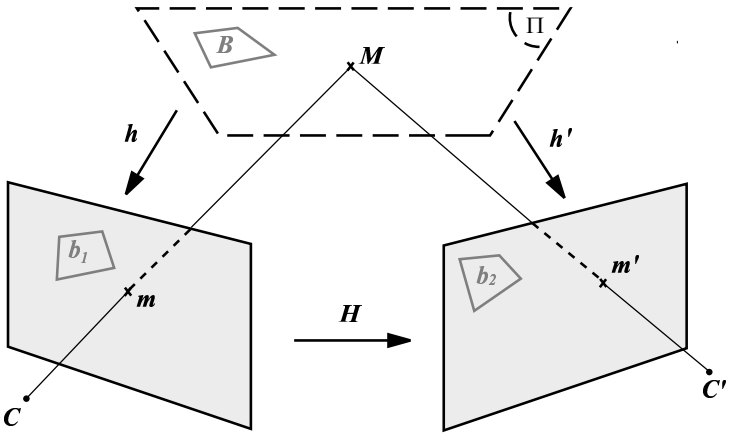


Figure 7.5 – *Points en correspondance homographique dans le plan projectif* $\mathcal{P}^2$.

Un minimum de quatre projections en correspondance et issues de points du plan de référence Π est nécessaire pour calculer cette homographie (cf. annexe E). Les méthodes d'estimation employées sont similaires à celles développées pour l'estimation de la matrice fondamentale (cf. section 7.6.1). Il est également possible de se limiter à trois points (suffisants pour définir le plan) et de la relation suivante exploitant la matrice fondamentale F :

$$H^T F + F^T H = 0 \tag{7.14}$$

7.4 Le cas de trois caméras

L'étude de la géométrie d'un système de trois caméras nous intéresse car les propriétés qui en découlent seront utiles pour calculer des points de vues intermédiaires à partir de deux images (cf. section 8.2 du chapitre 8). De plus, c'est la configuration générique avec le plus petit nombre de caméras. En effet, il a été montré [Beardsley et al., 1992] qu'aucun phénomène mathématique nouveau n'apparaît pour un nombre de caméras supérieur à 3.

[3]Le lecteur peut se référer à l'annexe E pour une introduction à la géométrie projective.

7.4.1 Les contraintes épipolaires

Pour un système trinoculaire, il y a trois matrices fondamentales, totalisant $3 \times 7 = 21$ paramètres. Ces paramètres ne sont pas indépendants, puisque le système trinoculaire possède réellement $11 \times 3 - 15 = 18$ paramètres indépendants (cf. équation 7.5). Il faut à partir de là considérer deux cas de figures, selon que les centres optiques des trois caméras sont alignés ou non. Le cas des centres optiques alignés se produit assez couramment, l'utilisateur pouvant être seulement amené à translater la caméra sur une ligne ou encore à lui appliquer une rotation autour de son centre optique.

7.4.1.1 Centres optiques non alignés

Si les trois centres optiques C_1, C_2 et C_3 ne sont pas alignés, alors les épipôles associés appartiennent tous au même plan défini par (C_1, C_2, C_3) et appelé *plan trifocal*. Cette notion est illustrée par la figure 7.6. Il existe alors dans ce plan les contraintes suivantes entre les épipôles [Faugeras et Robert, 1994] :

$$\begin{cases} F_{23}.e_{21} = e_{31} \wedge e_{32} \\ F_{31}.e_{32} = e_{12} \wedge e_{13} \\ F_{12}.e_{13} = e_{23} \wedge e_{21} \end{cases} \tag{7.15}$$

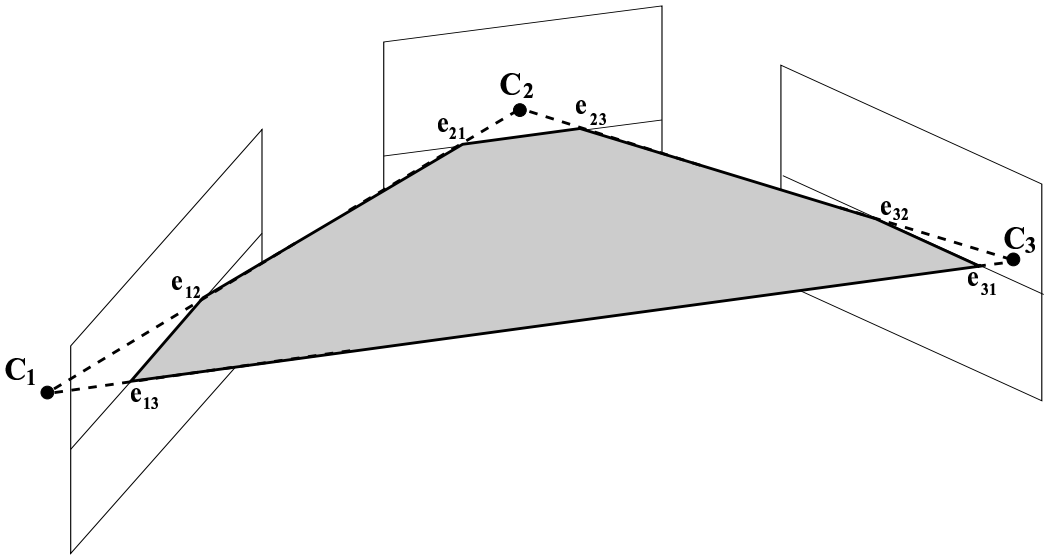


Figure 7.6 – *Contraintes épipolaires dans le plan trifocal.*

Considérons la première relation qui implique les projections e_{21} et e_{31} du centre optique C_1. La droite épipolaire de e_{21} dans la troisième image est $F_{23}.e_{21}$. Cette droite contient e_{31} et e_{32} ; elle peut donc également être exprimée par le produit vectoriel[4] $e_{31} \wedge e_{32}$. Il en est de même pour les deux autres centres optiques.

[4]Rappelons qu'en projectif, la droite (MN) a pour équation $M \wedge N$.

7.4.1.2 Centres optiques alignés

Dans le cas de centres optiques tous alignés, les contraintes sont simplifiées, puisqu'on obtient :

$$\begin{cases} e_{12} = e_{13} \\ e_{23} = e_{21} \\ e_{31} = e_{32} \end{cases} \tag{7.16}$$

La détermination d'un épipôle dans chacune des images fixe les trois autres. D'un point de vue géométrique, tous les épipôles sont sur la même droite (C_1, C_2, C_3). Les degrés de liberté du système sont alors bien moins nombreux. Dans ce cas, les matrices fondamentales ne contiennent pas toute l'information du système trinoculaire. Elles ne permettent pas de retrouver toute la structure projective de la scène.

Les matrices fondamentales ne semblent pas constituer un bon moyen pour caractériser un système composé de trois caméras. En effet, elles sont redondantes (avec 21 paramètres) et l'expression des contraintes de l'équation 7.15 est relativement complexe, car polynômiale d'ordre 3 ou 4. Dans certains cas, elles ne contiennent pas toute l'information du système trinoculaire. D'autres entités semblent davantage appropriées, comme les trilinéarités, que nous présentons à la section suivante.

7.4.2 Les trilinéarités

En 1994, Shashua a montré l'existence de quatre relations entre les points (x, y), (x', y') et $(x", y")$ en correspondance dans trois images. Ces relations sont données par le système suivant :

$$\begin{cases} \alpha_1 + \alpha_2 x + \alpha_3 x" + \alpha_4 y + \alpha_5 y' + \alpha_6 x x" + \\ \qquad \alpha_7 y y' + \alpha_8 x y' + \alpha_9 x" y + \alpha_{10} x" y' + \alpha_{11} x" y y' + \alpha_{12} x x" y' = 0 \\ \\ \alpha_{13} + \alpha_{14} x + \alpha_{15} y + \alpha_{16} y' + \alpha_3 y" + \alpha_{17} y y' + \\ \qquad \alpha_9 y y" + \alpha_{10} y' y" + \alpha_{18} x y' + \alpha_6 x y" + \alpha_{12} x y' y" + \alpha_{11} y y' y" = 0 \\ \\ \alpha_{19} + \alpha_{20} x + \alpha_5 x' + \alpha_{21} x" + \alpha_{22} y + \alpha_8 x x' + \\ \qquad \alpha_{23} x x" + \alpha_{10} x' x" + \alpha_7 x' y + \alpha_{24} x" y + \alpha_{11} x' x" y + \alpha_{12} x x' x" = 0 \\ \\ \alpha_{25} + \alpha_{26} x + \alpha_{16} x' + \alpha_{27} y + \alpha_{21} y" + \alpha_{18} x x' + \\ \qquad \alpha_{24} y y" + \alpha_{17} x' y + \alpha_{23} x y" + \alpha_{10} x' y" + \alpha_{12} x x' y" + \alpha_{11} x' y y" = 0 \end{cases} \tag{7.17}$$

Les 27 paramètres α_i sont les coefficients représentant la géométrie relative des trois caméras. Ils ne sont pas indépendants, puisque le système n'est fonction que de 18 paramètres. Ces quatre formes ont été démontrées dans [Shashua, 1994] mais aussi dans [Hartley, 1994]. Ce dernier a en effet montré que les trilinéarités de Shashua sont identiques aux relations issues du tenseur trifocal, de dimension $(3 \times 3 \times 3)$ et ayant pour équation :

$$T_{ijk} = a_{ji} b_{k4} - a_{j4} b_{ki} \qquad \forall i, j, k \in \{1, 2, 3\} \tag{7.18}$$

où les a_{ij} et b_{ij} représentent les coefficients des deux matrices de projection. Bon nombre d'auteurs expriment d'ailleurs les trilinéarités sous cette forme. Elle est l'équivalent pour trois images de la matrice fondamentale caractérisant deux images.

Dans ce qui suit, nous serons amenés selon le cas à employer l'expression des trilinéarités sous la forme introduite par Shashua à l'équation 7.17, ou bien sous la forme du tenseur trifocal de Hartley (équation 7.18). Selon la façon dont il aura été calculé, le tenseur pourra être caractérisé soit par une représentation *minimale* qui ne dépend que de 18 paramètres, c'est-à-dire pour laquelle les paramètres vérifient des contraintes définies entre eux, soit par une représentation qui dépend de plus de paramètres et qui ne vérifie aucune contrainte spécifique à la géométrie trinoculaire.

Autres contraintes

Notons $T(m_1, m_2, m_3)$ les trilinéarités du triplet de points (m_1, m_2, m_3). Il est facile de voir géométriquement que les trilinéarités vérifient :

$$\begin{cases} \forall m_1 & T(m_1, e_{21}, e_{31}) = 0 \\ \forall m_2 & T(e_{12}, m_2, e_{32}) = 0 \\ \forall m_3 & T(e_{13}, e_{23}, m_3) = 0 \end{cases} \tag{7.19}$$

Il est montré dans [Faugeras et Mourrain, 1995] que ces équations peuvent s'exprimer sous la forme de neuf équations différentes et indépendantes. On obtient ainsi les $27-18 = 9$ contraintes supplémentaires permettant de réduire l'espace du tenseur à un sous-espace de dimension 18, conduisant ainsi à une paramétrisation de T minimale. Dans [Papadopoulo et Faugeras, 1998], une nouvelle formulation du tenseur trifocal est proposée, permettant ainsi d'aboutir à une représentation minimale.

En réalité, les trilinéarités s'avèrent être équivalentes à une reconstruction projective *implicite* à partir de deux vues puis à une reprojection sur la troisième. Les paramètres α_i apparaissent comme étant des combinaisons des coefficients des trois matrices de projection perspective. Nous verrons au prochain chapitre qu'une reconstruction projective explicite se révèle plus facilement manipulable.

Signalons tout de même que contrairement aux contraintes basées sur les deux matrices fondamentales définies à la section précédente, les trilinéarités ont les avantages d'être d'une part valides quelle que soit la configuration géométrique des centres optiques et d'autre part, d'exploiter les trois vues en même temps.

Relations entre droites

Hartley a montré dans [Hartley, 1994] que trois droites en correspondance dans trois images vérifient une relation fonction du tenseur trifocal. En effet, connaissant l'image d'une droite dans deux caméras, il est possible de déduire son image dans une troisième caméra à partir de la formule suivante :

$$\lambda_3^i = \sum_{j,k} \lambda_1^j \lambda_2^k T_{ijk} \tag{7.20}$$

où λ_I^i est la $i$ème coordonnée du représentant projectif de la droite projetée dans

l'image I. Nous verrons plus loin que cette relation sera très utile pour l'estimation du tenseur ainsi que dans le chapitre suivant pour faire du transfert d'images.

7.5 Relations entre les différentes entités

Dans cette section, nous synthétisons les méthodes permettant de passer des bi- et trilinéarités aux matrices de projection et réciproquement. Nous allons notamment voir que le passage des matrices de projection vers les autres entités est une opération facile, alors que les opérations inverses sont beaucoup plus délicates à réaliser.

Les méthodes permettant de passer des bilinéarités aux trilinéarités et vice versa se seront pas passées en revue ici, puisqu'il est possible de les relier par les matrices de projection.

7.5.1 Relation entre la matrice fondamentale et les matrices de projection

7.5.1.1 De F vers P

Dans cette section, nous supposons connue la matrice fondamentale F_{12} reliant les deux images. Nous cherchons alors à partir de celle-ci un jeu possible de matrices de projection. Le système étant faiblement calibré, celles-ci seront bien sûr définies à une transformation projective de $\mathcal{P}^3$ près. Ce problème a été étudié par [Faugeras et al., 1992; Mundy et Zisserman, 1992; Luong et Vieville, 1994]. Il en ressort :

$$\begin{cases} P_1 = [I|0] \\ P_2 = [\widehat{e}_{21}.F_{12}|e_{21}] \begin{bmatrix} I & 0 \\ \alpha^T & \alpha_4 \end{bmatrix} .G \end{cases} \tag{7.21}$$

où $\widehat{e}_{21}$ représente la matrice antisymétrique associée à l'épipôle e_{21} de la seconde image. La matrice G est une matrice (4×4) arbitraire. Comme les matrices de projection sont définies à une transformation projective près, on a coutume de prendre l'identité pour G. Les scalaires α_i, $i \in \{1, .., 4\}$ sont également pris arbitrairement ; α_4 ne doit cependant pas être nul, pour que la matrice de projection composée par les α_i ne soit pas singulière. Ils représentent les $11 - 7 = 4$ paramètres manquants[5] permettant de déterminer P_2. En pratique, ces valeurs sont choisies de façon que la matrice de projection obtenue soit numériquement bien conditionnée.

La solution de l'équation 7.21 est bien sûr définie à un facteur d'échelle près. Elle est consistante avec la géométrie épipolaire du système. Remarquons qu'elle est asymétrique, l'une des matrices (P_1) est définie comme la matrice de projection canonique, alors que l'autre (P_2) contient toute l'information. Ces deux matrices vont permettre de construire un modèle projectif de la scène (cf. section 8.1.1 du chapitre 8).

[5]Rappelons que la matrice de projection perspective est caractérisée par 11 paramètres et la matrice fondamentale par 7.

Cas de trois caméras

Si le système possède plus de deux caméras, il est possible de mettre en place des relations entre les scalaires α_i des diverses matrices de projection. Par exemple, pour trois caméras, les quatre scalaires notés β_i relatifs à la matrice de projection P_3 s'expriment en fonction des α_i de P_2 selon une relation de la forme suivante [Laveau, 1996] :

$$\beta_i = \frac{u^T(\alpha_i v + w)}{u^T u} \tag{7.22}$$

où u,v et w sont des vecteurs de dimension 3 fonctions des matrices fondamentales et des épipôles entre les trois images. Cette relation n'est valable que si les centres optiques des trois caméras ne sont pas alignés. Il est nécessaire d'avoir à disposition des données supplémentaires dans le cas où ils le sont, comme par exemple la donnée d'une trilinéarité ou encore trois triplets de points en correspondance.

7.5.1.2 De P vers F

Il est extrêmement aisé de passer des matrices de projection à la matrice fondamentale. Si on décompose la matrice de projection P_i comme suit : $P_i = [Q_i|T_i]$, alors la matrice fondamentale engendrée est telle que :

$$F_{12} = \left[T_2 - \widehat{Q_2 Q_1^{-1} T_1}\right] Q_1 Q_2^{-1} \tag{7.23}$$

où $[\widehat{X}]$ dénote ici aussi la matrice antisymétrique associée à X. Cette relation est vraie si les matrices Q_i sont inversibles.

7.5.2 Relation entre les trilinéarités et les matrices de projection

7.5.2.1 De T vers P

Dans cette section, nous supposons connu le tenseur trifocal de l'équation 7.18. Nous nous plaçons ici dans une base projective telle que P_1 ait la forme canonique $P_1 = [I|0]$. Soient $P_2 = [a_{ij}]$ et $P_3 = [b_{ij}]$ les deux autres matrices de projection. Dans ce contexte, les épipôles associés au centre C_1 vérifient :

$$e_{21} = [a_{i4}] \quad \text{et} \quad e_{31} = [b_{i4}] \tag{7.24}$$

Ainsi, d'après le tenseur de l'équation 7.18, si les épipôles e_{21} et e_{31} sont connus, alors le problème du calcul de P_2 et P_3 en fonction des trilinéarités devient un problème linéaire de 27 équations à 18 inconnues, et donc peut être résolu aux moindres carrés.

Détermination des deux épipôles

Dans [Laveau, 1996], Laveau reformule le tenseur de l'équation 7.18 de la façon suivante :

$$T_i^T = a_i b_4^T - a_4^T b_i \quad \forall i \in \{1, 2, 3\} \tag{7.25}$$

où a_i représente la $i$ème colonne de la matrice P_2 et b_i la $i$ème colonne de la matrice P_3. Il remarque alors que $r_i = a_i \wedge a_4$ appartient au noyau de T_i et $s_i = b_i \wedge b_4$ au noyau de T_i^T. Il suffit donc de résoudre les systèmes suivants pour trouver r_i et s_i :

$$T_i r_i = 0 \qquad \text{et} \qquad T_i^T s_i = 0 \tag{7.26}$$

Par construction, on détermine également a_4 (resp. b_4), et donc e_{21} (resp. e_{31}) en résolvant :

$$r_i^T a_4 = 0 \qquad \text{et} \qquad s_i^T b_4 = 0 \tag{7.27}$$

7.5.2.2 De P vers T

Nous nous plaçons ici dans le même contexte qu'à la section précédente, à savoir $P_1 = [I|0]$, $P_2 = [a_{ij}]$ et $P_3 = [b_{ij}]$. Alors pour un triplet de points (m_1, m_2, m_3) en correspondance, Hartley exprime les trilinéarités en fonction des matrices de projection de la façon suivante :

$$\sum_j m_1^j (m_2^i m_3^k T_{jml} - m_2^m m_3^k T_{jil}) = \sum_j m_1^j (m_2^i m_3^l T_{jmk} - m_2^m m_3^l T_{jik}) \tag{7.28}$$

où T_{ijk} est le tenseur trifocal défini à l'équation 7.18 en fonction des coefficients de P_2 et de P_3 et où m_i^j représente la $j$ème coordonnée du point m_i, en coordonnées projectives (soit $j \in \{1, 2, 3\}$). En faisant varier les indices i,k,l et m avec $i < m$ et $l < k$, on obtient neuf équations non linéairement indépendantes. Si l'on normalise les points m_i en fixant les coordonnées $m_i^3 = 1$ et si l'on choisit $m = k = 3$, alors on retrouve les relations décrites par Shashua dans [Shashua, 1994].

7.6 Estimation des différentes entités

7.6.1 Estimation de F

Les premiers travaux conséquents sur le calcul de la géométrie épipolaire sont dus à Longuet-Higgins (1981) et portent sur l'estimation de la matrice essentielle. Depuis, des études similaires ont été menées pour l'estimation de la matrice fondamentale, les travaux les plus récents fournissant même des critères d'estimation beaucoup plus sophistiqués.

Les différentes méthodes existantes peuvent être séparées en deux grandes classes. La première classe est basée sur une minimisation de critères *linéaires* alors que la seconde exploite des critères *non linéaires* plus complexes. Ces deux classes font l'objet des deux prochaines sections. Des méthodes dites *robustes* sont également apparues. Elles utilisent selon le cas des critères linéaires ou non linéaires mais ont la particularité de tenir compte dans leur processus d'optimisation d'un éventuel bruit contenu dans les données, rendant ainsi l'estimation de la matrice fondamentale précise même en présence de données faussées. Ces approches sont détaillées à la section 7.6.1.3. Puis nous faisons un bilan des différentes méthodes passées en revue à la section 7.6.1.4. Enfin, le cas de trois caméras est envisagé à la section 7.6.1.4.

7.6.1.1 Critères linéaires

L'équation de base 7.12 reliant deux projections à l'aide de la matrice fondamentale peut être réécrite selon l'équation linéaire et homogène en les 9 coefficients de la matrice fondamentale F suivante :

$$u^T f = 0 \tag{7.29}$$

avec

$$\begin{aligned} m_i &= (x_i, y_i, 1)^T \\ u &= [x_1x_2, y_1x_2, x_2, x_1y_2, y_1y_2, y_2, x_1, y_1, 1]^T \\ f &= [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T \end{aligned}$$

Il existe quelques méthodes permettant de résoudre ce système, la plus connue étant bien sûr l'algorithme des huit points que l'on doit à Longuet-Higgins [Longuet-Higgins, 1981].

L'algorithme des huit points

F étant définie à un facteur d'échelle près, 8 appariements suffisent pour déterminer une solution unique. Cette approche a donné lieu à un grand nombre de travaux, citons entre autres [Faugeras et al., 1987; Spetsakis et Aloimonos, 1988; Weng et al., 1989; Lee, 1991; Luong, 1992; Zhang et al., 1995b; Hartley, 1997]. Il est en particulier connu que cet algorithme est très sensible au bruit.

Les travaux originaux de Longuet-Higgins n'utilisaient que 8 points pour déterminer F. En pratique, on dispose d'un nombre d'appariements (m_{1i}, m_{2i}) bien plus important. L'équation 7.29 qui correspond alors à un système surdéterminé, peut être résolue aux moindres carrés, ce qui revient à minimiser par rapport à F le critère suivant :

$$\min_F \sum_i (m_{2i}^T F m_{1i})^2 \tag{7.30}$$

que l'on a l'habitude de réécrire comme suit, n étant le nombre de mises en correspondances :

$$\min_f \|Uf\|^2 \quad \text{avec} \quad U = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} \tag{7.31}$$

La matrice F étant définie à un facteur d'échelle près, plusieurs contraintes peuvent être utilisées pour minimiser le critère 7.31. Les deux critères les plus classiques consistent d'une part à fixer à 1 l'une des neuf composantes de F. L'inconvénient de cette normalisation est que nous ne sommes pas à l'abri d'un mauvais choix en ce qui concerne le coefficient à fixer, étant donné que nous n'avons aucune information sur F. Il est d'autre part possible de choisir une contrainte moins restrictive en fixant la norme de f. Le critère 7.31 devient alors une minimisation sous contrainte :

$$\min_f \|Uf\|^2 \quad \text{avec} \quad \|f\| = 1 \tag{7.32}$$

La solution à ce classique problème d'algèbre linéaire est alors obtenue en considérant le vecteur propre associé à la plus petite valeur propre de la matrice U^TU. Dans la pratique, on diagonalisera directement cette matrice à l'aide de la méthode de Jacobi par exemple, ou bien on procédera à une décomposition en valeurs singulières.

La méthode linéaire présente l'avantage de conduire à une solution explicite par un calcul simple et non itératif. En revanche, elle est très sensible au bruit, c'est-à-dire à l'imprécision des appariements, même si ceux-ci sont en grand nombre. Les deux raisons principales dues à cela sont que :

- la contrainte $det(F) = 0$ n'est pas imposée, qui fait que la contrainte de rang[6] n'est pas exprimée. Cela entraîne une incohérence de la géométrie épipolaire au voisinage des épipôles ;
- le critère à minimiser n'est pas normalisé, ce qui conduit à un biais dans la localisation des épipôles.

Ces deux défauts entraînent des erreurs importantes dans l'estimation de la matrice fondamentale et des épipôles. Signalons qu'il existe une méthode permettant de tenir compte de la singularité de la matrice à l'issue de cette résolution linéaire. Elle consiste à modifier la matrice F obtenue à la résolution de l'équation 7.29 par une matrice F' qui minimise la norme de Frobenius $\|F - F'\|$ sous la condition $det(F') = 0$. Une implémentation utilisant la décomposition en valeurs singulières est proposée dans [Tsai et Huang, 1984] pour l'estimation de la matrice essentielle et dans [Hartley, 1997] pour celle de la matrice fondamentale. Cette solution donne de bons résultats, bien qu'elle ne soit pas parfaitement justifiée théoriquement.

Pour finir en ce qui concerne l'algorithme des huit points, citons les travaux de Hartley dans [Hartley, 1995] qui en propose une amélioration. Il remarque en effet que l'estimation de la matrice fondamentale dépend également du choix des coordonnées image dans lesquelles les appariements sont exprimés. Il montre alors qu'une simple transformation de ces coordonnées (une translation et un changement d'échelle) permet d'améliorer de façon conséquente l'algorithme des huit points, pour une complexité supplémentaire insignifiante. La méthode est comparée à plusieurs méthodes itératives (linéaires et non linéaires, parmi celles qui seront présentées ci-après). Les résultats montrent que ce prétraitement des données rend le critère linéaire plus robuste qu'un bon nombre des autres méthodes testées.

Autres méthodes

Il existe d'autres méthodes permettant de résoudre le système 7.29 à partir de 7 appariements seulement (la matrice ayant 7 paramètres indépendants). Ces méthodes fournissent au plus trois solutions qui respectent la contrainte $det(F) = 0$. Elles correspondent à des interpolations linéaires entre les deux matrices fondamentales associées aux vecteurs singuliers issus de la décomposition en valeurs singulières de U (cf. équation 7.31). Le lecteur peut se reporter notamment à [Torr et al., 1995] pour plus de détails concernant la méthode employée.

[6] Rappelons que la matrice fondamentale est seulement de rang 2.

7.6.1.2 Critères non linéaires

Plusieurs méthodes basées sur des critères non linéaires existent pour améliorer les critère linéaires. Nous présentons ici une méthode répandue qui exploite la distance aux droites épipolaires. D'autres méthode existent, comme la pondération par le gradient qui prend en compte l'incertitude dans la minimisation. Le critère obtenu est d'ailleurs fortement similaire à celui obtenu avec la distance aux droites épipolaires. Pour une présentation plus complète et des expérimentations concernant ces autres critères, le lecteur peut consulter [Luong, 1992; Zhang, 1996].

Distance aux droites épipolaires

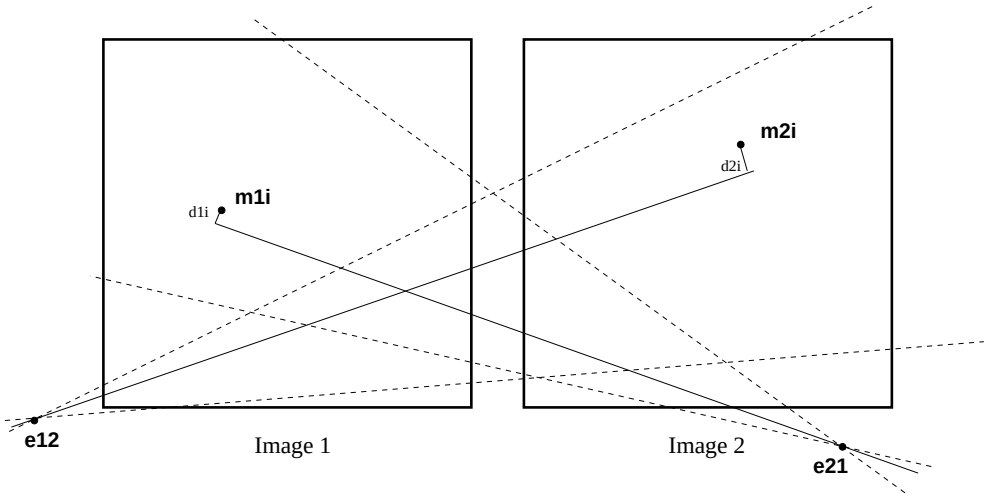


Figure 7.7 – *Distances d_{1i} et d_{2i} aux droites épipolaires Fm_{1i} et $F^T m_{2i}$.*

Il est possible de tenir compte de la distance séparant un point de la droite épipolaire de son correspondant, comme l'illustre la figure 7.7. Le critère doit être symétrique, c'est-à-dire opérer simultanément dans les deux images. On peut considérer :

$$\min_F \sum_i (d^2(m_{2i}, Fm_{1i}) + d^2(m_{1i}, F^T m_{2i})) \tag{7.33}$$

où $d(m, d)$ représente la distance euclidienne entre le point m et la droite d. En utilisant la relation $m_{2i}^T F m_{1i} = m_{1i}^T F^T m_{2i}$, ce critère peut s'écrire :

$$\min_F \sum_i D_i(F)(m_{2i}^T F m_{1i})^2 \tag{7.34}$$

avec

$$D_i(F) = \frac{1}{(Fm_{1i})_1^2 + (Fm_{1i})_2^2} + \frac{1}{(F^T m_{2i})_1^2 + (F^T m_{2i})_2^2} \tag{7.35}$$

où $(Fm_i)_j$ dénote la $j$ème composante du vecteur Fm_i. Contrairement au critère linéaire 7.30, ce critère représente une distance et a donc une interprétation plus physique. Il est normalisé, il ne dépend en effet pas du facteur d'échelle choisi pour F.

Etant donnée la similarité entre le critère qui vient d'être établi et le critère linéaire 7.30, il est possible de mettre en place une méthode linéaire itérative pour résoudre la minimisation. L'approche consiste à pondérer le critère linéaire avec la contrainte des distances aux droites épipolaires. Dans une première étape, une matrice fondamentale F_0 est estimée à partir de l'algorithme des huit points en minimisant le critère 7.30. Puis chacune des équations du système linéaire est multipliée par la racine carrée du poids $D_i(F_0)$ défini à l'équation 7.35. L'algorithme des huit points est alors relancé, et ainsi de suite. Malheureusement, cette variante des méthodes linéaires classiques n'améliore que très peu l'estimation de la matrice, toujours à cause de la contrainte de rang qui n'est pas respectée. On préfère procéder à une paramétrisation de la matrice, qui elle, tient compte du rang.

Paramétrisation

La première paramétrisation qui vient à l'esprit consiste à fixer un des coefficients de la matrice à 1, puisqu'elle est définie à un facteur d'échelle prés. La matrice résultante admet donc un paramétrage sur 8 valeurs, qui sont le quotient de 8 coefficients par le coefficient normalisateur. En pratique, le choix de ce dernier est important vis-à-vis de la stabilité numérique des résultats dans le cas des critères non linéaires. Il semble que les meilleurs résultats soient obtenus en choisissant comme coefficient celui qui a la plus grande valeur absolue [Luong, 1992].

Malheureusement cette approche ne tient toujours pas compte du fait que F est de rang deux avec un déterminant nul. Luong dans [Luong, 1992] propose alors d'exprimer la dernière ligne de F comme une combinaison linéaire des autres lignes, après avoir fixé un des coefficients à 1 et avant de faire appel à des techniques de minimisation non linéaires.

Il est également possible de tenir compte de la structure de la matrice fondamentale pour obtenir un paramétrage par les éléments de la transformation épipolaire. On obtient par exemple la matrice de rang deux suivante [Luong et al., 1993] :

$$F = \begin{pmatrix} a & b & -ax' - by' \\ c & d & -cx' - dy' \\ -ax - cy & -bx - dy & (ax' + by')x + (cx' + dy')y \end{pmatrix} \tag{7.36}$$

où (x, y) et (x', y') sont les coordonnées des deux épipôles et les coefficients ceux de la sous-matrice (2×2) extraite de l'homographie entre les deux faisceaux de droites épipolaires en supprimant la troisième ligne et la troisième colonne. La matrice obtenue est effectivement de rang deux et dépend de sept paramètres. Notons que la formulation proposée initialement par Luong n'est exploitable que pour des épipôles finis. Dans [Zhang, 1996], elle est améliorée afin d'être aussi valable pour des épipôles à l'infini.

7.6.1.3 Critères robustes

Les méthodes d'estimation *robuste* de la matrice fondamentale consistent à considérer que les données sont bruitées, c'est-à-dire que certains des appariements peuvent être

incorrects[7]. Elles ont même la capacité d'en éliminer une partie au cours du processus d'optimisation.

La première approche est une généralisation du critère linéaire. La méthode a été proposée par [Golub et Loan, 1989] et s'appelle méthode des moindres carrés total. Elle consiste à considérer que les données L sont bruitées puis à minimiser le critère ci-dessous, la norme considérée étant celle de Frobenius :

$$\min_{(L+E)x=b+r} \| [E|r] \|_F \tag{7.37}$$

Malheureusement Luong a montré dans sa thèse que cette méthode donne des résultats encore moins précis qu'avec le critère linéaire classique. Les matrices obtenues semblent meilleures avec les méthodes itératives qui permettent de disposer de critères non linéaires. Ces méthodes sont présentées ci-après.

Ajustement des points

Cette approche consiste à considérer un critère aux moindres carrés $\sum_i C(F, m_{1i}, m_{2i})^2$ parmi ceux présentés aux sections précédentes et à effectuer une correction optimale sur les appariements. A chaque itération, une matrice fondamentale F est estimée dans une première étape, en minimisant le critère $C(F, m_{1i}, m_{2i})$. La seconde étape consiste ensuite à déplacer les points dans une certaine mesure de façon à minimiser le critère suivant :

$$\sum_j C(F, m_{1j}^{(f)}, m_{2j}^{(f)})^2 + \lambda \sum_i \{d^2(m_{1j}^{(f)}, m_{1j}^{(i)}) + d^2(m_{2j}^{(f)}, m_{2j}^{(i)})\} \tag{7.38}$$

Le lecteur peut consulter [Luong, 1992] pour plus de détails concernant ces mesures. Le processus itératif est arrêté lorsque les points ne varient plus de manière significative. Cette procédure semble ne pas être très convaincante. Luong a tout de même retenu des expérimentations menées que l'introduction d'un seul point très faux peut modifier entièrement les résultats, et qu'un déplacement minime des autres points conduit à une géométrie épipolaire présentant une compatibilité élevée.

Les M-estimateurs

Soit $r_i = C(F, m_{1i}, m_{2i})$ la valeur résiduelle associée à la $i$ème donnée. Les méthodes traditionnelles basées sur les M-estimateurs consistent à remplacer la partie résiduelle au carré, utilisée dans les méthodes linéaires aux moindres carrés, par une autre fonction de cette partie. Le critère obtenu est de la forme :

$$\min \sum_i \rho(r_i) \tag{7.39}$$

où ρ est une fonction symétrique, définie positive avec un unique minimum à zéro. Les M-estimateurs peuvent également être vus comme un problème aux moindres carrés pondéré. Le critère à minimiser devient :

$$\min \sum_i \rho_\sigma(r_i) r_i^2 \tag{7.40}$$

[7] Un appariement est considéré comme faux si les points qu'il implique sont mal localisés ou encore si les deux points ne vérifient pas la contrainte épipolaire.

les poids $\rho_\sigma(r_i)$ étant estimés à partir des itérations précédentes. Chez plusieurs auteurs [Luong, 1992; Olsen, 1992], ils sont de la forme suivante :

$$\rho_\sigma(r_i) = \begin{cases} 1 & |r_i| \leq \sigma \\ \frac{\sigma}{|r_i|} & \sigma < |r_i| \leq c\sigma \\ 0 & c\sigma < |r_i| \end{cases} \tag{7.41}$$

où c désigne un seuil et σ l'écart-type des erreurs r_i. $\rho_\sigma(r_i)$ décroît avec l'éloignement de l'appariement i de la zone trouvée par les moindres carrés. Des expérimentations [Zhang, 1996; Torr et Murray, 1997] ont montré que cette méthode est robuste face aux points mal localisés mais en revanche sensible aux faux appariements, en partie à cause de la première estimation de la matrice fondamentale.

Les approches basées sur les moindres médianes des erreurs au carré (LMedS)

Par ce type de méthodes, les paramètres de la matrice sont estimés en résolvant le problème de minimisation non linéaire suivant :

$$\min \underset{i}{\text{med}}\, r_i^2 \tag{7.42}$$

Cette méthode est robuste face aux points mal localisés comme face aux faux appariements. Malheureusement, l'espace de recherche engendré par les données est beaucoup trop important et le problème ne peut pas être réduit à un problème aux moindres carrés. La solution proposée dans [Zhang et al., 1995b] consiste à minimiser le critère 7.42 à partir d'une partie des données. Les auteurs utilisent la technique de Monte Carlo pour déterminer m sous-ensembles de 8 appariements parmi les n appariements[8]. Puis une matrice fondamentale F_s linéaire est déterminée pour chacun de ces sous-ensembles. C'est ensuite le critère basé sur la distance aux droites épipolaires $r_i^2 = d^2(m_{2i}, F_s m_{1i}) + d^2(m_{1i}, F_s^T m_{2i})$ qui est utilisé pour la minimisation de l'équation 7.42, avec une paramétrisation basée sur les éléments de la transformation épipolaire (cf. équation 7.36). A partir de là, m médianes associées à chacune des matrices F_s sont calculées sur l'ensemble des données ($i = 1, ..., n$). La méthode LMedS s'avérant sensible en présence de bruit gaussien, il ne suffit pas de retenir simplement la matrice F_s associée à la plus petite de ces médianes. Une méthode aux moindres carrés pondérée est alors utilisée pour estimer la matrice fondamentale qui sera retenue. Le critère à minimiser est à nouveau celui basé sur la distance aux droites épipolaires. Les poids utilisés exploitent la plus petite médiane obtenue à l'étape précédente pour déterminer un seuil de distance. Ils permettent ainsi d'éliminer du calcul les appariements qui sont à une distance trop importante des droites épipolaires conjuguées, et par conséquent considérés comme incorrects. Cette variante rend la méthode plus robuste que les méthodes LMedS classiques, puisque seuls les *inliers* sont impliqués dans le calcul.

7.6.1.4 Synthèse des différentes approches présentées

Nous venons de passer en revue les principales méthodes d'estimation de la géométrie épipolaire d'un système stéréoscopique non calibré. En résumé :

[8]Dans leurs expérimentations, m est calculé en considérant que les données peuvent contenir jusqu'à 40% d'appariements incorrects. Une méthode de *bucketing* est également employée pour obtenir des sous-ensembles d'appariements correctement répartis dans les images.

Au moins 7 mises en correspondance de points sont nécessaires pour déterminer la matrice fondamentale entre les deux images. Avec seulement 7 appariements, on obtient trois solutions en résolvant une équation cubique, qui tient compte de la contrainte sur le déterminant de la matrice. Si plus de données sont disponibles, alors une solution unique existe et de multiples méthodes ont été mises en place pour l'estimer. Ces méthodes ont été décomposées en trois classes :

Les méthodes linéaires : elles sont simples à mettre en place mais ont le principal inconvénient d'être sensibles au bruit et de ne pas être stables, d'une part parce qu'elles ne tiennent pas compte de la singularité de la matrice fondamentale, et d'autre part parce que le critère minimisé n'a pas de réalité physique. Il est tout de même possible de les améliorer en faisant subir un pré-traitement aux appariements (travaux de Hartley). Les résultats restent tout de même moins bons qu'avec les meilleures méthodes non linéaires ou robustes ;

Les méthodes non linéaires : les critères non linéaires fournissant les meilleurs résultats sont ceux relatifs à la distance aux droites épipolaires et à la pondération du gradient (ils sont d'ailleurs très semblables). L'estimation est encore améliorée lorsque la matrice fondamentale est reparamétrée afin de tenir compte explicitement de la contrainte de rang deux ;

Les méthodes robustes : contrairement aux précédentes, ce sont des approches itératives qui tiennent compte du bruit contenu dans les données, c'est-à-dire d'éventuels faux appariements. Elles permettent même d'en éliminer une partie au cours de l'estimation. Les méthodes permettant d'aboutir à la meilleure estimation sont celles basées sur les M-estimateurs et sur les LMedS. Ce sont en effet les plus robustes car elles éliminent les faux appariements au cours du processus itératif, par un système de pondération des données. Les M-estimateurs peuvent être traités comme un problème aux moindres carrés pondéré. Les expérimentations ont cependant montré que cette approche est robuste face aux points mal localisés mais sensible aux faux appariements. Au contraire, les approches LMedS sont aussi bien robustes face aux points mal localisés que face aux faux appariements. Dans ce cas, la minimisation est non linéaire et il n'existe pas de méthode simple pour l'exprimer, ce qui rend malheureusement l'espace de recherche beaucoup trop grand. Zhang et son équipe ont alors proposé une variante de cette approche qui travaille sur un sous-espace des données, permettant ainsi de rendre cette approche réalisable ;

Le lecteur peut consulter [Luong et Faugeras, 1995; Zhang, 1996] pour une mise en œuvre et une comparaison de toutes ces méthodes et [Torr et Murray, 1997] spécifiquement pour les méthodes robustes.

Cas de trois caméras

Il est possible de généraliser le calcul des matrices fondamentales pour un nombre quelconque d'images. La méthode consiste à calculer les matrices F_{ij} en utilisant la même paramétrisation, mais en opérant une minimisation sous contrainte. Laveau montre dans [Laveau, 1996] que ce calcul s'avère très complexe dans le cas théorique de N caméras, puisque le nombre de contraintes varie entre $O(N^3)$ et $O(N^4)$. Dans le cas de trois caméras, on comptabilise trois contraintes, ce qui rend l'estimation encore réalisable dans la pra-

tique.

7.6.1.5 Décomposition de la matrice essentielle

Dans le cas où la matrice fondamentale et les matrices des paramètres intrinsèques sont connues, alors la relation 7.11 donne la matrice essentielle E. Les paramètres du mouvement entre les deux caméras peuvent alors être obtenus en décomposant cette matrice selon l'équation 7.10 [Longuet-Higgins, 1981]. Il est connu que la factorisation n'est pas unique, la translation t ne pouvant être déterminée qu'à un facteur d'échelle près et la rotation R à un angle π près. Ces ambiguïtés peuvent être levées à partir de la donnée d'une seule correspondance de points, le point reconstruit devant être localisé devant les caméras. La matrice E étant le plus souvent une approximation de la véritable matrice essentielle, les méthodes développées à la suite de Longuet-Higgins [Tsai et Huang, 1984; Faugeras et al., 1987; Hartley, 1992] consistent à décomposer cette approximation en vérifiant de manière seulement approchée les conditions de décomposabilité. Ces méthodes sont au nombre de deux :

Approche basée sur les quaternions

La propriété $E^T t = 0$ déduite de $\widehat{t}t = 0$ permet de déterminer t aux moindres carrés en minimisant le critère :

$$\min_{\|t\|=1} \|E^T t\|^2 \tag{7.43}$$

Puis la rotation R est estimée en minimisant le critère suivant [Faugeras et al., 1987] :

$$\|E - \widehat{t}R\|^2 \tag{7.44}$$

Une solution explicite peut être obtenue grâce à l'utilisation des quaternions pour représenter la rotation.

Décomposition en valeurs singulières

La seconde méthode est celle de Tsai et Huang [Tsai et Huang, 1984]. Elle consiste à utiliser la décomposition de E en valeurs singulières pour faire apparaître le produit d'une matrice antisymétrique avec une matrice orthogonale, représentant respectivement $\widehat{t}$ et R.

L'intérêt de cette approche réside dans le fait que la factorisation ainsi obtenue est optimale. En effet, si la matrice E décomposée n'est qu'une approximation, alors sa décomposition en t et R minimise $\|E - E'\|^2$, où $E' = \widehat{t}R$ représente la matrice essentielle la plus proche de E vérifiant les propriétés des matrices essentielles[9] [Hartley, 1992].

L'efficacité des deux méthodes présentées dépend bien sûr très fortement de la qualité de la matrice fondamentale estimée. L'estimation du mouvement étant en fait fondée sur ce calcul, il est également possible d'exploiter directement les méthodes d'estimation de la matrice fondamentale. On peut par exemple adapter le critère linéaire 7.30 ou encore

[9]Une de ces propriétés étant que la matrice essentielle possède deux valeurs singulières non nulles identiques et la troisième nulle.

le critère non linéaire 7.34 à partir des relations 7.11 et 7.10 existant entre t, R et F. La solution est alors donnée par la minimisation itérative du nouveau critère obtenu, selon les coefficients de t et de R [Luong, 1992]. Les expériences menées par Luong montrent que cette dernière méthode donne des résultats un peu plus précis qu'avec les deux premières approches, mais elle est tout de même fortement dépendante de l'initialisation du processus itératif.

7.6.2 Estimation de T

Les méthodes que nous venons de décrire pour l'estimation de la matrice fondamentale peuvent également être appliquées à l'estimation des trilinéarités, à quelques différences près que nous passons en revue ci-dessous :

7.6.2.1 Critères linéaires

Les trilinéarités fournissant les quatre relations du système 7.17 dépendent de 27 paramètres et définies à un facteur d'échelle près. Il est donc nécessaire d'avoir à notre disposition 7 triplets (m_{1i}, m_{2i}, m_{3i}) de mise en correspondance pour mettre en œuvre le calcul [Shashua, 1995]. Bien sûr de meilleurs résultats sont obtenus avec un plus grand nombre de points, en utilisant des méthodes de minimisation aux moindres carrés. Cependant, cette méthode souffre de plusieurs inconvénients :

- Les 27 paramètres sont estimés indépendamment les uns des autres alors qu'ils ne le sont pas, puisque le système n'est fonction que de 18 paramètres (cf. équation 7.5) ;
- Comme pour l'estimation de la matrice fondamentale, la quantité minimisée n'a aucune réalité physique. Cette inconvénient revêt encore plus d'importance ici, car le tenseur trilinéaire possède des coefficients qui diffèrent d'ordres de grandeur bien plus importants que ceux de la matrice fondamentale.

Cette minimisation basée sur une approche linéaire est bien sûr très sensible au bruit éventuellement contenu dans les données. Une première solution pour rendre la méthode plus robuste est d'introduire les équations trilinéaires basées sur les droites (cf. équation 7.20). Celles-ci sont en effet détectées de manière plus stable et rendent ainsi les équations moins sensibles au bruit. Chaque triplet de droites en correspondance fournissant deux équations linéaires, il est nécessaire d'avoir 13 appariements pour retrouver les 27 coefficients du tenseur trifocal [Hartley, 1994].

Il est également souhaitable de procéder à un pré-traitement des entrées, comme le préconise Hartley pour l'algorithme des huit points lors de l'estimation de la matrice fondamentale (cf. section 7.6.1.1). Laveau [Laveau, 1996] précise toutefois qu'elles demeurent très sensibles au bruit, même en prenant toutes les précautions nécessaires pour normaliser les données.

Une troisième amélioration consiste à faire la résolution à partir du système 7.28 qui exprime neuf équations, même si certaines sont des combinaisons linéaires des autres. Les résultats restent les mêmes en cas de bruit faible, mais se stabilisent mieux si le bruit contenu dans les données augmente.

7.6.2.2 Critères non linéaires

De manière similaire au critère non linéaire développé pour l'estimation de la matrice fondamentale qui exploite la distance aux droites épipolaires (cf. section 7.6.1.2), il est possible ici aussi de minimiser des distances physiques pour obtenir de meilleurs résultats.

La solution proposée dans [Laveau, 1996; Faugeras et Papadopoulo, 1998] considère les trilinéarités sous la forme 7.28 représentée par neuf équations. Etant donné un triplet de points (m_1, m_2, m_3) en correspondance, l'estimation consiste à minimiser la somme des carrés des distances entre m_k et les neuf droites générées à partir de ces équations et de la donnée de (m_i, m_j). Le critère obtenu est symétrique. Pour obtenir une paramétrisation minimale du tenseur, il est possible de tenir compte des neuf autres contraintes de l'équation 7.19 ou bien encore des relations entre le tenseur et les matrices de projection (cf. section 7.5.2). Le système à minimiser comporte donc 18 variables avec $3 \times 9 = 27$ équations par appariement présent dans les trois images. Cette approche a été testée et évaluée dans [Faugeras et Papadopoulo, 1998] ; les résultats obtenus sont bien meilleurs qu'avec les méthodes linéaires jusqu'alors employées.

7.6.3 Estimation de P

Nous présentons dans cette section les différentes méthodes permettant de construire les matrices de projection d'un système allant jusqu'à trois caméras, à partir des différentes entités qui viennent d'être présentées. Des approches directes, estimant P directement à partir des appariements, sont également rappelées.

7.6.3.1 A partir de F

Considérons le cadre d'un système binoculaire ou trinoculaire dans lequel la ou les matrices fondamentales ont été estimées selon une des méthodes présentées à la section 7.6.1. Les matrices de projection perspective associées peuvent alors être retrouvées à partir des relations 7.21 et 7.22 reliant matrices fondamentales et matrices de projection. Cette solution présente plusieurs inconvénients :

- Le calcul est extrêmement sensible à la qualité des matrices F_{ij} estimées ;
- Dans le cas de trois caméras, il subsiste le problème des centres optiques alignés, qui nécessite la donnée d'informations supplémentaires.

Quelques optimisations sont néanmoins envisageables. Il est par exemple possible de minimiser une distance entre les nouvelles matrices fondamentales engendrées par les matrices de projection estimées (selon la relation directe 7.23) et les matrices fondamentales initiales. Toute la difficulté réside ici dans le choix d'une métrique entre matrices fondamentales. Dans sa thèse [Luong, 1992], Luong propose d'étudier la distance relative entre les épipôles. Il est également possible de considérer la distribution des distances euclidiennes mesurées entre des appariements vérifiant la contrainte épipolaire associée à la matrice fondamentale initiale et les droites épipolaires engendrées par la matrice fondamentale recalculée. C'est en théorie la meilleure solution, elle a en effet une signification

physique dans l'image. Elle n'est cependant pas calculable simplement ; il faut avoir recours à des méthodes statistiques, ce qui la rend en pratique inutilisable à l'intérieur d'une minimisation. Laveau [Laveau, 1996] préfère quant à lui calculer la distance entre les sept paramètres indépendants de la matrice fondamentale. Ses expérimentations montrent qu'elle donne plus rapidement des résultats similaires à la méthode calculant la distance entre appariements et droites épipolaires recalculées.

7.6.3.2 A partir de T

Il suffit ici de reprendre la relation qui existe entre le tenseur trifocal et les matrices de projection, qui a été présentée à la section 7.5.2.1. Rappelons que dans la pratique, la qualité des matrices obtenues dépend fortement du tenseur trifocal estimé. En effet, si celui-ci n'est pas représenté par une paramétrisation minimale avec 18 paramètres, ce qui est le cas le plus courant, alors il est très difficile de dire si les matrices de projection engendrées sont les meilleures [Laveau, 1996].

En outre, le passage vers P implique la résolution d'un système linéaire aux moindres carrés. Ce système prend en compte les coefficients du tenseur trifocal qui sont connus pour ne pas avoir le même ordre de grandeur, rendant ainsi incertaine la minimisation.

7.6.3.3 Estimation directe

Nous venons de voir que la mise en place des méthodes d'estimation des matrices de projection à partir des entités bilinéaires et trilinéaires n'est pas évidente la plupart du temps, d'une part parce qu'elle dépend fortement de la qualité de ces entités, et d'autre part parce que le passage de ces entités vers les matrices de projection est souvent délicat à réaliser. Il existe donc des approches permettant leur calcul directement à partir des mises en correspondance.

Principe

Comme l'illustre l'équation 7.1, une matrice de projection correspond à une matrice de changement de base multipliée par une matrice de projection canonique, elle-même multipliée par une autre matrice de changement de base. Il suffit donc de considérer une base projective de l'espace projectif $\mathcal{P}^3$, et de connaître les images des cinq points la composant[10] dans toutes les vues. La donnée de ces cinq correspondances plus une sixième engendre alors 12 équations qui permettent de déterminer les matrices de projection dépendantes de 11 paramètres. On choisit typiquement pour la sixième donnée un épipôle, ce choix permettant de simplifier les équations ; un sixième appariement quelconque peut être considéré si la géométrie épipolaire est inconnue. La matrice de projection engendrée est de la forme :

$$P_i = \begin{pmatrix} \alpha_i x_i - 1 & 0 & 0 & 1 \\ 0 & \beta_i x_i - 1 & 0 & 1 \\ 0 & 0 & \gamma_i x_i - 1 & 1 \end{pmatrix} \tag{7.45}$$

[10] Une base projective de $\mathcal{P}^3$ est constituée de cinq points non coplanaires quatre à quatre.

où x_i est un paramètre facilement déterminé par la donnée de l'épipôle, et où le triplet $(\alpha_i, \beta_i, \gamma_i)$ est fonction des coordonnées d'un des cinq appariements (exprimé dans la base canonique du plan projectif $\mathcal{P}^2$ à partir des autres correspondances). Une démonstration complète peut être trouvée dans [Faugeras, 1992].

Méthode robuste

Comme le lecteur peut s'en douter, le principe que nous venons de développer ne s'applique que si les appariements choisis et la donnée de l'épipôle sont corrects, en supposant également que la base projective de $\mathcal{P}^3$ ainsi formée représente une base projective valide. La difficulté essentielle de l'estimation de la matrice P réside donc dans le choix de ces données. Sur le même principe que la matrice fondamentale ou encore le tenseur trifocal, la mise en place de méthodes d'estimation robuste nécessite la détermination d'une métrique, permettant ici de déterminer l'éloignement d'une base de la configuration dégénérée. Certains travaux à base d'invariants, comme [Zisserman et al., 1995], permettent de détecter les configurations dégénérées, mais ils ne proposent pas le type de métrique recherché. Dans sa thèse [Laveau, 1996], Laveau propose comme critère la distribution des distances séparant les appariements des nouveaux appariements formés après reprojection du point construit par triangulation[11] des appariements initiaux. Signalons que cette approche permet également, dans une certaine mesure, de détecter les appariements faux, ceux-ci présentant des distances supérieures aux autres. La procédure ensuite suivie est une minimisation classique de type LMedS similaire à celle présentée à la section 7.6.1.3 dans le cadre de l'estimation robuste de la matrice fondamentale.

7.6.3.4 Raffinement de la géométrie

Après avoir réalisé une première estimation des matrices de projection perspective selon une des approches proposées, il est classique de procéder à un raffinement de la géométrie obtenue. La technique employée porte le nom de "ajustement des rayons optiques", mieux connue dans la communauté de la Vision par Ordinateur sous sa dénomination en anglais de "bundle adjustment".

Son but consiste à obtenir à la fois les points de la scène et les matrices de projection qui vérifient au mieux les correspondances trouvées dans les images. Nous reprenons ici une distance similaire à celle introduite pour l'estimation robuste présentée plus haut, à partir de la notation adoptée par Hartley [Hartley, 1994]. Soient u_j^i le $j$ème point de la $i$ème image, P_i la matrice de projection et x_j le point tridimensionnel associés tels que $\widehat{u}_j^i = P_i x_j$ et tels :

$$\sum_{i,j} d(\widehat{u}_j^i, u_j^i)^2 \tag{7.46}$$

où $d(.,.)$ représente la distance euclidienne entre deux points du plan rétinien.

Le raffinement de la géométrie consiste à minimiser cette somme des erreurs au carré sur l'ensemble des matrices de projection. L'algorithme alors utilisé dans ce contexte est

[11] Le lecteur peut se reporter à la section 8.1 du prochain chapitre pour une revue des méthodes de triangulation.

classiquement un algorithme de Levenberg-Marquardt, bien connu pour effectuer des minimisations non linéaires.

7.7 Conclusion

Dans ce chapitre, nous avons tout d'abord présenté les différents modèles de caméra. Le modèle le plus utilisé est le modèle perspectif qui décrit notamment mieux que les autres les déformations dues à un changement de profondeur de la scène. Il est associé à un formalisme agréable, celui de la géométrie projective. A partir de là, nous avons vu qu'il est également possible soit d'intégrer des modèles de distorsion, soit de rectifier indépendamment chacune des images pour se ramener à un modèle perspectif pur.

A partir de ce modèle de la géométrie d'une caméra, nous avons rappelé les différentes entités qui peuvent être mises en place dans les systèmes binoculaires et trinoculaires. Pour les systèmes intégrant deux caméras, il existe une matrice 3×3 de déterminant nul et définie à un facteur d'échelle près, appelée matrice fondamentale. Celle-ci permet de caractériser la géométrie épipolaire du système, qui dépend de 7 paramètres. Cette donnée est la seule information géométrique disponible lorsque les caméras ne sont pas calibrées. En ce qui concerne les systèmes basés sur trois caméras, on retrouve bien sûr la notion de matrice fondamentale pour chaque paire de caméras. Nous avons cependant constaté que la géométrie pouvait être capturée ici par une entité plus adaptée : un tenseur de dimension $3 \times 3 \times 3$ défini lui aussi à un facteur d'échelle près. Dans sa version minimale, c'est-à-dire lorsqu'il vérifie certaines contraintes internes, ce tenseur est équivalent aux trilinéarités introduites par Shashua. Il traduit les relations linéaires existant entre les coordonnées de triplets de points en correspondance dans les trois images. Il permet ainsi de caractériser plus efficacement qu'avec la matrice fondamentale le système trinoculaire dépendant de 18 paramètres.

Nous avons également rappelé les différentes méthodes à la fois théoriques et pratiques permettant de passer d'une représentation (matrice fondamentale, trilinéarités/tenseur trifocal et matrice de projection) vers une autre. En résumé, le passage des matrices de projection vers les n-linéarités constitue une opération facile à réaliser. En revanche, les opérations inverses s'avèrent bien plus délicates à mettre en place.

Enfin, nous avons passé en revue les différentes méthodes permettant d'estimer ces principales entités. Les caméras n'étant pas calibrées, toutes les estimations sont issues de mises en correspondance de points entre les images. Quelle que soit l'entité concernée, ce sont les méthodes d'estimation robuste qui fournissent les meilleurs résultats. En effet, bien que simple d'utilisation, les méthodes linéaires ne respectent pas la plupart du temps les propriétés de l'entité à déterminer, comme la singularité de la matrice fondamentale par exemple. Les méthodes non linéaires donnent de meilleurs résultats en intégrant des critères plus robustes comme les distances euclidiennes entre points et droites, ou encore la paramétrisation de la matrice à déterminer. Les méthodes robustes, quant à elles, intègrent, en plus de critères non linéaires, un modèle de bruit dans les données en entrée, et ont même la capacité d'éliminer les faux appariements durant la minimisation.

En ce qui concerne l'estimation des matrices de projection, c'est la méthode directe à

partir des mises en correspondance qui semble la plus robuste. En effet, les approches à partir des matrices fondamentales et des trilinéarités sont très sensibles à l'estimation de ces dernières. Il est tout de même souvent nécessaire de raffiner la géométrie obtenue par des méthodes itératives.

Nous allons maintenant voir au prochain chapitre comment il est possible d'utiliser les entités estimées pour procéder à la reconstruction de scènes tridimensionnelles ou encore au transfert d'images.

Chapitre 8

Synthèse de vues intermédiaires

Ce chapitre est consacré à la présentation des méthodes permettant de synthétiser des images d'une scène sous différents points de vue, à partir d'images existantes. Ces approches peuvent être décomposées en deux classes : d'une part les approches procédant à une reconstruction tridimensionnelle de la scène, avant de la reprojeter sur le plan rétinien de l'image virtuelle, et d'autres part les méthodes, dites méthodes de transfert, qui génèrent l'image virtuelle uniquement à partir de l'information bidimensionnelle contenue dans les images de référence. Nous illustrons ce chapitre par des résultats de reconstruction tridimensionnelle et de transfert projectif, réalisés à partir de points appariés par la méthode qui a été présentée dans les deux premières parties de ce mémoire.

Sommaire

Les travaux exposés tout au long de ce mémoire ont permis de réaliser la mise en correspondance de points d'intérêt au sein d'un système binoculaire caractérisé par des caméras couleur non calibrées. A l'issue de la chaîne de traitements mise en place, nous avons à notre disposition un grand nombre d'appariements entre les images, ainsi que la matrice fondamentale décrivant le système. Au sein de ce chapitre, nous allons en premier exploiter ces informations pour procéder à la reconstruction tridimensionnelle de la scène. Nous rappelons à la section 8.1 les différentes méthodes de reconstruction existantes. La reconstruction projective est facilement réalisable à partir de la donnée de la géométrie épipolaire, mais a l'inconvénient de fournir des reconstructions dans une base projective quelconque. Nous passons donc brièvement en revue les méthodes permettant d'aboutir à des reconstructions affines, voire euclidiennes dans le meilleur des cas.

Dès l'instant où des mises en correspondance et la géométrie épipolaire sont disponibles, il est possible de mettre en œuvre une classe de méthodes permettant de synthétiser de nouvelles vues à partir de vues existantes, et ce sans passer par une reconstruction tridimensionnelle. Ces méthodes portent le nom de méthodes de *transfert d'images*. Elles sont présentées à la section 8.2. Nous allons voir principalement que ces méthodes exploitent les n-linéarités lorsqu'elles sont disponibles, ou encore certaines propriétés d'invariance de la géométrie projective.

Puis dans la section 8.3, nous présentons divers exemples de reconstruction euclidienne ainsi que de transfert projectif d'images. Dans tous les cas, les calculs ont été réalisés à partir des appariements et de la géométrie issus de notre algorithme complet de mise en correspondance. Les résultats exposés ici permettent d'illustrer concrètement la cohérence de l'ensemble de la chaîne de traitements.

Pour indication, la section traitant de la reconstruction tridimensionnelle utilise essentiellement les notions de géométrie binoculaire développées à la section 7.3 du chapitre 7, alors que celle portant sur le transfert d'images exploite les relations définies dans les systèmes trinoculaires, que nous avons présentées à la section 7.4 du même chapitre.

8.1 Reconstruction tridimensionnelle

La reconstruction tridimensionnelle du point correspondant à un appariement (m_1, m_2) s'effectue traditionnellement en deux étapes :

1. Il faut en premier lieu connaître les matrices de projection caractérisant les deux caméras. Selon les données que nous avons à notre disposition à cette étape du calcul, ces matrices contiennent plus ou moins d'information, et permettent d'aboutir à une reconstruction seulement projective, affine ou encore euclidienne lorsque la calibration des caméras est connue. Nous passons en revue aux sections 8.1.1, 8.1.2 et 8.1.3 les informations nécessaires pour obtenir ces trois types de reconstruction ;

2. A partir de ces matrices, on procède alors à la reconstruction du point 3D par *triangulation* de l'appariement (m_1, m_2). Les différentes méthodes de triangulation sont présentées à la section 8.1.4.

Cette classe de méthodes, qui nécessitent la donnée des matrices de projection, porte le nom de reconstruction *explicite*. Nous verrons notamment pour la reconstruction projective qu'il est possible de mettre en place des reconstructions *implicites*.

8.1.1 Reconstruction projective

Nous avons vu au chapitre 7 que lorsque le système est faiblement calibré, la seule information disponible est la géométrie épipolaire reliant les deux caméras. Il est alors possible d'obtenir les matrices de projection associées à ces deux caméras, selon les relations 7.5.1.1.

La reconstruction obtenue à l'issue de la triangulation est définie dans une base projective de $\mathcal{P}^3$ arbitraire. Il est clair que, utilisée telle quelle, elle ne respecte ni le parallélisme de la scène ni sa métrique. Notons cependant que l'utilisation des matrices de projection perspectives est particulièrement appropriée pour les applications dont le domaine d'action se limite aux images. En effet, s'il semble manquer des informations entre le monde 3D projectif et le monde euclidien ; en revanche, pour les images les matrices de projection contiennent autant d'information qu'une calibration complète.

Reconstruction projective implicite

Cette seconde classe d'algorithmes de reconstruction utilise des mesures de l'image basées sur des invariants projectifs qui relient les projections appariées. Ces invariants sont la plupart du temps issus de l'algèbre de Cayley ou encore de la géométrie projective avec les birapports[1]. Pour une comparaison de ces méthodes avec les approches explicites, le lecteur peut consulter [Rothwell et al., 1997]. Il y apparaît que ce sont les méthodes implicites qui fournissent les reconstructions les moins précises.

8.1.2 Vers une reconstruction affine

La reconstruction projective ne conserve pas le parallélisme. Par exemple, deux droites de l'image supposées être parallèles dans la scène peuvent ne pas l'être à l'issue de la triangulation. La solution permettant d'aboutir à une reconstruction affine consiste à déterminer *le plan à l'infini* Π_∞, c'est-à-dire le plan défini par les intersections des droites parallèles (qui sont donc de la forme $(X, Y, Z, 0)$ en coordonnées homogènes). L'équation de ce plan est donnée par le calcul des coordonnées de trois points non alignés, un point étant donné par l'intersection de deux droites parallèles dans la scène dont on connaît les projections dans les images (fournies de manière interactive par l'utilisateur). La méthode est la suivante :

Détermination de Π_∞

Soient M_∞ un point de la scène appartenant à Π_∞ et d_{ij} la projection dans I_j de la $i$ème droite passant par M_∞. Etant données les matrices de projection (en projectif) P_j,

[1] Le lecteur peut consulter l'annexe E pour une introduction à la géométrie projective.

le point M_∞ vérifie :

$$d_{ij}^T P_j M_\infty = 0 \tag{8.1}$$

Ce système comporte autant d'équations qu'il y a de lignes visibles dans les images. Il est linéaire et homogène en fonction des coordonnées de M_∞. Il peut être résolu par une décomposition en valeurs singulières. La donnée d'au moins trois points M_∞ permet de déterminer Π_∞.

Une fois Π_∞ déterminé, il reste à mettre en place une base affine de la scène telle que Π_∞ corresponde effectivement à un plan à l'infini, c'est-à-dire vérifie l'équation $T = 0$. La reconstruction alors obtenue dans cette base conserve le parallélisme. Pour plus de détails concernant la mise en place du plan à l'infini et de la base affine, le lecteur peut consulter [Faugeras, 1995; Zeller, 1996; Pollefeys et VanGool, 1997].

L'homographie H_∞

On associe très souvent à Π_∞ l'homographie de plan H_∞. Comme toutes les homographies (cf. section 7.3.3 du chapitre 7), il est possible d'estimer celle-ci à partir de quatre points (ou trois points avec la géométrie épipolaire). Ce calcul est cependant délicat car il nécessite la donnée de points à l'infini, dont on connaît la projection dans l'image. Mais sa connaissance permet de savoir facilement si deux droites à reconstruire doivent être parallèles ou non.

8.1.3 Vers une reconstruction euclidienne

La méthode de reconstruction affine qui vient d'être présentée ne garantit ni la conservation des angles ni celle des rapports de longueur, permettant d'aboutir à une reconstruction euclidienne.

Si les caméras sont disponibles, il est bien sûr possible de procéder à un étalonnage fort des caméras à partir d'une mire 3D, selon une des méthodes proposées par Tsai [Tsai, 1986].

Si seuls les paramètres intrinsèques sont disponibles (en plus de la matrice fondamentale), alors le mouvement relatif entre les deux caméras peut être déterminé en décomposant la matrice essentielle selon une des méthodes présentées à la section 7.6.1.5 du chapitre 7. On obtient alors la calibration complète du système, qui permet de conduire à une reconstruction euclidienne, les objets reconstruits étant exprimés dans le repère de la première caméra.

Si aucune de ces données n'est disponible, alors trois types de solutions sont envisageables :

- On procède à l'estimation des matrices de paramètres intrinsèques à partir de la matrice fondamentale F. Cette méthode de résolution utilise les équations de Kruppa, qui sont la traduction algébrique de l'utilisation d'un invariant projectif important : la conique absolue [Kruppa, 1913]. Le lecteur peut consulter notamment [Luong, 1992; Heyden et Astrom, 1996; Zeller, 1996; Lourakis et Deriche, 2000] pour une étude détaillée de ces équations. On procède ensuite à une décomposition de la ma-

trice essentielle pour retrouver le mouvement ;

- Il est également possible d'estimer les paramètres intrinsèques à partir de l'homographie du plan à l'infini H_∞ [Vieville et al., 1996] ;
- Lorsque c'est possible, on peut se servir de la donnée d'angles et de rapports de longueurs connues. Ainsi dans [Boufama, 1994; Laveau, 1996] notamment, l'homographie permettant de passer de la reconstruction affine vers une reconstruction euclidienne est estimée en sélectionnant dans les images des segments représentant des arêtes perpendiculaires dans la scène, et en fixant leurs longueurs respectives.

Bien qu'il soit plus aisé d'estimer F que H_∞ (F nécessite uniquement la donnée de correspondances dans les images), il s'avère beaucoup plus difficile d'estimer les paramètres intrinsèques à partir F qu'à partir de H_∞. La troisième solution, quant à elle, n'est pas automatique et impose des contraintes métriques sur la scène.

Rappelons que dans tous les cas, la reconstruction euclidienne est obtenue à un facteur d'échelle près. En effet, il subsiste toujours une ambiguïté fondamentale entre la distance de prise de vue et la taille de la scène observée. Si la scène est plus grande et observée de plus loin, alors les images formées restent inchangées. Cette observation a par exemple pour conséquence qu'il ne suffit pas simplement de combiner les différents repères mis en jeux pour ramener dans le même repère plusieurs reconstructions de la même scène issues de plusieurs couples d'images. Un exemple de reconstruction tridimensionnelle à partir de trois vues est présenté à la section 8.3.

8.1.4 Triangulation

A ce stade des traitements, nous avons à notre disposition une certain nombre de mises en correspondance de points ainsi que les matrices de projection associées aux caméras (qu'elles soient projectives, affines ou euclidiennes). Il est donc possible de reconstruire les points 3D par triangulation. Les appariements localisés dans les images pouvant ne pas être exempts de bruit, les lignes de vue associées peuvent ne pas être sécantes. La triangulation revient donc à la minimisation d'un problème sous contraintes.

La méthode linéaire classique

La méthode classique de triangulation consiste alors à calculer une solution aux moindres carrés sur toutes les équations de projection, comme l'illustre la figure 8.1.

Pour un point M de l'espace, les équations de projection (cf. équation 7.1) pour trois caméras sont les suivantes :

$$\begin{cases} m_1 = P_1 M \\ m_2 = P_2 M \\ m_3 = P_3 M \end{cases} \tag{8.2}$$

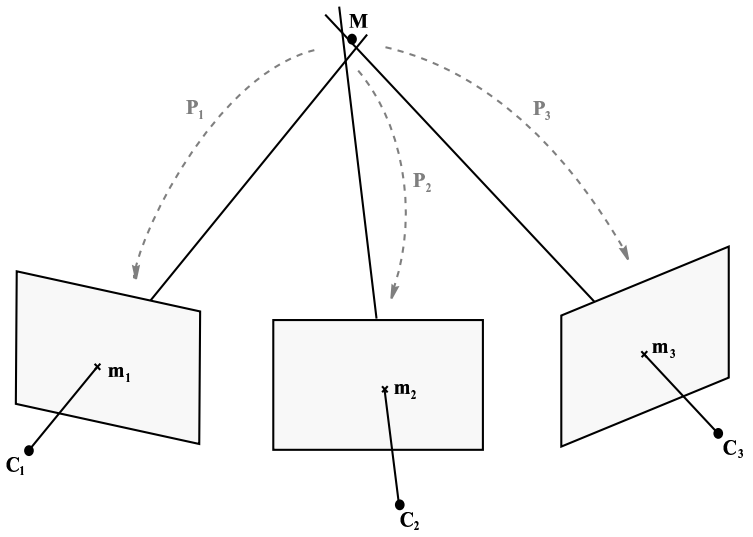


Figure 8.1 – *Reconstruction d'un point 3D par triangulation.*

Et le système linéaire correspondant à résoudre :

$$\begin{pmatrix} P^1_{31}m^x_1 - P^1_{11} & P^1_{32}m^x_1 - P^1_{12} & P^1_{33}m^x_1 - P^1_{13} & p^1_{34}m^x_1 - P^1_{14} \\ P^1_{31}m^y_1 - P^1_{21} & p^1_{32}m^y_1 - P^1_{22} & P^1_{33}m^y_1 - P^1_{23} & P^1_{34}m^y_1 - P^1_{24} \\ P^2_{31}m^x_2 - P^2_{11} & P^2_{32}m^x_2 - P^2_{12} & P^2_{33}m^x_2 - P^2_{13} & P^2_{34}m^x_2 - P^2_{14} \\ P^2_{31}m^y_2 - P^2_{21} & P^2_{32}m^y_2 - P^2_{22} & P^2_{33}m^y_2 - P^2_{23} & P^2_{34}m^y_2 - P^2_{24} \\ P^3_{31}m^x_3 - P^3_{11} & P^3_{32}m^x_3 - P^3_{12} & P^3_{33}m^x_3 - P^3_{13} & P^3_{34}m^x_3 - P^3_{14} \\ P^3_{31}m^y_3 - P^3_{21} & P^3_{32}m^y_3 - P^3_{22} & P^3_{33}m^y_3 - P^3_{23} & P^3_{34}m^y_3 - P^3_{24} \end{pmatrix} \cdot \begin{pmatrix} M^x \\ M^y \\ M^z \\ M^t \end{pmatrix} = 0_{6\times 1} \tag{8.3}$$

Le système ci-dessus de six équations à quatre inconnues permet de traiter le cas des lignes de vue toutes parallèles, pour lesquelles le point d'intersection se situe à l'infini, soit $M^t = 0$. Dans [Rothwell et al., 1997], les auteurs préconisent de résoudre ce système à l'aide d'une décomposition en valeurs singulières. La reconstruction alors obtenue est plus stable qu'avec les autres méthodes.

Ce type de reconstruction n'a en théorie de sens que dans un espace euclidien, la solution correspondant au point M le plus proche de toutes les lignes de vue simultanément, au sens de la distance euclidienne. Pour des reconstructions projective et affine pour lesquelles cette métrique n'a pas de sens, il est nécessaire d'employer d'autres approches :

Dans [Hartley et Sturm, 1994], une première solution basée sur une méthode linéaire itérative est proposée pour la reconstruction affine binoculaire. Elle consiste à introduire des poids dans la minimisation, qui tiennent compte des erreurs entre les points appariés et le projections exactes du point M. La méthode obtenue fournit des résultats légèrement meilleurs que la méthode linéaire classique. Elle utilise des contraintes invariantes aux transformations affines, et peut donc être employée dans le cadre d'une reconstruction affine.

Ajustement des mises en correspondance

Toujours dans le cadre d'un système binoculaire, les mêmes auteurs proposent également une méthode permettant d'aligner les appariements précisément sur les droites épipolaires, afin de garantir que les lignes de vue se coupent. Pour un appariement donné, la méthode consiste à déterminer les droites épipolaires conjuguées les plus proches des deux points impliqués, puis à projeter ces points sur les droites trouvées. Le nouvel appariement formé respecte exactement la géométrie épipolaire, ce qui garantit d'avoir deux lignes de vue sécantes. Cette approche a de nombreux avantages : elle est très rapide, la géométrie épipolaire utilisée peut être relativement imprécise, et elle peut être employée pour n'importe quel type de reconstruction. Rappelons que si elle n'a pas été calculée, la géométrie épipolaire est facilement obtenue à partir des matrices de projection par l'équation 7.23 présentée au chapitre 7.

Méthode robuste pour N oculaires

Lorsque le nombre de caméras est supérieur à deux, la minimisation aux moindres carrés, qui donne une solution satisfaisant au mieux toutes les contraintes simultanément, peut être complètement faussée à cause d'un appariement incorrect dans la $i$ème image. Laveau propose dans sa thèse [Laveau, 1996] une solution robuste basée sur l'approche LMedS par tirages aléatoires (similaire à celle utilisée pour l'estimation de la matrice fondamentale à la section 7.6.1.3 du chapitre 7). Pour chaque appariement, le point M est reconstruit aux moindres carrés uniquement à partir de deux points parmi les N points en correspondance, puis il est reprojeté dans toutes les images. Les N-uplets de points qui vérifient une trop grande erreur entre un des points les composant et le point reprojeté sont éliminés du processus. La reconstruction aux moindres carrés est alors réitérée à partir des correspondances restantes complètes.

8.2 Synthèse de nouvelles vues sans modèle tridimensionnel

Les techniques permettant de générer des nouvelles images à partir de vues existantes portent souvent le nom de *transfert d'images*. Elles permettent de supprimer l'étape de modélisation tridimensionnelle de la scène, parfois fastidieuse ou inutile, en la définissant non pas par un modèle tridimensionnel, mais par un ensemble de vues (bidimensionnelles) réelles. Plus formellement, connaissant les N projections d'un point de l'espace dans les N images, il s'agit de le transférer dans une $(N+1)$ème image, c'est-à-dire de calculer la position où il se projetterait dans cette nouvelle image, et ceci sans passer par une reconstruction tridimensionnelle du point. Ces techniques sont issues de la stéréovision trinoculaire et de la reconnaissance d'objets à partir de vues multiples. Nous faisons essentiellement référence ici aux travaux réalisés par [Laveau, 1996; Avidan et Shashua, 1998; Blanc, 1998].

Nous nous plaçons ici dans la catégorie des systèmes de transfert qui fournissent des vues *géométriquement* valides, c'est-à-dire semblables à celles qui seraient réellement vues par une caméra placée à l'endroit donné. Nous ne développons donc pas les approches comme le *morphing interpolation* [Beier et Neely, 1992], bien connu du grand public ou encore la méthode de Werner [Werner et al., 1995] basée sur des combinaisons linéaires entre projections.

8.2.1 Principe

Soient I_1 et I_2 les deux images servant de référence et I_3 l'image transférée. Etant donnée la correspondance (m_1, m_2) de points dans les deux images de référence, le problème du transfert consiste à déterminer dans l'image virtuelle la position du point m_3, telle que le triplet (m_1, m_2, m_3) soit issu du même point tridimensionnel. Une fois trouvée, il suffit d'affecter à m_3 une combinaison des couleurs des pixels m_1 et m_2.

Correspondances denses/éparses

Ces méthodes sont appliquées selon le cas sur des mises en correspondance denses ou éparses. Dans le cas de mises en correspondances denses, on obtient en théorie en sortie du transfert l'image virtuelle dans son intégralité. Nous verrons qu'il est nécessaire dans la pratique de raffiner l'image obtenue lorsque la méthode de transfert utilisée est directe. Dans le cas de mises en correspondances éparses, il est classique de construire un maillage sur les points d'intérêt transférés, puis de calculer les couleurs des pixels non renseignés par mapping (affine ou mieux, projectif) des textures des images de référence. Dans [Laveau, 1996], Laveau utilise une triangulation de Delaunay. Dans [Blanc, 1998], Blanc construit quant à lui une triangulation plus sophistiquée, basée sur des critères de coplanarité. Celle-ci permet notamment de gérer le problème du recouvrement des zones occultées.

Positionnement de l'image virtuelle

Lorsque la calibration des caméras est inconnue, une des difficultés majeures du transfert est de définir la position de l'image virtuelle par rapport aux images de référence. En effet, nous n'avons aucune information sur la position des caméras employées. La seule solution possible consiste à décrire la troisième caméra en fixant certains points dans les images de référence. Ainsi :

- Le centre optique virtuel C_3 est défini par deux épipôles e_{13} et e_{23} dans les deux images de référence ;
- Le plan rétinien virtuel est défini par trois appariements dans les deux images de référence.

Les quatre paires de points ainsi définies ne sont pas obligatoirement la projection de réels points de la scène mais elles doivent au moins vérifier la seule contrainte que nous ayons à notre disposition, à savoir la géométrie épipolaire. La notion de perpendicularité n'existant pas en projectif, il est impossible de définir un plan rétinien virtuel qui soit perpendiculaire à la direction de vue choisie.

Nous allons maintenant passer en revue dans les sections qui viennent les différentes méthodes permettant de localiser m_3 dans l'image virtuelle.

8.2.2 Morphing exact

Lorsque les images respectent la contrainte d'ordre, il est possible d'obtenir des vues interpolées physiquement valides [Seitz et Dyer, 1997]. Leur idée est de rectifier les deux

images de référence avant d'interpoler la vue intermédiaire linéairement par morphing. La nouvelle vue est alors obtenue en ramenant l'image interpolée dans le plan de la caméra virtuelle. Cette technique requiert une mise en correspondance dense, qui est réalisée ici par programmation dynamique pour être en cohérence avec la contrainte d'ordre[2]. Les résultats obtenus sont corrects géométriquement, mais le déplacement de la caméra virtuelle reste tout de même limité au cône de vision entre les deux images de référence.

8.2.3 Utilisation de la géométrie épipolaire

Supposons connus l'appariement (m_1, m_2) dans les deux images de référence I_1 et I_2 ainsi que les matrices fondamentales[3] F_{13} et F_{23} reliant respectivement les images I_1 avec I_3 et I_2 avec I_3. Le point m_3 transféré dans I_3 se trouve à la fois sur la droite épipolaire conjuguée de m_1 dans I_3 et sur la droite épipolaire conjuguée de m_2 dans I_3, comme l'illustre la figure 8.2. Il est donc défini tel que [Faugeras et Robert, 1994] :

$$m_3 = F_{13}m_1 \wedge F_{23}m_2 \tag{8.4}$$

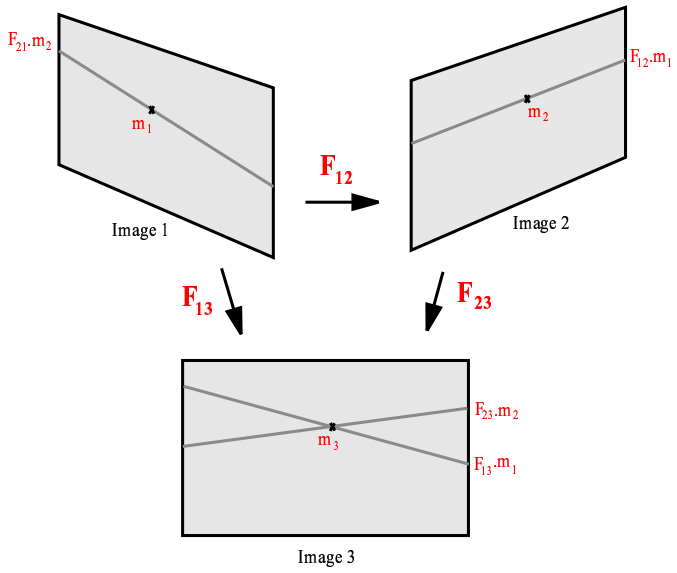


Figure 8.2 – *Transfert à partir de la géométrie épipolaire.*

Il est donc possible de transférer dans I_3 tous les points des deux images de référence, à condition d'avoir procédé à une mise en correspondance dense entre I_1 et I_2 et de connaître les géométries épipolaires associées aux couples stéréo (I_1, I_3) et (I_2, I_3). Aucun étalonnage fort n'est nécessaire, puisque les matrices fondamentales suffisent.

Cette méthode simple a de nombreux inconvénients :

[2]Se reporter à la section 5.3.3 du chapitre 5 pour une présentation de la mise en correspondance dense par programmation dynamique.

[3]Se reporter à la section 7.3.2 du chapitre 7 pour une définition.

- Il faut connaître les matrices fondamentales F_{13} et F_{23}. Cela implique la donnée d'un certain nombre de correspondances entre une des images de référence et l'image virtuelle pour estimer ces matrices selon une des méthodes présentées au chapitre 7 ;
- La méthode est dégénérée pour les points appartenant au plan trifocal, les droites épipolaires engendrées dans I_3 étant confondues dans cette configuration. Mais Laveau résout ce problème dans [Laveau, 1996] à l'aide d'une construction impliquant des points éloignés du plan trifocal ;
- Elle l'est aussi lorsque les centres optiques sont alignés, le plan trifocal n'étant plus défini ;
- La transfert est direct : on part des points des images de référence pour construire l'image virtuelle. A cause d'une possible distribution irrégulière dans I_3 des pixels issus de I_1 et I_2, des zones de I_3 peuvent ne pas être renseignées si l'on essaie de prédire tous les points. Une solution possible consiste à employer des méthodes de régularisation dans l'image obtenue, ou encore à procéder à un transfert épars de points d'intérêt, avant de construire l'image virtuelle par interpolation des niveaux de gris au sein d'un maillage triangulaire par exemple.

8.2.4 Utilisation des trilinéarités

Si les trilinéarités[4] qui caractérisent le système trinoculaire sont connues, alors le transfert est immédiat puisque l'on dispose d'équations permettant de déterminer les coordonnées de m_3 directement en fonction de celles de m_1 et de m_2.

Avec les trilinéarités, le problème des centres optiques alignés est éliminé. Les résultats semblent aussi plus précis qu'à partir de l'estimation des F_{ij}, selon une étude menée par Shashua [Shashua, 1994].

Il subsiste tout de même le problème des pixels non renseignés dans l'image virtuelle générée. En outre, les quatre relations définies en 7.17 sont algébriquement équivalentes mais il a été montré dans [Bobet et al., 1996] qu'elles ne le sont pas numériquement. Le choix d'une forme précise semble dépendre des images, et plus précisément de la configuration épipolaire. Par exemple, il s'avère que la troisième relation ne permet pas de calculer la géométrie du système de façon stable si les droites épipolaires sont proches de la verticale. En effet, cette relation ne fait pas intervenir la coordonnée y' qui représente pourtant une information plus pertinente que x' le long de ces droites. Il serait peut-être nécessaire d'envisager la combinaison des quatre formes.

Déplacement de la caméra virtuelle

Dans [Avidan et Shashua, 1998], Avidan et Shashua mettent en place une méthode permettant de paramétrer les coefficients du tenseur trifocal en fonction du déplacement de la caméra virtuelle. Ils définissent un opérateur qui décrit la transformation permettant de passer d'un tenseur trifocal donné vers un autre tenseur correspondant au déplacement de cette caméra. Il suffit alors de définir ensuite un tenseur de base $< 1, 2, 3 >$ à partir des vues disponibles et d'appliquer cet opérateur pour synthétiser une nouvelle vue I_ψ, à partir

[4] Se reporter à la section 7.4.2 du chapitre 7 pour une définition.

du tenseur obtenu $< 1, 2, \psi >$. Si trois vues de référence sont disponibles, alors le tenseur $< 1, 2, 3 >$ est estimé de façon robuste à partir de mises en correspondances dans les trois images. Si seulement deux vues sont disponibles, un tenseur spécial $< 1, 2, 2 >$ est proposé, dont les paramètres dépendent de la matrice fondamentale estimée entre les deux vues. Cette méthode a l'avantage de rendre hors-ligne l'estimation du tenseur entre les images de référence. Notons que les auteurs ne spécifient pas la position de la caméra virtuelle par des points de contrôle, mais par la donnée d'un déplacement (R, t) par rapport à l'une des caméras de référence.

Reconstruction projective explicite

Principalement à cause de l'instabilité numérique des trilinéarités, certains auteurs, citons notamment [Bobet et al., 1996; Blanc, 1998], préfèrent mettre en œuvre une reconstruction projective *explicite*, en estimant les matrices de projection des deux premières caméras à partir de la matrice fondamentale, avant de reprojeter les points dans la troisième. Cette solution permet également d'envisager facilement le déplacement de la caméra virtuelle lors du transfert, l'estimation des matrices de projection des deux premières caméras pouvant être réalisée hors-ligne.

8.2.5 Transfert projectif

Nous passons en revue ici deux méthodes de transfert d'images basées sur les invariants projectifs et similaires à celles proposées par Laveau dans [Laveau et Faugeras, 1994]. Comme toutes les approches qui viennent d'être présentées, la première méthode est directe, c'est-à-dire transfère les appariements des images de référence vers l'image virtuelle. Au contraire, la seconde approche part de l'image virtuelle pour rechercher les pixels correspondants dans les images de référence ; c'est pourquoi elle porte le nom de "lancer de rayon".

8.2.5.1 Transfert projectif direct

Nous montrons dans cette section comment il est possible d'exploiter la conservation du birapport (cf. annexe E) pour obtenir les coordonnées du point à transférer dans l'image virtuelle. La méthode requiert en premier lieu la donnée d'une base projective du plan projectif $\mathcal{P}^2$.

Construction d'une base projective de $\mathcal{P}^2$

En ajoutant aux trois paires de points définissant le plan rétinien virtuel (cd. section 8.2.1) une quatrième paire de telle sorte que le point 3D correspondant appartienne aussi au plan virtuel, on obtient une base projective. La construction de ces deux points s'obtient en utilisant la propriété de conservation de la colinéarité en géométrie projective ainsi que la contrainte épipolaire, comme l'illustre la figure 8.3. Le point c' est construit tel que (a, a'), (b, b'), (c, c') et (d, d') soient les projections de quatre points appartenant au plan rétinien virtuel [Morin, 1993].

Cette construction est assez facile à réaliser. Cependant, si le choix des points des

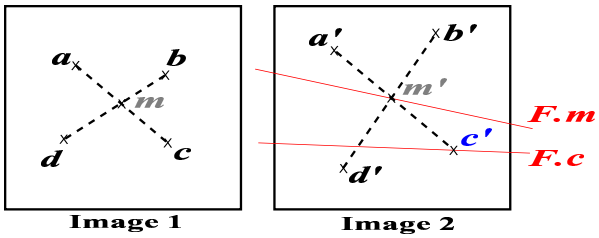


Figure 8.3 – *Construction d'une base projective de $\mathcal{P}^2$.*

bases projectives dans les images de référence se fait sous certaines contraintes (contrainte épipolaire, colinéarité), il n'en existe a priori aucune pour le choix des points concernant la base projective du plan virtuel. Il est donc important de noter que le transfert sera réalisé à une homographie planaire inconnue près.

Une fois la base établie, il est alors possible d'exprimer les projections dans leurs coordonnées projectives. La méthode de transfert est illustrée par la figure 8.4. B est la base projective, elle est contenue dans le plan rétinien virtuel et ses projections dans les images de référence sont b_1 et b_2. Les points p_1 et p_2 du plan virtuel ont pour projections respectives dans les deux images de référence les mises en correspondances m_1 et m_2. Leurs positions sont facilement exprimées en utilisant la conservation du birapport (cf. annexe E). Notons aussi que l'épipôle e_{31} (resp. e_{32}) a les mêmes coordonnées projectives que e_{13} (resp. e_{23}). Les transformations projectives conservant la colinéarité, la position du point m_3 recherché est donc donnée par l'intersection des droites (e_{31}, p_1) (trace de (C_1, m_1)) et (e_{32}, p_2) (trace de (C_2, m_2)).

Le principal avantage de cette solution est qu'il suffit de placer les quatre points définissant la caméra virtuelle et de connaître la géométrie épipolaire F_{12} entre les images de référence, contrairement aux autres méthodes qui requièrent la donnée de correspondances dans les trois images. Malheureusement, elle ne fonctionne pas si les centres optiques sont alignés ou encore pour les points contenus dans le plan trifocal, et comme les autres approches directes, produit également des "trous" dans l'image synthétisée.

La méthode est illustrée à l'aide des images de la figure 8.5. Les deux images qui ont servi de référence sont la première et la dernière de la figure. Ces deux images synthétiques sont issues d'une scène 3D contenant un cube en fil de fer dans lequel une centaine de points aléatoires ont été générés. Les appariements entre les deux images de référence sont connus avec exactitude, la calibration des deux caméras étant disponible. Bien qu'elle soit connue, la matrice fondamentale a été estimée à partir de ces appariements à l'aide d'une méthode robuste de type LMedS. Les quatre points permettant de définir la position de la caméra virtuelle ont pu être fixés de façon réaliste, le système étant calibré. Nous avons ici simulé une interpolation du mouvement entre les deux caméras de référence, en s'assurant que la configuration trinoculaire obtenue ne soit pas dégénérée. Six vues intermédiaires transférées sont présentées. Elles ont été comparées aux véritables vues qui auraient été calculées si la scène avait été projetée sur le plan rétinien de la caméra virtuelle. Nous n'avons observé aucune différence.

Variante de Laveau

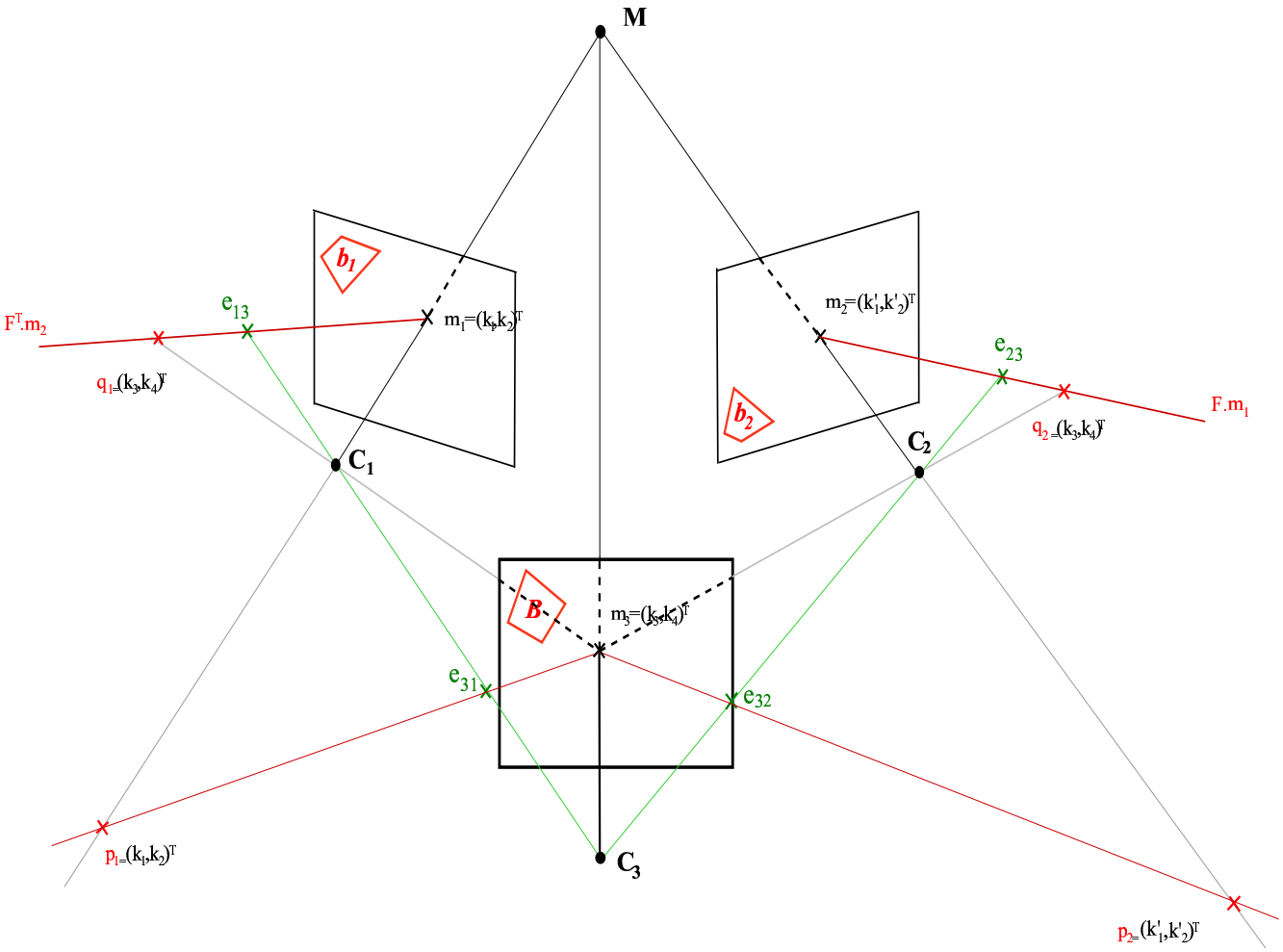


Figure 8.4 – *Une méthode de transfert projectif directe.*

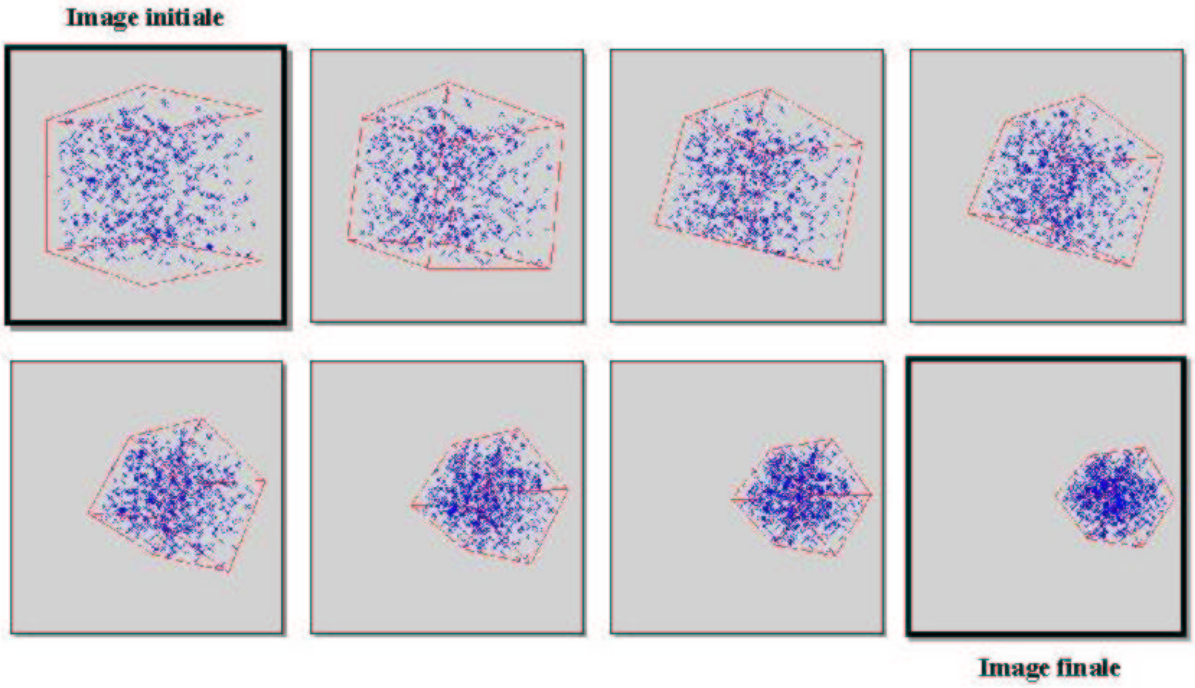


Figure 8.5 – *Un exemple de transfert projectif direct.*

Il existe une variante de cette approche : dans [Laveau et Faugeras, 1994], Laveau cherche à estimer les projections de m_3 dans les deux images de référence, qui doivent alors avoir les mêmes coordonnées projectives (q_1 et q_2 sur la figure 8.4). Pour cela, il estime l'homographie H_{12} entre les deux images de référence à partir des bases projectives. La position de la projection dans la première image est alors donnée par l'intersection des droites (e_{13}, m_1) et $H_{21}(e_{23}, m_2)$, et celle de la deuxième image par l'intersection des droites (e_{23}, m_2) et $H_{12}(e_{13}, m1)$.

8.2.5.2 Lancer de rayon (méthode indirecte)

Avec cette méthode, Laveau [Laveau et Faugeras, 1994] a souhaité pallier l'inconvénient que l'on retrouve dans toutes les méthodes présentées jusqu'à ici, à savoir la génération de "trous" dans l'image virtuelle. L'idée est de parcourir l'image virtuelle, c'est-à-dire de fixer le point m_3, puis de mettre en œuvre une solution pour retrouver les positions de m_1 et de m_2 dans les deux images de référence. C'est une approche dite *indirecte* qui s'apparente au fameux lancer de rayon utilisé en synthèse d'images. Comme pour la méthode projective directe, elle nécessite seulement la donnée des mises en correspondances bien sûr et celle de la base projective. La figure 8.6 illustre la technique employée. Pour chaque pixel m_3 du plan virtuel, il s'agit de :

1. Déterminer ses coordonnées projectives (k_1, k_2) dans la base projective B ;
2. Positionner p_1 et p_2, ses projections dans les deux images de référence. Elles doivent avoir les mêmes coordonnées projectives dans les bases projectives associées b_1 et b_2.

Remarquons que (e_{13}, p_1) (resp. (e_{23}, p_2)) est la droite épipolaire conjuguée de m_3 dans la première (resp. seconde) image ;

3. Déterminer m_1 comme étant l'intersection de (e_{13}, p_1) avec la carte des disparités de (e_{23}, p_2), et m_2 comme l'intersection de (e_{23}, p_2) avec la carte des disparités de (e_{13}, p_1). Ces cartes sont représentées en bleu sur la figure.

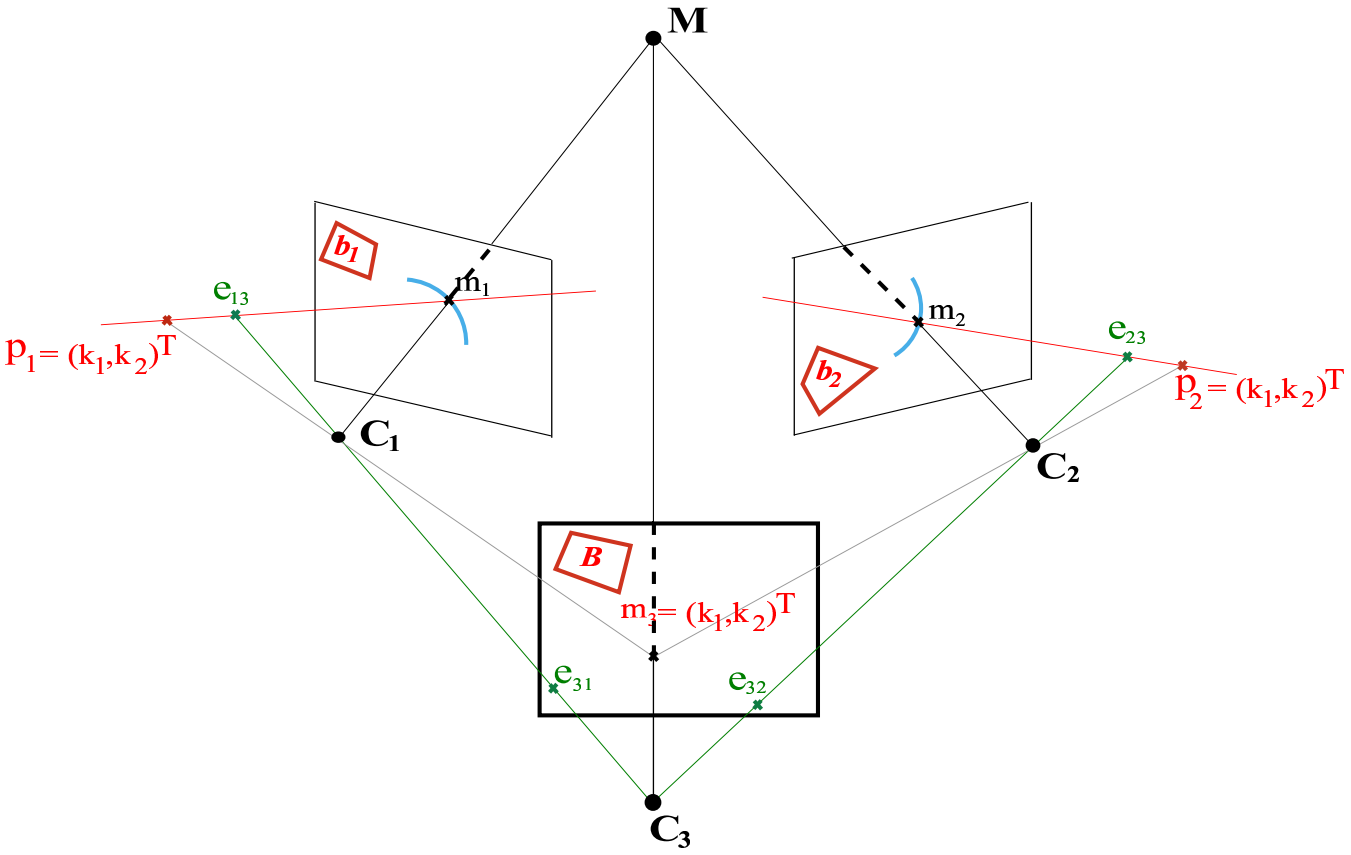


Figure 8.6 – *Une méthode de transfert projectif indirecte.*

Cette solution permet de renseigner chaque pixel de l'image virtuelle. Elle ne règle cependant pas le problème des centres optiques alignés. Un autre inconvénient est que la carte des disparités n'étant pas obligatoirement une droite, plusieurs solutions peuvent être trouvées dans chaque image de référence. Laveau apporte une solution pour éliminer ces ambiguïtés en travaillant avec un plus grand nombre de caméras de référence et en utilisant des notions de géométrie projective *orientée* [Laveau, 1996]. Ces contraintes supplémentaires lui permettent de trouver la solution qui convient le mieux, c'est-à-dire celle qui est la plus proche du plan virtuel sur le rayon (C_3M).

Variante utilisant les trilinéarités

Si les trilinéarités sont disponibles, alors en fixant la donnée m_3 dans le système d'équations 7.28, on obtient deux équations indépendantes, qui indiquent que les coordonnées de m_2 sont des fonctions linéaires des coordonnées de m_1. Il suffit alors de trouver

dans les images de référence les points vérifiant à la fois ces contraintes trilinéaires et les contraintes de disparité. Parmi les points trouvés, Laveau préconise l'utilisation de la géométrie projective orientée pour extraire l'appariement correspondant au point tridimensionnel visible.

8.2.6 Mosaïque d'images

Le principe de la construction de mosaïques à partir d'images de référence est un domaine très actif en Vision par Ordinateur. Il s'agit de "recoller" toutes les images de référence dans un même repère, de manière à reconstituer une seule vue de la scène. Cette technique est considérée comme faisant partie des méthodes de transfert d'images, puisqu'elle consiste à transférer plusieurs images de référence vers une autre image définie dans un nouveau repère. Les techniques de "mosaïquage" imposent cependant certaines contraintes sur la scène : celle-ci doit être plane ou bien doit contenir des objets suffisamment éloignés des caméras de référence pour être supposés plans, ou encore le mouvement entre les caméras doit être limité à une rotation du centre optique.

Principe

Nous nous limitons ici à l'étude de deux images de référence pour simplifier les explications, la méthode étant identique pour N images. Le calcul de mosaïques est basé sur l'estimation d'une homographie entre les deux images de référence. Nous avons vu à la section 7.3.3 du chapitre 7 qu'il existe une homographie reliant les projections des points de la scène lorsque celle-ci est plane (ou très éloignée des caméras). Il est facile de montrer qu'il existe le même type de relation lorsque le mouvement entre les deux caméras est limité à une rotation autour du centre optique [Mohr, 1993]. Rappelons qu'au moins quatre points appariés entre les deux images sont nécessaires pour estimer cette homographie. Il est ensuite possible d'utiliser cette relation homographique pour reprojeter sur l'une des images les pixels de l'autre image. Faire la reprojection dans un repère quelconque s'avère plus délicat lorsqu'on a aucune information sur la géométrie des caméras de référence. Il faudrait en effet fixer quatre points de façon totalement arbitraire, ce qui ne garantirait pas la cohérence géométrique de l'image ainsi calculée.

La technique que nous venons de décrire est notamment utilisée dans le logiciel QuickTime VR® d'Apple [Chen, 1995] et a même fait l'objet d'un brevet. Ici on dispose d'une connaissance approximative des paramètres optiques des caméras, qui sont contraintes de se déplacer selon un mouvement panoramique. Plus récemment, la société Realviz[5] a commercialisé le logiciel Stitcher®. Ici seule la donnée approximative de la focale est requise et le processus d'appariement est automatisé.

8.3 Résultats

Nous présentons dans cette section divers résultats de reconstruction euclidienne et de transfert d'images, illustrant certaines des méthodes présentées depuis le début de ce

[5] http ://www.realviz.com.

chapitre. Les images utilisées sont trois vues en couleur de la *Maison Carrée*, représentées à la figure 8.7.

(1) (2) (3)

Figure 8.7 – *Vues de la Maison Carrée utilisées.*

Ces images ont été prises à l'aide d'un appareil-photo Canon F1 et numérisées par Kodak. Les seules informations mises à notre disposition sont la donnée constructeur de la focale de l'objectif (50mm). Les paramètres intrinsèques sont calculés à partir de cette donnée ; quant aux paramètres extrinsèques, ils sont inconnus.

Nous présentons à la section 8.3.1 des résultats de reconstruction euclidienne de points d'intérêt détectés dans ces images et à la section 8.3.2 des résultats de transfert direct à partir de correspondances éparses.

Les points d'intérêt utilisés dans les deux approches sont les points de Harris Couleur présentés au chapitre 1, car ce sont eux qui présentent les meilleures mesures de localisation euclidienne (cf. section 2.2.5 du chapitre 2). Ils sont calculés en précision demi-pixel. L'étape de mise en correspondance est réalisée à partir des invariants différentiels couleur (cf. chapitre 4), selon le processus incrémental (cf. chapitre 6). La ou les matrices fondamentales sont estimées à l'aide d'une méthode LMedS (cf. section 7.6.1.3 du chapitre 7).

8.3.1 Reconstruction euclidienne

La première étape de ce travail consiste à mettre en correspondance les points détectés entre les images (1) et (2) puis (2) et (3). Le processus de mise en correspondance incrémental nous permet d'apparier efficacement plusieurs milliers de points pour chaque couple d'images mis en jeu. Le positionnement relatif des trois caméras est ensuite estimé à partir des paramètres intrinsèques et des matrices fondamentales, à l'aide d'une décomposition en valeurs singulières des matrices essentielles (cf. section 7.6.1.5 du chapitre 7). Une fois les caméras fortement calibrées, deux reconstructions euclidiennes S_{12} et S_{23} sont calculées par triangulation aux moindres carrés (cf. section 8.1.4) à partir des appariements respectifs. Ces deux ensembles doivent maintenant être "fusionnés". La méthode utilisée est présentée ci-après.

Regroupement des points 3D reconstruits

Bien que la reconstruction effectuée soit euclidienne, les deux ensembles S_{12} et S_{23} obtenus diffèrent d'un changement d'échelle et sont exprimés dans deux repères relatifs

différents. Il faut donc estimer ce facteur d'échelle ainsi que la transformation qui lie les deux repères, pour pouvoir fusionner les deux ensembles de points 3D.

Ces paramètres sont déterminés à partir des éléments communs à S_{12} et à S_{23} (il suffit de considérer les points de la deuxième image qui ont des appariements dans la première et dans la troisième). Le facteur d'échelle est calculé en considérant le rapport des distances inter-points dans chacun des deux sous-ensembles. Puis le changement de repère est estimé en minimisant les distances entre les points 3D se correspondant dans ces mêmes ensembles. On applique ensuite le changement d'échelle, la translation et la rotation obtenus sur l'un des ensembles complets, de façon à rassembler les données. Avec cette méthode, nous avons réussi à réunir approximativement 12500 points de la scène. Quelques vues de la reconstruction obtenue sont visibles à la figure 8.8.

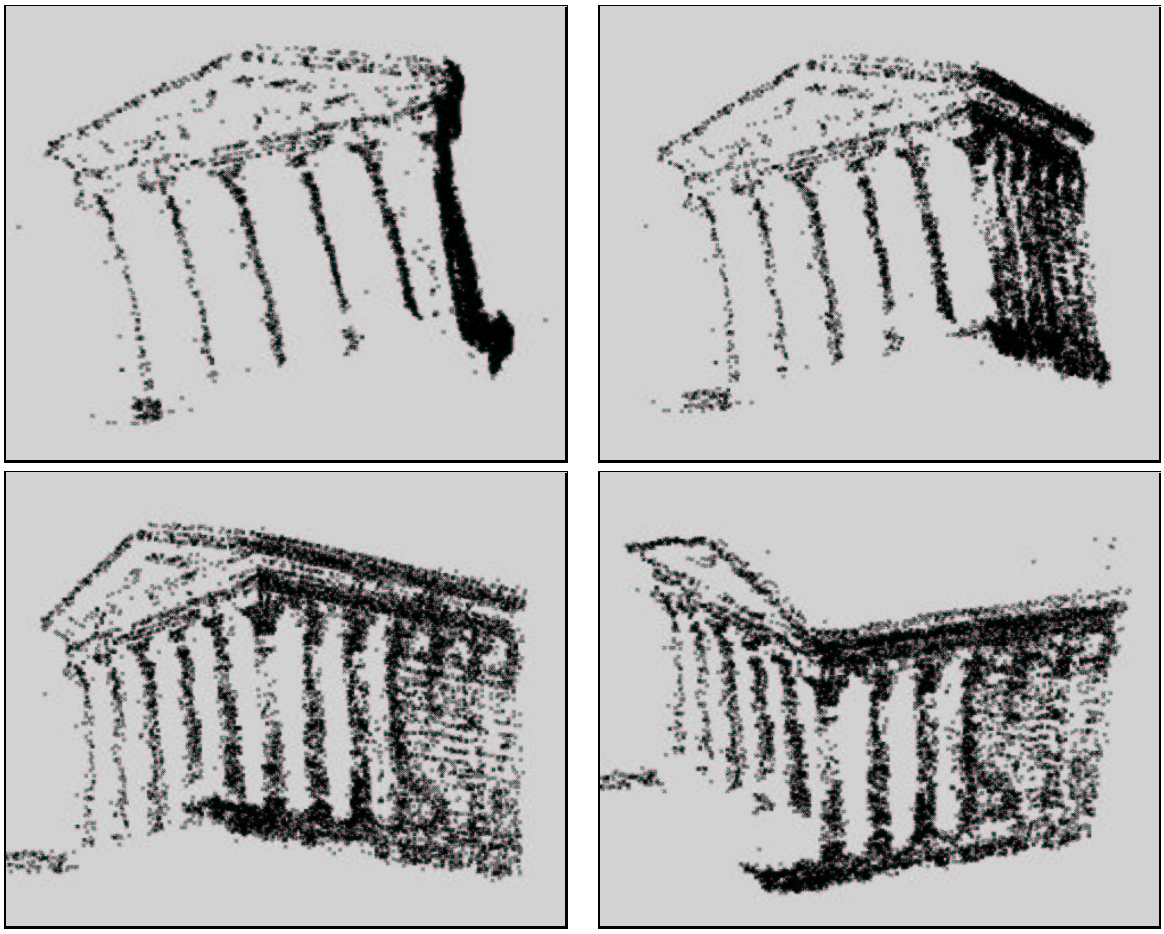


Figure 8.8 – *Reconstruction tridimensionnelle de la Maison Carrée.*

Nous nous sommes "contentés" ici de reconstruire uniquement les points d'intérêt mis en correspondance. Notre but n'était pas d'aboutir à un modèle tridimensionnel complet (avec facettes texturées par exemple) mais uniquement de valider visuellement la cohérence de la chaîne de traitements qui fait l'originalité de cette thèse, depuis l'extraction des points dans les images en couleur jusqu'à l'estimation de la géométrie épipolaire.

Nous constatons que les résultats de reconstruction sont excellents, les façades étant

parfaitement reconnaissables. L'algorithme incrémental de mise en correspondance a permis d'apparier rapidement plus d'une dizaine de milliers de points, mais aussi très précisément, grâce aux contraintes géométriques employées et grâce à la caractérisation par les invariants différentiels couleur.

8.3.2 Transfert projectif

Nous avons implémenté ici la méthode de transfert direct présentée à la section 8.2.5.1. Quelques points ont été mis en correspondance à la l'aide du processus incrémental. A l'issue de ce traitement, nous avons obtenu un ensemble de points appariées ainsi qu'une triangulation des appariements dans chacune des images de référence. Ce sont ces triangles qui sont transférés sur le plan rétinien de la caméra virtuelle. Les pixels présents dans les triangles de l'image virtuelle sont ensuite calculés par interpolation à partir des couleurs des images de référence, par transformation affine ("texture mapping").

Le transfert vers une troisième image nécessite le positionnement de la caméra virtuelle par rapport aux deux caméras de référence. Comme pour la reconstruction euclidienne, il est donc également nécessaire d'estimer le mouvement relatif entre les deux caméras de référence, selon la méthode de décomposition en valeurs singulières de la matrice essentielle. Il est alors possible de spécifier la position et l'orientation de la caméra virtuelle dans le repère de ces deux caméras.

Les images de référence utilisées ici pour réaliser le transfert sont les vues (1) et (2) de la figure8.7. Les triangulations impliquées sont visibles à la figure 8.9. Quelques vues des séquences engendrées sont présentées aux prochaines sections.

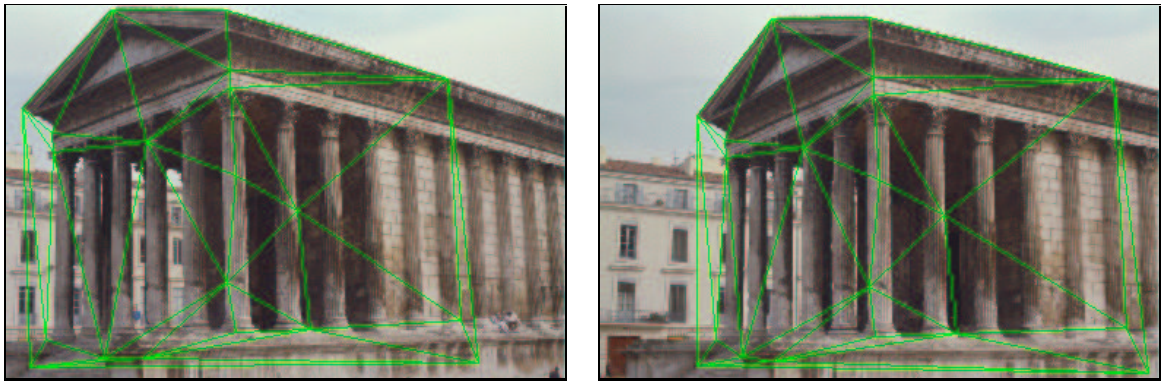


Figure 8.9 – *Triangulations calculées dans les deux images de référence.*

8.3.2.1 Déplacement vers l'avant

Pour cette première séquence, nous avons simulé un déplacement vers l'avant de la première caméra. Ce mouvement correspond à une translation selon l'axe des z du repère de cette caméra. La séquence comporte 30 images, mais seulement 6 sont représentées à la figure 8.10.

Figure 8.10 – *Transfert de 6 images. Le mouvement est une translation vers l'avant à partir de la première caméra.*

8.3.2.2 Vue de la scène en plongée

Dans cette séquence, la caméra virtuelle est déplacée vers le haut et inclinée vers l'avant de façon à visualiser le dessus de la scène. La séquence comporte également 30 images, nous en présentons 6 à la figure 8.11.

Figure 8.11 – *Transfert de 6 images. Vue plongeante de la Maison Carrée.*

Discussion

Au vu des images extraites des deux séquences et présentées ici, quelques imperfections sont notables. En effet, la modélisation en triangles n'est pas exacte d'un point de vue théorique puisqu'elle suppose que les zones définies par les triangles correspondent à des parties planes de la scène, ce qui n'est pas toujours le cas. Ceci explique par exemple les

déformations obtenues pour l'arrière-plan des façades situées à gauche de la Maison Carrée à la figure 8.11. Cette méthode basée sur les triangles suppose également que la contrainte d'ordre est toujours vérifiée. Ceci est bien évidemment faux pour cette scène, témoins en sont les façades de l'arrière-plan entre les colonnes de la Maison Carrée. Notre méthode de plaquage de texture suppose quant à elle que la transformation affine permettant de texturer les triangles dans l'image virtuelle est une bonne approximation de l'homographie associée au plan du triangle, ce qui rarement le cas.

Toutes ces approximations expliquent les petites déformations que l'on peut voir dans les images engendrées de ces séries. Cependant, le rendu visuel reste bon et la simulation du déplacement de la caméra virtuelle parfaitement crédible malgré ces simplifications et malgré un déplacement conséquent de la caméra virtuelle par rapport aux caméras initiales.

En résumé, le transfert de deux images de référence vers une troisième image aura nécessité ici :

1. La détection et la mise en correspondance d'un petit nombre de points dans les deux images ;
2. La triangulation de ces points ;
3. L'estimation de la matrice fondamentale permettant, avec les paramètres intrinsèques, de calculer le mouvement relatif entre les deux caméras ;
4. Le positionnement de la caméra virtuelle par rapport au repère des caméras initiales ;
5. Le transfert vers la troisième image des sommets des triangles contenus dans les images initiales ;
6. Le plaquage de texture dans chacun des triangles générés.

8.4 Conclusion

Dans ce chapitre, nous avons passé en revue les différentes approches permettant de synthétiser des vues intermédiaires, à partir de vues existantes. Il est tout d'abord possible de reconstruire la scène en trois dimensions, avant de la reprojeter sur le plan rétinien de la caméra virtuelle. Trois types de reconstruction sont envisageables selon les informations disponibles. La reconstruction projective nécessite la donnée de la géométrie épipolaire du système, c'est-à-dire d'un certain nombre de points mis en correspondance. Elle permet d'obtenir une reconstruction à une homographie de l'espace près, et ne conduit pas à un modèle tridimensionnel réaliste. Les reconstructions affines et euclidiennes, quant à elles, tiennent compte du parallélisme, des mesures d'angles et de longueurs de la scène. Elles sont cependant beaucoup plus délicates à obtenir. En effet, les méthodes rencontrées nécessitent ou bien la donnée manuelle d'informations (droites parallèles, segments de droite, etc) ; elles ne sont donc pas automatiques et supposent que la scène contient ces informations de structure. Ou bien elles nécessitent la mise en œuvre de méthodes d'estimations relativement difficiles, comme celles utilisant les équations de Kruppa. Lorsque les paramètres des caméras sont inconnus, il est donc encore actuellement délicat d'arriver jusqu'à une reconstruction tridimensionnelle réaliste.

Nous avons ensuite passé en revue les diverses solutions développées ces dernières années, qui permettent de se passer de cette reconstruction. Elles consistent à générer l'image uniquement à partir des relations existant entre les caméras de référence et la caméra virtuelle. Certaines de ces méthodes utilisent les n-linéarités, qui supposent au moins la connaissance de correspondances de points entre les images. D'autres exploitent certaines propriétés d'invariance bien connues en géométrie projective. Ces approches fournissent globalement de bons résultats, dans la mesure où la caméra virtuelle est correctement positionnée. Or tout le problème réside dans le choix des points définissant le placement relatif de cette caméra par rapport aux autres caméras. Nous avons en effet constaté que lorsque les caméras ne sont pas calibrées, il n'existe pas de méthode permettant de placer ces points de façon à être certain de calculer une image réaliste, c'est-à-dire non projectivement déformée. Il s'avère nécessaire de mener une étude pour essayer de contraindre au maximum leurs positions.

Dans la dernière section de ce chapitre, nous avons présenté divers résultats de reconstruction tridimensionnelle et de transfert d'images obtenus à partir de plusieurs images. Nous avions seulement à notre disposition une estimation des paramètres intrinsèques des caméras. En ce qui concerne la reconstruction tridimensionnelle, un très grand nombre de points ont pu être mis en correspondance efficacement grâce à l'algorithme d'appariement présenté au chapitre 6, avant d'être reconstruits. La précision de la reconstruction obtenue valide tout la chaîne de traitements. Nous avons également présenté des résultats de transfert projectif basée sur une approche directe et mettant en jeu des correspondances couleur éparses. Les séquences d'images générées restent parfaitement crédibles malgré la simplicité de la méthode mise en œuvre.

Conclusion

Cette thèse a été consacrée au développement d'une méthode d'appariement entre images capable de s'affranchir d'une grande partie des limitations inhérentes aux méthodes existantes. La première originalité de notre approche a résidé dans le choix d'exploiter l'information supplémentaire disponible dans les images en couleur. Notre apport a également consisté à la mise en place d'un algorithme d'appariement robuste aux grands nombres de points.

Une méthode de mise en correspondance robuste

Nous résumons ici les principaux points que nous avons abordés dans ce mémoire ainsi que les contributions apportées.

Les points d'intérêt couleur

Notre méthode d'appariement est basée sur l'utilisation de points d'intérêt. Elle permet donc de manipuler des familles d'images très générales, ceux-ci pouvant être extraits à partir de n'importe quel type de scène. Nous avons présenté deux nouveaux détecteurs dont la spécificité est d'exploiter l'information couleur pour extraire les points d'intérêt. En outre, ces détecteurs permettent d'extraire des points en précision sub-pixel. Au sein du chapitre 2, il a alors été démontré par de nombreuses expérimentations que ces détecteurs sont plus stables que les détecteurs en niveau de gris. Plus précisément, ils ont été évalués selon un critère de répétabilité sur plusieurs séquences différant des principales transformations de l'image, et aussi selon deux critères de localisation. Ces mesures ont également été réalisées sur le détecteur de Harris dans sa version précise, qui représente à l'heure actuelle le détecteur le plus stable en niveau de gris. Exception faite du cas particulier des changements affines de luminosité, les détecteurs couleur fournissent de meilleurs résultats que le détecteur de Harris précis, pour chacune des transformations de l'image appliquées. L'étude des changements affines de luminosité s'est avérée délicate à réaliser. En effet, les plans des images couleur ont été modifiés synthétiquement selon une transformation affine des niveaux de gris et les images en niveau de gris correspondantes ont été calculées par moyenne de ces plans modifiés. Il en résulte des images qui ne respectent pas les mêmes conditions que celles en couleur. Nous sommes donc plutôt réservés quant aux résultats obtenus.

La comparaison entre les deux détecteurs couleur doit être beaucoup plus nuancée. Nous sommes tout de même arrivés à la conclusion que le détecteur de Harris couleur possède une répétabilité et une localisation euclidienne supérieures à celles obtenues pour le détecteur de coins couleur. Nous préconisons donc son utilisation pour la mise en correspondance ainsi que pour la reconstruction tridimensionnelle. Au contraire, l'extracteur de coins possède une meilleure localisation projective, ce qui le destine par exemple à l'étalonnage de caméras. Soulignons enfin que ces deux détecteurs n'extraient pas les mêmes points, le détecteur de Harris couleur ne calcule en effet pas exclusivement des coins de l'image.

Les invariants différentiels couleur

La méthode de caractérisation que nous avons présentée au chapitre 4 utilise une caractérisation locale du signal. Elle étend à la couleur l'approche proposée par Schmid [Schmid, 1996] et est basée sur les invariants différentiels de Hilbert. Nous avons montré que l'apport de deux nouveaux invariants inter-canaux, conjugués aux invariants sur les trois plans couleur, permet de pallier un des principaux inconvénients qui faisait défaut à la caractérisation correspondante en niveau de gris, à savoir sa sensibilité au bruit. Elle réduit en effet à l'ordre un le calcul des invariants différentiels, tout en restant aussi riche avec huit invariants.

Cette caractérisation est invariante aux rotations de l'image. Les dérivées étant calculées sur un support gaussien, elle peut être intégrée dans un contexte multi-échelle qui lui permet d'être robuste aux changements d'échelle. Elle est donc invariante au groupe des similitudes, qui absorbe les variations dues à des changements de point de vue lors d'une projection perspective. Cette caractérisation est donc robuste à de telles transformations.

Invariance aux changements d'illumination

La deuxième étape de notre travail de caractérisation des points d'intérêt a ensuite consisté à traiter le problème de la constance des couleurs. Nous avons adopté comme modèle d'illumination le modèle de Finlayson enrichi d'une translation (caractérisé par six paramètres au total) et implémenté localement. Il s'avère être le plus performant dans le cadre de la caractérisation locale d'images. Pour rendre les invariants différentiels invariant aux changements d'illumination, les approches traditionnelles consistent à les normaliser en considérant les rapports de certains d'entre eux. Cette solution n'est pas envisageable dans notre approche, d'une part parce que le vecteur d'invariants couleur qui en résulterait serait trop pauvre avec deux invariants seulement, et d'autre part parce que l'emploi de quotients de dérivées ne ferait que renforcer la sensibilité au bruit de la caractérisation.

Nous avons donc opté pour un pré-traitement de l'image qui permet de conserver nos huit invariants couleur au travers du modèle d'illumination à six paramètres. Il consiste en une normalisation locale de l'image qui permet d'éliminer les paramètres du modèle. Cette normalisation est robuste au bruit puisqu'elle travaille à l'ordre zéro. Elle est implémentée localement pour préserver les propriétés locales des pixels. Les résultats obtenus sur plusieurs séries d'images montrent que cette approche est robuste aux changements d'illumination internes. Elle a également été testée sur des images ayant subi des changements d'illumination externes, liés au déplacement des sources lumineuses de la scène. Les résultats se

sont également avérés satisfaisants. Une méthode de normalisation face à ces changements externes a aussi été présentée et comparée à notre approche locale, mais elle n'est pas exploitable avec les invariants différentiels couleur qui y perdraient en richesse.

Robustesse aux grands ensembles de points

A partir des points d'intérêt caractérisés, nous avons mis en place au chapitre 6 une méthode de mise en correspondance robuste aux grands ensembles de points. La métrique que nous avons utilisée pour la comparaison des invariants n'est pas la plus optimale, au regard de la distance de Mahalanobis qui intègre un modèle de bruit. Cependant elle se comporte bien avec notre caractérisation, dans la mesure où celle-ci n'implique que les dérivées de l'image d'ordre peu bruitées. Nous avons également développé un algorithme de relaxation basé sur des contraintes semi-locales de voisinage et des contraintes géométriques invariantes au groupe de similitudes. La méthode de mise en correspondance ainsi développée est invariante aux principales transformations de l'image incluant le changement de point de vue.

Puis nous avons mis en place un algorithme incrémental basé sur la géométrie épipolaire et sur une contrainte de semi-planarité locale utilisant la triangulation de Delaunay. Cette méthode permet de réduire considérablement la complexité des méthodes de mise en correspondance traditionnelles, mais aussi limiter les faux appariements. Les tests menés sur de grands ensembles de points ont montré que nous sommes capables d'obtenir des taux d'appariements corrects de l'ordre de 95%, pour des temps de calcul très réduits. A l'issue de toute la chaîne de traitement, nous avons à notre disposition la géométrie épipolaire, qui a été estimée de manière robuste (LMedS) à partir des points appariés, ainsi qu'une triangulation de Delaunay. Ces données permettent alors d'envisager la reconstruction tridimensionnelle ou encore le transfert d'images.

Perspectives

Une première amélioration directe des travaux qui viennent d'être présentés est l'intégration de la distance de Mahalanobis pour la comparaison des invariants, qui suppose d'abord l'estimation empirique de la matrice de covariance des composantes du vecteur d'invariants. La méthode de mise en correspondance en sera encore améliorée.

A plus long terme, d'autre améliorations sont envisageables :

Autres espaces couleur

Les deux opérateurs de détection mis en place pour la couleur ainsi que la caractérisation utilisant les invariants différentiels couleur sont basés sur l'espace des couleurs RVB. Il serait peut-être intéressant de développer une méthode de caractérisation dans d'autres espaces couleur et de la comparer à celle que nous avons présentée. L'étude pourra être notamment réalisée en fonction des caméras utilisées et de l'échantillonnage couleur réalisé selon les capteurs employés.

En ce qui concerne notre méthode caractérisation basée les invariants différentiels couleur, une étude pourrait être menée sur leur utilisation jusqu'à l'ordre deux, en ayant pris soin au préalable de normaliser localement les images pour obtenir l'invariance aux chan-

gements d'illumination. Cette étude permettrait de vérifier si les calculs supplémentaires impliqués apportent un gain significatif dans la caractérisation des points d'intérêt.

Etude comparative des deux détecteurs couleur plus approfondie

Les deux détecteurs mis en place dans la première partie ont été comparés par rapport aux principales transformations de l'image. Nous avons constaté que leur utilisation doit être fonction des applications à envisager. Il s'avère cependant que le détecteur de Harris couleur extrait des points à l'endroit où le signal possède un contenu informatif riche, ce qui n'implique pas obligatoirement que le point extrait corresponde à un coin. Il est donc possible que ce détecteur s'avère mieux adapté à des vues issues de scènes texturées. Au contraire, le détecteur de coins couleur pourrait se révéler plus efficace sur des scènes au fort contenu géométrique. Une étude reste à mener dans ce sens.

Contraintes géométriques supplémentaires pour l'appariement

Le nombre de contraintes géométriques aidant à la mise en correspondance de points d'intérêt est très limitée. Dans nos travaux, nous avons exploité la géométrie épipolaire ainsi qu'une contrainte de semi-planarité. Il pourrait être intéressant d'introduire également des primitives de plus haut niveau, comme des jonctions ou des segments. On procéderait ainsi à une mise en correspondance structurelle [Alquier, 1998], qui permettrait de réduire encore plus l'espace de recherche des correspondants d'un point d'intérêt.

Changement d'échelle

Nous pensons que le processus de mise en correspondance incrémental, basé sur la géométrie épipolaire et une triangulation de Delaunay, pourrait permettre d'améliorer la mise en correspondance dans le contexte d'un changement d'échelle entre les deux images. Nous sommes actuellement en train de mettre en place une méthode incrémentale qui tient compte de cette transformation et qui exploite les invariants calculés dans un contexte multi-échelle.

Transfert d'images

Les résultats de transfert d'images présentés au chapitre 8 sont basés sur une approche projective directe utilisant une mise en correspondance éparse. Ces résultats sont assez satisfaisants malgré les approximations faites, mais sont fortement dépendants de la structure de la scène et en conséquence de la triangulation mise en œuvre. Pour mieux tenir compte de sa géométrie, une première amélioration consisterait à utiliser une triangulation contrainte sur les contours présents dans les images.

Comme l'a fait Laveau dans sa thèse [Laveau, 1996], il serait également intéressant de mettre en place la méthode de transfert indirect présentée à la section 8.2.5.2 du chapitre 8. Celle-ci exploite la carte de disparité des images de référence. C'est la raison pour laquelle nous sommes actuellement en train de mettre en œuvre une méthode de mise en correspondance dense qui exploite la caractérisation basée sur les invariants différentiels couleur et les contraintes géométriques de la mise en correspondance incrémentale. Nous espérons arriver à une solution plus efficace que les traditionnelles approches qui utilisent des mesures de corrélation.

D'autre part, lorsque la calibration des caméras est totalement inconnue, l'inconvénient majeur des techniques de transfert projectif réside dans la détermination des points permettant le positionnement de la caméra virtuelle. Une étude reste à mener pour essayer de contraindre au maximum les points à fixer. Remarquons néanmoins que si le transfert est réalisé dans le but de faire de la compression vidéo, alors l'image à synthétiser est connue à l'avance, rendant ainsi aisée la détermination de ces points.

Annexes

Annexe A

Notations utilisées

$\mathbf{I}$	luminance d'une image en niveau de gris.
$\{\mathbf{R}, \mathbf{V}, \mathbf{B}\}$	plans Rouge, Vert et Bleu d'une image couleur.
$\mathbf{I_x}$ et $\mathbf{I_y}$	dérivées premières en x et en y de l'image I.
$\mathbf{I_{xx}}$, $\mathbf{I_{xy}}$ et $\mathbf{I_{yy}}$	dérivées secondes en x et en y de l'image I.
$\nabla\mathbf{I}$	gradient de l'image I.
σ	taille du filtre gaussien.
$\mathbf{u}.\mathbf{v}$	produit scalaire de u par v.
$\mathbf{u} \wedge \mathbf{v}$	produit vectoriel de u par v.
$\mathbf{dist}(\mathbf{m_1}, \mathbf{m_2})$	distance euclidienne entre les points m_1 et m_2.
$\widehat{\mathbf{t}}$	matrice antisymétrique de $t = (t_x, t_y, t_z)^T$ telle que $\widehat{t}x = t \wedge x$ pour tout vecteur x. $$\widehat{t} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$$

$\mathcal{A}$ matrice d'auto-corrélation de l'image, définie pour le déplacement (t_1, t_2) telle que :

$$\mathcal{A} = \frac{\sum_{x,y} \left\{ \left[I(x,y) - \overline{I(x,y)} \right] \times \left[I(x-t_1, y-t_2) - \overline{I(x-t_1, y-t_2)} \right] \right\}}{\sqrt{\sum_{x,y} \left[I(x,y) - \overline{I(x,y)} \right]^2 \times \sum_{x,y} \left[I(x-t_1, y-t_2) - \overline{I(x-t_1, y-t_2)} \right]^2}} \quad \text{(A.1)}$$

où $\overline{I(x,y)}$ désigne la moyenne des niveaux de gris $I(x,y)$.

F matrice fondamentale d'un système binoculaire (quelques fois F_{ij} pour la matrice fondamentale du système (I_i, I_j)).

E matrice essentielle d'un système binoculaire (quelques fois E_{ij} pour la matrice essentielle du système (I_i, I_j)).

C centre optique d'une caméra (quelques fois C_i pour la $i$ème caméra).

A matrice des paramètres intrinsèques.

$\mathbf{e_{ij}}$ épipôle généré par l'image I_j dans l'image I_i.

H homographie (quelques fois H_{ij} pour l'homographie reliant les images I_i et I_j).

P matrice de projection (quelques fois P_i pour la matrice de projection associée à l'image I_i).

M point de la scène tridimensionnelle.

$\mathbf{m_i}$ $i$ème point de l'image.

Annexe B

Matériel photographique utilisé

Ce mémoire est illustré par divers résultats de détection de contours et de coins, de mise en correspondance et de reconstruction tridimensionnelle. Ceux-ci ont été calculés à partir de différentes séries d'images, en niveau de gris comme en couleur. Il est bon de noter que les images qui ont servi de référence proviennent de sources différentes. La table B.1 ci-après présente les caractéristiques des capteurs qui ont permis de les obtenir.

Séquence	Matériel	Capteur
Toys	Sony DXC-750P + scanner Sony	tri-CCD
Lézard	-	tri-CCD
Salle robotique	-	mono-CCD 24 bits
Bureau	Sony DFW-V500 8.5mm	mono-CCD 16 bits
Bonzaï	Sony DFW-V500 8.5mm	mono-CCD 16 bits
Maison Carrée	Canon F1 50mm + scanner Nikon	mono-CCD 24 bits

Table B.1 – *Matériel photographique utilisé selon l'image.*

Les images en niveau de gris

Toutes les images de la table B.1 sont disponibles en couleur. Nos travaux de détection (cf. chapitres 1 et 2) ont nécessité d'avoir, pour certaines d'entre elles, l'image équivalente en niveau de gris. Nous avons obtenu les images monochromatiques correspondantes en considérant la moyenne des trois plans couleur. Les vues de la Maison Carrée n'ont pas nécessité de correspondances en niveau de gris.

Il nous a parfois été nécessaire de bruiter quelques unes des images en niveau de gris et en couleur (cf. chapitre 2). Pour les images en niveau de gris, le bruit a été superposé à l'image calculée par moyenne des plans couleur non bruités. En couleur, le bruit a été superposé sur les trois plans couleur simultanément.

Annexe C

Filtrage sub-pixel et segmentation contours en niveau de gris

Cette annexe a pour but de présenter les notions essentielles de filtrage employées en traitement d'images et utilisées tout au long de ce mémoire. Nous rappelons à la section C.1 les principales méthodes permettant de dériver une image, ainsi que le principe du lissage en précision sub-pixel, qui autorise un gain de précision conséquent par rapport au méthodes "classiques". La section C.2 rappelle brièvement le principe de la détection de contours dans une image en niveau de gris.

C.1 Filtrage, précision pixel et précision sub-pixel

C.1.1 Méthodes linéaires pour la détection de contours

Les méthodes linéaires pour la détection de contours font partie des méthodes de segmentation les plus utilisées en traitement d'images.

Une image est représentée par une fonction bi-dimensionnelle échantillonnée. Les contours des objets sont généralement localisés au zones de transition d'intensité lumineuse de l'image (dans le cas d'images non texturées). Il est donc nécessaire de dériver celle-ci afin d'expliciter les fortes variations d'intensité lumineuse. Les points de contours sont alors repérés par les maxima locaux du gradient de l'image dans la direction du gradient.

Une image étant un signal échantillonné bruité, il est impossible d'effectuer de simples différences finies afin d'estimer ses dérivées de manière fiable. Généralement, les dérivées des images sont estimées à l'aide de filtres de convolution. De nombreuses études en segmentation d'images ont montré qu'il était seulement nécessaire de déterminer un filtre de lissage [Canny, 1983; Deriche, 1987; Castan et Shen, 1989]. Le filtre de dérivation s'obtient

alors simplement en dérivant le filtre de lissage, grâce à la propriété suivante du produit de convolution :

$$\begin{cases} I_x(x,y) = \frac{\partial}{\partial x}(I(x,y) * L(x,y)) = I(x,y) * \frac{\partial}{\partial x}L(x,y) \\ I_y(x,y) = \frac{\partial}{\partial y}(I(x,y) * L(x,y)) = I(x,y) * \frac{\partial}{\partial y}L(x,y) \end{cases} \quad \text{(C.1)}$$

où $I(x,y)$ représente l'image initiale, $L(x,y)$ représente le filtre de lissage et $I_x(x,y)$, $I_y(x,y)$ représentent respectivement les images dérivées selon les axes x et y.

Selon la modélisation de Canny [Canny, 1983], on recherche un filtre optimal selon trois critères : rapport signal sur bruit, localisation et réponse unique à un contour unique. Canny approxime l'opérateur idéal répondant à ces trois critères, par la dérivée première d'une gaussienne. Dans la bibliographie, on trouve essentiellement deux familles de filtres, des filtres basés sur la gaussienne et ses dérivées [Canny, 1983; Deriche et Giraudon, 1993], et des filtres basés sur le filtre exponentiel et ses dérivées [Deriche, 1987; Castan et Shen, 1989; Montesinos et Dattenny, 1997]. Nous utilisons plus spécialement dans ce mémoire le filtre gaussien qui présente la meilleure stabilité pour les dérivées d'ordre supérieur à un.

C.1.2 Filtrage gaussien en précision pixel

Les dérivées de l'image peuvent être approximées par convolution avec les dérivées du filtre gaussien. Ce filtre, exprimé par l'équation C.2, est contrôlé par un paramètre de lissage : l'écart-type σ de la gaussienne. Lorsque σ est grand le lissage est important, à l'inverse lorsque σ est petit le lissage est faible.

$$G_\sigma(x) \;=\; C_0\, e^{-\frac{x^2}{2\sigma^2}} \quad \text{(C.2)}$$

où C_0 représente un coefficient de normalisation.

Deux propriétés intéressantes du filtre gaussien sont sa séparabilité et son isotropie, c'est à dire que d'une part, le filtrage peut être décomposé selon les deux directions x et y de l'image et d'autre part que le filtrage réalisé est indépendant de l'orientation des structures de l'image considérée.

L'expression du filtre gaussien pour un signal bi-dimensionnel est la suivante :

$$G_\sigma(x,y) \;=\; C_0\, e^{-\frac{x^2}{2\sigma^2}}\, e^{-\frac{y^2}{2\sigma^2}} \;=\; C_0\, e^{-\frac{r^2}{2\sigma^2}} \quad \text{(C.3)}$$

avec $r \;=\; \sqrt{x^2+y^2}$.

Une opération de filtrage bi-dimensionnel va donc se réduire à deux opérations successives de filtrage mono-dimensionnel.

$$I(x,y) *_{2D} G_\sigma(x,y) \;=\; (I(x,y) *_{1D,x} G_\sigma(x)) *_{1D,y} G_\sigma(y) \quad \text{(C.4)}$$

où $*_{1D,x}$, $*_{1D,y}$ et $*_{2D}$ représentent respectivement les convolutions $1D$ selon les axes x et y, et $*_{2D}$ la convolution $2D$.

Cependant lorsque σ est grand, l'augmentation de la taille des noyaux de convolution rend coûteuse l'opération de filtrage. Nous préférons alors une méthode approchée basée sur une approximation récursive à l'ordre quatre de la gaussienne et de ses dérivées [Deriche et Giraudon, 1993]. La gaussienne et ses dérivées peuvent alors être approximées par l'opérateur suivant :

$$h(x) = (a_0 \cos(\frac{\omega_0}{\sigma}x) + (a_1 \sin(\frac{\omega_0}{\sigma}x))e^{-\frac{b_0}{\sigma}x} + (c_0 \cos(\frac{\omega_1}{\sigma}x) + (c_1 \sin(\frac{\omega_1}{\sigma}x))e^{-\frac{b_1}{\sigma}x} \quad \text{(C.5)}$$

où les paramètres des filtres sont déterminés aux moindres carrés par la méthode de Prony.

C.1.3 Les problèmes d'échantillonnage, vers la précision sub-pixel

Il est montré dans [Montesinos et Dattenny, 1997] que les implantations traditionnelles des filtres de détection de contours sont incapables d'extraire toute l'information contenue dans une image (au sens de Shannon), et que ce problème provient de l'échantillonnage des filtres utilisés. Afin de remédier à ce problème, les auteurs définissent des filtres décalés d'une quantité ϵ (avec $0 \leq \epsilon < 1$). Nous donnons ici les expressions des filtres décalés de lissage et de dérivation jusqu'à l'ordre deux (pour certains traitements, il peut être nécessaire d'utiliser des dérivées jusqu'à l'ordre trois, par exemple pour le calcul d'invariants différentiels, les expressions données ici se généralisant aisément aux dérivées d'ordre supérieur).

$$\begin{cases} f_{\epsilon,0}(n) = C_{\epsilon,0}\Gamma_{\epsilon,0}(n) \\ f_{\epsilon,1}(n) = C'_{\epsilon,1}\Gamma_{\epsilon,1}(n) + C_{\epsilon,1}f_{\epsilon,0}(n) \\ f_{\epsilon,2}(n) = C''_{\epsilon,2}\Gamma_{\epsilon,2}(n) + C'_{\epsilon,2}f_{\epsilon,1}(n) + C_{\epsilon,2}f_{\epsilon,0}(n) \end{cases} \quad \text{(C.6)}$$

Pour les filtres gaussiens, les fonctions de base s'écrivent :

$$\begin{cases} \Gamma_{\epsilon,0}(n) = e^{-\frac{(n-\epsilon)^2}{2\sigma^2}} \\ \Gamma_{\epsilon,1}(n) = -\frac{(n-\epsilon)}{\sigma^2}e^{-\frac{(n-\epsilon)^2}{2\sigma^2}} \\ \Gamma_{\epsilon,2}(n) = \frac{(n-\epsilon)^2}{\sigma^4}e^{\frac{-(n-\epsilon)^2}{2\sigma^2}} - \frac{e^{\frac{-(n-\epsilon)^2}{2\sigma^2}}}{\sigma^2} \end{cases} \quad \text{(C.7)}$$

En deux dimensions, ces filtres basés sur la gaussienne et ses dérivées sont séparables mais non récursifs. Cependant pour des valeurs courantes du paramètre de lissage σ les noyaux de convolution (1D) sont de petite taille, ce qui assure une complexité algorithmique faible. L'estimation précise des dérivées des images se fait alors en combinant plusieurs filtres avec des décalages différents, fournissant ainsi une information localisée entre les pixels. En précision demi-pixel, il est nécessaire de combiner quatre filtres, avec les valeurs de (ϵ_x, ϵ_y) suivantes : $(0,0)$, $(\frac{1}{2},0)$, $(0,\frac{1}{2})$ et $(\frac{1}{2},\frac{1}{2})$. L'idée est illustrée par la figure C.1.

Coefficients de normalisation

Un filtrage en précision sub-pixel combine plusieurs filtres de décalages différents. Une attention toute particulière doit donc être portée aux coefficients de normalisation afin

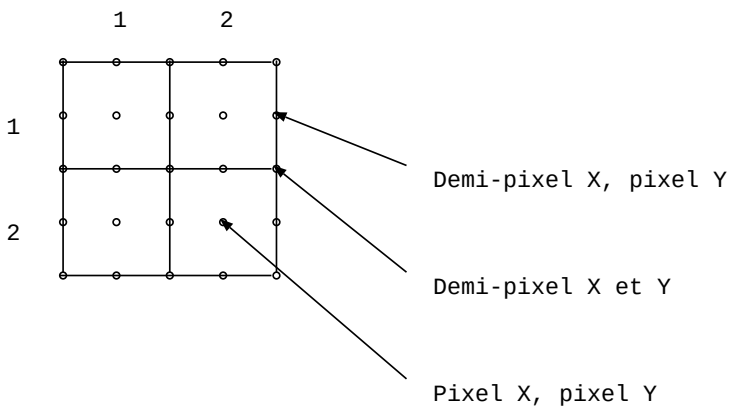


Figure C.1 – *Filtrage demi-pixel : localisation des réponses des différents détecteurs en fonction du décalage.*

d'obtenir des dérivées correctes. Pour cela, les coefficients de normalisation des filtres sont calculés de manière à obtenir des dérivées exactes pour des fonctions polynômiales. Ces coefficients (jusqu'à l'ordre deux) doivent donc vérifier les équations suivantes :

$$\begin{aligned}
&\sum_{n=-\infty}^{+\infty} f_{\epsilon,0}(n) = 1 && \sum_{n=-\infty}^{+\infty} f_{\epsilon,1}(n) = 0 \\
&\sum_{n=-\infty}^{+\infty} (n-\epsilon) f_{\epsilon,1}(n) = -1 && \sum_{n=-\infty}^{+\infty} f_{\epsilon,2}(n) = 0 \\
&\sum_{n=-\infty}^{+\infty} (n-\epsilon) f_{\epsilon,2}(n) = 0 && \sum_{n=-\infty}^{+\infty} \frac{(n-\epsilon)^2}{2} f_{\epsilon,2}(n) = 1
\end{aligned} \tag{C.8}$$

C.2 Segmentation contours

Les contours sont extraits par maximisation de la norme du gradient dans sa direction, le gradient et sa direction proviennent directement de l'étape de filtrage où les dérivées sont estimées par exemple en précision sub-pixel. Le calcul des contours nécessite les opérations suivantes :

1. Calcul de la norme du gradient :

$$\| \vec{\nabla} I \| = \sqrt{I_x^2 + I_y^2} \tag{C.9}$$

2. Calcul de la direction du gradient par rapport à l'axe des x :

$$\theta = \arctan\left(\frac{I_y}{I_x}\right) \tag{C.10}$$

3. Suppression des non maxima locaux de la norme du gradient dans sa direction ;
4. Seuillage par hystérésis.

Annexe D

Résultats complémentaires de segmentation couleur

Cette annexe donne des résultats de segmentation contour couleur et de détection de points d'intérêt couleur, qui sont complémentaires à ceux présentés au chapitre 1.

D.1 Contours couleur en précision sub-pixel

D.1.1 Opérateur de Di-Zenzo

Des contours calculés à partir de l'opérateur de Di-Zenzo et en précision quart de pixel sont présentés aux figures D.1 jusqu'à D.7. Pour toutes ces images, les dérivées sont estimées à l'aide de filtres gaussiens sub-pixel par convolution ($\sigma = 1$). Dans tous les cas, les contours obtenus sont obtenus par un seuillage par hystérésis (seuil bas=0.005, seuil haut=0.05).

D.1.2 Opérateur MAX

Au quart de pixel de précision, le lecteur peut consulter les images des figures D.8 jusqu'à D.14 obtenues à partir de l'opérateur MAX. On peut constater des dédoublements (voire des triplements) de certains contours qui correspondent à de légers décalages sub-pixelliques entre les différents plans de l'image couleur.

D.2 Points d'intérêt en précision sub-pixel

En ce qui concerne la détection en précision sub-pixel, les dérivées sont estimées à l'aide de filtres gaussiens sub-pixel par convolution ($\sigma = 1$). Les paramètres de détection des coins sont : seuil=0.05, maximisation des coins dans une fenêtre circulaire de diamètre 15 pixels. Les coins détectés sont présentés des figures D.15 jusqu'à D.19. D'autres exemples de détections à partir des deux opérateurs couleur sont disponibles aux figures D.20 et D.21.

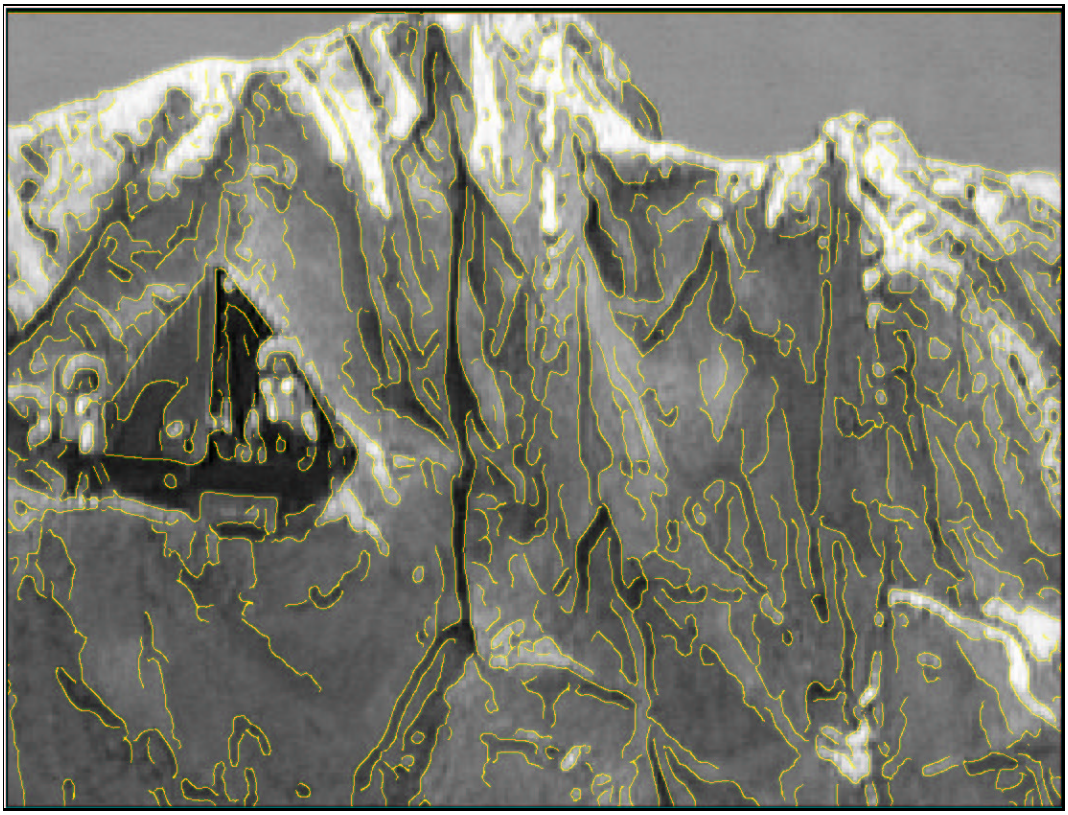


Figure D.1 – *Détection de contours couleur en précision quart de pixel par la méthode de Di-Zenzo : détail des "collines".*

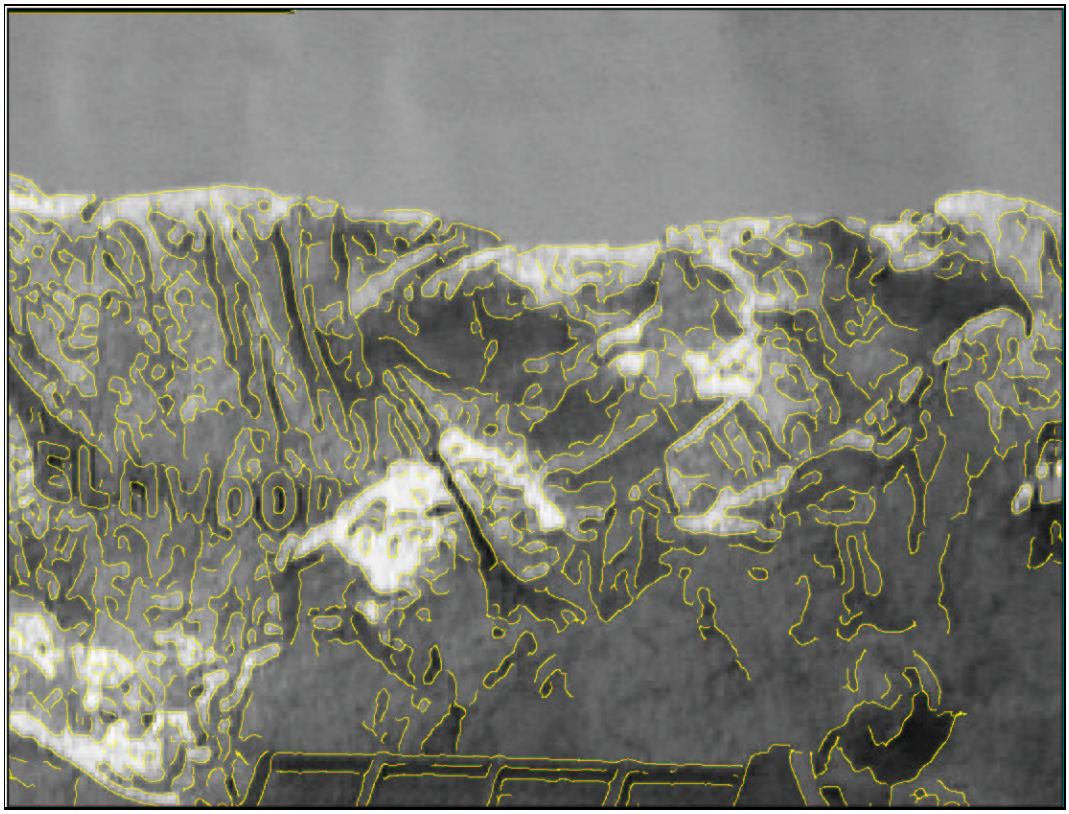


Figure D.2 – *Détection de contours couleur en précision quart de pixel par la méthode de Di-Zenzo : détail des "collines".*

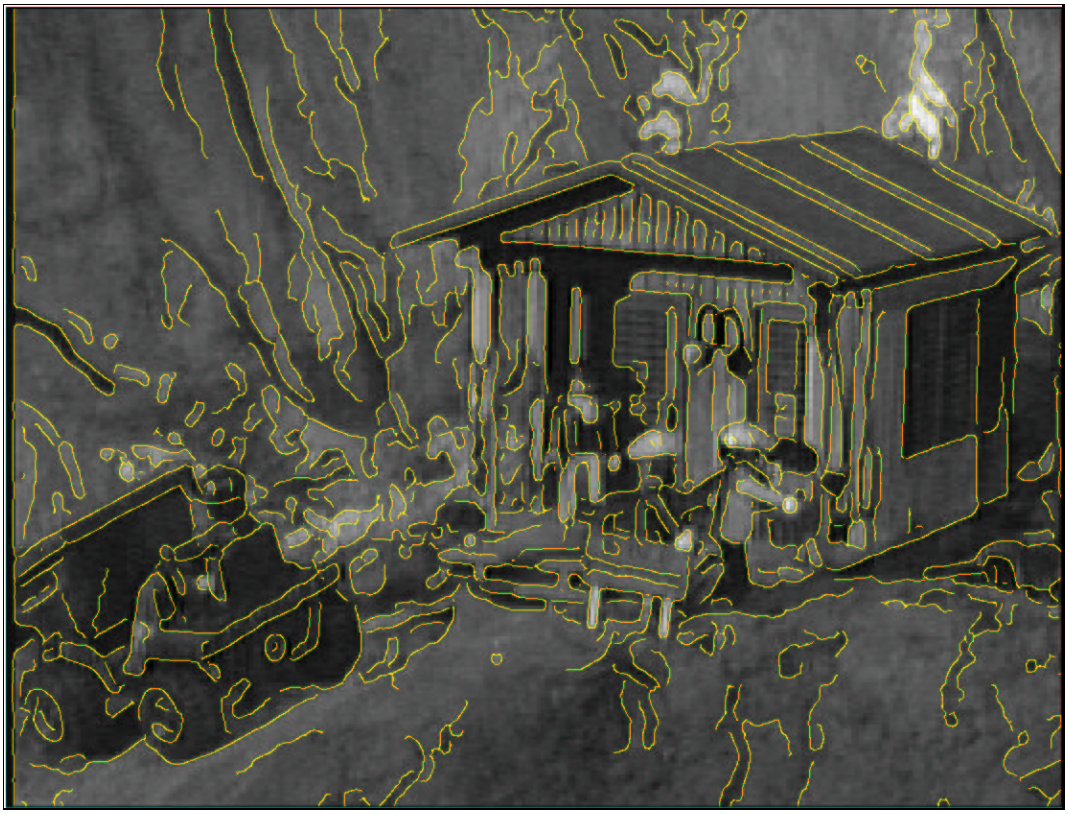


Figure D.3 – *Détection de contours couleur en précision quart de pixel par la méthode de Di-Zenzo : détail des "maisons".*

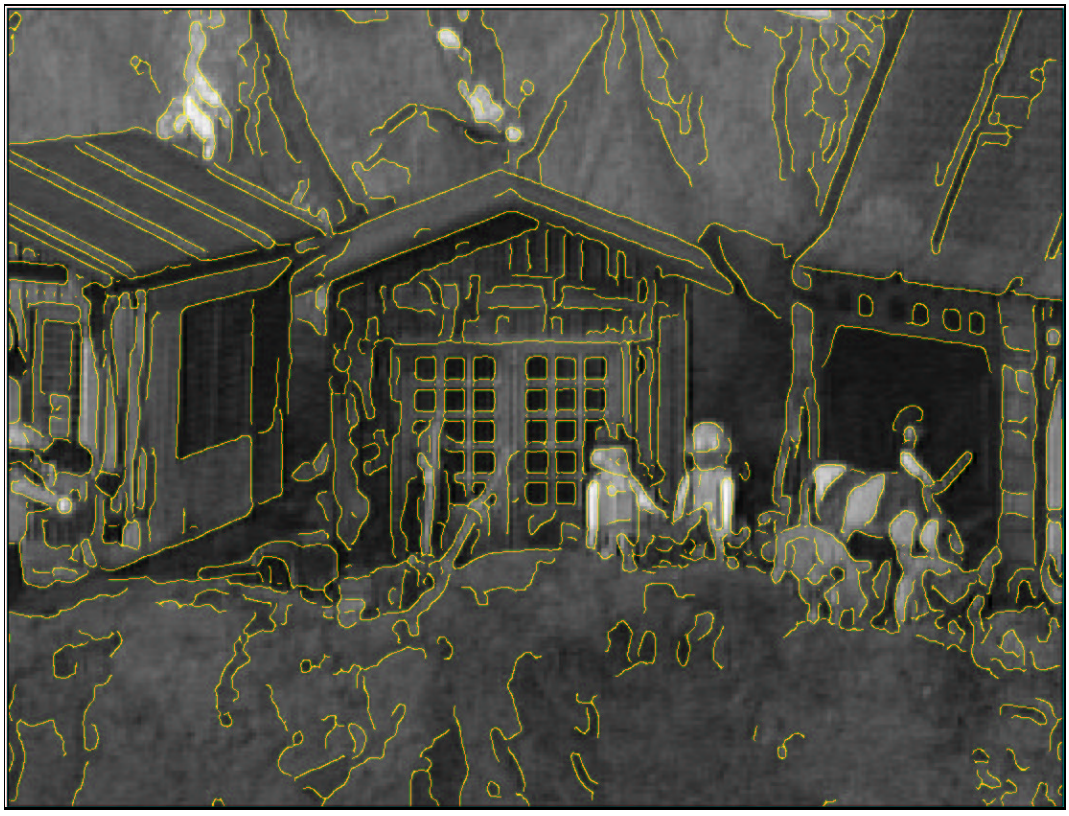


Figure D.4 – *Détection de contours couleur en précision quart de pixel par la méthode de Di-Zenzo : détail des "maisons".*

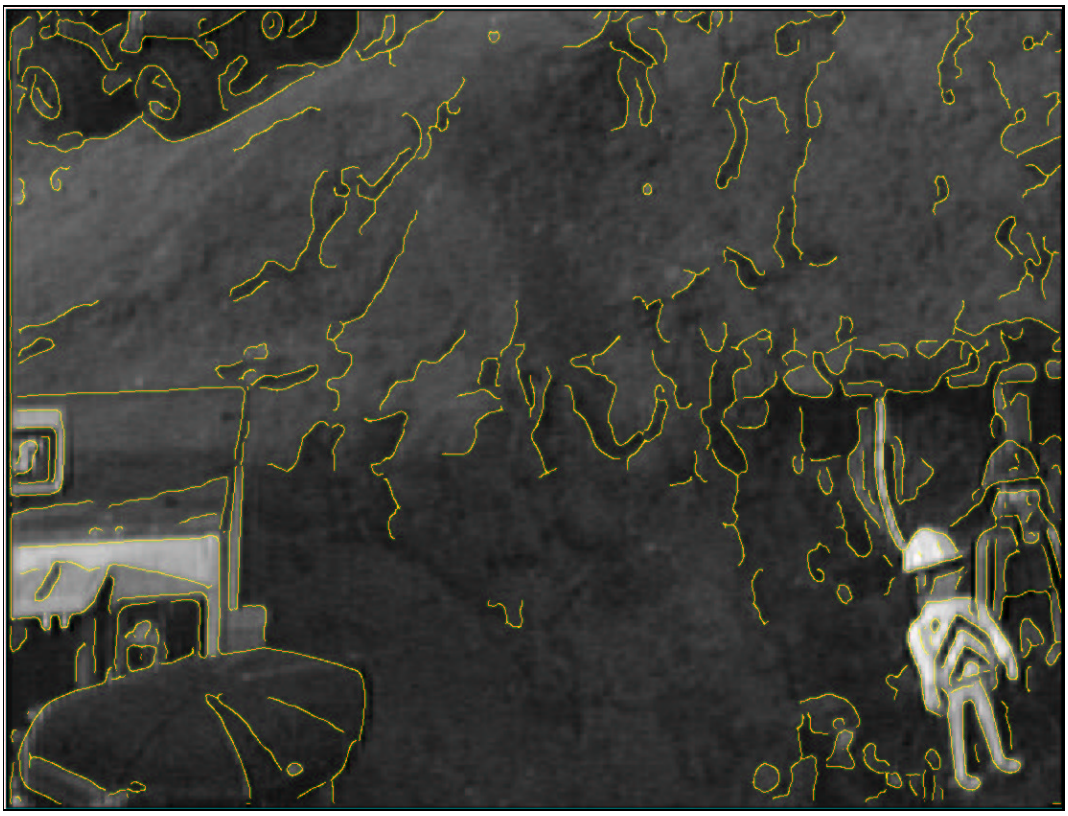


Figure D.5 – *Détection de contours couleur en précision quart de pixel par la méthode de Di-Zenzo : détail du "parasol".*

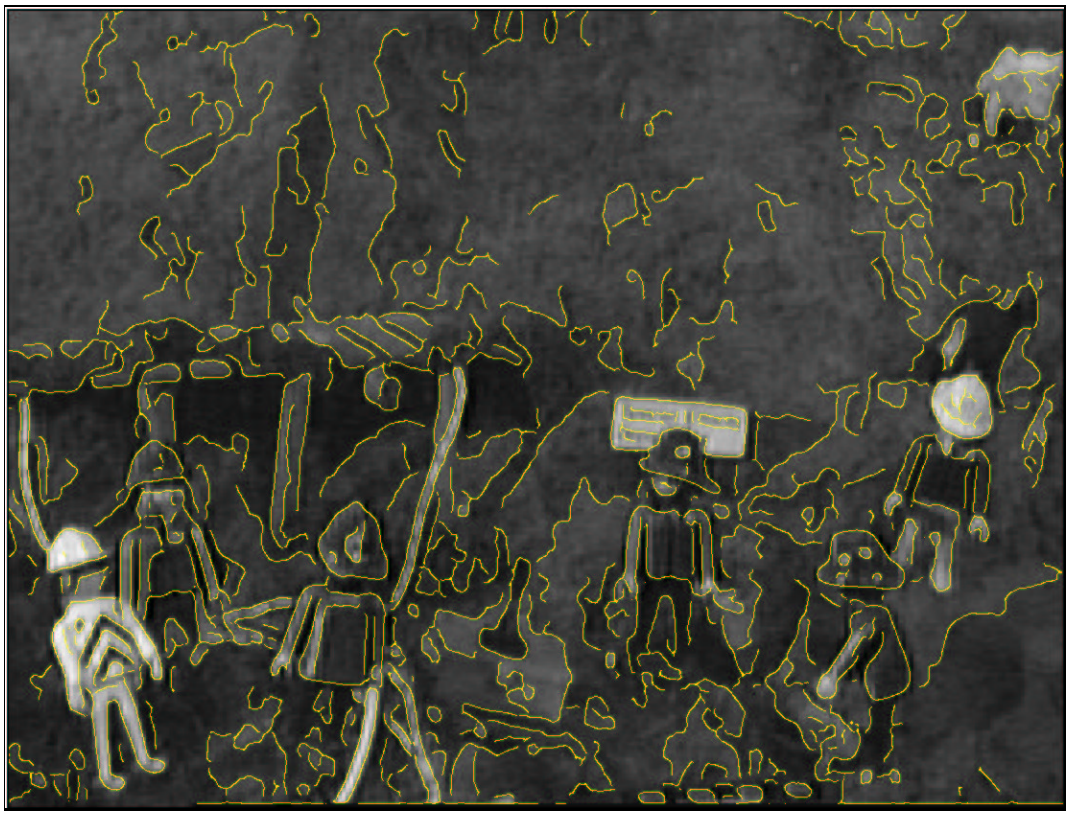


Figure D.6 – *Détection de contours couleur en précision quart de pixel par la méthode de Di-Zenzo : détail des "personnages".*

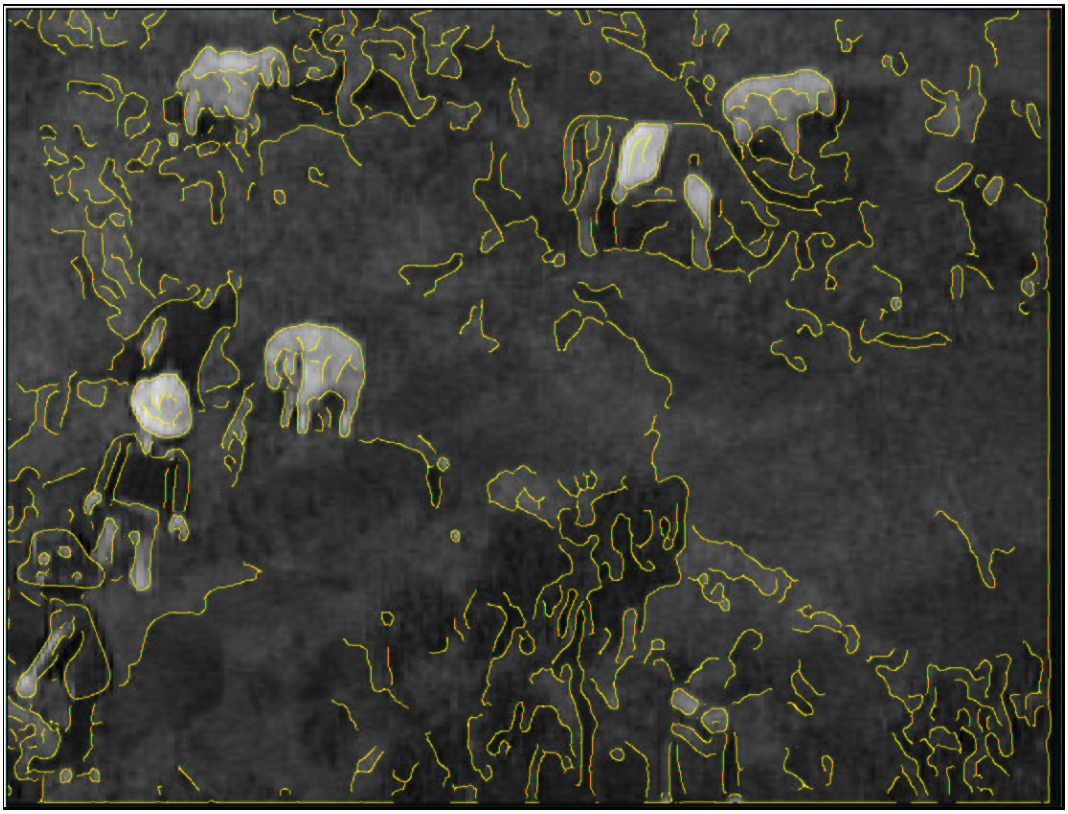


Figure D.7 – *Détection de contours couleur en précision quart de pixel par la méthode de Di-Zenzo : détail des "animaux".*

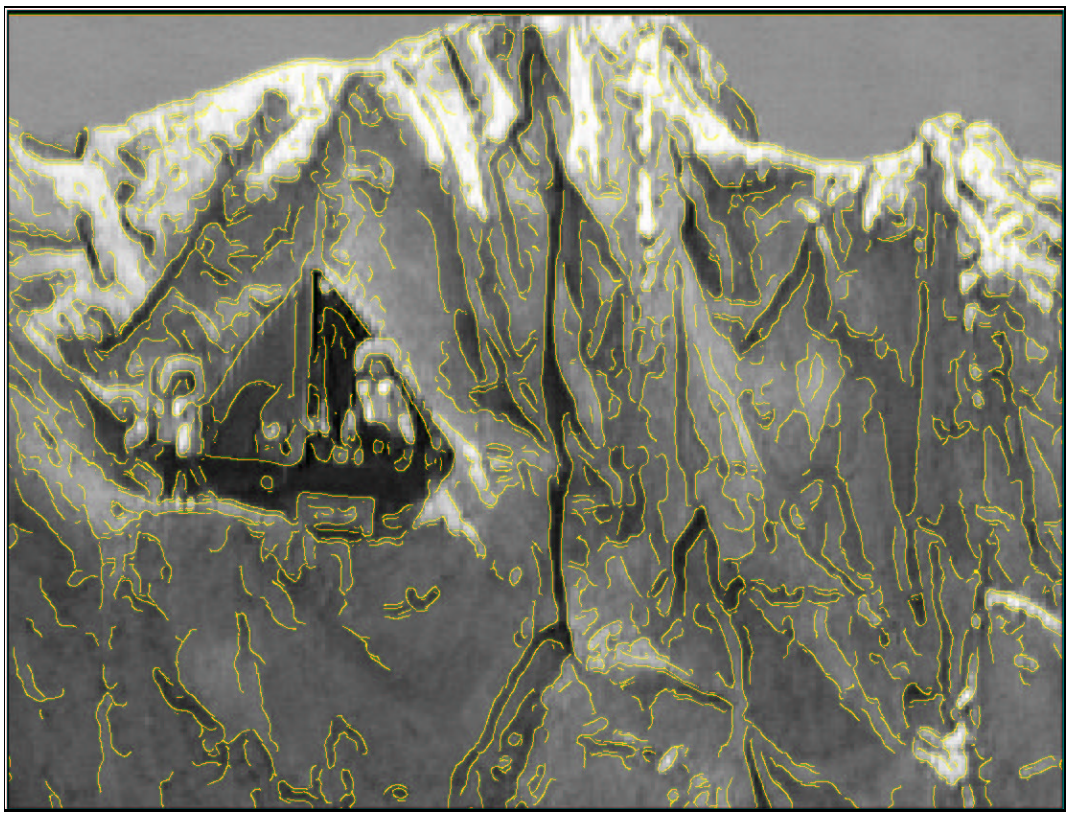


Figure D.8 – *Détection de contours couleur en précision sub-pixel par la méthode* MAX *: détail des "collines".*

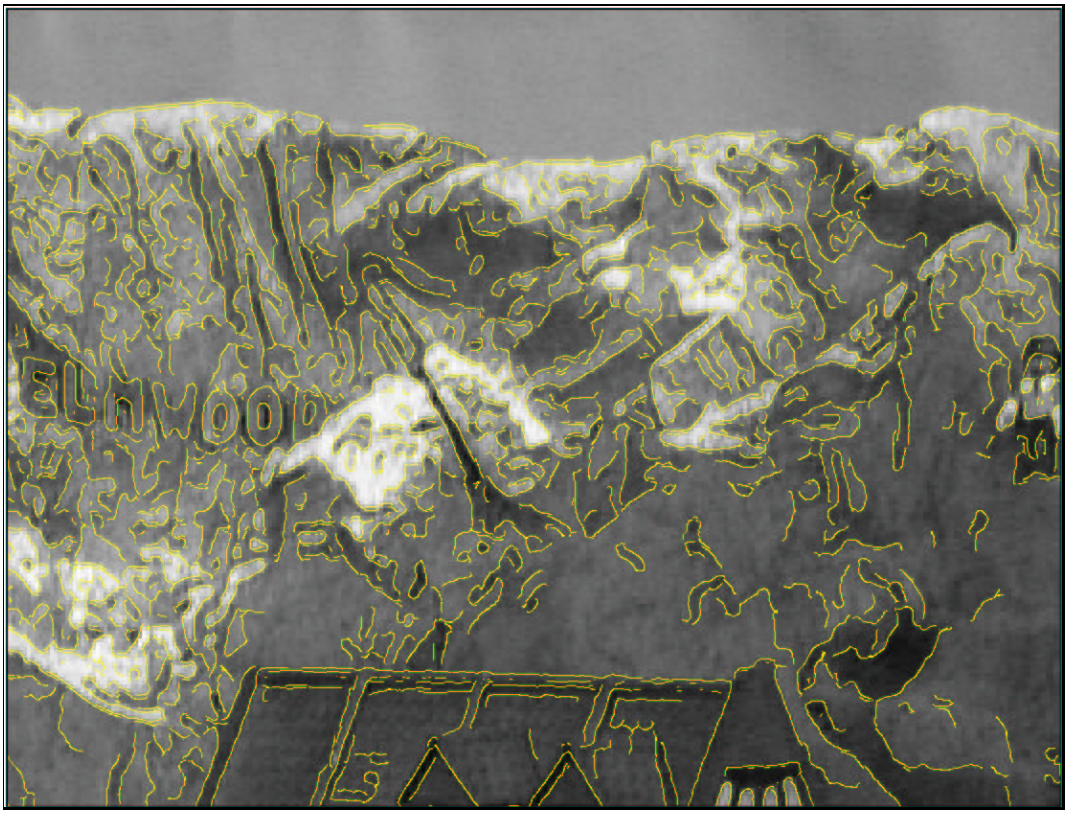


Figure D.9 – *Détection de contours couleur en précision sub-pixel par la méthode* MAX *: détail des "collines".*

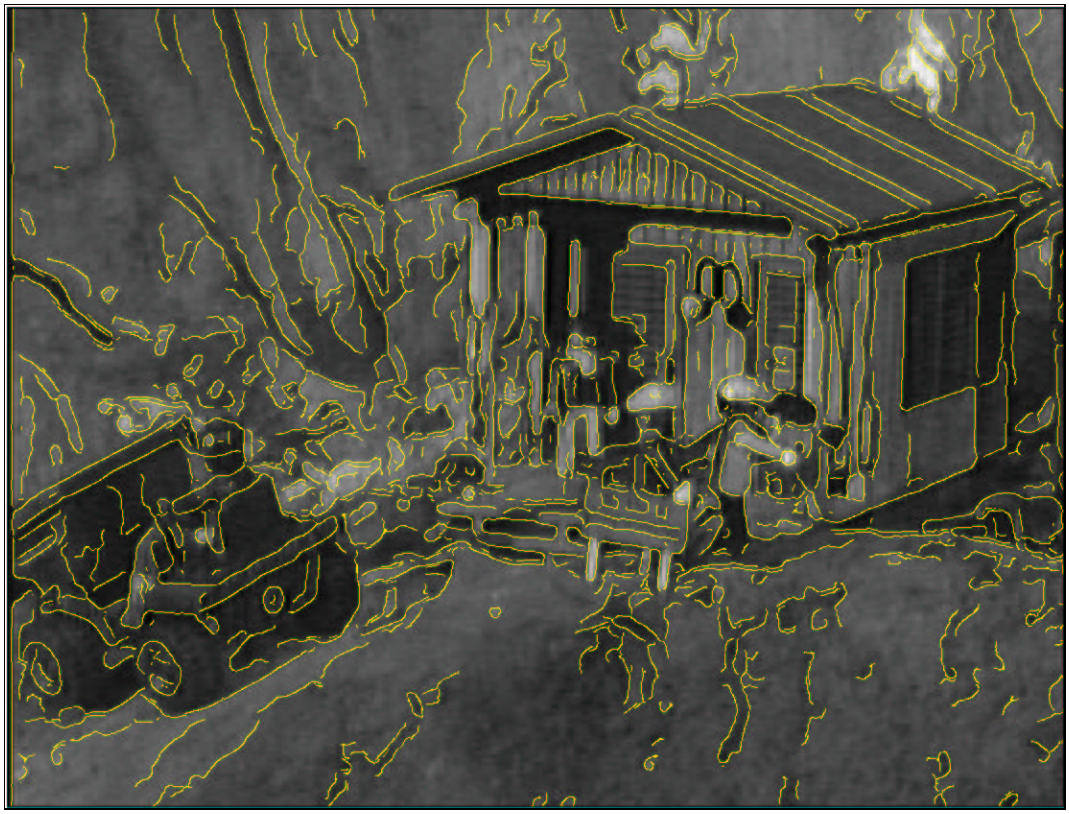


Figure D.10 – *Détection de contours couleur en précision sub-pixel par la méthode* MAX *: détail des "maisons".*

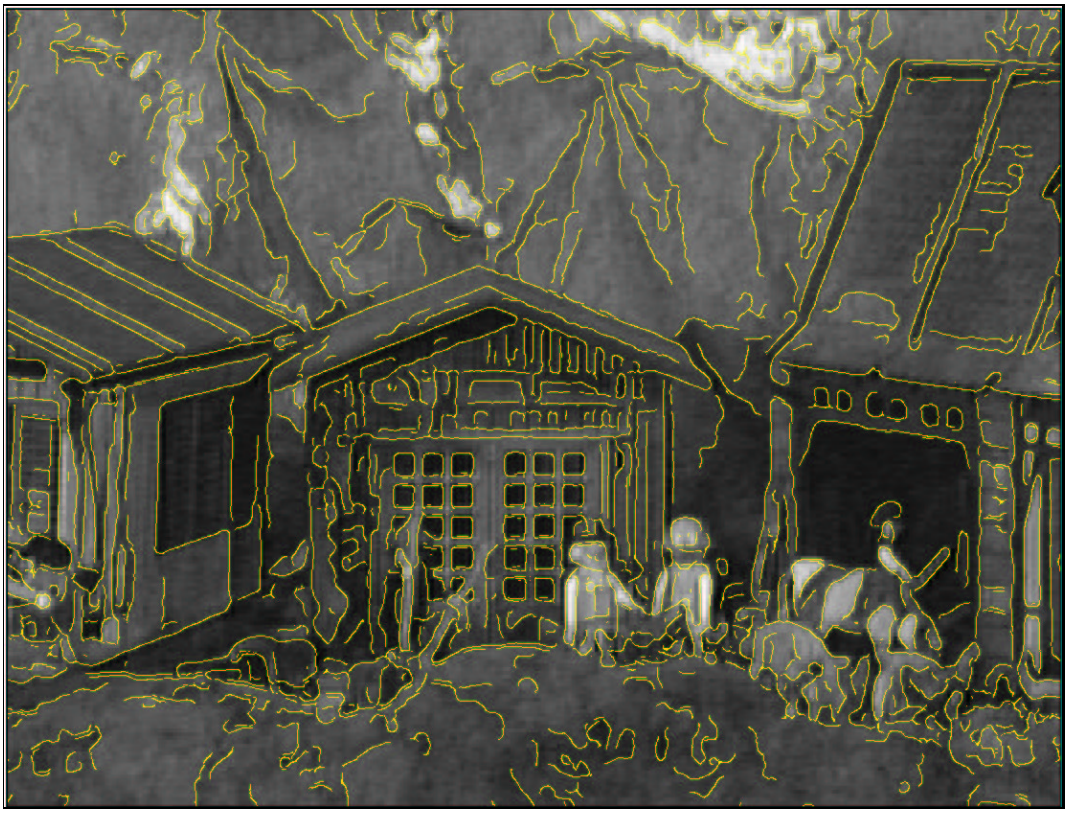


Figure D.11 – *Détection de contours couleur en précision sub-pixel par la méthode* MAX *: détail des "maisons".*

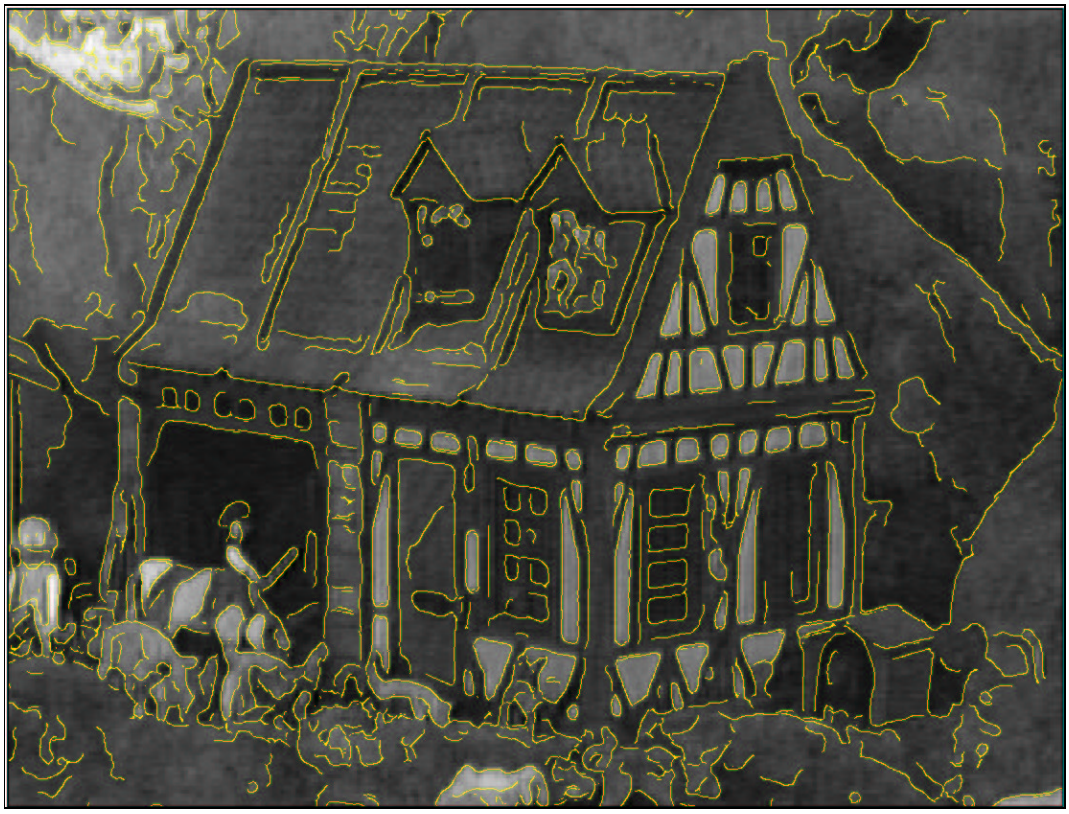


Figure D.12 – *Détection de contours couleur en précision sub-pixel par la méthode* MAX *: détail des "maisons".*

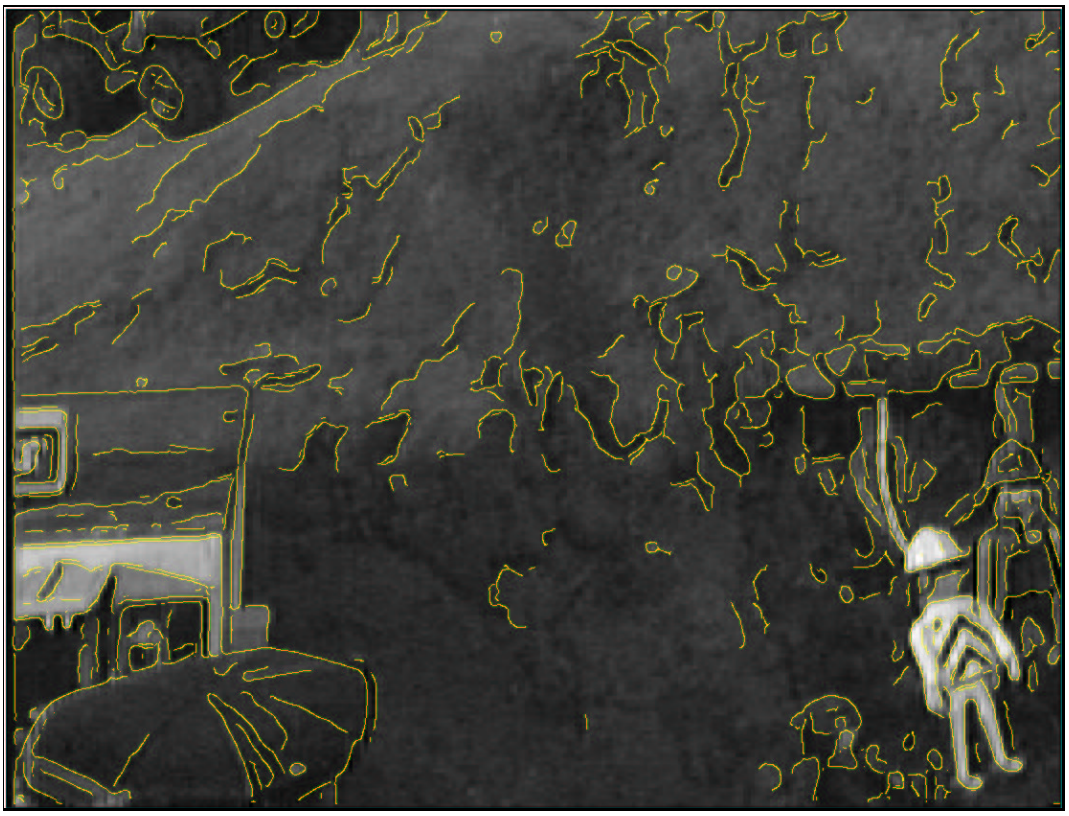


Figure D.13 – *Détection de contours couleur en précision sub-pixel par la méthode* MAX *: détail du "parasol".*

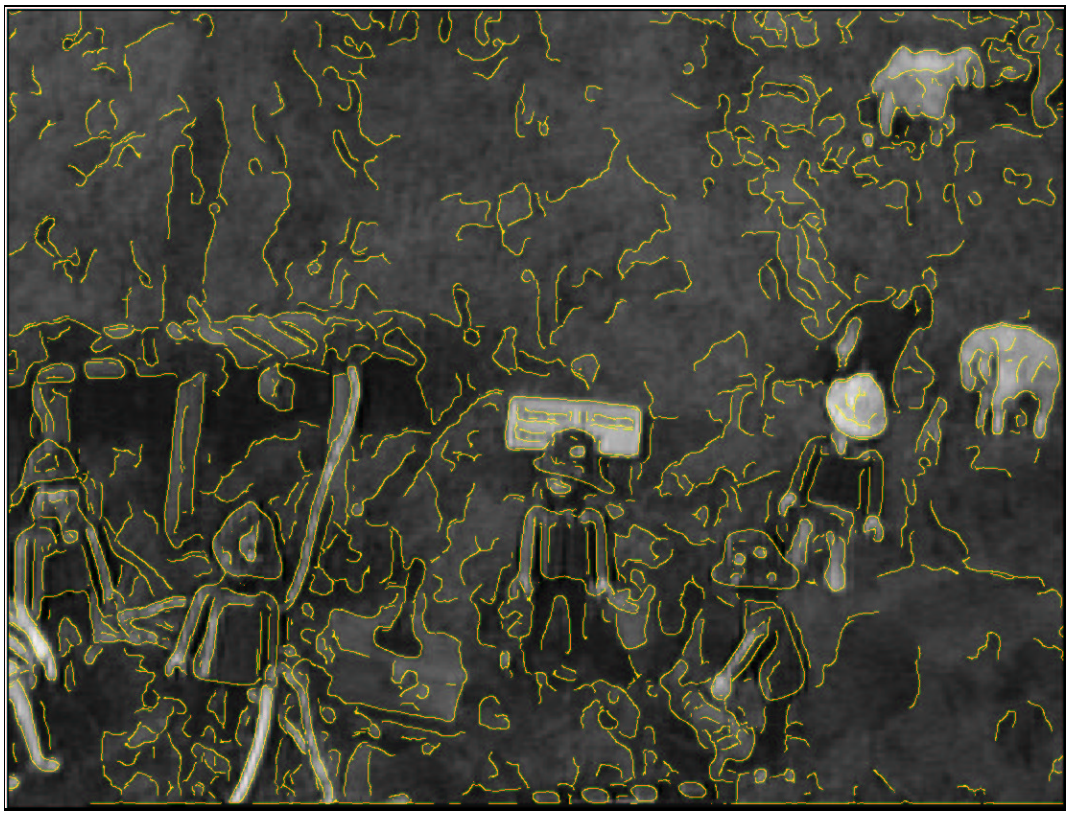


Figure D.14 – *Détection de contours couleur en précision sub-pixel par la méthode* MAX *: détail des "personnages".*

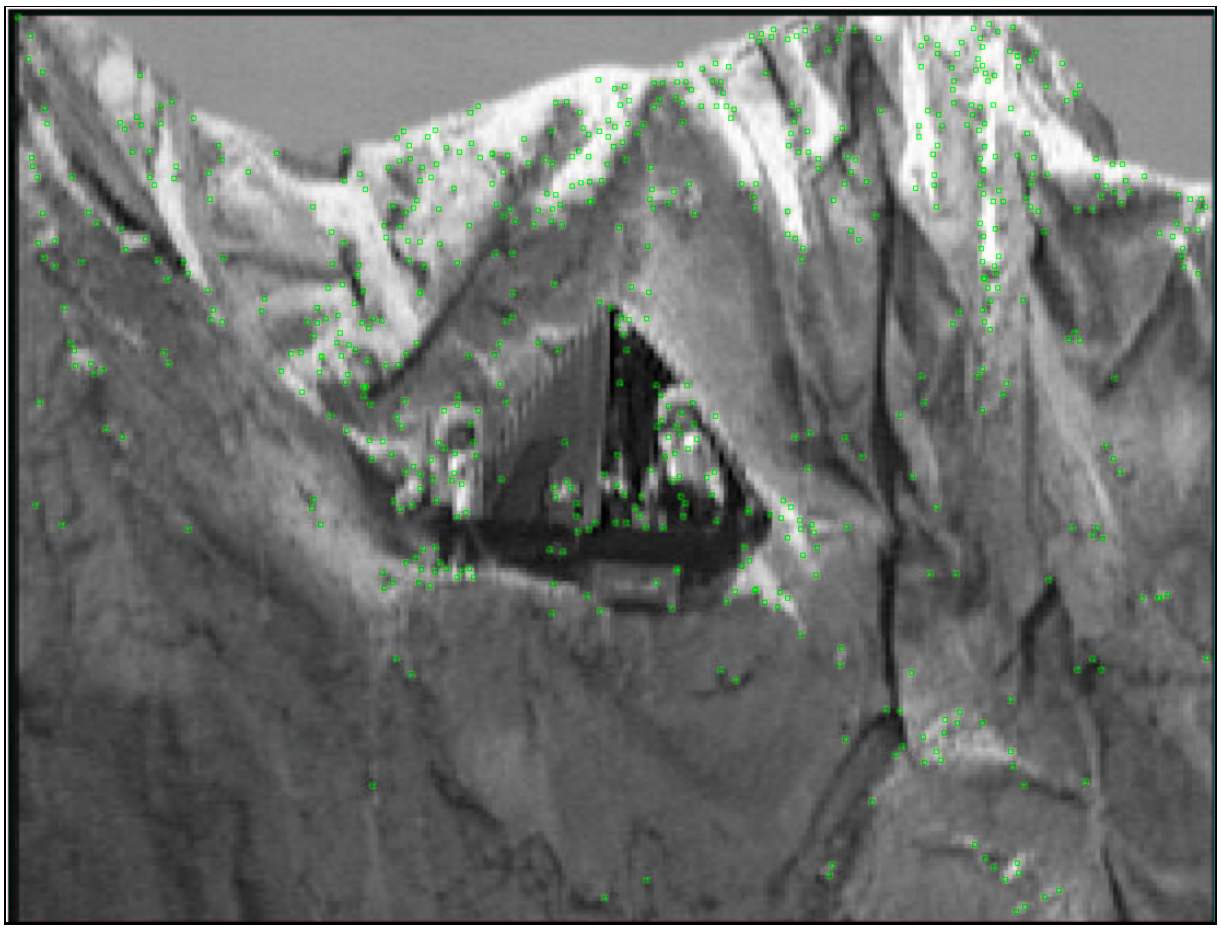


Figure D.15 – *Détection de coins couleur en précision quart de pixel (détail des "collines").*

Figure D.16 – *Détection de coins couleur en précision quart de pixel (détail des "maisons").*

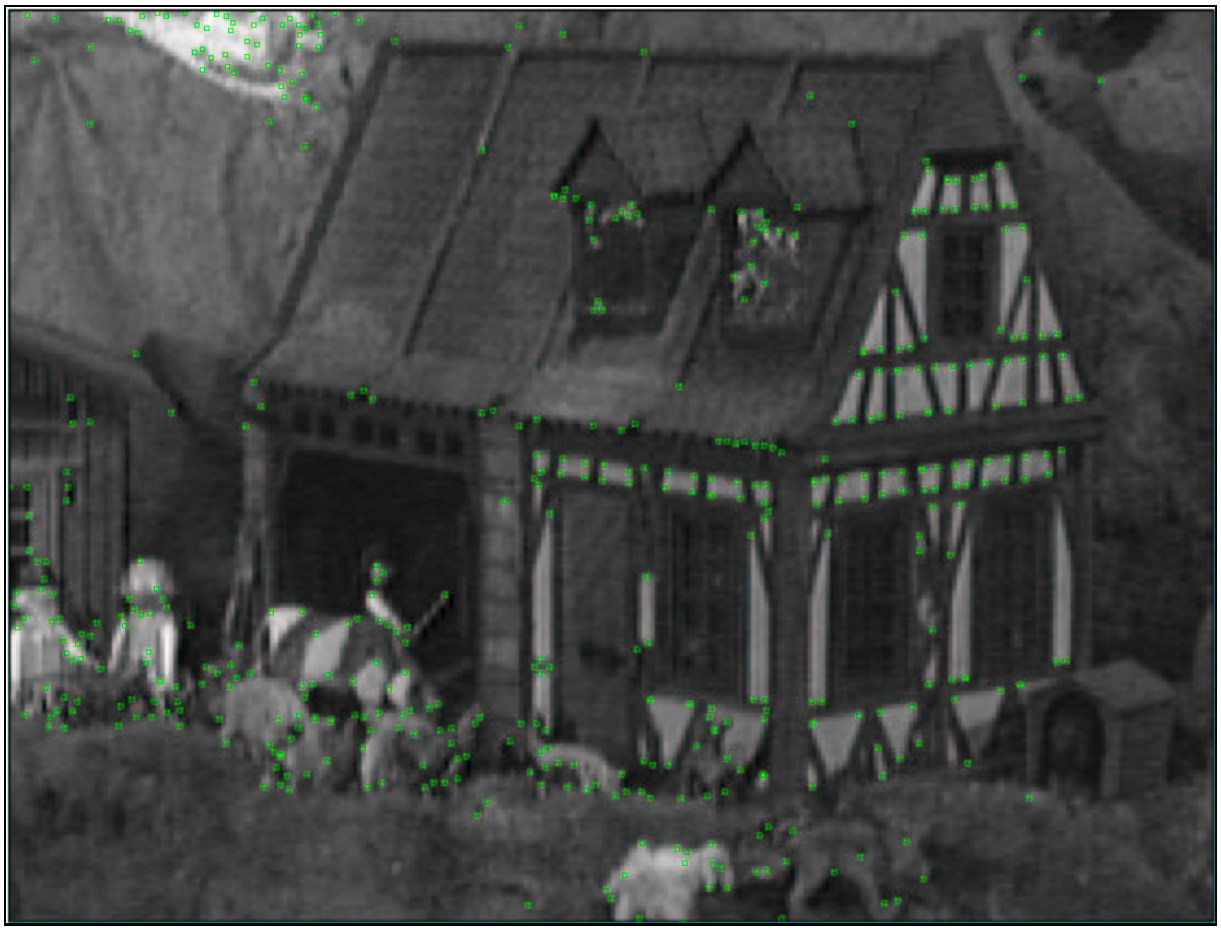


Figure D.17 – *Détection de coins couleur en précision quart de pixel (détail des "maisons").*

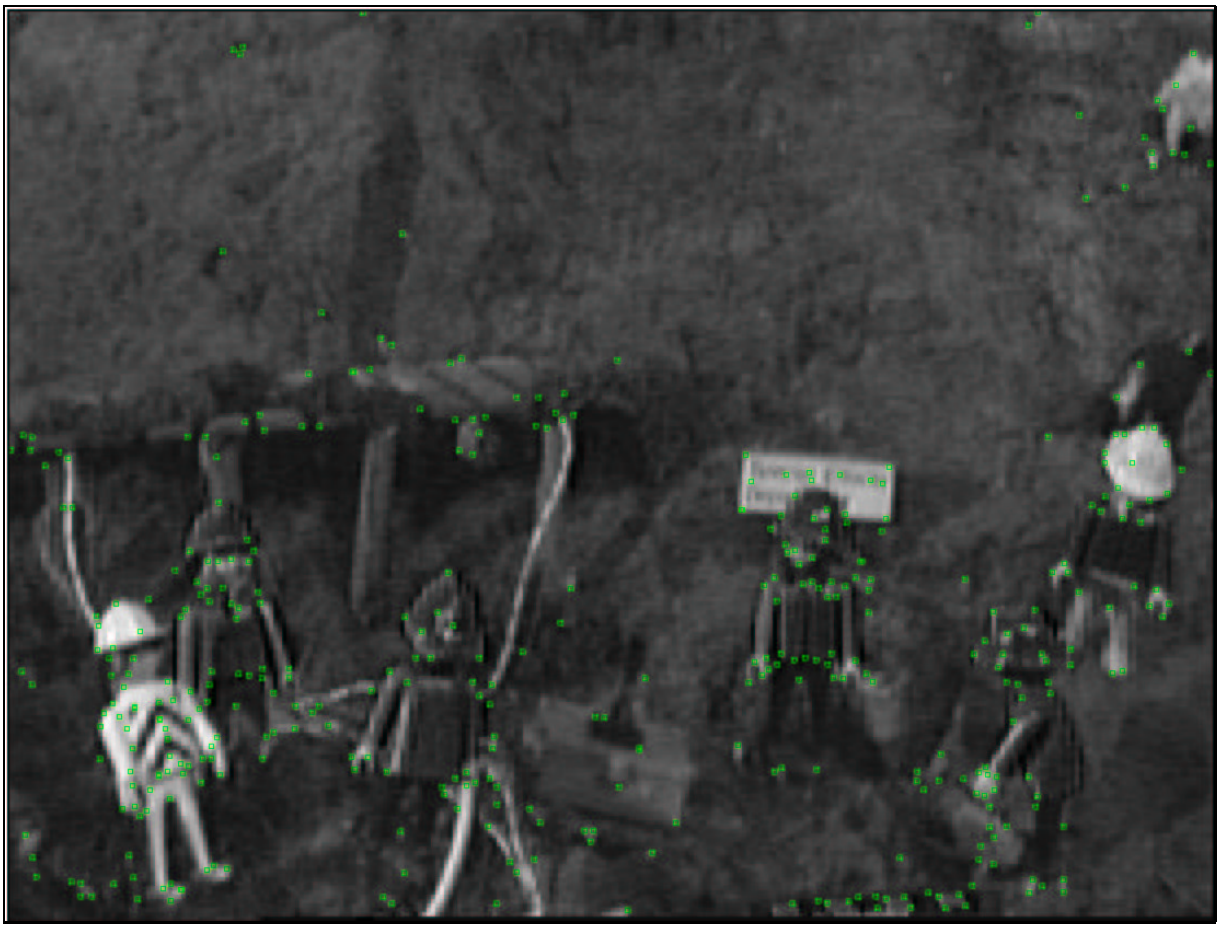


Figure D.18 – *Détection de coins couleur en précision quart de pixel (détail des "personnages").*

Figure D.19 – *Détection de coins couleur en précision quart de pixel (détail des "animaux").*

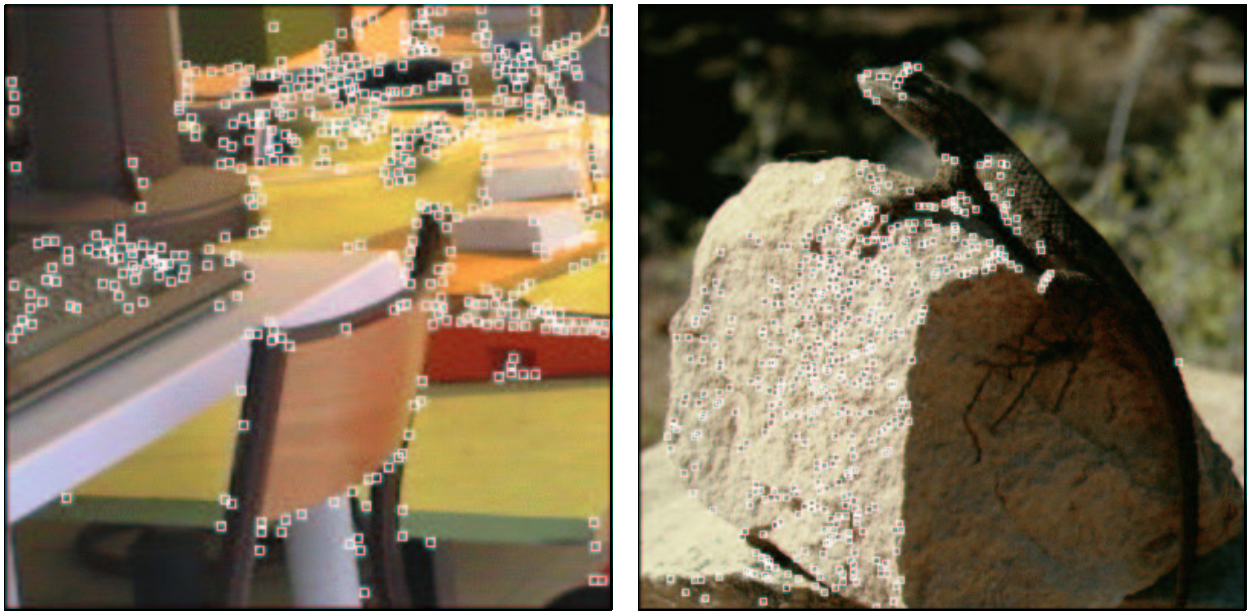


Figure D.20 – *Un exemple de points Harris couleur sur deux images : l'image "Lézard" est caractérisée par beaucoup de texture (472 points détectés) alors que l'image "Salle robotique" présente un grand nombre de contours (493 points détectés).*

Figure D.21 – *Un exemple de coins couleur sur les mêmes images qu'à la figure D.20. 457 coins sont détectés sur "Salle robotique" et 438 sur "Lézard".*

Annexe E

Géométrie projective : propriétés et notations

Dans cette annexe, nous rappelons brièvement les notions et propriétés de base définies et utilisées en géométrie projective. Celles-ci sont nécessaires à la bonne compréhension des travaux présentés dans ce mémoire, notamment aux chapitres 2 et 6 où nous mettons en place des critères définis à partir de l'homographie du plan projectif; et également au chapitre 8 où nous présentons des méthodes de transfert d'images basées sur l'invariance du birapport. Pour plus de détails, le lecteur peut par exemple se référer à [Coelho et al., 1992; Gros et Quan, 1992; Mohr, 1993; Morin, 1993; Laveau, 1996; Zeller, 1996; Gros et al., 1997a].

E.1 Notion d'espace projectif

Soit $(x_1, ..., x_{n+1})$ un point de $\mathcal{R}^{n+1}$. On définit l'espace projectif $\mathcal{P}^n$ comme l'espace quotient de $\mathcal{R}^{n+1} - \{0, ..., 0\}$ par l'équivalence suivante :

$$X = (x_1, ..., x_{n+1}) \equiv Y = (y_1, ..., y_{n+1}) \Leftrightarrow \exists \lambda \neq 0 \ / \ X = \lambda Y \tag{E.1}$$

X et Y représentent le même point de l'espace projectif. Les coordonnées $(x_1, ..., x_{n+1})$ définies à un facteur près sont les coordonnées homogènes du point. L'espace $\mathcal{P}^2$ est appelé *plan projectif* et l'espace $\mathcal{P}^1$ est appelé *droite projective*.

Prolongement canonique de $\mathcal{P}^n$ dans $\mathcal{R}^n$

Un point de $\mathcal{P}^n - \{x_1, ..., x_n, 0\}$ est identifié à un point de $\mathcal{R}^n$ par l'application suivante :

$$(x_1, ..., x_{n+1}) \mapsto (\frac{x_1}{x_{n+1}}, ..., \frac{x_n}{x_{n+1}}) \tag{E.2}$$

E.2 Les transformations projectives

On appelle trnasformation projective toute application linéaire des coordonnées homogènes. Une transformation projective de $\mathcal{P}^n$ dans $\mathcal{P}^m$ s'exprime donc sous la forme matricielle suivante :

$$Y^T = AX^T \tag{E.3}$$

où A est une matrice $(m+1) \times (n+1)$.

Notons que A et λA définissent la même application projective. Comme elle est définie à un facteur multiplicatif près, elle possède $(m+1) \times (n+1) - 1$ degrés de liberté. Toute transformation projective de $\mathcal{P}^n$ dans lui-même est une *homographie*.

E.3 Notion de base projective

Une base projective de $\mathcal{P}^n$ est formée de $n+2$ points tels qu'aucun sous-ensemble de $n+1$ points n'appartient à un hyperplan. Dans le plan projectif, tout quadruplet de points non colinéaires 3 à 3 définit une base projective.

Considérons deux bases projectives B_1 et B_2 de $\mathcal{P}^n$. La donnée de ces deux bases permet de déterminer une homographie de $\mathcal{P}^n$ dans lui-même. En effet, l'homographie recherchée est de dimension $(n+1) \times (n+1)$ donc possède $(n+1)^2 - 1 = n^2 + 2n$ inconnues et les $n+2$ points des deux bases (décrits par n coordonnées) permettent d'avoir $n^2 + 2n$ équations. Dans le plan projectif, la donnée de quatre paires de points définissant deux bases projectives permet de calculer l'homographie du plan.

Exemple de la base projective canonique

Le $(n+2)$-uplet $\{(1,0,...,0)^T, (0,1,0,...,0)^T, ...(0,...,0,1)^T, (1,...,1)^T\}$ représente la base canonique de l'espace projectif $\mathcal{P}^n$. Cela revient à prendre n axes (directions), le point $(0,...,0,1)$ pour marquer l'origine et le point $(1,...,1)$ pour marquer l'unité. On peut toujours se ramener à la base canonique par une homographie de dimension $(n+1)\times(n+1)$.

E.4 Le birapport

Etant donnés quatre points (A,B,C,D) d'une droite projective, le birapport de ces points, noté $[A,B,C,D]$ est défini comme suit :

$$[A,B,C,D] = \frac{\overline{CA}}{\overline{CB}} \times \frac{\overline{DB}}{\overline{DA}} \tag{E.4}$$

avec

$$\frac{\infty}{\infty} = 1 \qquad \frac{a}{\infty} = 0 \qquad \frac{\infty}{a} = \infty \tag{E.5}$$

Le birapport ainsi défini est un invariant projectif. Pour toute transformation projective f, on a en effet :

$$[A,B,C,D] = [f(A), f(B), f(C), f(D)] \tag{E.6}$$

Ceci est un théorème essentiel, c'est cette notion d'invariance qui permet de mettre en place les méthodes de transfert projectif développées au chapitre 8. L'invariance du birapport est illustrée par la figure E.1 pour la droite projective.

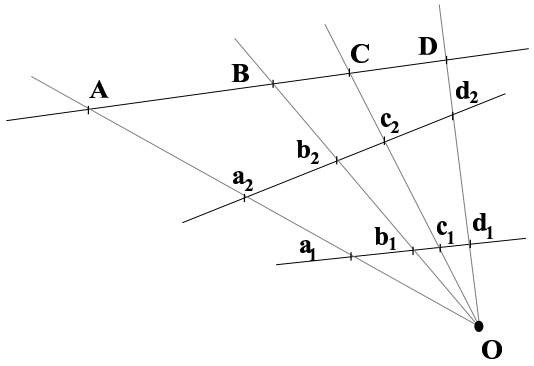


Figure E.1 – *Conservation du birapport sur la droite projective. On a :* $[A, B, C, D] = [a_1, b_1, c_1, d_1] = [a_2, b_2, c_2, d_2]$.

On définit de la même façon le birapport de faisceaux de droites. Ainsi dans le cas de la droite projective de la figure E.1, on a :

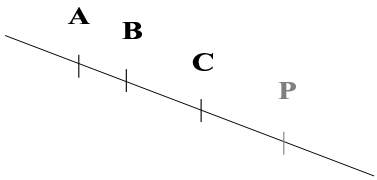
$$[(0A), (0B), (0C), (0D)] = [A, B, C, D] \tag{E.7}$$

Le birapport d'un faisceau de droites est indépendant du choix de la droite qui coupe ce faisceau.

E.5 Les coordonnées projectives

Soit un espace projectif $\mathcal{P}^n$ muni d'une base projective. Le concept de coordonnées projectives permet de localiser tout point de cet espace dans l'espace lui-même. Elles sont définies à l'aide des birapports.

Cas de la droite projective $\mathcal{P}^1$



Soient trois points alignés A, B et C définissant un repère projectif d'une droite (A, B). A tout point P de cette droite, on associe λ_P le birapport de ces points. Cette valeur détermine de manière unique la position de P sur la droite. λ_P est définie telle que :

$$\lambda_P = [A, B, C, P] \tag{E.8}$$

Cas du plan projectif $\mathcal{P}^2$

Soient quatre points (A, B, C, D) qui définissent la base projective du plan projectif. A partir de chacun de ces points, il est possible de définir un faisceau de droites et d'en calculer le birapport, comme l'illustre la figure E.2.

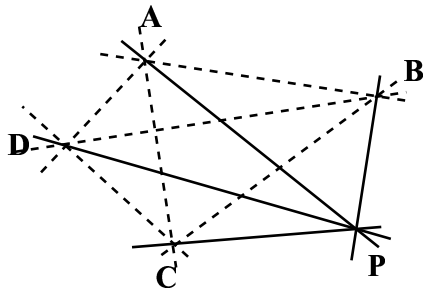


Figure E.2 – *Localisation du point P dans le plan projectif défini par la base projective (A, B, C, D).*

Deux birapports définissent deux droites dont l'intersection donne P. Les coordonnées projectives de P sont donc déterminées par la donnée de deux birapports (k_1, k_2). Par exemple :

$$\begin{cases} k_1 = [(AB), (AC), (AD), (AP)] \\ k_2 = [(BA), (BC), (BD), (BP)] \end{cases} \tag{E.9}$$

En pratique, on considère la donnée de trois birapports. Cela permet de pallier les cas dégénérés. Par exemple, on a $k_1 = \infty$ lorsque le point P appartient à la droite (A, B).

Liste des publications

- **Conférences internationales avec comité de lecture**

  Stereo matching of color images using differential invariants.
  V. Gouet, P. Montesinos, D. Pelé
  International Conference on Image Processing, Chicago, Etats-Unis, Octobre 1998.

  Differential invariants for color images.
  P. Montesinos, V. Gouet, R. Deriche
  International Conference on Pattern Recognition, Brisbane, Australie, Août 1998.

  A fast matching method for color uncalibrated images using differential invariants.
  V. Gouet, P. Montesinos, D. Pelé
  British Machine Vision Conference, Southampton, vol. I, pages 367-376, Septembre 1998.

- **Conférences nationales avec comité de lecture**

  Evaluation de détecteurs de points d'intérêt pour la couleur.
  V. Gouet, P. Montesinos, R. Deriche, D. Pelé
  Reconnaissance des Formes et Intelligence Artificielle, Paris, vol. II, pages 257-266, Février 2000.

- **Revue**

  Matching color uncalibrated images using differential invariants.
  P. Montesinos, V. Gouet, R. Deriche, D. Pelé
  Image and Vision Computing 18(9), pages 659-671, Juin 1999.

- **Tutoriel**

  Differential invariants for color images.
  P. Montesinos, V. Gouet, R. Deriche, D. Pelé
  CVOnline 1999.
  http ://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/MONTESSINOS/cv_online.html

Références

Alquier, L. (1998). *Analyse et représentation de scènes complexes par groupement perceptuel. Application à la perception de structures curvilignes.* Thèse de doctorat, Université Montpellier II, France.

Armande, N. (1997). *Caractérisation de Réseaux Fins dans les Images 2D et 3D. Application aux images satellites et médicales.* Thèse de doctorat, Paris Sud centre d'Orsay, France.

Avidan, S. et Shashua, A. (1998). Novel view synthesis by cascading trilinear tensors. IEEE *Transactions on Visualization and Computer Graphics*, 4(4).

Bauckhage, C. et Schmid, C. (1996). Evaluation of keypoint detectors. Rapport technique, INRIA.

Beardsley, P., Murray, D., et Zisserman, A. (1992). Camera calibration using multiple images. Dans Sandini, G., editor, *Proceedings of Computer Vision (ECCV '92)*, volume 588 of *LNCS*, pages 312–320, Berlin, Germany. Springer.

Beaudet, P. (1978). Rotationally invariant image operators. Dans *Proceedings of 4th International Joint Conference on Pattern Recognition.* Kyoto, Japan.

Beier, T. et Neely, S. (1992). Feature-based image metamorphosis. Dans Catmull, E. E., editor, *Computer Graphics (SIGGRAPH '92 Proceedings)*, volume 26, pages 35–42, Chicago, Illinois, Etats-Unis.

Beyer, H. (1992). Accurate calibration of ccd cameras. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 96–101, Urbana-Champaign, Illinois, Etats-Unis.

Bhat, D. et Nayar, S. (1996). Ordinal measures for visual correspondence. Dans *International Conference on Computer Vision and Pattern Recognition*, pages 351–357, San Francisco, California, Etats-Unis.

Biernacki, C. et Mohr, R. (1999). Indexation et appariement d'images par modèles de mélanges gaussien des couleurs. Rapport technique, INRIA Rhône-Alpes.

Binford, T. et Lewitt, T. (1993). Quasi-invariants : theory and exploitation. Dans *Proceedings of DARPA Image Understanding Workshop*, pages 819–829.

Blanc, J. (1998). *Synthèse de nouvelles vues d'une scène 3D à partir d'images existantes.* Thèse de doctorat, Institut National Polytechnique de Grenoble.

Blaszka, T. (1997). *Approches par modèles en vision précoce.* Thèse de doctorat, Université de Nice Sophia Antipolis.

Bobet, P., Blanc, J., et Mohr, R. (1996). Aspects cachés de la tri-linéarité. Dans *Actes du 10ème Congrès AFCET de Reconnaissance des Formes and Intelligence Artificielle, Rennes, France*, pages 137–146. LIFIA–IMAG–INRIA Rhône-Alpes.

Boissonnat, J. et Teillaud, M. (1986). A hierarchical representation of objects : the delaunay tree. *Second ACM Symposium on Computational Geometry in YorkTown Heights.*

Boufama, B. (1994). *Reconstruction tridimensionnelle en vision par ordinateur : cas des caméra non étalonnées.* Thèse de doctorat, Institut National Polytechnique de Grenoble.

Bracewell, R. N. (1978). *The Fourier Transform and its Applications.* Electrical and Electronic Engineering Series, McGraw-Hill, New York.

Brainard, D., Wandell, B., et Cowan, W. (1989). Black light : how sensors filter spectral variation of the illuminant. IEEE *Transactions on Biomedical Engineering*, 36 :140–149.

Brand, P. (1995). *Reconstruction tridimensionnelle d'une scène à partir d'une caméra en mouvement : de l'influence de la précision.* Thèse de doctorat, Université Claude Bernard, Lyon I.

Brockelbank, D. C. et Yang, Y. H. (1989). An experimental investigation in the use of color in computational stereopsis. IEEE *Transactions on Systems, Man, and Cybernetics*, 19(6) :1365–1383.

Brown, D. (1966). Decentering distorsion of lenses. *Photogrammetric Engineering and Remote Sensing*, pages 444–462.

Burns, J., Weiss, R., et Riseman, E. (1990). View variation of point set and line segment features. Dans *Proceedings of DARPA Image Understanding Workshop*, pages 650–659.

Canny, J. (1983). Finding edges and lines in images. Rapport technique, MIT technical report 720.

Canny, J. (1986). A computational approach to edge detection. *PAMI*, pages 679–698.

Casasent, D. et Psaltis, D. (1980). Hybrid processor to compute invariant moments for pattern recognition. *Optics Lett.*, 5 :395–397.

Castan, S. et Shen, J. (1989). Une famille de détecteurs de contours basée sur le filtre exponentiel optimal. *Congrés AFCET : "Reconnaissance Des Formes Et Intelligence Artificielle".*

Chen, S. E. (1995). Quicktime VR - an image-based approach to virtual environment navigation. Dans Cook, R., editor, *SIGGRAPH 95 Conference Proceedings*, Annual Conference Series, pages 29–38. ACM SIGGRAPH, Addison Wesley. held in Los Angeles, California, 06-11 Août 1995.

Chou, C. et Chen, Y. (1990). Moment preserving pattern matching. *Pattern Recognition*, 23(5) :461–474.

Coelho, C., Heller, A., Mundy, J., Forsyth, D., et Zisserman, A. (1992). An experimental evaluation of projective invariants. Dans *Geometric Invariance in Computer Vision, the MIT Press, Cambridge, MA*, pages 87–104.

Coutance, V., Baron, T., et Briot, M. (1989). Segmentation of color images in robotics. *Reconnaissance Des Formes Et Intelligence Artificielle*, pages 1115–1122.

Crouzil, A., Massip-Pailhes, L., et Castan, S. (1996). A new correlation criterion based on gradient fields similarity. Dans *Proceedings of the International Conference on Pattern Recognition*, pages 632–636.

Cumani, A., Grattoni, P., et Guiducci, A. (1990). An edge-based description of color images. *CVGIP : Graphical Models and Image Processing*, 53(4) :313–323.

Deriche, R. (1987). Optimal edge detection using recursive filtering. Dans *Proceedings First International Conference on Conputer Vision, London.*

Deriche, R. et Giraudon, G. (1991). Acurate corner detection : An analytical study. Rapport technique 1420, INRIA Sophia-Antipolis, France.

Deriche, R. et Giraudon, G. (1993). A computational approach for corner and vertex detection. *International Journal of Computer Vision*, 10(2) :101–124.

Devernay, F. (1997). *Vision stéréoscopique et propriétés différentielles des surfaces.* Thèse de doctorat, Ecole Polytechnique, France.

Devernay, F. et Faugeras, O. (1995). Automatic calibration and removal of distorsion from scenes of structured environment. Dans Rudin, L. et Bramble, S., editors, *Proceedings of* SPIE *Investigative and Trial Image Processing*, volume 2567, San Diego, CA.

Di Zenzo, S. (1986). A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33 :116–125.

Dreschler, L. et Nagel, H. H. (1982). Volumetric model and 3-D trajectory of a moving car derived from monocular TV frame sequences of a street scene. *Computer Graphics and Image Processing*, 20(3) :199–228.

Drew, M. S., Wei, J., et Li, Z.-N. (1998). Illumination-invariant color object recognition via compressed chromaticity histograms of normalized images. Dans *International Conference on Computer Vision*, pages 533–540.

Dufournaud, Y., Schmid, C., et Horaud, R. (2000). Appariement d'images à des échelles différentes. Dans *Actes du 12ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Paris, France*, volume 2, pages 327–336.

Dyer, C. R. (1987). Multiscale image understanding. Dans *Parallel Computer Vision*, pages 171–213.

Ennesser, F. et Medioni, G. (1993). Finding waldo, or focus of attention using local color information. *International Conference on Computer Vision and Pattern Recognition*, pages 711–712.

Ennesser, F. et Medioni, G. (1995). Finding waldo, or focus of attention using local color information. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(8) :805–809.

Falzon, F. (1994). *Analyse multi-échelle, détection des singularités et caractérisation de la régularite des images.* Thèse de doctorat, INRIA Sophia-Antipolis, France.

Faugeras, O. (1995). Stratification of 3-dimensional vision : Projective, affine, and metric representations. *Journal of the Optical Society of America A*, 12(3) :465–484.

Faugeras, O. et Berthod, M. (1981). Improving consistency and reducing ambiguity in stochastic labeling : an optimization approach. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 3(4) :412–424.

Faugeras, O. et Mourrain, B. (1995). On the geometry and algebra of the point and line correspondences between n images. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 951–956, Boston, MA. IEEE Computer Society Press.

Faugeras, O. et Papadopoulo, T. (1998). A nonlinear method for estimating the projective geometry of three views. Dans *International Conference on Computer Vision*, pages 477–484.

Faugeras, O. et Robert, L. (1994). What can two images tell us about a third one ? *Lecture Notes in Computer Science*, 800 :485–494.

Faugeras, O. D. (1992). What can be seen in three dimensions with an uncalibrated stereo rig. Dans Sandini, G., editor, *Proceedings of the 2nd European Conference on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 563–578, Santa Margherita Ligure, Italy. Springer-Verlag.

Faugeras, O. D., Luong, Q.-T., et Maibank, S. J. (1992). Camera self–calibration : Theory and experiments. Dans Sandini, G., editor, *Proceedings of the 2nd European Conference on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 321–334, Santa Margherita Ligure, Italy. Springer-Verlag.

Faugeras, O. D., Lustman, F., et Toscani, G. (1987). Motion and structure from motion from point and line matches. Dans *First International Conference on Computer Vision*, pages 25–34, London, England. IEEE Computer Society Press.

Fellah, S. (1994). Détection de contours sur des images couleur. Rapport technique, Rapport de DEA CASIAM, Université de Montpellier II, France.

Fidrich, M. et Thirion, J. (1998). Stability of corner points in scale space : The effects of small nonrigid deformations. *Proceedings of Computer Vision and Image Understanding*, 72(1) :72–83.

Finlayson, G. (1995). *Coefficient Color Constancy*. Thèse de doctorat, School of Computing Science, Université Simon Fraser, Vancouver, Canada.

Finlayson, G., Drew, M., et Funt, B. (1994). Color constancy : Generalized diagonal transforms suffice. *Journal of the Optical Society of America A*, 11(11) :3011–3019.

Finlayson, G. D., Chatterjee, S. S., et Funt, B. V. (1996). Color angular indexing. Dans *Proceedings of the European Conference on Computer Vision*, volume II, pages 16–27.

Fleet, D. J., Jepson, A. D., et Jenkin, M. R. M. (1991). Phase-based disparity measurement. *Computer Vision, Graphics, and Image Processing. Image Understanding*, 53(2) :198–210.

Florack, L., ter Haar Romeny, B., Koenderink, J., et Viergever, M. (1994). General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4(2) :171–187.

Flusser, J., Saic, S., et Suk, T. (1994). Registration of images with affine geometric distortion by means of moment invariants. Dans *Image and Signal Processing for Remote Sensing*, pages 843–852, Bellingham, WA, Etats-Unis. SPIE Optical Engineering Press.

Forstner, W. (1994). A framework for low level feature extraction. Dans *Proceedings of the 3rd European Conference on Computer Vision*.

Forsyth, D. (1990). A novel algorithm for color constancy. *International Journal of Computer Vision*, 5 :5–36.

Freeman, M. et Saleh, B. (1988). Moment invariants in the space and frequency domains. *Journal of the Optical Society of America A*, 5(7) :1073–1084.

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

Förstner, W. et Gülch, E. (1987). A fast operator for detection and precise location of distinct points, corners and centers of circular features. Dans *Proceedings of the Intercommision Workshop of the International Society for Photogrammetry and Remote Sensing*, Interlaken, Suisse.

Fua, P. (1991). Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities. Dans Myopoulos, John ; Reiter, R., editor, *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 1292–1298, Sydney, Australia. Morgan Kaufmann.

Funt, B. et Finlayson, G. (1995). Color constant color indexing. IEEE *Transactions on Pattern Analysis and Machine Intelligence 17*, pages 522–529.

Gabor, D. (1946). Theory of communication. *Journal of the Inst. Elec. Eng.*, 93(26) :429–457.

Golub, G. H. et Loan, C. F. V. (1989). *Matrix Computations*, volume 3 of *Johns Hopkins Series in the Mathematical Sciences*. The Johns Hopkins University Press, Baltimore, MD, Etats-Unis, second edition.

Gouet, V., Montesinos, P., et Pelé, D. (1998a). A fast matching method for color uncalibrated images using differential invariants. Dans *Proceedings of the British Machine Vision Conference*, volume I, pages 367–376, Southampton, UK.

Gouet, V., Montesinos, P., et Pelé, D. (1998b). Stereo matching of color images using differential invariants. Dans *Proceedings of the* IEEE *International Conference on Image Processing*, Chicago, Etats-Unis.

Gouet, V., Montesinos, P., Reriche, R., et Pelé, D. (2000). Evaluation de détecteurs de points d'intérêt pour la couleur. Dans *Reconnaissance des formes et Intelligence Artificielle (RFIA'2000)*, volume II, pages 257–266, Paris, France.

Grace, A. E. et Spann, M. (1991). A comparison between fourier-mellin descriptors and moment based features for invariant object recognition using neural networks. *Pattern Recognition Letters*, 12(10) :635–643.

Gros, P., Hartley, R., Mohr, R., et Quan, L. (1997a). How useful is projective geometry ? *Computer Vision and Image Understanding*, 65(3) :442–446.

Gros, P., Mclean, G., Delon, R., Mohr, R., Schmid, C., et Mistler, G. (1997b). Utilisation de la couleur pour l'appariement et l'indexation d'images. Rapport technique 3269, INRIA.

Gros, P. et Quan, L. (1992). Projective invariants for vision. Rapport technique RT 90 IMAG - 15 LIFIA, LIFIA–IRIMAG, Grenoble, France.

Gros, P. et Quan, L. (1993). 3D projective invariants from two images. Dans *Proceedings of the* SPIE *Conference on Geometric Methods in Computer Vision II, San Diego, California, Etats-Unis*, pages 75–86.

Grossmann, A. et Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4) :723–736.

Hancock, J., Baddeley, R., et Smith, L. (1992). The principal components of natural images. *Network 3*, pages 61–70.

Hannah, M. J. (1989). A system for digital stereo image matching. *Photogrammetric Engineering and Remote Sensing*, 55(12) :1765–1770.

Harris, C. et Stephens, M. (1988). A combined corner and edge detector. *Proceedings of the* 4^{th} *Alvey Vision Conference*, pages 147–151.

Hartley, R. (1992). Calibration of cameras using the essential matrix. Dans *Proceedings of* ARPA *Image Understanding Workshop*, pages 911–915. Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc.

Hartley, R. (1994). Lines and points in three views - an integrated approach. Dans *Proceedings of* ARPA *Image Understanding Workshop*. Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc.

Hartley, R. (1995). In defence of the 8-point algorithm. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 1064–1070. IEEE Computer Society Press, Boston, MA.

Hartley, R. (1997). In defense of the eight-point algorithm. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(6) :580–593.

Hartley, R. et Sturm, P. (1994). Triangulation. Dans *Proceedings of* ARPA *Image Understanding Workshop*, pages 957–966, Monterey, CA, Etats-Unis. Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc.

Hartley, R. et Zisserman, A. (2000). *Multiple View Geometry*. Cambridge University Press, Cambridge, UK.

Healey, G. et Slater, D. (1994). Global color constancy : Recognition of objects by use of illumination-invariant properties of color distributions. *Journal of the Optical Society of America A*, 11(11) :3003–3010.

Healey, G. et Slater, D. (1997). Computing illumination-invariant descriptors of spatially filtered color image regions. IEEE *Transactions on Image Processing*, 6(7) :1002–1013.

Heyden, A. et Astrom, K. (1996). Algebraic varieties in multiple view geometry. Dans *Proccedings oh the 4th European Conference on Computer Vision*, volume II, pages 671–682.

Heyden, A. et Rohr, K. (1996). Evaluation of corner extraction schemes using invariance methods. Dans *Proceedings of the* 13^{th} *Internat. Conf. on Pattern Recognition, Vienna, Austria*, volume I, pages 895–899.

Hilbert, D. (1890). *Theory of Algebraic Invariants*. Cambridge Mathematica Library, Cambridge University Press.

Horaud, R., Christy, S., et Dornaika, F. (1995). Object pose : the link between weak perspective, para perspective, and full perspective. Dans *Proceedings of the 5th International Conference on Computer Vision*, Cambridge, Massachusetts, Etats-Unis.

Horaud, R. et Monga, O. (1993). *Vision par ordinateur : outils fondamentaux*. Éditions Hermès, Paris.

Hu, M. K. (1962). Visual pattern recognition by moment invariants. IEEE *Trans. Infor. Theory*, 8 :179–187.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W.-J., et Zabih, R. (1999). Spatial color indexing and applications. *International Journal of Computer Vision*, 35(3) :245–268.

Huttenlocher, D., Klanderman, D., et Rucklige, A. (1993). Comparing images using the Hausdorff distance. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 15(9) :850–863.

Huttenlocher, D. et Rucklidge, W. (1993). A multi-resolution technique for comparing images using the hausdorff distance. *International Conference on Computer Vision and Pattern Recognition*, pages 705–706.

Jones, G. (1997). Constraint, optimization, and hierarchy : Reviewing stereoscopic correspondence of complex features. *Computer Vision and Image Understanding*, 65(1) :57–78.

Jordan, J. et Bovik, A. (1988). Computational stereo vision using color. IEEE *Control systems magazine*, pages 31–36.

Kankanhalli, M., Mehtre, B., et Wu, J. (1996). Cluster based colour matching for image retrieval. *Pattern Recognition*, 29(4) :701–708.

Kim, W.-Y. et Yuan, P. (1994). A practical pattern RecognitiQn system for translation, scale, and rotation invariance. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 391–396, Los Alamitos, CA, Etats-Unis. IEEE Computer Society Press.

Kitchen, L. et Rosenfeld, A. (1982). Gray-Level corner detection. *Pattern Recognition Letters*, pages 95–102.

Koenderink, J. (1984). The structure of images. *Biological Cybernetics*, 50 :363–396.

Koenderink, J. et Van Doorn, A. (1987). Representation of local geometry int the visual system. *Biological Cybernetics*, 55 :367–375.

Kondepudy, R. et Healey, G. (1994). Use of invariants for recognition of three-dimensional color textures. *Journal of the Optical Society of America A*, 11 :3037–3049.

Kruppa, E. (1913). Zur Ermittlung eines Objecktes aus zwei Perspektiven mit innerer Orientierung. *Sitz.-Ber. Akad. Wiss., Wien, Math. Naturw., Kl. Abt. IIa*, 122 :1939–1948.

Kuglin, C. D. et Hines, D. C. (1975). The phase correlation image alignment method. Dans IEEE *International Conference on Cybernetics and Society.*, pages 163–165.

Lan, Z. et Mohr, R. (1997a). Non-parametric invariants and application to matching. Rapport technique 3246, INRIA Sophia-Antipolis, France.

Lan, Z. et Mohr, R. (1997b). Robust location based partial correlation. Dans *International Conference on Computer Analysis of Images and Patterns*, pages 313–320.

Land, E. H. (1977). The retinex theory of color vision. *Scientific American*, 237(6) :108–120, 122–123, 126, 128.

Laveau, S. (1996). *Géométrie d'un système de N caméras. Théorie, estimation et applications.* Thèse de doctorat, Ecole Polytechnique, France.

Laveau, S. et Faugeras, O. (1994). 3-d scene representation as a collection of images. Dans *Proceedings of International Conference on Pattern Recognition*, pages 689–691.

Lee, C.-H. (1991). Time-varying images : The effect of finite resolution on uniqueness. *Computer Vision, Graphics, and Image Processing. Image Understanding*, 54(3) :325–332.

Lemestre, G. (1997). *Analyse de séquences d'images pour la construction de vues intermédiaires*. Thèse de doctorat, Université de Rennes I, CCETT-DIH/ATI Rennes, France.

Lew, M. S. et Huang, T. S. (1999). Optimal multi-scale matching. Dans *International Conference on Computer Vision and Pattern Recognition*, pages 88–93.

Li, J. et Schenk, T. (1990). An accurate camera calibration for the aerial image analysis. Dans *Proceedings of* 10^{th} *International Conference on Pattern Recognition*, pages 207–209, Atlantic City, New Jersey, Etats-Unis.

Li, Z.-N., Zaiane, O. R., et Tauber, Z. (1999). Illumination invariance and object model in content-based image and video retrieval. *Journal of Visual Communication and Image Representation*, 10 :219–244.

Lindeberg, T. (1993). On scale selection for differential operators. Dans *Proceedings of* 8^{th} *Scandinavian Conference on Image Analysis, Tromso, Norway*, pages 857–866.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, 293 :133–135.

Lotti, J. L. et Giraudon, G. (1993). Adaptive window algorithm for aerial image stereo. Technical Report 2121, INRIA Sophia-Antipolis, France.

Lourakis, M. et Deriche, R. (2000). Camera self-calibration using the singular value decomposition of the fundamental matrix. Dans *Asian Conference on Computer Vision*, volume I, pages 403–408.

Luong, Q. (1992). *Matrice fondamentale and calibration visuelle sur l'environnement, vers une plus grande autonomie des systèmes robotiques.* Thèse de doctorat, Université de Paris-Sud centre d'Orsay, France.

Luong, Q. T. (1991). La couleur en vision par ordinateur : Une revue. *Revue Traitement du Signal*, 8(1) :3–34.

Luong, Q.-T., Deriche, R., Faugeras, O., et Papadopoulo, T. (1993). On determining the fundamental matrix : analysis of different methods and experimental results. Dans *Israelian Conf. on Artificial Intelligence and Computer Vision*, Tel-Aviv, Israel. A longer version is INRIA Research Report RR-1894.

Luong, Q.-T. et Faugeras, O. (1995). The fundamental matrix : theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1) :43–76.

Luong, Q.-T. et Vieville, T. (1994). Canonic representations for the geometries of multiple projective views. Dans Eklundh, J.-O., editor, *Proceedings of the 3rd European Conference on Computer Vision, volume 800 of Lecture Notes in Computer Science*, volume 1, pages 589–599, Stockholm, Sweden.

Machuca, R. et Phillips, K. (1983). Application of vector fields to image processing. IEEE *Transaction on Pattern Analysis and Machine Intelligence*, 5(3).

Mallat, S. (1989). A theory for multiresolution signal decomposition : The wavelet representation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(7) :674–693.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

Maloney, L. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A*, 3 :1673–1683.

Marr, D. (1982). *Vision*. W. H. Freeman and Company, San Francisco, 1982.

Marr, D. et Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, 194(4262) :283–287.

Mehtre, B., Kankanhalli, M., et Narasimhalu, A. (1995). Color matching for image retrieval. *Pattern Recognition Letters*, 16 :325–331.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge Studies in Advanced Mathematics 37. Cambridge University Press. Translated to English by D. H. Salinger.

Mindru, F., Moons, T., et Gool, L. V. (1999). Recognizing color patterns irrespective of viewpoint and illumination. Dans *Proceedings* IEEE *Conference on Computer Vision and Pattern Recognition*, volume 1, pages 368–373, Fort Collins, Colorado, Etats-Unis.

Mohr, R. (1993). Projective geometry and computer vision. Dans Chen, C., L.F.Pau, et Wang, S., editors, *Handbook of Pattern Recognition and Computer Vision*. World Scientific Pub.

Montesinos, P. et Dattenny, S. (1997). Sub-pixel accuracy using recursive filtering. *Proceedings of The* 10^{th} *Scandinavian Conference on Image Analysis*, 1(10).

Montesinos, P., Gouet, V., et Deriche, R. (1998). Differential Invariants for Color Images. Dans *Proceedings of* 14^{th} *International Conference on Pattern Recognition*, Brisbane, Australia.

Montesinos, P., Gouet, V., Deriche, R., et Pelé, D. (2000). Matching color uncalibrated images using differential invariants. *Image and Vision Computing*, 18(9) :659–672.

Moravec, H. (1977). Towards automatic visual obstacle avoidance. Dans *Proceedings of the* 5^{th} *International Joint Conference on Artificial Intelligence*, page 584, Cambridge, Massachusetts, Etats-Unis.

Morin, L. (1993). *Quelques contributions des invariants projectifs à la vision par ordinateur*. Thèse de doctorat, Institut National Polytechnique de Grenoble.

Mundy, J. et Zisserman, A. (1992). *Geometric Invariance in Computer Vision*. The MIT Press, Cambridge, MA, Etats-Unis.

Murase, H. et Nayar, S. (1995). Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, 14(1) :5–24.

Nagel, H.-H. (1983). Displacement vectors derived from second-order intensity variations in image sequences. *Computer Vision, Graphics, and Image Processing*, 21(1) :85–117.

Nagel, H.-H. (1987). On the estimation of optical flow : Relations between different approaches and some new results. *Artificial Intelligence*, 33(3) :299–324.

Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D., Yanker, P., Faloutsos, C., et Taubin, G. (1993). The QBIC project : querying images by content using color, texture and shape. Research Report, IBM Research Division, Almaden Research Center, San Jose, California.

Nishihara, K. (1983). Prism : a practical real-time imaging stereo matcher. *Proceedings of* 3^{st} *International Conference on Robot Vision and Sensory Controls.*

Noble, J. A. (1988). Finding corners. *Image and Vision Computing*, 6 :121–128.

Olsen, S. I. (1992). Epipolar line estimation. Dans Sandini, G., editor, *Proceedings of Computer Vision (ECCV '92)*, volume 588 of *LNCS*, pages 307–311, Berlin, Germany. Springer.

Ouali, M., Lange, H., et Laurgeau, C. (1996). An energy minimization approach to dense stereovision. Dans *Proceedings of the* IEEE *International Conference on Image Processing*, pages 841–846, Chicago, Etats-Unis.

Papadopoulo, T. et Faugeras, O. (1998). A new characterization of the trifocal tensor. Dans *European Conference on Computer Vision.*

Pass, G. et Zabih, R. (1996). Histogram refinement for content-based image retrieval. Dans IEEE *Workshop on Applications of Computer Vision*, pages 96–102.

Perez, F. et Koch, C. (1994). Towards color image segmentation in analog vlsi : Algorithm and hardware. *International Journal of Computer Vision*, 12(1) :17–42.

Poelman, C. J. et Kanade, T. (1994). A paraperspective factorization method for shape and motion recovery. *Proceedings of the 3rd European Conference on Computer Vision*, 800 :97–110.

Pollard, S., Maihew, J., et Frisby, J. (1985). PMF : A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14 :449–470.

Pollefeys, M. et VanGool, L. (1997). A stratified approach to metric self-calibration. Dans *Computer Vision and Pattern Recognition*, pages 407–412.

Prokop, R. J. et Reeves, A. P. (1992). A survey of moment-based techniques for unoccluded object representation and recognition. *Computer Vision, Graphics, and Image Processing. Graphical Models and Image Processing*, 54(5) :438–460.

Quan, L. et Mohr, R. (1995). Projective reconstruction from multiple uncalibrated images. Dans Bunke, H., Kanade, T., et Noltemeier, H., editors, *Modelling and Planning for Sensor Based Intelligent Robot Systems*, volume 21, pages 236–256. World Scientific Pub.

Rao, R. et Ballard, D. (1995). Object indexing using an Iconic Sparse Distributed Memory. *Proceedings of* 5^{th} *International Conference on Computer Vision*, pages 24–31.

Reddy, B. S. et Chatterji, B. (1996). An fft-based technique for translation, rotation, and scale-invariant image registration. IEEE *Transactions on Image Processing*, 5(8) :1266–1271.

Reiss, T. H. (1993). Recognizing planar objects using invariant image features. *Lecture Notes in Computer Science*, 676 :1–171.

Robert, L. et Deriche, R. (1996). Dense depth map reconstruction : a minimization and regularization approach which preserves discontinuities. Dans *Proceedings of the 4th European Conference on Computer Vision*, pages 439–451, Cambridge, England.

Rohr, K. (1992). Modelling and identification of characteristic intensity variations. *Image and Vision Computing*, 10 :66–76.

Rothwell, C., Faugeras, O., et Csurka, G. (1997). A comparison of projective reconstruction methods for pairs of views. *Computer Vision and Image Understanding*, 68(1) :37–58.

Rucklidge, W. (1997). Efficiently locating objects using the hausdorff distance. *International Journal of Computer Vision*, 24(3) :251–270.

Salden, A., ter Haar Romeny, B., Florack, L., Viergever, M., et Koenderink, J. (1992). A complete and irreductible set of local orthogonally invariant features of 2-dimensional images. *Proceedings of* 11^{th} *International Conference on Pattern Recognition*, pages 180–184.

Sanger, T. (1988). Stereo disparity computation using gabor filters. *Biological Cybernetics*, 59 :405–418.

Sapiro, G. et Ringach, D. L. (1996). Anisotropic diffusion of multivalued images with application to color filtering. IEEE *Transaction on Image Processing*, 9(11) :1582–1586.

Schettini, R. (1993). A segmentation algorithm for color images. *Pattern Recognition Letters*, 14 :499–506.

Schiele, B. (1997). *Reconnaissance d'objets utilisant des histogrammes multidimensionnels de champs réceptifs*. Thèse de doctorat, Institut National Polytechnique de Grenoble, GRAVIR – IMAG.

Schmid, C. (1996). *Appariement d'images par invariants locaux de niveaux de gris*. Thèse de doctorat, Institut National Polytechnique de Grenoble, GRAVIR – IMAG – INRIA Rhône–Alpes.

Schmid, C. et Mohr, R. (1996). Mise en correspondance par invariants locaux. *Traitement du Signal*, 13(6) :591–606.

Schmid, C., Mohr, R., et Bauckhage, C. (1998). Comparing and evaluating interest points. Dans *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*. IEEE Computer Society Press.

Schmid, C., Mohr, R., et Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2) :151–172.

Schulz-Mirbach, H. (1994). Constructing invariant features by averaging techniques. Dans *Proceedings of the 12th International Conference on Pattern Recognition*, volume II, pages 387–390, Jerusalem, Israel.

Seitz, S. et Dyer, C. (1997). View morphing : Uniquely predicting scene appearance from basis images. Dans *DARPA97*, pages 881–888.

Sethi, I. et Patel, N. (1995). A statistical approach to scene change detection. SPIE *Conference on Storage and Retrieval for Image and Video Databases V*, 2420 :381–392.

Shah, M. A. et Jain, R. (1984). Detecting time-varying corners. *Computer Vision, Graphics, and Image Processing*, 28(3) :345–355.

Shashua, A. (1994). Trilinearity in visual recognition by alignment. Dans *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Suisse*, pages 479–484. Springer Verlag.

Shashua, A. (1995). Algebraic functions for recognition. IEEE *Trans. on Pattern Analysis and Machine Intelligence*, 17(8) :779–789.

Shekarforoush, H., Berthod, M., et Zerubia, J. (1996). Subpixel image registration by estimating the polyphase decomposition of cross power spectrum. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, Etats-Unis*, pages 532–537.

Sheng, Y. et Lejeune, C. (1991). Invariant pattern recognition using fourier-mellin transforms and neural networks. *Journal of Optics*, 22(5) :223–228.

Siggelkow, S. et Schael, M. (1999). Fast estimation of invariant features. Dans Förstner, W., Buhmann, J., Faber, A., et Faber, P., editors, *Mustererkennung, DAGM 1999, Informatik aktuell*, Bonn. Springer.

Smith, J. et Chang, S. (1996). Tools and techniques for color image retrieval. *SPIE Proc.*, 2670 :1630–1639.

Spetsakis, M. E. et Aloimonos, J. (1988). Optimal computing of structure from motion using point correspondences in two frames. Dans *Second International Conference on Computer Vision*, pages 449–453, Tampa, FL. Computer Society Press.

Stricker, M. et Orengo, M. (1995). Similarity of color images. Dans *Storage and Retrieval for Image and Video Databases III*, volume 2420 of *SPIE Proceedings Series*, pages 381–392.

Stricker, M. et Swain, M. (1994). The capacity and the sensitivity of color histogram indexing. Rapport technique TR-94-05, Department of Computer Science, University of Chicago.

Swain, M. (1993). Interactive indexing into image databases. *Storage and Retrieval for Image and Video Databases, San Jose, Etats-Unis*, pages 95–103.

Swain, M. et Ballard, D. (1991). Color indexing. *International Journal of Computer Vision*, 7(1) :11–32.

Taubin, G. et Cooper, D. B. (1992). *Object Recognition Based on Moment (or Algebraic) Invariants*, chapter 19. MIT Press, Cambridge, Massachusets.

Teague, M. (1980). Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70(8) :920–930.

Teh, C.-H. et Chin, R. T. (1988). On image analysis by the methods of moments. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10(4) :496–513.

ter Haar Romeny, B. (1996). Introduction to scale-space theory : multiscale geometric image analysis. Rapport technique ICU-96-21, Utrecht University.

Thai, B. et Healey, G. (1998). Modeling and classifying symmetries using a multiscale opponent color representation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 20(11) :1224–1235.

Torr, P. et Murray, D. (1997). The development and Comparison of Robust Methods for Estimating the Fundamental Matrix. *International Journal of Computer Vision*, 24(3) :271–300.

Torr, P., Zisserman, A., et Maibank, S. (1995). Robust detection of degenerate configurations for the fundamental matrix. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 1037–1042, Boston, MA, Etats-Unis. IEEE Computer Society Press.

Triggs, B. (1994). The geometry of projective reconstruction I : Matching constraints and the joint image. Rapport technique, LIFIA-IMAG-INRIA Rhône-Alpes.

Tsai, R. Y. (1986). An efficient and accurate camera calibration technique for 3D machine vision. Dans *Proceedings, CVPR '86 (*IEEE *Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Publ.86CH2290-5, pages 364–374, Miami Beach, FL, Etats-Unis.

Tsai, R. Y. et Huang, T. S. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 6 :13–27.

Van Gool, L., Moons, T., Pauwels, E., et Oosterlinck, A. (1995). Vision and lie's approach to invariance. *Image and Vision Computing Journal*, 13(4).

Van Gool, L., Moons, T., et Ungureanu, D. (1996). Affine/photometric invariants for planar intensity patterns. Dans *Proceedings of the 4th European Conference on Computer Vision*, pages 642–651, Cambridge, England.

Vieville, T., Zeller, C., et Robert, L. (1996). Using collineations to compute motion and structure in an uncalibrated image sequence. *International Journal on Computer Vision*, 20(3) :213–242.

Wang, H. et Brady, J. (1992). Corner detection with sub-pixel accuracy. Rapport technique OUEL 1925/92, Dept. Engineering Science, Oxford University.

Wang, L. et Healey, G. (1998). Using zernike moments for the illumination and geometry invariant classification of multispectral texture. IEEE *Transactions on Image Processing*, 7(2) :196–203.

Weiss, I. (1993). Review : Geometric invariants and object recognition. *International Journal of Computer Vision*, 10(3) :207–231.

Weng, J., Huang, T. H., et Ahuja, N. (1989). Motion and structure from two perspective views : Algorithms, error analysis, and error estimation. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(5) :451–476.

Werner, T., Hersch, R., et Hlavac, V. (1995). Rendering real-world objects using view interpolation. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 957–962, Cambridge, Massachussets, Etats-Unis.

Witkin, A. P. (1983). Scale-space filtering. Dans Bundy, A., editor, *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 1019–1022, Karlsruhe, West Germany. William Kaufmann.

Wolff, L. (1994). On the relative brightness of specular and diffuse reflection. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 369–376, Los Alamitos, CA, Etats-Unis. IEEE Computer Society Press.

Wolfson, H. J. et Rigoutsos, I. (1997). Geometric hashing : An overview. IEEE *Computational Science & Engineering*, 4(4) :10–21.

Wood, J. (1996). Invariant pattern recognition : a review. *Pattern Recognition*, 29(1) :1–17.

Wu, X. et Bhanu, B. (1995). Gabor wavelets for 3d object recognition. Dans *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachussets, Etats-Unis*, pages 537–542.

Zabih, R. et Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. Dans *3rd European Conference on Computer Vision*, pages 151–158, Stockholm, Sweden.

Zeller, C. (1996). *Calibration projective, affine et euclidienne en Vision par Ordinateur et application à la perception tridimensionnelle*. Thèse de doctorat, Ecole Polytechnique, France.

Zhang, H., Gong, Y., Low, C. Y., et Smoliar, S. (1995a). Image retrieval based on color features : an evaluation study. Dans *SPIE Conference on Storage and Retrieval for Image and Video Databases*, volume 2606, pages 381–392. SPIE-Int. Soc. Opt. Eng.

Zhang, Z. (1993). Le problème de la mise en correspondance : l'état de l'art. Rapport technique 2146, INRIA.

Zhang, Z. (1996). Determining the epipolar geometry and its uncertainty : A review. Rapport technique 2927, INRIA, Sophia-Antipolis, France.

Zhang, Z., Deriche, R., Faugeras, O., et Luong, Q.-T. (1995b). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78 :87–119.

Zhu, S. C. et Yuille, A. (1996). Region competition : Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. IEEE *Transaction on Pattern Analysis and Machine Intelligence*, 18(9).

Zisserman, A., Forsyth, D., Mundy, J., Rothwell, C., Liu, J., et Pillow, N. (1995). 3d object recognition using invariance. *Artificial Intelligence*, 78(1-2) :239–288.

Zuniga, O. et Haralick, B. (1983). Corner detection using the facet model. Dans *Proceedings* IEEE *Conference on Computer Vision and Pattern Recognition*, pages 30–37.