Une nouvelle méthode d'appariement de points d'intérêt pour la mise en correspondance d'images

Jean-Louis Palomares

Université de Monpellier II

Ecole doctorale I2S

Thèse

présentée pour obtenir le titre de DOCTEUR en
INFORMATIQUE DE L'UNIVERSITÉ DE MONPELLIER II

Une nouvelle méthode d'appariement de points d'intérêt pour la mise en correspondance d'images

Jean-Louis Palomares

Soutenue publiquement le 22/05/2012 devant un jury composé de :

Rapporteurs	Pierre Bonton	Université Blaise Pascal, Clermont-Ferrand.
	Ludovic Macaire	Université de Lille I, Villeneuve d'Ascq.
Examinateurs	René Zapata	Université de Montpellier II.
	Olivier Strauss	Université de Montpellier II.
	Dro Désiré Sidibe	Université de Bourgogne.
Directeur de thèse	Philippe Montesinos	Ecole des Mines d'Alès, LGI2P Nîmes.

Thèse réalisée au Laboratoire LGI2P de Nîmes
Ecole des Mines d'Alès, Site EERIE
Parc Georges Besse
30000 NIMES cedex 01

Tél : +33 (0) 4 66 38 70 88
Fax : +33 (0) 4 66 38 70 74

Web : http ://www.lgi2p.ema.fr

Sous la direction de Philippe Montesinos Philippe.Montesinos@mines-ales.fr

Adresses mails Palomares jean-louis palomares.jean-louis@voila.fr
jldenimes@gmail.com
Financement Allocation de recherche de l'école des mines d'Alès

A mon fils, Bastien.

Résumé

Ce mémoire de thèse traite de la mise en correspondance d'images pour des applications de vision stéréoscopique ou de stabilisation d'images de caméras vidéo. Les méthodes de mise en correspondance reposent généralement sur l'utilisation de points d'intérêt dans les images, c'est-à-dire de points qui présentent de fortes discontinuités d'intensité lumineuse. Nous présentons tout d'abord un nouveau descripteur de points d'intérêt, obtenu au moyen d'un filtre anisotropique rotatif qui délivre en chaque point d'intérêt une signature mono-dimensionnelle basée sur un gradient d'intensité. Invariant à la rotation par construction, ce descripteur possède de très bonnes propriétés de robustesse et de discrimination. Nous proposons ensuite une nouvelle méthode d'appariement invariante aux transformations euclidiennes et affines. Cette méthode exploite la corrélation des signatures sous l'hypothèse de faibles déformations, et définit une mesure de distance, nécessaire à l'appariement de points. Les résultats obtenus sur des images difficiles laissent envisager des prolongements prometteurs à cette méthode.

Mots clés : stéréovision, stabilisation, points d'intérêt, descripteur, mise en correspondance, appariement.

Abstract

This thesis addresses the issue of image matching for stereoscopic vision applications and image stabilization of video cameras. Methods of mapping are generally based on the use of interest points in the images, i.e. of points which have strong discontinuities in light intensity. We first present a new descriptor of points of interest, obtained by means of an anisotropic rotary filter which delivers at each point of interest a one-dimensional signature based on an intensity gradient. Invariant to rotation by construction, this descriptor has very good properties of robustness and discrimination. We then propose a new matching method invariant to Euclidean and affine transformations. This method exploits the correlation of signatures subject to moderate warping, and defines a distance measure, necessary for the matching of points. The results obtained on difficult images augur promising extensions to this method.

Keywords : stereoscopic vision, camera stabilization, interest point, feature descriptor, mapping, matching.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Introduction

Cette thèse à été initiée par l'entrepise XAP, et s'est déroulée dans le cadre d'un partenariat avec le laboratoire LGI2P de l'école des mines d'Alès. L'entreprise XAP développe des systèmes d'acquisition vidéo embarqués et plus généralement des équipements électroniques, pour la course automobile. La problématique initiale était la stabilisation de vidéos fortement perturbées, cependant en cours d'étude le sujet a pris une orientation vers la vision stéréoscopique et la 3D. Dans ce contexte nous nous sommes rapidement tournés vers la mise en correspondance d'images, qui apparait comme le principal verrou scientifique et technique des applications envisagées.

Enjeux

Dans ce travail, l'enjeu scientifique est clairement d'obtenir une méthode de mise en correspondance robuste et rapide, pour laquelle l'hypothèse de petits déplacement des structures contenues dans les images n'est pas valide. En stéréo, le déplacement des structures peut être important, en vidéo embarquée en course automobile, la vitesse du véhicule entraine de fortes modifications entre les images successives d'une séquence vidéo. De plus, en vidéo, le flux d'image nous impose un rythme de traitement qui doit rester compatible avec le temps réel.

La mise en correspondance d'images est une étape fondamentale dans le domaine de l'analyse d'images. Parmi les nombreuses applications qui utilisent cette technique, on peut citer la vision stéréoscopique et la reconstruction de scènes en 3D, la synthèse de nouveaux points de vue pour la réalité virtuelle ou la réalité augmentée, la reconnaissance d'objets, l'analyse du mouvement, ou encore l'assemblage d'images de grande taille (cartes, vues panoramiques). Etant données deux ou N images d'une même scène tridimensionnelle, prises selon des points de vue différents, le problème de la mise en correspondance consiste à trouver un ensemble de points dans une image qui peuvent être identifiés comme les mêmes points dans une autre image. Les images peuvent provenir de différents points de vue, ou bien être prises à des instants différents, et les objets de la scène peuvent être en mouvement par rapport à la caméra. Les méthodes

de mise en correspondance utilisent généralement des caractéristiques locales associées à ces points d'intérêt. D'autres approches existent, basées sur une description globale, elles sont cependant sensibles aux changements d'arrière-plans, aux occultations et aux principales transformations de l'image.

On distingue deux approches pour rechercher des points et les apparier :
La première consiste à trouver des points et les caractériser pour ensuite les suivre par des techniques de recherche locales telles que la corrélation ou les moindres carrés (tracking). Cette approche convient pour des images proches (points de vue proches ou images se succédant rapidement, comme en vidéo).
La seconde approche consiste à détecter de façon indépendante des points d'intérêt dans chaque image, pour ensuite les apparier. Cette seconde approche, plus générale, recherche des correspondances dans des images très différentes, et convient par exemple, pour assembler plusieurs images dans une même vue panoramique.

Compte tenu des objectifs visés, nous nous situerons plutôt dans la seconde approche, qui suit habituellement un schéma en plusieurs étapes :

- Tout d'abord, la détection de points d'intérêt sélectionne des points facilement repérables dans une image. Pour être utilisable dans un appariement, un point d'intérêt doit présenter un haut degré de répétabilité, c'est-à-dire être stable lors de perturbations locales ou globales qui peuvent affecter l'image, telles que les transformations projectives ou les variations d'intensité lumineuse.
- Dans une deuxième phase d'extraction de caractéristiques, le voisinage autour des points d'intérêt est analysé, de manière à fournir un descripteur stable, compact et invariant, qui pourra être comparé à d'autres descripteurs. Parmi les méthodes de détection et de description, la méthode "SIFT" et ses variantes ("SURF", "ASIFT", "GLOH",$\cdots$) ont permis un saut qualitatif permettant de les utiliser dans des cas difficiles de changement d'illumination ou de changement de point de vue important.
- Enfin, la phase d'appariement établit une correspondance entre les points de deux images présentant des caractéristiques similaires. Pour limiter la combinatoire et réduire le temps de calcul, en stéréo-vision, on introduit en général des contraintes issues de la géométrie épipolaire : si un point est donné sur une image, alors le point en correspondance sur l'autre image est à rechercher sur une droite. Ensuite, des méthodes d'élimination des fausses correspondances telles que RANSAC permettent de filtrer les erreurs pendant l'estimation de la matrice fondamentale qui décrit la géométrie épipolaire.

Contributions

Le point de départ de ce travail a été d'étudier un nouveau descripteur de points d'intérêt, autorisant une mise en œuvre suffisamment simple et des calculs suffisamment rapides pour être utilisable en stéréovision et reconstruction 3D, aussi bien que

pour traiter des flux vidéo. Notre choix s'est porté vers un descripteur à base de filtres anisotropes tournants développés dans notre laboratoire pour la segmentation d'images, et c'est l'objet de notre première contribution. L'idée est d'appliquer en un point d'une image une série de "demi-filtres" selon des directions orientées, c'est-à-dire de 0° à 360° selon un pas angulaire arbitraire. Nous obtenons de cette façon une famille de filtres basés sur une fonction semi-gaussienne : filtres de lissages, filtres de dérivation, qui vont permettre de faire une description fine de l'image. C'est ainsi que nous proposons de caractériser un point d'intérêt par une signature, résultat du filtrage de dérivation à l'ordre 1 dans toutes les directions.

Dans une seconde étape, nous nous sommes intéressés à la mise correspondance des points d'intérêt ainsi caractérisés. Notre deuxième contribution a donc porté sur une méthode de mise en correspondance capable de prendre en compte naturellement les rotations et les déformations projectives d'une image à une autre. Nous avons défini une nouvelle méthode d'appariement basée sur d'une part, la corrélation et d'autre part, la déformation ("warping") des signatures obtenues. Contrairement à des méthodes comme "SIFT", notre méthode est mono-échelle, elle est en cela bien adaptée aux applications visées. Dans le cadre de nos applications, cette méthode donne souvent des résultats supérieurs à "SIFT", qui laissent envisager une extension au cas multi-échelle très prometteuse.

Organisation de ce mémoire

Ce mémoire est divisé en quatre chapitres : deux chapitres bibliographiques, et deux chapitres consacrés à l'apport scientifique de ce travail.

Le deuxième chapitre présente un état de l'art de la mise en correspondance, il est dédié d'une part aux problèmes géométriques liés aux applications (mouvement, stéréovision) et d'autre part aux problèmes photométriques liés aux éventuelles variations d'éclairage. Dans ce chapitre, sont abordées, différentes méthodes d'appariement et les primitives sur lesquelles elles s'appuient.

Le troisième chapitre se focalise sur les méthodes basées sur des points d'intérêt : type de primitive et appariement. Nous présentons les méthodes actuelles de détection de points et les principes utilisés dans la représentation des caractéristiques et l'appariement.

Au chapitre quatre, nous introduisons un nouveau descripteur de points, basé sur une technique de filtrage anisotrope. Les filtres utilisés sont construits à partir de demi noyaux gaussiens anti-aliasés par une fonction sigmoïde, et génèrent au niveau des points d'intérêt une signature circulaire fortement discriminante.

Le dernier chapitre présente la nouvelle méthode d'appariement que nous avons développée au cours de cette thèse. Cette méthode est basée sur la corrélation et le "warping"

des signatures décrites au chapitre trois. De nombreux résultats concrets de mise en correspondance, de reconstruction 3D et d'analyse du mouvement dans le plan image illustrent ce chapitre et montrent la validité de cette approche.

Publications

Ce travail a donné lieu à deux publications :

P. Montesinos, B. Magnier, JL. Palomares. A New Perceptually Edge Detector. *IWIA2010. Proceedings of the third International Workshop on Image Analysis.* pp. 185-192. Presse des Mines 2011, ISBN 978-2911256-55-4.

JL. Palomares, P. Montesinos, D. Diep. A New Affine Invariant Method for Image Matching. *IST Society for Imaging Science and Technology - SPIE international society for optics and photonics, 2012.*

Chapitre 2

Introduction à la mise en correspondance

Ce chapitre présente un état de l'art concernant des applications nécessitant une "reconnaissance" d'image ou de parties d'images dans une autre image. Nous nous intéresserons ici à des applications comme la stéréo-vision, la reconnaissance d'objet, l'indexation d'images, la stabilisation vidéo, etc. Pour toutes ces applications, la mise en correspondance d'images représente une composante essentielle. Parmi les méthodes employées, certaines effectuent une mise en correspondance directe mais la plupart du temps la mise en correspondance est réalisée en plusieurs étapes : segmentation, description, interprétation. Nous distinguerons des méthodes de "mise en correspondance éparse" qui font correspondre seulement certains points (par exemple des points d'intérêt) mais aussi des méthodes effectuant une "mise en correspondance dense" qui essaient de mettre en correspondance tous les pixels ou la plupart des pixels.

Plus précisément, nous pouvons citer de manière non exhaustive un certain nombre d'applications concernées par la mise en correspondance :

- La stéréo-vision/reconstruction 3D consiste à recréer une vue 3D par appariement puis reconstruction à partir de deux images (ou plus) d'une scène prise selon des points de vue différents. Dans ce contexte nous distinguerons principalement deux cas particuliers : dans le premier cas, la matrice fondamentale est connue et la géométrie du système impose alors des contraintes pour la mise en correspondance, dans le deuxième cas, la calibration est inconnue alors, pour le processus d'appariement, le correspondant d'un pixel de la première image doit être recherché parmi tous les pixels de la seconde. Le paragraphe 2.1 présente la géométrie d'un système de prises de vues stéréoscopiques soit : la géométrie épipolaire et les contraintes qui en découlent vis-à-vis d'un processus d'appariement.

- La réalité virtuelle et la réalité augmentée, mélangent des images réelles et des images synthétiques ou des informations additionnelles [43]. Nous sommes ici dans un con-

texte de mouvement dans une séquence vidéo et de mise en correspondance temporelle. Les contraintes naturelles imposées à ce type d'application sont un mouvement "lent" dans les images successives. Le processus de mise en correspondance doit effectuer un suivi de points particuliers ou de points de contours.

- Le morphing de visages ou la déformation d'images de visage, l'image d'un visage doit être transformée en celle d'un autre. Dans un cas tel que celui-ci on cherchera plutôt à mettre en correspondance certains points particuliers d'un visage, par exemple les coins des yeux ou de la bouche, certains points de contours, etc. [95].

- Des applications médicales dans lesquelles il s'agit de mettre en relation des images provenant de sources différentes (scanner, d'IRM, etc.), ou encore de mettre en relation temporelle des clichés successifs pour le suivi de l'évolution de la maladie d'un patient, etc. [107, 3]. On retrouvera ce type d'application sous la dénomination de : "registration" dans la bibliographie, il s'agit à la fois de mise en correspondance et de déformations géométriques des différentes images à mettre en relation.

- Des applications de vidéo surveillance liées par exemple à l'analyse du mouvement dans des vidéos [37][57][49]. Plus précisément on trouvera des application liées à la sécurité comme l'analyse du mouvement au sein d'une foule, d'une file d'attente ou encore du suivi de véhicules sur autoroutes, mais encore des applications "commerciales" comme le suivi de personnes dans un centre commercial.

- On pourra citer encore des applications liées à la stabilisation de vidéo. Jusqu'à présent ce type d'application a été principalement concerné par le domaine militaire. Les réponses apportée à ce type de problème comportent souvent une composante mécanique et une composante informatique. Cependant compte tenu de la baisse des coûts des caméras vidéos et des appareils photographiques numériques grand public, une solution purement informatique à ce problème est d'un grand intérêt. De même on touve de plus en plus fréquement des caméras embarquées dans des véhicules où encore juste fixées sur un casque (exemple pour des sportifs), et là encore un besoin en stabilisation existe.

- Dans un tout autre domaine, on a vu apparaître des problèmes d'indexation d'image dans des bases de données. En effet, avec la numérisation de bases d'images, de nouveaux besoins sont apparus. Considérons une base de données comportant plusieurs dizaines ou centaines de milliers d'images, savoir rapidement si une image donnée a déjà été stockée dans la base est un problème important. De nombreuses solutions faisant intervenir une mise en correspondance ont été proposées.

- De même pour la reconnaissance d'objets [93, 73], un objet est souvent stocké sous la forme d'une liste d'images le représentant sous plusieurs orientations et points de

vue. Nous avons alors encore affaire à une base de donnée d'images qu'il est nécessaire d'indexer pour reconnaître un objet (l'image d'un objet).

- etc.

Ces applications multiples, et non limitatives, montrent l'importance de pouvoir mettre en relation des images entre elles. Les techniques utilisées pour cette mise en correspondance dépendent du problème et des contraintes qui lui sont propres. Dans cette thèse, nous nous intéresserons plus particulièrement à deux applications de la mise en correspondance : la stéréo-vision dans le cas non-calibré et la stabilisation vidéo. La section suivante va donc naturellement décrire ces applications et leurs contraintes spécifiques.

2.1 Contraintes géométriques liées aux applications

Avant d'entrer dans les détails des contraintes liées aux applications, nous allons tout d'abord examiner le modèle des caméras utilisées dans toutes les applications de type stéréo-vision ou mouvement : le modèle "pinhole".

2.1.1 Modèle de caméra, le modèle "pinhole"

Le modèle de caméra le généralement utilisé est le modèle "pinhole" ou modèle sténopé (figure 2.1). Ce modèle est constitué d'une boite noire comportant un petit trou ponctuel pour laisser passer la lumière provenant de la scène observée. Ce système projette donc l'image du monde 3D sur le fond de la boite recouvert d'une rétine plane. Mathématiquement, ce modele se représente par une projection perspective du monde 3D sur la rétine.

Après acquisition de l'image, nous obtenons un fichier contenant une matrice de pixels. Dans un cas idéal, le centre de la rétine (le centre de l'image) coincide avec l'axe optique de la caméra. Mais en général dans toute réalisation physique ce n'est pas le cas, et l'intersection de l'axe optique et de la rétine n'est donc pas exactement le pixel central de l'image. Un modèle plus complet de caméra est présenté à la figure 2.2.

- La rétine est placée en avant du centre optique de la caméra de manière à avoir l'image dans le mème sens que la scène.
- le pixel $(0,\ 0)$ est situé en haut à gauche de l'image.
- $(u_0,\ v_0)$ sont les coordonnées en pixels de l'intersection de l'axe optique avec la rétine.
- f est la focale de l'objectif, lorsque l'objectif réglé sur l'infini, la distance focale est égale à la distance centre optique-rétine.
- Le repère $(O,\ X,\ Y,\ Z)$ est un repère lié à la caméra, l'axe Z est orienté dans le sens de l'axe optique.

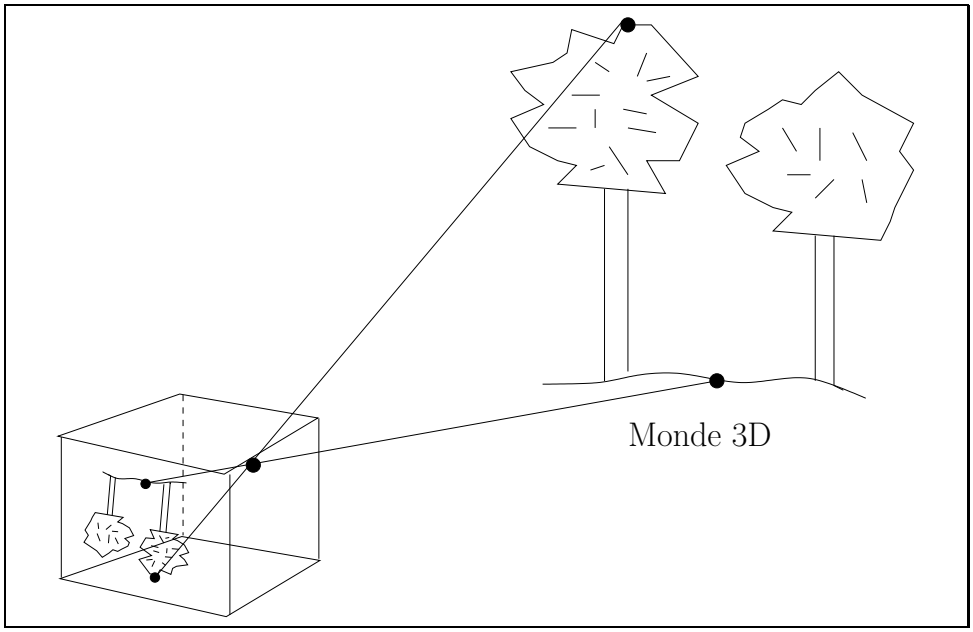


FIGURE 2.1: Modèle de caméra “pinhole”.

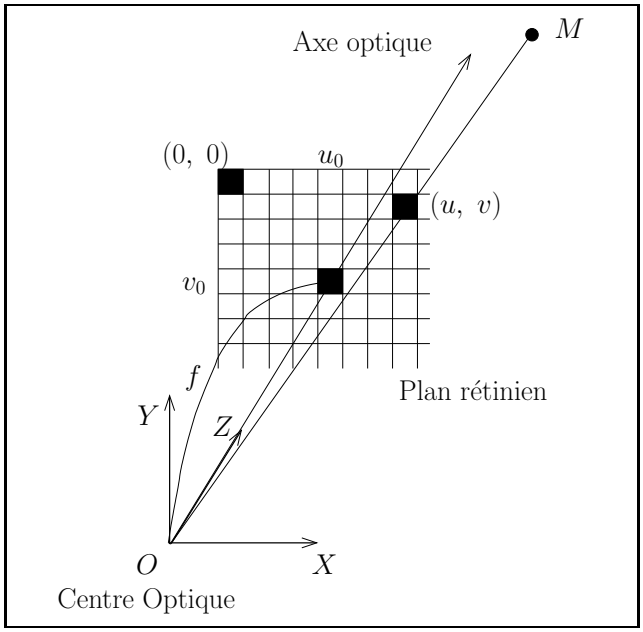


FIGURE 2.2: Modèle sténopé de caméra.

D'un point de vue algébrique, la caméra est modélisée par une matrice de paramètres intrinsèques généralement notée A.

$$A = \begin{pmatrix} \alpha_u & \epsilon & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.1}$$

Dans laquelle :

- α_u représente le grandissement en pixels sur l'axe X,
- α_v représente le grandissement en pixels sur l'axe Y,
- ϵ est généralement un petit coefficient, il rend compte d'une éventuelle non orthogonalité des axes de la rétine.
- $(u_0,\ v_0)$ sont les coordonnées en pixels de l'intersection de l'axe optique avec la rétine.

En coordonnées homogènes, un pixel projeté sur la rétine en ${}^t(x,\ y,\ 1)$ sera vu en coodonnées pixel ${}^t(u,\ v,\ 1)$ en :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = A \,.\, \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2.2}$$

2.1.2 Mise en correspondance pour la stéréo-vision

Avec deux yeux, l'être humain possède la capacité à percevoir le monde en trois dimensions. La stéréo-vision en vision par ordinateur tente de reproduire cette capacité en utilisant deux (ou plusieurs) caméras. La figure 2.3 présente un système de vision stéréoscopique.

En stéréo-vision calibrée, l'acquisition est réalisée par un ensemble de caméras (deux caméras en stéréo binoculaire, trois pour un système trinoculaire, etc.) fixées sur une base stéréoscopique et d'un système de numérisation d'images (par exemple un ordinateur comportant une carte d'acquisition). La calibration, ici, consiste d'une part à modéliser et à mesurer, les optiques des caméras ou paramètres intrinsèques (matrices des paramètres intrinsèques et les distortions géométriques éventuelles des optiques utilisées) ainsi que la géométrie du système stéréoscopique où paramètres extrinsèques (coordonnées du centre optique de la caméra C_2 dans un repère lié à la caméra C_1, rotation 3D reliant les orientations des axes optiques des différentes caméras). En revanche en stéréo-vision non calibrée, la connaissance des paramètres intrinsèques et extrinsèques n'est pas disponible, un banc stéréoscopique n'est généralement pas utilisé et les différentes images sont souvent acquises à l'aide d'une unique caméra pointée manuellement. Nous pourrons encore distinguer le cas semi-calibré où seule la connaissance des paramètres

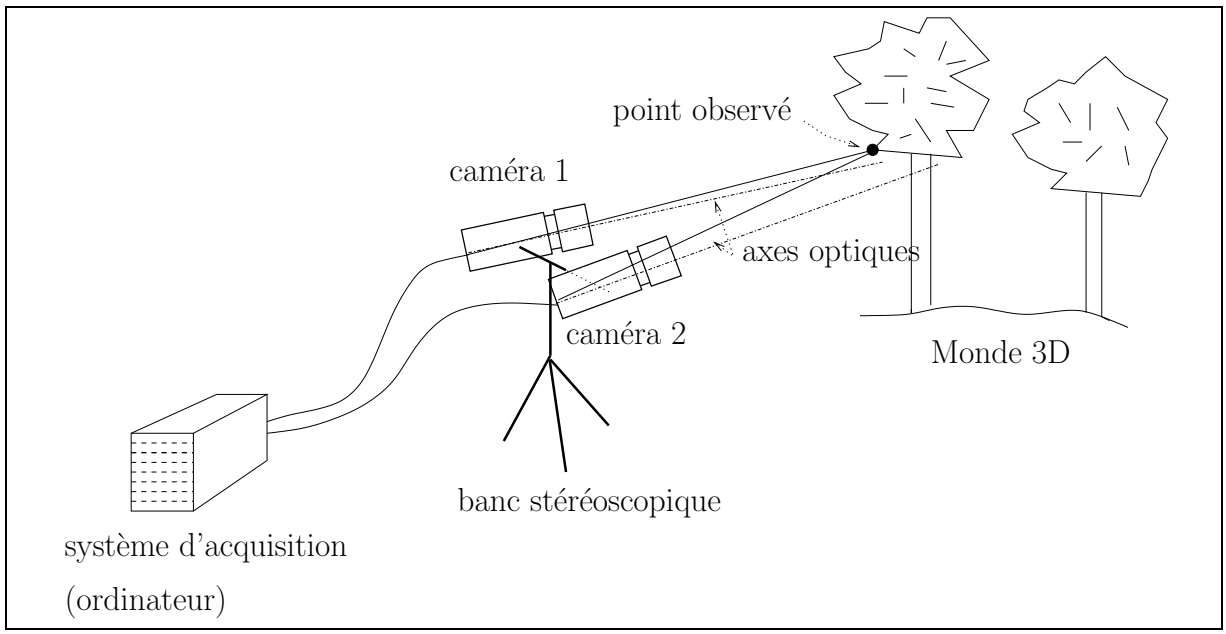


FIGURE 2.3: Système d'acquisition stéroscopique à deux caméras.

intrinsèques est disponible.

Connaissant la géométrie du système d'acquisition, un point vu dans les différentes images peut alors être reconstruit en 3D par triangulation. La figure 2.4 illustre la reconstruction 3D à partir de deux caméras. Il est évidement possible de multiplier le nombre de caméras et de points de vue afin d'obtenir des informations 3D plus riches, nous nous limiterons ici au cas de deux caméras. En effet, nous nous intéressons uniquement au problème de la mise en correspondance entre paires d'images.

Le paragraphe suivant se focalise sur la géométrie d'un système de prise de vues stéréoscopique. Nous verrons que dans le cas calibré, cette géométrie impose des contraintes fortes pour la mise en correspondance. Cependant, dans le cadre de cette thèse, nous nous placerons plutôt dans le cas où cette géométrie est inconnue et nous verrons que la méthode proposée résout élégamment ce problème complexe d'appariement. Ayant obtenu des appariements (hors contraintes), la géométrie pourra alors être estimée de manière robuste.

2.1.2.1 Géométrie épipolaire

La géométrie d'un système stéréoscopique est appelée géométrie épipolaire. La figure 2.5 présente la géométrie d'un système stéréoscopique à deux caméras.

Soit un point M de la scène qui vu dans les deux images. Ce point se projette en m_1 sur la première image, et en m_2 dans la seconde. Pour le point m_1 sur la première image,

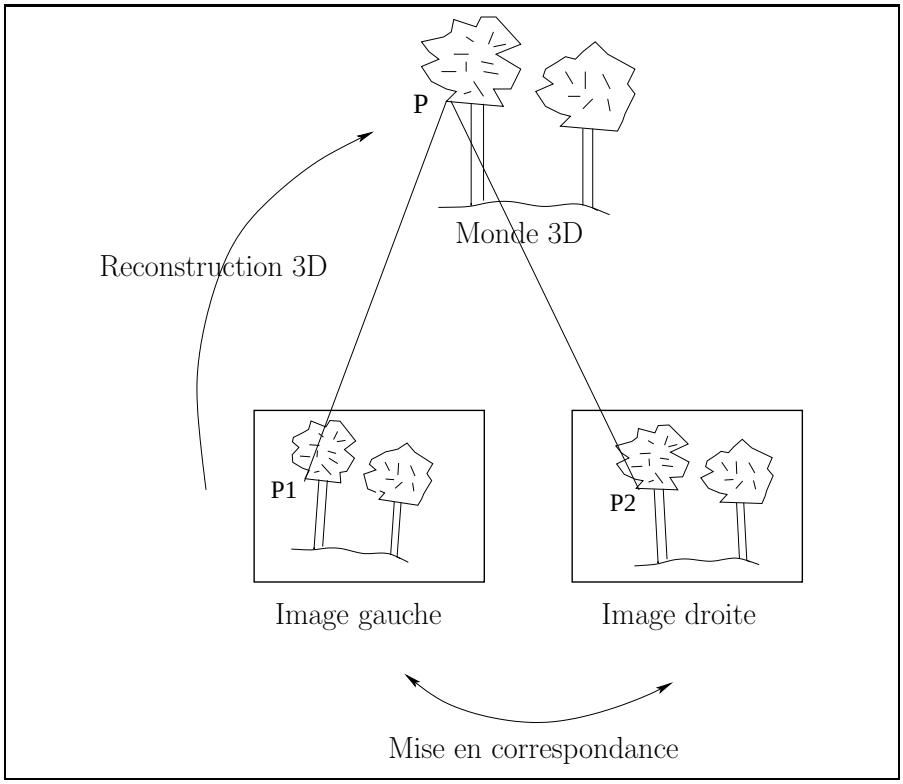


FIGURE 2.4: Reconstruction 3D à partir de deux images stéréoscopiques : un point P du monde 3D se projette en P_1 et en P_2 dans les deux images. Si les points P_1 et P_2 sont appariés, alors connaissant la géométrie du système de prises de vues, le point P peut être reconstruit en 3D.

le point m_2 en correspondance se trouve sur une droite épipolaire dans la seconde image (et inversement). Ces droites correspondent aux intersections entre les plans rétiniens et le plan de vue formé par les centres optiques des deux caméras et le point 3D M considéré.

Les épipoles sont alors les projections des centres optiques de la caméra 2 dans la caméra 1 (et inversement), ils définissent l'origine des faisceaux de droites épipolaires dans les deux images.

La connaissance de cette géométrie peut être résumée à la connaissance de la matrice fondamentale qui contient les paramètres intrinsèques des caméras (focale nonbre de pixels, etc.) et extrinsèques purement géométriques (distance entre les centres optiques, angles de rotations entre les axes optiques).

La matrice fondamentale généralement notée F est une matrice 3×3 de rang 2 et définie à un facteur d'échelle près. Elle transforme les coordonnée d'un pixel dans la première image en équation de droite (en pixels) dans la seconde (la figure 2.6 illustre l'obtention des droites épipolaires) :

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = F \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{2.3}$$

La figure 2.7 illustre la géométrie épipolaire sur un couple stéréoscopique réel. Connaissant la matrice fondamentale, partant de l'image gauche un point permet de tracer une droite épipolaire dans l'image droite. Ces droites épipolaires passent bien par les points correspondants de l'image gauche dans l'image droite. Partant de l'image droite, un point quelconque d'une droite épipolaire permet de tracer une droite épipolaire dans l'image gauche en utilisant la transposée de la matrice fondamentale. Nous avons donc des droites en correspondance, n'importe quel point d'une droite de l'image gauche est en relation avec n'importe quel point de la droite correspondante dans l'image droite.

2.1.3 Autres contraintes couramment utilisées en stéréo-vision

Contrainte d'unicité

Lorsque l'on veut mettre en correspondance deux images, il est fréquent d'imposer la contrainte d'unicité. En effet un point d'une image doit correspondre à un point unique dans la deuxième image et inversement.

Contrainte d'ordre

En stéréo-vision, dans le cas d'un éloignement faible des deux caméras (stéréo à petite base), la contrainte d'ordre est fréquement utilisée.

Considérons deux droites épipolaires en correspondance dans les deux images (D_1 dans l'image 1 et D_2 dans l'image 2), alors si deux points de D_1 sont vus dans un certain

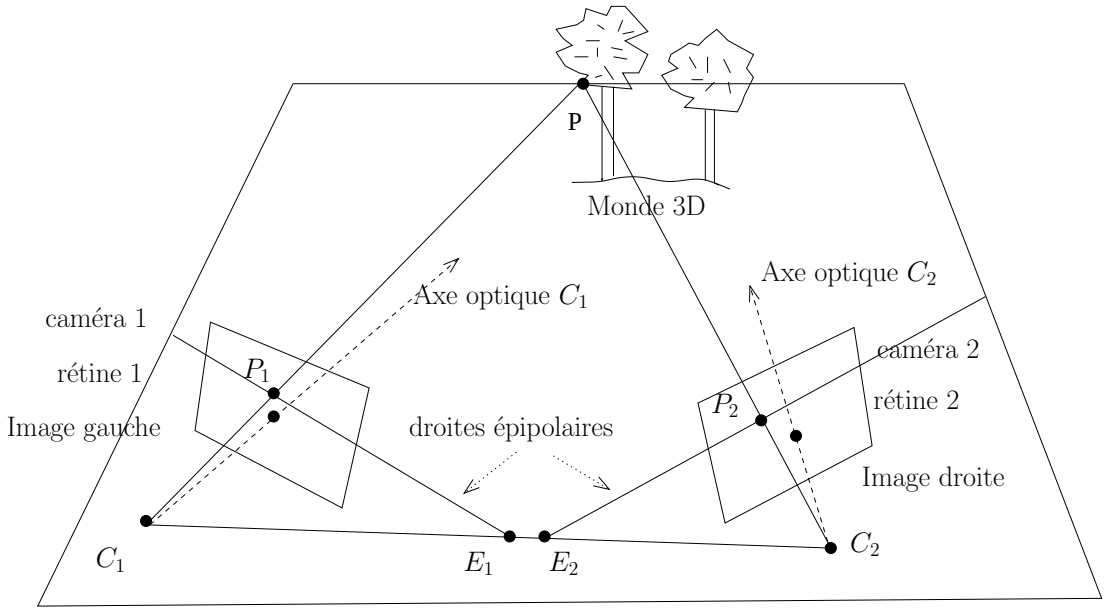


FIGURE 2.5: La géométrie épipolaire entre deux vues. Les rayons P_1 et P_2 sont respectivement les projections du point P sur les rétines des caméras 1 et 2, C_1 et C_2 sont les centres optiques des caméras 1 et 2, E_1 et E_2 sont les épipoles : E_1 est la projection du centre optique C_2 dans la caméra 1 (et inversement).

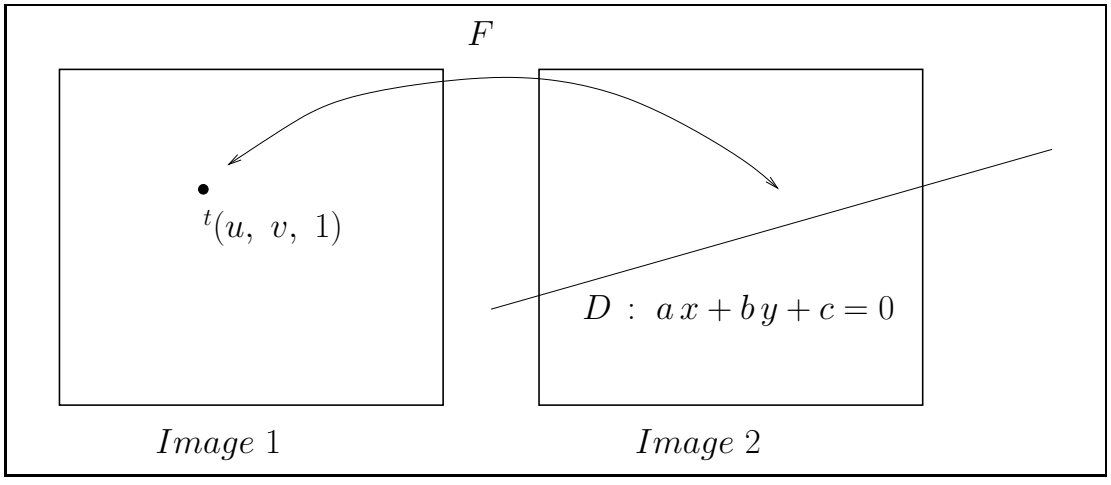


FIGURE 2.6: Obtention des droites épipolaires grâce à la matrice fondamentale.


a)

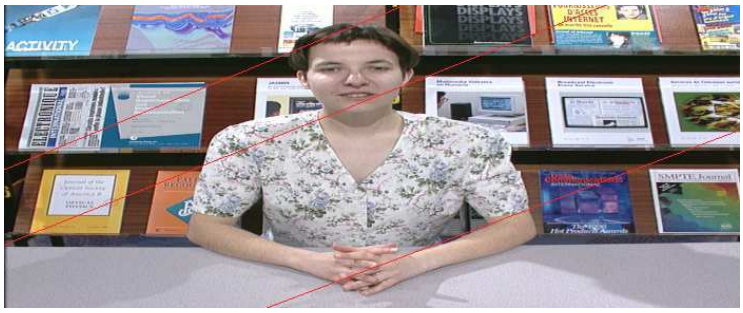

b)

FIGURE 2.7: Illustration de la géométrie éopipolaire sur un couple stéréoscopique réel. a) image gauche, b) image droite. Trois points ont été sélectionnés sur l'image gauche les droites épipolaires correspondantes sont tracées sur l'image droite. Sur l'image gauche sont tracées les droites épipolaires correspondant aux points des droites épipolaires de l'image droite (ce sont des droites en correspondance).

ordre par rapport à l'epipole, alors leus correspondants dans l'image 2 sur la droite D_2 sont vus dans le même ordre.

Le paragraphe suivant introduit la notion de flot optique, qui décrit la relation entre le mouvement dans les images et les contours des objets.

2.1.4 Contraintes pour le mouvement

Nous sommes ici dans un contexte de mouvement, nous pourrons distinguer plusieurs types de mouvements possibles :

- Dans le cas de scènes rigides : seule la caméra se déplace dans la scène. Ce déplacement de la caméra va entrainer des mouvements apparents des objets et du fond (de la scène) dans la vidéo. Cependant, l'aspect 3D des scènes observées entrainera des mouvements apparents différents pour des objets proches ou loin de la caméra. Si nous considérons deux images différentes d'une vidéo contenant des structures communes, alors, ce cas s'apparente au cas de la stéréo-vision.
- Dans le cas général les objets de la scène peuvent posséder un mouvement propre. le mouvement observé sera alors une combinaison du déplacement des objets, du déplacement de la caméra et d'effets 3D.

Concernant le déplacement de la caméra, il est encore possible de distinguer deux cas :

- La caméra avance dans la scène selon l'axe optique, alors nous allons observer un effet "centrifuge" les objets (immobiles) de la scène ont un mouvement apparent dans la vidéo du centre vers la périphérie de l'image et finissent par sortir du champ.
- La caméra effectue un "Travelling" dans la scène. Alors nous observons globalement un mouvement de translation du fond (et des objets immobiles) dans la vidéo par exemple de la droite vers la gauche, etc.

Dans le cas de scènes non rigides, nous n'avons plus de contrainte géométrique externe nous permettant de restreindre le domaine de recherche d'une image à l'autre, comme par exemple en stéréo-vision (donnée par la position des caméras), il est nécessaire d'introduire une contrainte temporelle ou contrainte d'acquisition. L'acquisition doit se faire à une cadence élevée entrainant de faibles déplacements apparents des structures dans les images successives de la vidéo. La recherche de correspondants peut donc se ramener à une recherche locale dans des fenêtres de taille réduite.

2.1.4.1 Méthodes différentielles, flot optique

La notion de flot optique a été énoncée dans les années 1950 par James J. Gibson, psychologue américain [31]. Le flot optique décrit le mouvement apparent des frontières

des objets de la scène dans une séquence vidéo.

Tout d'abord, il est nécessaire de définir deux notions :

- Le champ de vitesses $\vec{v}$ est la projection dans le plan image des vitesses (3D) des objets de la scène. Evidement, le champ de vitesse est une combinaison des différents mouvements présents : objets, caméra.

- Le flot optique, quant à lui, est défini par les variations spatio-temporelles de la fonction intensité lumineuse de l'image. Le flot optique décrit la composante du mouvement normale aux contours des objets : $\vec{v_\eta}$. La figure 2.8 prérente le repère de Frénet lié à la surface image, la figure 2.9 quant à elle, représente la composante normale aux contours du vecteur mouvement.

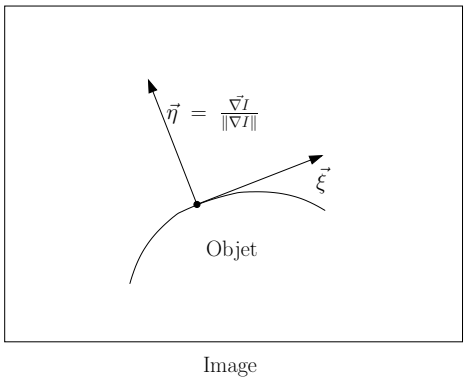


FIGURE 2.8: Repère de Frénet lié à la surface image $\vec{\eta}$ est le vecteur unitaire dans la direction du gradient de l'image, $\vec{\xi}$ est le vecteur unitaire dans la direction du contour.

D'un point de vue mathématique, on suppose dans ces méthodes que l'image n'a subi que de petites variations spatiales entre les temps t et $t + dt$. Une séquence d'images peut être modélisée par une fonction comportant deux variables d'espace et une variable temporelle : $I(x,\ y,\ t)$.

Considérons un point de la scène qui se projette dans l'image, sous l'hypothèse que l'intensité lumineuse de ce point n'a pas changé, nous pouvons écrire :

- Au temps t, on observe un point de la scène avec l'intensité lumineuse :

$$I(x,\ y,\ t)$$

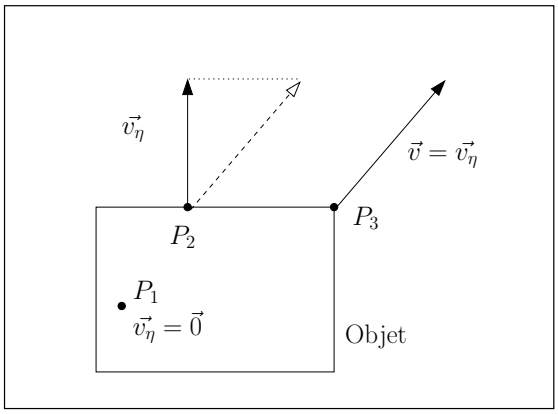


FIGURE 2.9: Le flot optique $\vec{v_\eta}$ décrit la composante du mouvement normale aux contours des objets. Au niveau des coins des objets nous mesurons directement la vitesse des objets dans la scène.

- Au temps $t + dt$, on observe ce même point à une position différente :

$$I(x + dx,\ y + dy,\ t + dt)$$

- Alors :

$$I(x + dx,\ y + dy,\ t + dt) \ = \ I(x,\ y,\ t) + \frac{\partial I}{\partial x}dx \ + \ \frac{\partial I}{\partial y}dy \ + \ \frac{\partial I}{\partial t}dt$$

Nous obtenons alors l'équation fondamentale du flot optique :

$$\frac{\partial I}{\partial t} + \vec{v}\,.\,\vec{\nabla} I \ = \ 0 \tag{2.4}$$

Où :

$\vec{v} \ = \ \left(\frac{\partial x}{\partial t},\ \frac{\partial y}{\partial t}\right)^t$ est la vitesse d'un point dans le plan image.

$\vec{\nabla} I \ = \ \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right)^t$ est le vecteur gradient de l'image.

La détermination du flot optique, comme de nombreux problèmes de vision par ordinateur est un problème mal posé. Un problème mal posé est un problème pour lequel l'équation (ou les équations) le représentant ont une infinité de solutions. Le flot optique n'est défini qu'au niveau des contours des objets, et la direction du vecteur vitesse n'est pas connue, nous avons le produit scalaire avec le gradient de l'image (figure 2.9). La détermination du champ de vitesses : $\vec{v}$ par résolution de l'équation du flot optique, est un problème difficile et nécessite l'introduction de contraintes de régularisation. En revanche le mouvement est connu au niveau des coins (figure 2.9), les méthodes de mise

en correspondance de coins ou points d'intérêt permettent, elles, de calculer directement la composante du mouvement dans le plan image.

2.2 Contraintes photométriques et invariance colorimétrique

Nous avons vu dans les paragraphes précédents que, selon certaines applications, il est possible d'obtenir des contraintes géométriques qui vont aider le processus de mise en correspondance (géométrie épipolaire en stéréo-vision, contrainte de faible déplacement pour certaines applications liées au mouvement, etc.). Cependant nous allons voir encore qu'il est encore possible d'imposer de nouvelles contraintes de nature photométriques ou colorimétriques en définissant des modèles de tranformation.

2.2.1 Conservation "naïve" des intensités lumineuses

Lorsque deux images sont proches, étant donné un point vu à la fois dans deux images, l'idée d'imposer la conservation des intersités lumineuses est l'une des premières qui vient à l'esprit.

Une image couleur est une image dans laquelle les pixels sont des quantités vectorielles. Une image couleur peut être assimilée à la superposition de trois images en niveau de gris (image "rouge", image "vert", image "bleu") :

$$\vec{I}(x,\ y)\ =\ \begin{pmatrix} R(x,\ y) \\ V(x,\ y) \\ B(x,\ y) \end{pmatrix}$$

Dans ce cas, il est naturel d'imposer la conservation de chaque canal.

Cependant, dans la réalité, la conservation des intensités lumineuses ou des couleurs est souvent mise en défaut. Par exemple en stéréo-vision, le simple fait de déplacer légèrement une même caméra peut changer l'intensité lumineuse recue d'un point donné, par exemple dans le cas de réflexions spéculaires (cf. section 2.2.3).
Dans le cas du mouvement les fluctuations de l'éclairage vont aussi beaucoup influencer l'acquisition des images successives, nous n'aurons alors pas conservation des intensités lumiseuses d'un point de vue temporel.

Il est alors nécessaire de modéliser les changements de luminosité qui peuvent survenir au cours des prises de vue, de manière à "normaliser" les mesures que l'on pourra effectuer sur les images. Nous nous placerons dans le cas "raisonnable" où la majorité des surfaces observées présentent peu de spécularités. Nous allons tout d'abord nous focaliser sur les conditions d'illumination, nous développerons ensuite les propriétés des surfaces d'un point de vue physique.

2.2.2 Conditions d'illumination

Une scène peut être soumise à plusieurs types l'illumination :

- Lorsque la source lumineuse fluctue d'un point de vue temporel, nous parlerons de changement interne de la source. L'éclairage de la scène est modifié dans sa globalité.

- Lorsque la source lumineuse se déplace dans la scène, nous parlerons de changement de luminosité externe. L'énergie reçue en un point d'une surface dépend de la distance du point considéré à la source (fonction en $1/r^2$). Cela est une première approximation qui ne tient pas compte d'éventuelles réflexions secondaires. Nous allons alors obtenir des variations locales d'éclairage dans la scène dues aux variations de distance de la source aux objets.

Ces deux types de transformation peuvent être rencontrés à la fois en stéréo-vision et en mouvement. En mouvement, lorsque la source fluctue et/ou se déplace, en stéréo, lorsque les images ne sont pas acquises au même instant, ce qui en pratique est souvent le cas en stéréo-vision non calibrée (les images ont acquises par une même caméra qui s'est déplacée).

2.2.3 Différents types de réflexions

D'un point de vue physique, une lumière incidente sur une surface va être en partie absorbée et en partie réfléchie, ceci se traduit par des coefficients de réflexion et d'absopsion.

Evidement ces coefficients dépendent de nombreux facteurs, comme par exemple de la longueur d'onde de la lumière émise par la source, des propriétés de la surface des objets, etc.

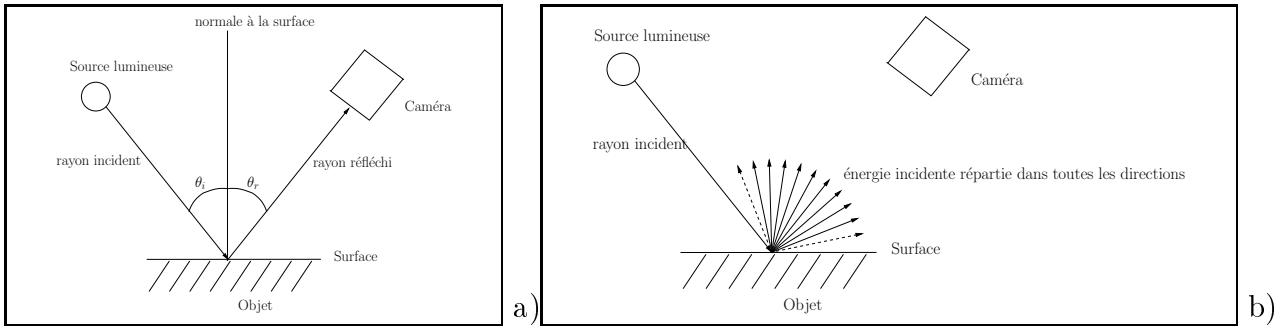


FIGURE 2.10: Réflexion et diffusion d'une source lumineuse sur une surface. a)Réfléxion spéculaire. b) Diffusion lambertienne.

Considérons les propriétés de la surface des objets :

- Certaines surfaces réfléchissent entièrement la lumière reçue selon un angle de réflexion égal à l'angle d'incidence, nous parlerons de surfaces spéculaires. C'est le cas en particulier des miroirs.

- Certaines surfaces diffusent isotropiquement la lumière reçue, nous parlerons alors de surfaces lambertiennes.

- Enfin dans la majorité des cas, les surfaces se comportent en partie comme des surfaces lambertiennes, et en partie comme des surfaces spéculaires.

Dès lors que l'on s'intéresse aux propriétés physiques des illuminants et des surfaces (éclairage spectral, réflexion, diffusion spectrale, etc.), les modélisations obtenues peuvent vite devenir extrêmement complexes et inutilisables dans notre contexte de mise en correspondance.

Il est donc nécessaire d'une part de considérer des scènes plus simples et d'autre part d'effectuer une simplifications des modèles (approximations linéaires) [58]. Au niveau des scènes observées, on ne considèrera que des scènes de type Mondrian [32] dans lesquels on ne rencontre que des surfaces planes, lambertiennes, illuminées par une source ponctuelle émettant de manière isotropique.

2.2.4 Cas de la diffusion lambertienne

Dans ce cas, la fonction de réflectance spectrale de la surface, ne dépendra que de la longueur d'onde de la lumière incidente. Alors la lumière reçue par le capteur k de la caméra ($k \in (R, V, B)$) pour un point donné dun objet de la scène, peut être modélisée par l'équation suivante :

$$I_k = \int_\lambda S_\lambda \,.\, E_\lambda \,.\, R_{\lambda,k}\, d\lambda$$

Où :

- E_λ : représente la lumière incidente, (au point considéré)
- S_λ : représente la réfléctance spectrale de la surface (au point considéré)
- $R_{\lambda,k}$: représente la réponse spectrale du capteur k.

Dans la section suivante nous présentons une modélisation des changements internes de la source.

2.2.5 Modélisation des changements internes de la source

Dans la cas de surfaces lambertiennes, il a été montré [66] qu'une approximation linéaire de ce modèle pouvait rendre compte correctement des changements interne d'il-

lumination. Alors si $\vec{I}$ et $\vec{I'}$ ($\vec{I} = {}^t(R,\ V,\ B)$ et $\vec{I'} = {}^t(R',\ V',\ B')$) représentent respectivement la couleur d'un point perçue par un capteur avant et après changement interne d'illumination :

$$\vec{I'} \ = \ M \,.\, \vec{I} \tag{2.5}$$

Dans le cas couleur, la matrice M est une matrice 3×3 à neuf paramètres.

D'autres variantes ont été étudiées dans [27], [26], [25], le modèle diagonal à trois paramètres. Dans [35], onze modèles d'illumination interne ont été évalués.

Dans [32], le modèle affine (modèle diagonal plus un vecteur de translation des couleurs) est étudié, c'est un modèle à six paramètres :

$$\begin{pmatrix} R' \\ V' \\ B' \end{pmatrix} \ = \ \begin{pmatrix} \alpha_R & 0 & 0 \\ 0 & \alpha_V & 0 \\ 0 & 0 & \alpha_B \end{pmatrix} \ . \ \begin{pmatrix} R \\ V \\ B \end{pmatrix} \ + \ \begin{pmatrix} \beta_R \\ \beta_V \\ \beta_B \end{pmatrix} \tag{2.6}$$

Où :

- ${}^t(R,\ V,\ B)$ et ${}^t(R',\ V',\ B')$: représentent les couleurs avant et après transformation.
- $\alpha_R,\ \alpha_V,\ \alpha_B$: représentent les facteurs d'échelle respectivement pour les canaux rouge, vert et bleu.
- $\beta_R,\ \beta_V,\ \beta_B$: représentent des translations d'intensité respectivement pour les canaux rouge, vert et bleu.

Dans nos travaux nous utiliserons le modèle à six paramètres défini ci-dessus (eq. 2.6).

2.2.6 Invariance colorimétrique affine

Dans [32], le modèle de transformation colorimétrique affine est utilisé. Afin d'éliminer cette tranformation, les auteurs proposent d'utiliser des normalisations locales de l'image au niveau de points d'intérêt, plusieurs méthodes de normalisations sont testées et évaluées. Les mesures géométriques qui sont alors effectuées au niveau des points d'intérêt ne dépendent plus de la transformation colorimétrique.

2.3 Techniques de mise en correspondance

La mise en correspondance d'images s'effectue généralement en deux étapes :

- D'abord, le calcul d'une mesure de similarité, soit entre les images elles mêmes, soit entre parties d'images. Cette mesure de similarité permet alors d'obtenir un score ou une liste de scores de similarité.
- Enfin, une étape de décision ou de seuillage des scores fixe les correspondances.

Lorsque la méthode travaille sur les images entières, nous parlerons de mise en correspondance globale, en revanche, l'orsqu'il s'agit de parties d'images nous parlerons de mise en correspondance locale.
Dans ce deuxième cas (qui est le cas le plus courant), on cherche généralement à établir des correspondances entre primitives des images (régions, contours, segments, points d'intérêt, etc.). Dans [2] les auteurs cherchent à mettre en correspondance des segments de droite obtenus par approximation polygonale de contours, dans [13] l'auteur utilise à la fois des régions et des segments pour effectuer la mise en correspondance. La primitive "Segment" à surtout été utilisée pour des scènes d'intérieur rigides donc des scènes dans lesquelles la géométrie est prépondérante. Dans la plupart des applications récentes, ce sont des points d'intérêt qui sont utilisés. Les points d'intérêt ont la propriété d'être moins ambigüs que des primitives d'étendue plus importante, par exemple, un coin dans une image est plus facilement retrouvé dans une autre image qu'un segment ou une région, et ceci pour plusieurs raisons :

- Les algorithmes de segmentation pour des primitives de type coin comportent moins de paramètres. Il existe de plus des détecteurs invariants à l'échelle ou invariants à certaine tranformations géométriques de l'image qui seront détaillés plus loin.
- Des primitives de grande taille comme par exemple des région seront plus rapidement occultées lors d'un déplacement de caméra, rendant les identification plus difficile.

Enfin, les points d'intérêts présentent des caractéristiques de répétabilité même dans des zones présentant une géométrie peu significative par exemple dans des scènes naturelles, ou des images comportant des zones texturées.

2.3.1 Calcul d'attributs

Nous avons vu que la majorité des méthodes de mise en correspondances travaillent de manière locale. Les primitives dans une image sont porteuses d'information. Il est donc nécessaire de calculer un vecteur d'attribut à chaque primitive de manière à pouvoir les comparer avec celles d'une autre image. Il existe de nombreuses méthodes permettant de caractériser des primitives, dans une image, ces méthodes seront détaillées dans le chapitre suivant.

2.3.2 Calcul des scores

La comparaison entre un vecteur de caractéristiques (de dimension n) d'une primitive i d'une image I_1 : $\vec{v_{1,i}}$ et un vecteur de caractéristiques d'une primitive j d'une image I_2 : $\vec{v_{2,j}}$ découle simplement de fonctions distance, le tableau suivant présente, de mani'ere non exhaustive, les distances les plus courament utilisées pour la mise en correspondance :

Par exemple la distance de Manhattan est très utilisée (souvent dénommée SAD : Sum of absolute differences) : $\sum_{k=1}^{n} |v_{1,i}^k - v_{2,j}^k|$. Dans certains cas cette distance présente un intérêt en terme de temps calcul, en effet, elle évite le calcul de la racine carrée de la distance euclidienne. De nombreuses distances ont été définies pour des mesures de

Nom	Distance entre $\vec{v}_{1,i}$ et $\vec{v}_{2,j}$ (n composantes)
Distance de Manhattan :	$\sum_{k=1}^{n} \lvert v_{1,i}^k - v_{2,j}^k \rvert$
Distance Euclidienne :	$\sqrt{\sum_{k=1}^{n} \left(v_{1,i}^k - v_{2,j}^k\right)^2}$
Distance de Chebychev :	$\max_{k=1,.,n} \lvert v_{1,i}^k - v_{2,j}^k \rvert$
Divergence de Kullback-Leibler :	$\sum_{k=1}^{n} v_{1,i}^k \log \frac{v_{1,i}^k}{v_{2,j}^k}$
Divergence de Jeffrey :	$\sum_{k=1}^{n} v_{1,i}^k \log \frac{v_{1,i}^k}{\frac{v_{1,i}^k + v_{2,j}^k}{2}} + v_{2,j}^k \log \frac{v_{2,j}^k}{\frac{v_{1,i}^k + v_{2,j}^k}{2}}$
Distance quadratique (A matrice de similarité) :	$\sqrt{{}^t(v_{1,i}^k - v_{2,j}^k).A.(v_{1,i}^k - v_{2,j}^k)}$
Distance de Mahalanobis (C matrice de covariance) :	$\sqrt{{}^t(v_{1,i}^k - v_{2,j}^k).C^{-1}.(v_{1,i}^k - v_{2,j}^k)}$

TABLE 2.2: Exemple de calcul de distances entre vecteurs de caractéristiques $\vec{v}_{1,i}$ et $\vec{v}_{2,j}$ de deux primitives i et j prises chacune dans deux images différentes.

similarité entre primitives : Une liste de distances classiques est visible dans le tableau 2.2. On notera que la divergence de Kullback-Leibler n'est pas exactement une distance (elle n'est pas symétrique), on trouvera une variante symétrique dans la divergence de Jeffrey. Ces deux mesures sont plutôt utilisées pour la recherche d'images de contenu similaire dans un cadre d'indexation, que pour effectuer de la mise en correspondance exacte [34].

On notera qu'il existe encore de multiples mesures de distance, par exemple des mesures plus statistiques qui s'effectuent sur des histogrammes. De nombreux travaux ont été menés sur la recherche et l'identification de texture ou d'objets dans une image en utilisant des distances sur des distributions cumulées, par exemple les mesures de Kolmogorov Smirnov, le test de Cramer Von Mises, Distance de Baddeley, etc. Dans [14], les auteurs utilisent comme vecteur de mesure l'image entière, ils définissent une mesure de similarité statistique entre les couleurs des images basée sur une distance de Baddeley. Mais encore on trouvera des distances plus élaborés comme l' EMD (Earth Mover Distance EMD) qui s'appuie sur un problème d'optimisation [90] [91].

2.3.3 Mise en correspondance directe

Etant donné une distance entre vecteurs de caractéristiques, la mise en correspondance directe consiste à affecter comme correspondant d'une primitive i de l'image I_1, la primitive j de l'image I_2 dont la distance est minimale (ou le score est maximal).

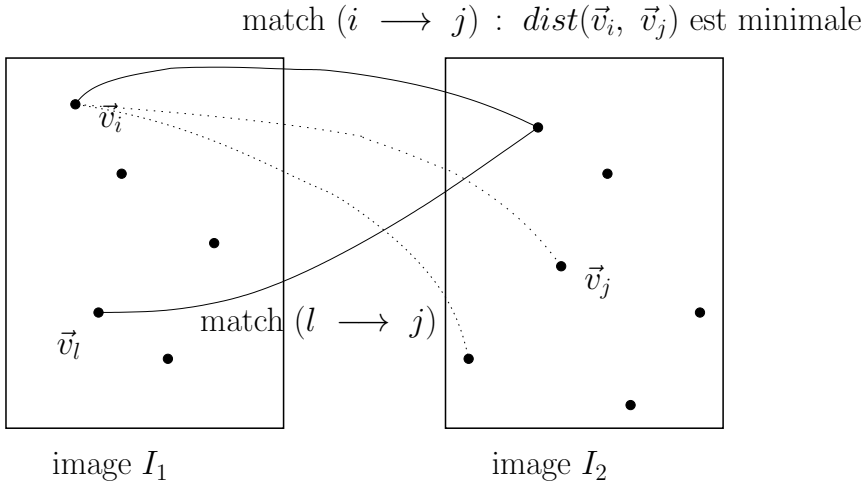


FIGURE 2.11: Mise en correspondance directe (image I_1 vers image I_2).

Cette méthode ne vérifie pas la contrainte d'unicité enoncée au paragraphe 2.1.3, plusieurs primitives de l'image I_1 peuvent être mises en correspondance avec la même primitive de l'image I_2. La figure 2.11 illustre cette propriété.

2.3.4 Mise en correspondance par vérification croisée

La méthode de mise en correspondance par vérification croisée consiste à mettre en correspondance des régions ou des primitives d'une image I_1 vers l'image I_2 et inversement de l'image I_2 vers I_1 (Figure 2.12). En effet l'ensemble des matches obtenus doit vérifier la contrainte d'unicité. Cette méthode de mise en correspondance donne de bons résultats lorsque l'ambiguité de la scène n'est pas trop importante (peu de structures répétitives) et que les descritions utilisées sont suffisament informatives.

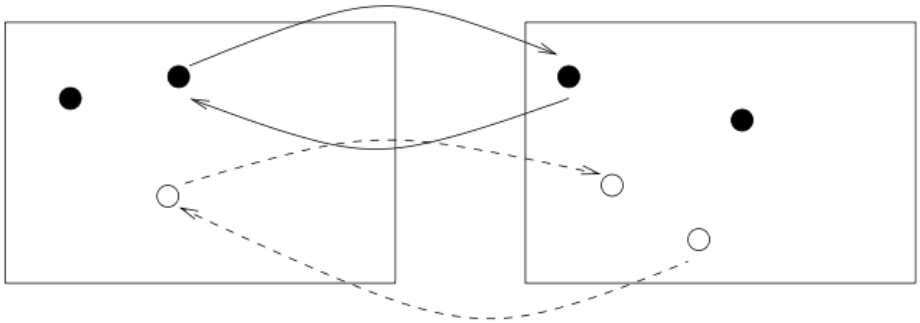


FIGURE 2.12: Mise en correspondance croisée des primitives entre deux images.
[Droit Ill. images de [97]]

2.3.5 Mise en correspondance par relaxation

Les techniques de relaxation sont basées sur des méthodes d'étiquetage stochastique [89], [88], [39]. Le schéma de relaxation est extrêmement général, il s'agit d'un schéma de mise en correspondance "floue" de graphes, ou étiquetage de graphes.
La relaxation tente d'attribuer des étiquettes aux noeuds d'un graphe (les étiquettes provenant des noeuds d'un autre graphe). Nous présentons ici la version probabiliste la plus simple de [24], on notera que de nombreux travaux de mise en correspondance ont été réalisés par relaxation, par exemple : [52], [108], [50], [33], [96].

Problème :
Assigner un ensemble L de l labels à un ensemble N de n noeuds :

$$L = (l_1, \ \ldots, \ l_j, \ \ldots, \ l_l)$$
$$N = (n_1, \ \ldots, \ n_i, \ \ldots, \ n_n)$$

- Pour chaque noeud n_i, un vecteur de probabilité d'assignation (de dimension l) est défini : $P(n_i = l_j)$. Ce vecteur de probabilité représente la probabilité d'assignation du noeud n_i par l'étiquette l_j.
- A chaque noeud n_i, correspond un certain voisinage $V(n_i)$ défini par les arêtes du graphe (des noeuds).

- On se donne enfin une relation de compatibilité entre les assignations de noeuds voisins. Cette relation est généralement modélisée par un ensemble de probabilités conditionnelles :

$$P(n_i = l_j | n_k = l_m)$$

c'est à dire la probabilité que le noeud n_i soit étiqueté par l'étiquette l_j sachant que le noeud voisin n_k ($\in V(n_i)$) est étiqueté par l'étiquette l_m.

La mise en correspondance par relaxation est un schéma itératif :

- à chaque itération les probabilités d'assignation sont mises à jour grâce aux probabilités d'assignation des noeuds voisins. La probabilité que le noeud n_i soit étiqueté par l'étiquette l_j augmente si les vecteurs de probabilité de ses voisins sont compatibles (et inversement).

$$P^{(t+1)}(n_i = l_j) = \frac{P^{(t)}(n_i = l_j)\ Q_i^{(t)}(j)}{\sum_{k=1}^{l} p^{(t)}(n_i = l_k)\ Q_i^{(t)}(k)} \tag{2.7}$$

$P_i^{(t)}(k)$ représente le vecteur de probabilité d'assignation : il représente la probabilité d'assignation du noeud n_i par létiquette l_k à l'itération t. Etant donné $P_i^{(0)}(k)$, une estimation initiale de la mise en correspondance, cette probabilité initiale est mise à jour gràce à la fonction de compatibilité $Q_i^{(t)}(k)$ calculée à partir des probabilités conditionnelles.

$$Q_i^{(t)}(k) = \sum_{j \in V(n_i)} w_{ij} \left[\sum_{k'=1}^{l} p(i = k | j = k') p^{(t)}(j = k') \right] \tag{2.8}$$

Où :

w_{ij} sont des poids qui indiquent l'influence du voisin n_j du noeud n_i.

D'un point de vue de la mise en correspondance, le grand intérêt de la relaxation est de pouvoir faire intervenir deux types d'informations :

- Des mesures locales sur les primitives : les probabilités d'assignation découlent directement des mesures de distance entre vecteurs (cf. paragraphe 2.3.2).
- Des mesures de nature géométrique : les probabilités conditionnelles et la propagation bayesienne de ces probabilités intègrent naturellement les propriétes géométriques de la scène.

2.3.6 Techniques de votes

Les techniques de votes dérivent des méthodes de détection de formes basées sur la transformation de Hough [38], [21]. Considérons une scène rigide contenant plusieurs

objets, observée par une caméra en mouvement, alors pour un objet donné se déplaçant avec une vitesse propre nous allons observer un ensemble vecteurs mouvement identiques (ou proches), correspondant aux parties de l'objet mises en correspondance (par exemple des points d'intérêt).

Il est alors possible d'utiliser un tableau accumulateur de vecteurs mouvement, après une mise en correspondance, pour tenter d'éliminer des appariements isolés dans l'espace des vitesses. En pratique ceci est réalisé par simple seuillage de l'accumulateur.

2.3.7 Mise en correspondance dense

La mise en correspondance dense consiste à essayer de mettre en relation tous les pixels de deux images en stéréo-vision ou en mouvement.

- En stéréo-vision, la connaissance de la géométrie épipolaire, permet d'effectuer une mise en correspondance dense. Il existe pour cela de nombreuses méthodes basées sur la corrélation [36], des méthodes par programmation dynamique [104] [82] (graph cuts) ou encore des méthodes par minimisation d'énergie [1].
- En mouvement, on trouvera de nombreux travaux liés au flot optique [37].

2.4 Changements de point de vue et transformations géométriques de l'image

Nous avons examiné au paragraphe 2.2 les principales transformation colorimétriques qui pouvaient survenir dans les problèmes de mise en correspondance d'image, nous allons maintenant nous focaliser sur les différentes transformations géométriques que nous pourrons rencontrer.

Soient I_1 et I_2 deux vues d'une même scène prises dans des conditions différentes : Une transformation géométrique de l'image est une fonction des coordonnées de l'image :

$$I_2(x, y) = I_1(\vec{g}(x, y)) \tag{2.9}$$

dans laquelle la fonction $\vec{g}$ est une fonction bi-dimentionnelle des coordonnées image. Il existe de nombreux types de transformations linéaires ou non linéaires comme par exemple une simple rotation de l'image ou une transformation plus complexe comme un morphing. Nous retiendrons principalement les transformations linéaires, qui sont les plus courantes en stéréo-vision ou en mouvement : rotation, translation, transformation affines ou perspectives.

2.4.1 Rotation image

La transformation géométrique de rotation de l'image est définie par :

$$I_\theta(x,y) = I\left(\left(\begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right)^t\right) \tag{2.10}$$

Où θ, représente l'angle de rotation appliqué à l'image

2.4.2 Translation image

La transformation géométrique de translation de l'image est définie par :

$$I_2(x,y) = I_1(x+T_x, y+T_y)$$

2.4.3 Changement d'échelle

Le changement d'échelle est lié à la taille d'observation des structures, un zoom est un exemple typique de changement d'échelle, mais dans notre cas, nous rencontrerons plus fréquemment un rapprochement (ou un éloignement) de la caméra dans la scène. On notera tout de même que d'un point de vue physique, le zoom ou le rapprochement de la caméra n'est pas équivalent : les différents plans ne sont pas traités de la même manière.

Nous considérons ici des scènes planes perpendiculaires à l'axe optique de la caméra, alors le changement d'échelle s'écrit simplement par un coefficient de proportionnalité entre les coordonnées de l'image I_1 et celles de l'image I_2 :

$$I_2(x,y) = I_1\left(\left(\begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right)^t\right)$$

2.4.4 Transformations affines et projectives

Nous avons vu que le changement d'échelle permettait d'approximer en partie certains changements de point de vue (zoom ou rapprochement de la caméra), lorsque la caméra tourne ou se déplace par exemple latéralement dans la scène (ou les deux), il est alors nécessaire de faire appel à des transformations algébriques plus complexes pour rendre compte de certains mouvements de caméra.

Dans un changement de point de vue ne sont plus conservées, ni les longueurs, ni les angles. Par exemple la longueur d'une arête d'un objet n'est pas conservée, l'angle relatif entre deux arêtes d'un objet, n'est pas non plus conservé, notament le paralléllisme ne l'est pas. Evidement, cette transformation dépend de la distance de l'objet à la caméra, un objet lointain sera moins affecté qu'un objet proche.

Dans le cas de scènes planaires, les transformations projectives vont permettre de modéliser ce type de déplacement de caméra, en effet, la caméra effectue une projection perspective (qui est une transformation projective particulière) et la combinaison

d'une tranformation projective et d'une transformation perspective est une transformation projective (un plan est transformé en un plan, mais ni les distances ni les angles ne sont conservés).

Dans ce cas de scènes planaires, connaissant quatre points (coplanaires) non alignés vus à la fois dans les deux images, ces points forment une base projective du plan. Les méthodes de transfert basées sur les invariants projectifs et bases projectives sont abondamment décrits dans [32]. Il est alors possible de calculer des coordonnées projectives d'un point quelconque du plan (données par trois invariants projectifs : k_1, k_2, k_3). Si les points de la base projective sont appariés dans les deux images les coordonnées projectives calculées dans la première image peuvent être utilisées pour effectuer le transfert du point dans la seconde image. La figure 2.13 illustre une transformation projective entre deux images.

Dans le cas général où la scène n'est pas planaire, il n'est évidement pas possible de trouver une transformation projective transformant la première image en la deuxième. Il n'existe pas d'invariant dans le cas général.

D'un point de vue pratique, par exemple au niveau d'un simple point d'intérêt, il est impossible de retrouver une éventuelle transformation projective (trop de degrés de liberté). Dans ce cas il est fréquent de modéliser localement cette transformation par une transformation affine plus simple. C'est la transformation la plus complexe pouvant être estimée et éliminée localement, par exemple, la courbure affine est invariante par toute transformation affine.

FIGURE 2.13: Exemple transformation projective entre deux images, ni les distances ni les angles ne sont conservés.

2.4.4.1 Importance de l'invariance affine

Dès lors que le changement de point de vue est important, les transformations de l'image au niveau local ne sont plus assimilables à de simples transformations euclidiennes

(rotation, translation), nous verrons que ces transformations influencent fortement les descripteurs des primitives. Au niveau d'une face plane d'un objet, nous aurons une transformation projective. D'un point de vue global sur la scène, nous chercherons à approximer localement ces transformations complexes par des transformations affines. La recherche de l'invariance affine dans un détecteur/descripteur de point est donc une qualité importante abondamment discutée dans la littérature.

2.5 Quelques techniques de mise en correspondance

2.5.0.2 Mise en correspondance globale

Ces méthodes sont principalement utilisées en indexation d'images pour retrouver des images aux couleurs ou aux contenus proches. Nous avons déja vu au paragraphe 2.3.2 certaines mesures de distances qui s'appliquent aux images entières [14]. Nous n'entrerons pas ici dans les détails de l'indexation d'image, dans la mesure où nous nous intéressons à des méthodes de mises en correspondances exactes et précises des structures des images.

2.5.0.3 Mise en correspondance de régions

Dans le contexte qui nous intéresse, nous trouvons les travaux de Matas et al. [67], [68], il définissent la notion de régions stables par changement de point de vue : "Maximally Stable Extremal Regions" (MSER).

Régions :

Les régions sont obtenues de la manière suivante :

1. Les pixels de l'image sont tout d'abord triés par intensité décroissante.
2. Ces pixels sont alors replacés dans l'image par intensité décroissante et les régions sont alors étiquetées par composantes connexes.
3. Les régions les plus stables dans l'espace des luminosités sont retenues.

Ces régions peuvent être obtenues en temps quasi-linéaire [79]. L'algorithme est donc très efficace en temps de calcul.

A partir de ces régions, des zones de mesure sont déterminées. Pour une région donnée les zones de mesure à différentes échelles sont déterminée à partir de l'enveloppe convexe de cette région (1.5, 2 et 3 fois la taille de l'enveloppe convexe).

Descripteurs :

Des descripteurs invariants par rotation sont calculés àprès avoir appliqué aux régions une transformation qui diagonalise la matrice de covariance. Il s'agit d'une description

affine des régions.

Mise en correspondance :

La méthode de mise en correspondance compare les mesures aux différentes échelles des régions des deux images puis utilise une technique de vote pour éliminer les faux appariements. La figure 2.14 présente un résultat de mise en correspondance tiré de [67].

FIGURE 2.14: Mise en correspondance de régions par MSER.
[Résultat d'après l'article [67]]

2.5.0.4 Mise en correspondance de contours ou segments

Les contours dans deux images différentes peuvent difficilement être mis en correspondance directement, en effet, les contours sont relativement sensibles aux conditions d'illumination et contrairement aux périmètres de régions (cf. paragraphe 2.5.0.2) sont rarement fermés. La plupart des travaux se sont donc tournés vers la mise en correspondance de segments de droite provenant d'approximation polygonale de contours.

La plupart de ces travaux ont eu lieu dans un contexte de robotique et de scènes d'intérieur par exemple [51].

D'une part, ces approches sont difficilement utilisables en stéréo non calibrée, la détection de segments manque de précision, il est donc difficile de remonter à la géométrie épipolaire connaissant un appariement de segments.

D'autre part dès que l'on s'intéresse à des scènes peu géométriques, les segments manquent de stabilité et cette fois la mise en correspondance devient difficile.

2.5.0.5 Appariement par corrélation

Dans ces méthodes, les descripteurs de primitives sont directement formés par les pixels de l'image. Le principe est de considérer une fenêtre de pixels en général carrée,

centrée sur un point de l'image et de calculer sa corrélation avec une fenêtre de même taille dans la deuxième image. On trouvera dans [32] une description exhaustive des méthodes de calcul de scores de correlation, on notera que certaines mesures sont invariantes aux changements de luminosité décrits au paragraphe 2.2. Nous présentons ici une méthode classique de mise en correspondance dense par correlation, le cas de points d'intérêt sera abordé au paragraphe 2.5.1.

Mise en correspondance dense par correlation :

Connaissant la géométrie épipolaire entre deux images, la correlation permet d'obtenir une mise en correspondance dense : l'algorithme 1 présente cette méthode.

Algorithme 1 *: Stéréo dense par corrélation*

pour chaque pixel de l'image 1 **faire**
- Extraire un bloc de référence centré sur ce pixel
- Calculer la droite épipolaire correspondante dans l'image 2
- Rechercher le bloc le plus proche du bloc de référence le long de la doite épipolaire par correlation

fin pour

En général on effectue une mise en correspondance croisée afin de satisfaire la contrainte d'unicité, et éliminer les faux appariements. La figure 2.15 présente un résultat obtenu avec ce type de méthode.

Cependant, ces méthodes sont difficilement utilisables lorsque les transformations géométriques entre les images sont importantes, notamment la corrélation n'est pas invariante par rotation.

2.5.1 Mise en correspondance de points d'intérêt

2.5.1.1 Correlation

Nous venons de voir au paragraphe 2.5.0.5 qu'un point dans une image est porteur d'information, par exemple une petite fenêtre centrée sur ce point va décrire l'information locale de l'image. Cette information nous permet par simple correlation de retrouver ce point dans une autre image. Contrairement au cas décrit au paragraphe 2.5.0.5 (algorithme 1), où il s'agissait de mise en correspondance dense en stéréo-vision connaissant la géométrie épipolaire, nous avons ici des points dans les deux images, et la recherche d'appariement est réalisée sur les points d'intérêt.


a)


b)

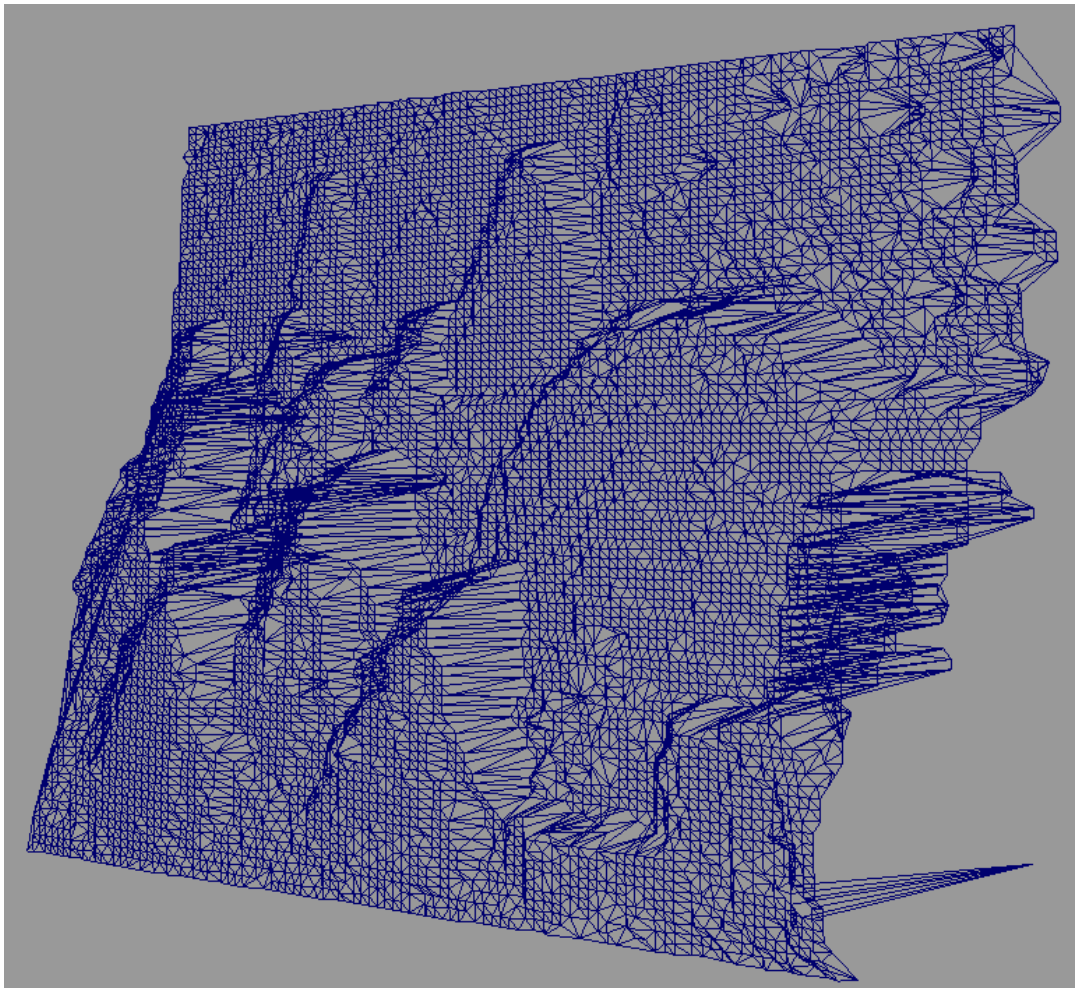

c)

FIGURE 2.15: Mise en correspondance dense par corrélation. Les images a) et b) sont respectivement l'image droite et gauche du couple stéréoscopique, la figure c) représente les résultats obtenus après appariement et recostruction 3D (1 pixel sur 64 est reconstruit pour raison d'affichage)

Ce type de méthode est utilisable à la fois dans un cadre de stéréo-vision non calibrée et dans un cadre de mouvement (cf. figure 2.9) afin d'estimer directement le mouvement sans passer par le flot optique. En stéréo-vision, dans les cas où la calibration est inconnue, le but est en général d'estimer la géométrie épipolaire (inconnue) [59] par appariement de points particuliers : des points d'intérêt.

2.5.1.2 Invariants différentiels euclidiens

Dans [32] les points d'intérêt utilisés sont calculés à l'aide d'une variante couleur de la méthode de Harris [10]. Les points sont décrits par un ensemble d'invariants euclidiens (8 invariants à l'ordre 1, 17 invariants à l'ordre 2) La mise en correspondance est effectuée en deux étapes :

- Calcul de scores d'appariement entre les points de l'image 1 et ceux de l'image 2. Ces scores sont obtenus à l'aide d'un calcul de distance entre vecteurs d'invariants (de dimension 8 ou 17).

- Mise en correspondance par relaxation pour d'une part imposer la contrainte d'unicité et d'autre part éliminer les faux appariements.

2.5.1.3 Descriptions invariantes

Les méthodes précédentes décrites paragraphes 2.5.0.5, 2.5.1.1 et 2.5.1.2, sont difficilement utilisables dans des cas complexes, notamment lorsque le changement de point de vue entre les deux images est important.

- La corrélation de fenêtres n'est pas invariante par rotation.

- Les méthodes par invariants différentiels euclidiens, d'une part, ne sont pas invariantes par transformation affine, et d'autre part, les caractérisations obtenues sont relativement pauvres et nécessitent en général une étape de relaxation pour obtenir des résultats et cela uniquement dans les cas où la transformation peut être approximée par une transformation euclidienne.

- Enfin le détecteur de points d'intérêt doit présenter des propriétés de répétabilité. ni l'opérateur de Harris [10], ni sa variante couleur [32] ne possèdent cette propriété par changement de point de vue important. Cela est une caractéristique fondamentale pour pouvoir comparer les mêmes points vus dans deux images.

Dans le cadre de l'indexation d'images et de la recherche d'objet (un objet peut être décrit par un ensemble d'images), de nombreux travaux ont donc été consacrés à la

recherche de détecteurs et de caractérisations affines. Ces travaux visent à s'affranchir des problèmes de déformations dùs aux changements de point de vue, aux changements d'échelle, mais aussi aux variations d'intensité lumineuse ou de couleurs.

2.6 Conclusion

Nous avons abordé dans ce chapitre le problème de la mise en correspondance entre deux ou plusieurs images. Nous avons abordé les problèmes des contraintes géométriques liées aux applications, les problèmes d'invariance géométriques et colorimétriques, puis les différentes techniques utilisées pour la mise en correspondance.

Nous avons pu établir que les méthodes de mise en correspondance par points d'intérêt présentaient des caractéristiques essentielles pour notre problème. Ces méthodes d'appariement seront décrites et développées au chapitre suivant.

Chapitre 3

Extraction de points d'intérêt

Nous abordons dans ce chapitre la détection et la description de points d'intérêt dans une image. Cette méthode est de loin la plus utilisée pour la mise en correspondance d'images ; elle consiste à extraire un ensemble limité de points reconnaissables dans une image, parfaitement localisés et identifiables individuellement (on parle de répétabilité), pour pouvoir ensuite apparier le même point d'intérêt sur deux images différentes.

Les méthodes de mise en correspondance distinguent en général l'étape de détection de points d'intérêt de celle de la description. Dans une première étape, les points d'intérêt sont détectés ainsi généralement qu'une région autour de ces points, puis est associé à chaque région un descripteur. Détecteurs et descripteurs doivent être aussi invariants que possible. Pour cette raison, on les dénomme parfois invariants locaux (local invariant features), bien que ce terme désigne plus particulièrement le descripteur.

Ce que le descripteur représente réellement ne présente pas forcément d'intérêt en soi, du moment que la position du point dans l'image est connue de façon précise et stable. C'est cette particularité qui va être utilisée pour des opérations de calibration de caméra ou de reconstruction 3D, ou encore pour l'alignement ou la fusion d'images.

Mentionnons par ailleurs un autre type d'application, qui consiste à utiliser un ensemble de points d'intérêt comme la représentation robuste d'une image. Ceci permet la reconnaissance directe d'objets ou de scènes sans passer par une étape de segmentation ni de mise en correspondance. Là encore, ce que représente le descripteur importe peu, ni même sa localisation, le but étant d'analyser les statistiques de l'ensemble des points d'intérêt [92], et d'opérer une détection d'objets ou une recherche d'images par leur contenu.

Idéalement, un point d'intérêt devrait être un point unique dans une image. En pratique, il n'est pas toujours possible ou souhaitable de se limiter à des coordonnées géométriques ponctuelles rigoureuses, d'une part en raison de la nature discrète des images, qui pose des problèmes d'échantillonnage et de localisation, et d'autre part en raison de la nécessité de caractériser non seulement le point mais aussi son voisinage.

Dans certaines applications comme la calibration de caméras, seul le point d'intérêt est utilisé, mais la plupart du temps il est nécessaire d'obtenir aussi une description de son voisinage local, comme par exemple de sa taille ou de sa forme. Dans ce cas, la notion de point d'intérêt est étendue à une région d'intérêt, ou un *blob* , ou plus généralement une caractéristique locale.

Pour être jugés *intéressants* et utilisables, les points détectés doivent fournir des caractéristiques locales invariantes :

- locales : les points doivent posséder des propriétés locales remarquables, identifiables d'une image à l'autre (coins, pointes, jonctions, etc.)
- invariantes à des changements de point de vue : translation, rotation, changement d'échelle, transformation affine
- invariantes à des changements d'illumination : changement affine d'intensité

La notion intuitive d'invariance demande toutefois à être précisée. En toute rigueur, si la caractéristique d'un objet est invariante, alors sa valeur reste inchangée lorsque l'objet subit une transformation. Par exemple, la surface d'un objet dans une image est clairement invariante, restant identique lorsque l'on fait subir à l'image une rotation. En revanche, l'orientation de l'objet a varié au cours de la rotation, et de façon identique à la rotation de l'image : on parle alors de covariance. Les invariances à des transformations géométriques (figure 3.1) sont donc en réalité des covariances. D'un autre côté, la plupart des descripteurs utilisent l'opération de normalisation, ce qui ramène à la notion d'invariance.

Dans la suite, nous ne ferons pas la distinction entre covariance et invariance, le terme invariance étant le plus largement utilisé

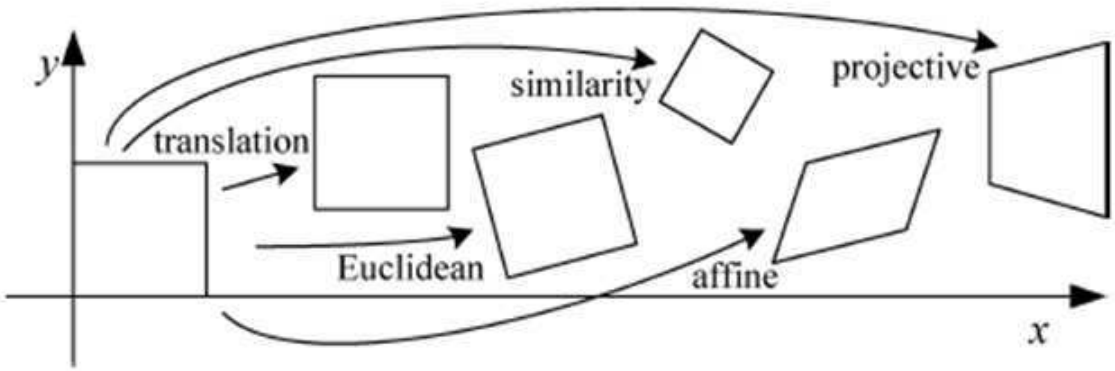


FIGURE 3.1: Transformations géométriques 2D.

Un état de l'art extrêmement complet et relativement détaillé a été mené par Tuyte-

laars et Mikolajczyk [71]. Le lecteur intéressé pourra s'y référer utilement.

3.1 La détection de points d'intérêt

Historiquement, les méthodes de détection de points d'intérêt se sont fondées sur l'analyse des contours et des arêtes, c'est-à-dire les zones où l'intensité de l'image comporte une discontinuité. Les détecteurs de contours (par exemple Canny, Sobel) effectuent une dérivation de l'image, vue comme une fonction de deux variables x et y, à valeurs dans $\Re$ pour les images en niveau de gris, ou dans $\Re^3$ pour les images couleur. Les algorithmes de détection de points d'intérêt se focalisent sur des points de contours particuliers.

Ainsi, les coins sont des points de l'image où le contour change brutalement de direction et sont des points particulièrement stables et intéressants pour la répétabilité.

Une méthode simple pour détecter des coins mais coûteuse en calculs et non-optimale est d'utiliser la corrélation. Aujourd'hui, la méthode la plus répandue est celle de Harris (ou Harris-Stephens), qui repose sur une analyse locale de l'image à l'ordre 2. D'autres techniques similaires se différencient par l'opérateur de dérivation utilisé, par exemple l'opérateur DoG (Difference of Gaussians), LoG (Laplacian of Gaussians), ou DoH (Difference of Hessians).

Nous nous limiterons ici à la présentation du détecteur de Harris et celui de la méthode SUSAN. Ces deux détecteurs sont invariants à la rotation (invariance euclidienne).

3.1.1 Le détecteur de Harris

Soit I une image d'intensité en niveau de gris. Le principe utilisé à l'origine par Moravec [77] consiste à calculer en chaque point l'auto-corrélation locale :

$$E(x,y) = \sum_{u,v} G(u,v)\,|I(x+u,y+v) - I(u,v)|^2$$

où $G(u,v)$ définit un voisinage autour du point (x,y), classiquement $G = 1$ dans un carré autour de(x,y). Un coin (ou plus généralement un point d'intérêt) est caractérisé par une grande variation de E dans toutes les directions.

Utilisant un développement de Taylor, Harris et Stephens [10] ont montré que la fonction E peut s'exprimer sous la forme quadratique :

$$E(x,y) = (x,y)M(x,y)^t$$

avec :

$$M = G * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{3.1}$$

où I_x et I_y sont les dérivées premières de l'image, obtenues après un filtrage gaussien, et G une fenêtre gaussienne et isotrope.

M est la matrice des moments d'ordre 2, aussi appelée matrice d'auto-corrélation, elle représente la variation locale de l'image en un point.

La détection d'un point d'intérêt s'appuie alors sur le calcul des valeurs propres de la matrice M, qui sont les courbures principales ; est considèré comme point d'intérêt un point possédant des courbures fortes dans les 2 directions orthogonales, donc des valeurs propres importantes. On évitera cependant le calcul explicite des valeurs propres en formant l'opérateur :

$$R = Det(M) - k\,trace^2(M) \tag{3.2}$$

$Det(M)$ et $trace(M)$ sont en effet le produit et la somme des 2 valeurs propres, k est une constante empirique, typiquement k est compris entre $0,04$ et $0,06$.

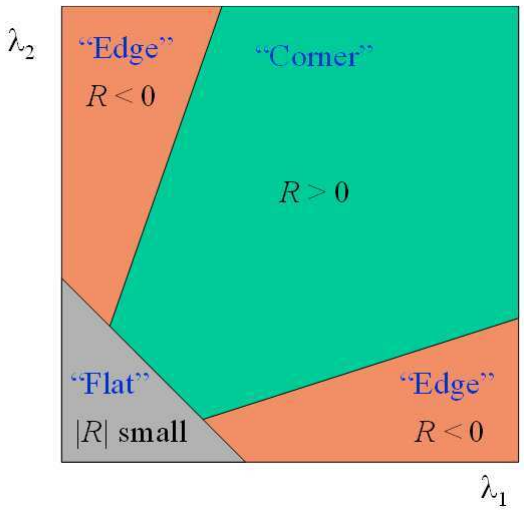


FIGURE 3.2: Mesure de Harris aux contours, aux coins et aux régions homogènes selon les valeurs de R et des deux valeurs propres.

Les valeurs de R sont alors positives au voisinage d'un coin, négatives au voisinage d'un contour, et proches de 0 dans une région homogène, la figure 3.2 illustre les résultats

de la classification dans le plan des deux valeurs propres. Les coins ou points d'intérêt sont les maxima locaux de la fonction R.

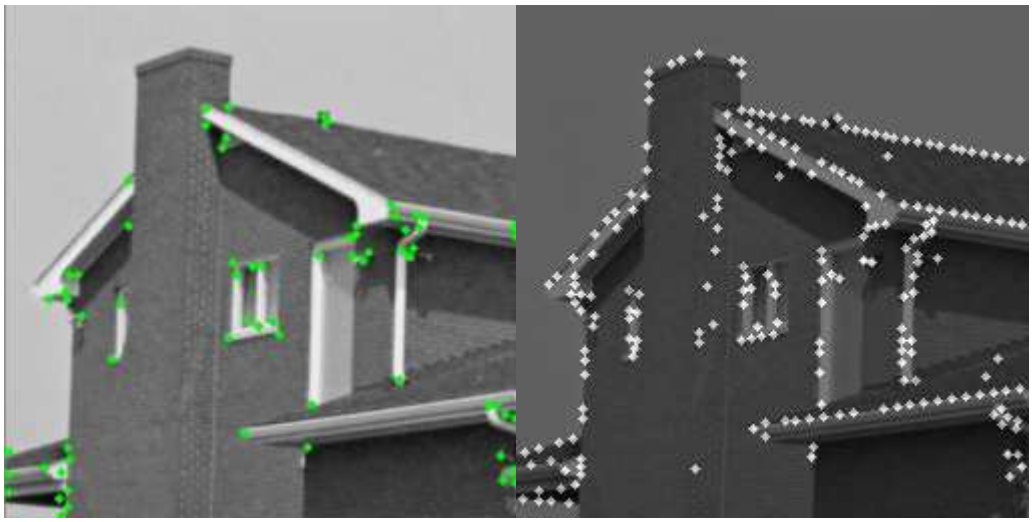


FIGURE 3.3: Détection de points d'intérêt avec les détecteur de Harris et SUSAN.

Une évaluation du détecteur de Harris dans une version améliorée [94] fait ressortir l'invariance du détecteur à une rotation, son invariance partielle à un changement affine d'illumination, et enfin sa faible robustesse à un changement d'échelle.

Cette dernière propriété a suscité des travaux pour adapter le détecteur au cas multi-échelle et aboutir au détecteur Harris-Laplace, présenté plus loin au paragraphe 3.2.1.

3.1.2 SUSAN

La méthode SUSAN (Smallest Univalue Segment Assimilating Nucleus) [99] consiste à placer sur un pixel à tester (le noyau) un masque circulaire. Pour chaque point m du masque autour de m_0 on calcule une fonction de forme rectangulaire :

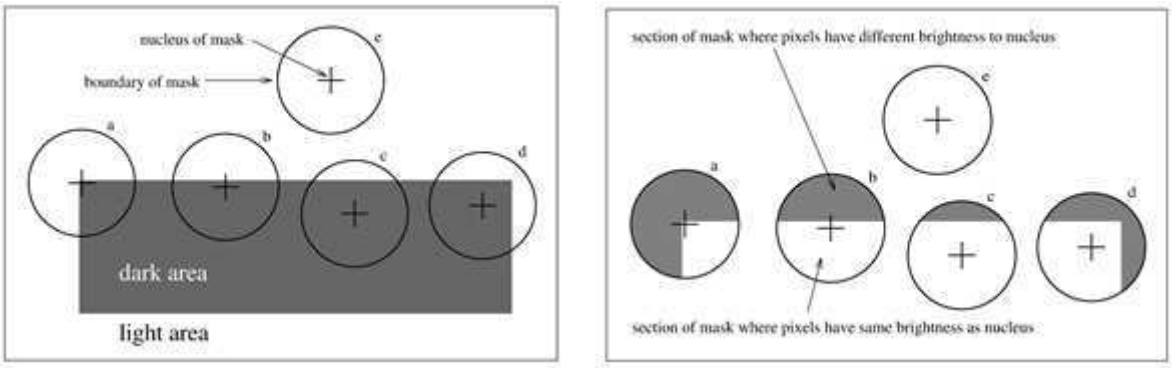
$$C(m) = e^{-\left(\frac{I(m)-I(m_0)}{t}\right)^6}$$

où I est l'intensité, et t le rayon du disque.

Soit $n(m_0)$ la somme de la fonction sur les points du masque (région *USAN* de points de même intensité), alors la réponse du détecteur est la quantité :

$$n(m) \;=\; \sum_m c(m)$$

Le paramétre *g, seuil géométrique*, détermine la taille minimale du segment recherché. Si g est grand (de l'ordre de 50% de la taille du masque), SUSAN détecte un contour. Avec un g plus faible (25%), un coin est détecté, sa position est calculée comme celle du centroïde de la région USAN détectée.



a b

FIGURE 3.4: Méthode SUSAN : a) exemple de points d'intérêt sur une image synthétique, b) les régions USAN à l'intérieur d'un masque sont les zones claires

3.2 Les méthodes à invariance d'échelle

Les méthodes multi échelles prennent en compte des changements d'échelle identiques dans toutes les directions, avec une certaine robustesse à de faibles déformations affines. Leur principe est de rechercher des extrema locaux d'une fonction dans un espace à 3 dimensions (x, y, échelle), selon l'approche proposée par Lindeberg [53]. Pour cela on construit un espace échelle en lissant l'image avec des noyaux gaussiens de différentes tailles. Plusieurs fonctions ont été proposées : le détecteur Harris-Laplacien [70], qui associe au détecteur de Harris un opérateur Laplacien multi échelle, le détecteur SIFT [54], où l'opérateur DoG (Difference of Gaussians) est utilisé à la fois dans le domaine spatial et la dimension échelle.

3.2.1 Le détecteur Harris-Laplace

La méthode due à Mikolajczyk et Schmid [70] part d'une formulation du détecteur de Harris adaptée aux changements d'échelle :

$$M(\sigma_{l,}\sigma_D) = \sigma_D^2 G(\sigma_l) * \begin{bmatrix} I_x^2(\sigma_{\mathbf{D}}) & I_x I_y(\sigma_{\mathbf{D}}) \\ I_x I_y(\sigma_{\mathbf{D}}) & I_y^2(\sigma_{\mathbf{D}}) \end{bmatrix}$$

où l'on distingue cette fois :

σ_D paramètre de la fenêtre gaussienne de calcul des dérivées (paramètre d'échelle de dérivation)

σ_l paramètre de lissage gaussien de la matrice (paramètre d'échelle d'intégration)

L'idée est alors de rechercher le paramètre d'échelle caractéristique pour lequel une certaine fonction est maximisée.

Pour les auteurs, la fonction LoG (Laplacian of Gaussians) s'avère donner les meilleurs résultats. Pour un paramètre de noyau Gaussien σ_n, on définit la fonction LoG :

$$|LoG(\sigma_n)| = \sigma_n^2 |I_{xx}(\sigma_n) + I_{yy}(\sigma_n)| \tag{3.3}$$

La recherche du point d'intérêt se fait de façon itérative, en calculant la fonction LoG sur un ensemble discret de valeurs :

$$\sigma_n = \xi^n \sigma_0,\ \sigma_1 = \sigma_n,\ \sigma_D = s\sigma_l,$$

avec typiquement $\xi = 1.4$ et $s = 0.7$ et en recherchant le maximum local dans un voisinage du point considéré.

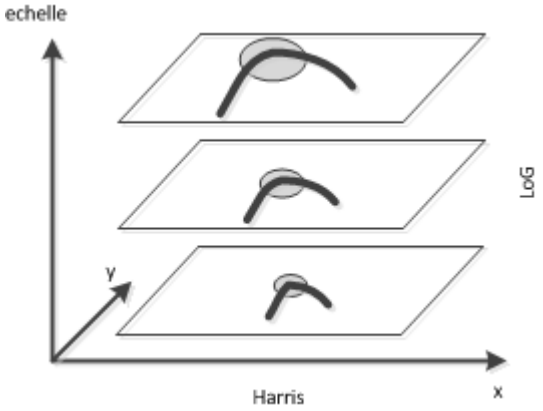


FIGURE 3.5: Détecteur multi échelle Harris-Laplace.

Le détecteur Harris Laplace est bien adapté pour la localisation de régions d'intérêt (blobs, ou binary large objects) ou de structures comme les coins, contours, jonctions.

3.2.2 SIFT

La méthode SIFT (Scale Invariant Feature Transform) a été élaborée par D. Lowe [54][55] afin de résoudre la mise en correspondance de différentes images d'une même scène. La méthode comporte ainsi deux parties : un algorithme de détection de points d'intérêt et de calcul de descripteurs, et un algorithme de mise en correspondance proprement dit. Les points d'intérêt détectés sont invariants aux rotations et aux changements d'échelles, et partiellement invariants aux changements d'illumination et de point de vue 3D. Selon le principe adopté précédemment, la détection de points d'intérêts recherche des extrema dans l'espace à 3 dimensions (x, y, échelle).

Considérons à nouveau une image I filtrée par une fenêtre gaussienne : $l(\sigma) = G(\sigma) * I$

L'opérateur DoG, ou différence de gaussiennes, est défini par :

$$D(\sigma) = l(k\sigma) - l(\sigma)$$

k étant une constante multiplicative. Cet opérateur fournit une bonne approximation de l'opérateur LoG, ou Laplacien de gaussiennes normalisé, tout en offrant une simplicité des calculs.

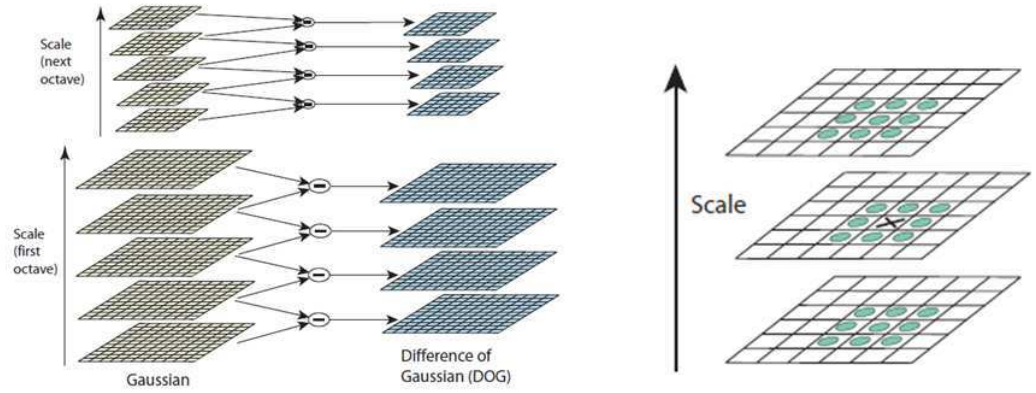
Pour évaluer l'opérateur DoG, l'espace échelle est discrétisé en octaves, correspondant à un doublement du facteur d'échelle σ, chaque octave étant divisée en s intervalles, de sorte que l'on a $k = 2^{1/s}$. En pratique, Lowe utilise la valeur $s = 3$.

Les extrema locaux de la fonction autour d'un point sont ensuite recherchés parmi les 8 voisins de la même échelle, ainsi que les 18 voisins que comptent les échelles immédiatement inférieure et supérieure (Figure 3.6).

Un ajustement consiste ensuite à éliminer les points d'intérêt présentant une valeur faible (points de faible contraste) ou situés sur des contours. En chaque point d'intérêt détecté, on calcule ensuite l'amplitude et l'orientation du gradient. La dernière étape consiste à former le descripteur de chaque point d'intérêt, elle sera décrite au paragraphe 3.2.2.

L'approche SIFT est aujourd'hui l'une des plus utilisées, donnant de bonnes performances dans de nombreuses applications. Sa robustesse lui permet de tolérer des changements de points de vue importants, bien que ses descripteurs n'aient pas la propriété d'invariance affine.

Depuis les travaux de Lowe, de nombreuses variantes et extensions de la méthode ont vu le jour, telles PCA-SIFT [44], GLOH [72] ou SURF [4], dans le but d'améliorer sa robustesse et faire décroître sa complexité. Le paragraphe 3.3 présente ces améliorations,



a b

FIGURE 3.6: Détection dans l'espace échelle : a) images DoG issues d'une pyramide de Gaussiennes par octaves b)obtention des extrema locaux de la fonction DoG par comparaison d'un point (marqué X) à ses 26 voisins dans l'espace échelle. D'après [54]

nous décrirons au paragraphe 3.4 les descripteurs associés à ces méthodes.

3.3 Les méthodes à invariance affine

Dans une transformation affine, le changement d'échelle n'est pas identique dans toutes les directions, et les méthodes multi échelle introduisent des erreurs lors de la mise en correspondance des points d'intérêt. Différentes approches ont été proposées pour préserver l'invariance affine.

3.3.1 Le détecteur Harris Affine

Mikolajczyk et Schmid [71] ont étendu leur détecteur Harris Laplace pour l'adapter aux transformations affines. La méthode est itérative et comporte les étapes suivantes :

1. Initialisation des points d'intérêt à l'aide du détecteur Harris Laplace
2. Normalisation de la région autour de chacun des points par une adaptation affine
3. Estimation itérative de la région affine : localisation, paramètre échelle d'intégration, paramètre échelle de dérivation
4. Répétition de l'étape 3 jusqu'à la condition d'arrêt. Une approche un peu plus heuristique pour obtenir l'invariance affine consiste à exploiter la géométrie des contours que l'on trouve généralement à proximité d'un point de Harris. Une telle méthode a été proposée par Tuytelaars et Van Gool [71] [103] sous le nom de Edge

Based Regions. Le principe est de considérer un coin de Harris et deux points de contour proches, extraits à l'aide d'un détecteur de contour de Canny [11] (Figure 3.7). Ces points définissent une famille de régions en forme de parallélogrammes.

L'étape 1 prend comme points de départ les points d'intérêt du détecteur de Harris multi échelle. Ces points sont calculés à l'aide de changements d'échelle isotropiques, cependant dans le cas de transformations affines le changement d'échelle est anisotropique, ce qui occasionne des erreurs dans la localisation des points. Il convient alors de calculer la matrice M des moments d'ordre 2 (éq. 3.3) dans l'espace échelle où le voisinage circulaire du point d'intérêt est remplacé par une ellipse.

Etape 2 : calculer les paramètres de l'ellipse est équivalent à modifier la matrice M pour rendre égales ses 2 valeurs propres λ_{min} et λ_{max} . C'est ce que fait cette deuxième étape en normalisant la matrice autour du point d'intérêt courant, transformant la région affine en région circulaire.

L'étape 3 réajuste successivement :

- le paramètre d'échelle d'intégration σ_l, par maximisation du laplacien normalisé,
- le paramètre d'échelle de dérivation σ_D, en sélectionnant celui qui maximise le rapport $Q = \frac{\lambda_{min}}{\lambda_{max}}$,
- la position du point d'intérêt, par maximisation de la mesure de Harris (eq. 3.2)

Enfin à l'étape 4 on évalue si la méthode a convergé, c'est-à-dire si la matrice M est assez proche d'une pure matrice de rotation. Dans ce cas, les valeurs propres sont égales, et l'on teste l'inégalité suivante, avec par exemple $\varepsilon_c = 0.05$.

$$l - \frac{\lambda_{min}}{\lambda_{max}} \tag{3.4}$$

Le détecteur Harris affine donne de très bons résultats en détection de points d'intérêt, ainsi que le montrent des exemples où les changements d'échelle et de point de vue sont importants.

3.3.2 Les détecteurs EBR et IBR

Une approche un peu plus heuristique pour obtenir l'invariance affine consiste à exploiter la géométrie des contours que l'on trouve généralement à proximité d'un point de Harris. Une telle méthode a été proposée par Tuytelaars et Van Gool [71] [103] sous le nom de Edge Based Regions. Le principe est de considérer un coin de Harris et deux points de contour proches, extraits à l'aide d'un détecteur de contour de Canny [11] (Figure 3.7).

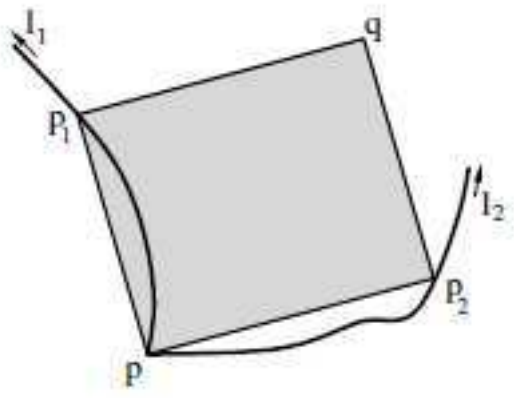


FIGURE 3.7: Méthode EBR. Les points P, P1 et P2 définissent une région d'intérêt. D'après [102].

La région retenue est celle qui maximise une fonction des moments jusqu'à l'ordre 2. On note que les régions déterminées par cette méthode sont en forme de parallélogrammes. Pour plus de commodité, les auteurs proposent de remplacer les parallélogrammes par des ellipses possédant les mêmes moments du premier et du second ordre, ce qui est une méthode de construction covariante affine.

La seconde méthode proposée par les mêmes auteurs, Intensity Based Regions, exploite l'information photométrique de l'image. Elle prend comme points de départ les extrema d'intensité, détectés sur une échelle multiple, et explore la région autour de chaque extremum de façon radiale, délimitant des régions de forme arbitraire, qui sont ensuite remplacées par des ellipses.

Etant donné un extremum local d'intensité, on étudie une fonction $f(t)$ le long de chaque rayon issu de ce point (Figure 3.8) :

$$f(t) = \frac{abs(I(t) - I_0)}{max(\frac{1}{t}\int_0^t abs(I_t) - I_0), d)} \tag{3.5}$$

où t représente un paramètre arbitraire le long du rayon,$I(t)$ l'intensité à la position t, I_0 l'intensité à l'extremum, et d une constante faible pour éviter une division par zéro.

La fonction f(t) atteint un extremum lorsque l'intensité lumineuse le long d'un rayon varie très rapidement. Tous les points correspondant aux maxima de la fonction f sont reliés et forment la région invariante qui est ensuite approximée par une ellipse.

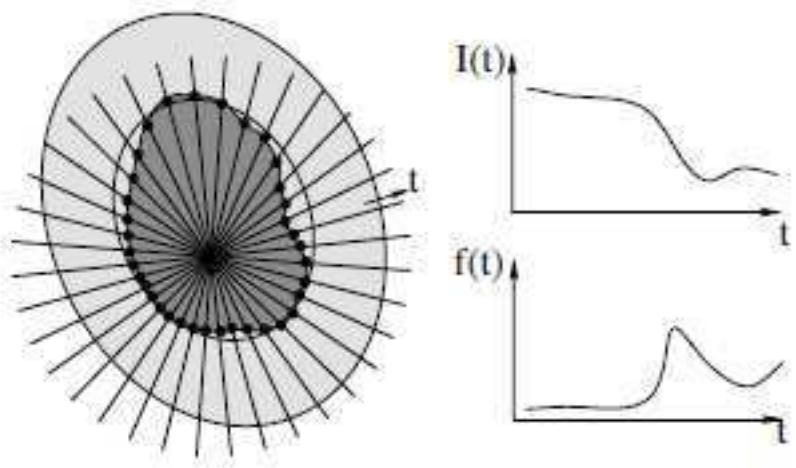


FIGURE 3.8: Méthode IBR : évaluation d'une fonction d'intensité f(t). D'après [102].

3.3.3 Le détecteur MSER

De même que le détecteur IBR mentionné précédemment, le détecteur MSER (Maximally Stable Extremal Regions) proposé par Matas et al. [68] s'attache à détecter des régions. Une MSER est un ensemble connexe de pixels qui possèdent la propriété d'être tous plus clairs ou plus foncés (d'où le terme *extremal*) que les pixels situés sur la frontière.

La méthode procède par seuillage, le terme "maximally stable" fait référence au fait que les régions détectées sont celles qui sont stables dans un large intervalle de seuils. Cette méthode se prête à des implémentations efficaces utilisant des algorithmes rapides [19] [81].

Les régions MSER ont les propriétés suivantes :

- invariance à une transformation affine d'intensité d'image
- covariance aux transformations géométriques préservant l'adjacence
- stabilité : seules les régions de même support sur une gamme de seuils sont sélectionnées
- détection multi-échelle : les structures de toutes tailles sont détectées sans aucun lissage
- l'ensemble des MSER varie au plus linéairement avec le nombre de pixels de l'image

De même que précédemment, les régions peuvent être remplacées par des ellipses possédant les mêmes moments d'ordre 1 et 2.

3.3.4 Le détecteur de régions saillantes

Le détecteur proposé par Kadir et Brady [42] est inspiré par la théorie de l'information. L'idée est de rechercher dans une image les caractères saillants (salient features), la saillance étant définie comme complexité locale ou imprédictibilité. Elle est mesurée par l'entropie des intensités d'une région locale de l'image. L'entropie seule ne permet pas de localiser précisément à différentes échelles, aussi on lui adjoint une fonction d'auto-dissimilarité dans l'espace échelle, qui favorise la localisation de structures complexes.

La détection procède en deux étapes :

- tout d'abord pour chaque pixel on évalue l'entropie $\mathcal{H}$ à l'aide de la densité de probabilité de l'intensité $p(I)$, pour un ensemble de valeurs d'échelles s.

$$\mathcal{H} = -\sum_{l} p(I) \log p(I) \tag{3.6}$$

p(I) est estimée empiriquement à partir de la distribution des intensités dans un voisinage circulaire de rayon s autour du point considéré.

- On considère ensuite l'ensemble des maxima locaux de $\mathcal{H}$. Pour ces points, on calcule la quantité $\mathcal{W}$:

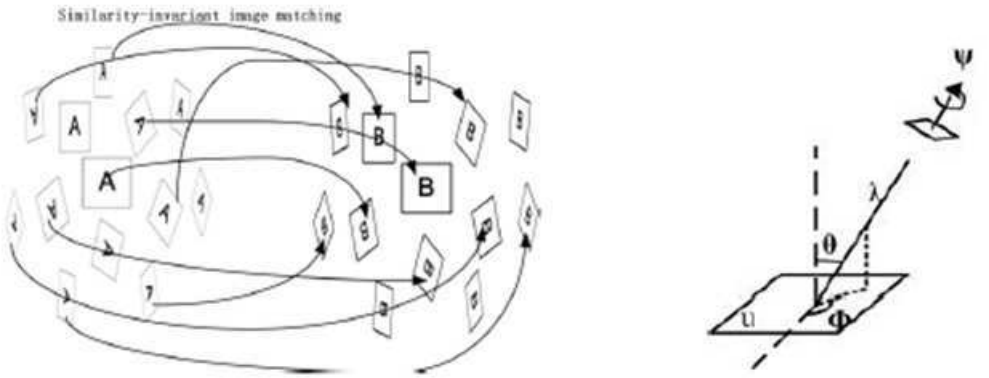
$$\mathcal{W} = \frac{s^2}{2s-1} \sum_{l} \left| \frac{\partial p(I;s)}{\partial s} \right| \tag{3.7}$$

Pour finir, la mesure de saillance est définie par $\mathcal{Y} = \mathcal{H}\mathcal{W}$. La méthode retient uniquement les régions dont la saillance est supérieure à un seuil donné.

3.3.5 ASIFT

ASIFT ou Affine-SIFT est une amélioration de la méthode SIFT pour la mise en correspondance d'images, ses auteurs Morel et Yu [78] se sont attachés à en démontrer la propriété d'invariance affine, tant mathématiquement que par des essais expérimentaux. L'idée principale est de combiner simulation et normalisation : ASIFT simule toutes les distorsions apportées par un changement de direction de l'axe optique de la caméra (Figure 3.9 - a), puis applique la méthode SIFT.

Plus précisément, ASIFT simule 3 paramètres : l'échelle, la longitude, la latitude, et normalise les 3 autres (translation et rotation). L'algorithme de comparaison se décompose en trois étapes :



a b

FIGURE 3.9: ASIFT : a) Simulation de distorsions affines et comparaison de 2 images A et B. b) Modèle géométrique de la caméra. D'après [78].

- Chaque image est transformée en simulant la distorsion affine due à un changement de direction de l'axe optique de la caméra, à partir de la position frontale. Les distorsions dépendent de deux paramètres : la longitude ϕ et la latitude θ (Figure 3.9 - b). L'image subit une rotation d'un angle ϕ suivie d'un tilt (inclinaison) de paramètre $t = |1/\cos\theta|$.
- Rotations et tilts sont discrétisés de la façon suivante : le tilt est échantillonné selon une suite géométrique $1, a, a^2, \cdots, a^n$, avec comme valeurs typiques $a = 2, n = 5$. La longitude ϕ est fixée pour chaque valeur de tilt par la suite $0, b/t, \cdots, kb/t$, avec $b \simeq 1°$, et k dernier entier tel que $kb/t < 180°$.
- Toutes les images simulées sont ensuite comparées à l'aide de SIFT.

Les résultats d'ASIFT ont été évalués en termes de mise en correspondance, et montrent d'excellentes performances en présence de changements de points de vue importants, allant jusqu'à 80°.

3.4 La description des points d'intérêt

Une fois détectés les points ou les zones d'intérêt, on cherche à les caractériser de façon discriminante au moyen d'un descripteur. Notons que certaines méthodes telles que SIFT comprennent à la fois une phase de détection et une phase d'extraction de caractéristiques.

De même que pour la détection, le détecteur doit présenter une certaine robustesse aux transformations géométriques, à des changements d'illumination, et d'autres sources de *bruit* telles que la compression.

Il existe différentes méthodes pour décrire une région dans une image, et chaque descripteur caractérise différentes propriétés telles que la couleur, la texture, les contours, etc.

La méthode de description la plus simple consiste à stocker dans un vecteur les niveaux d'intensité des pixels du voisinage, la comparaison de deux descripteurs pouvant se faire par une mesure de corrélation. C'est le cas par exemple de la méthode MOPS [9] qui utilise une description multi échelle. Cependant ces méthodes sont rapidement limitées par le volume des calculs, et ne disposant pas de l'invariance affine, elles ne fonctionnent pas pour des changements de perspective importants.

Descripteurs basés sur des distributions. Ces techniques utilisent des histogrammes pour représenter des caractéristiques d'apparence ou de forme, par exemple la distribution des intensités de pixels. Un de ces descripteurs les plus utilisés est le descripteur SIFT.

Techniques spatio-fréquentielles. Les filtres de Gabor et les transformées en ondelettes sont fréquemment utilisées pour explorer les fréquences contenues dans une image, notamment à des fins de classification de texture.

Invariants différentiels. L'ensemble des dérivées d'une région à des ordres différents (*local jet*) décrit de façon approchée le voisinage d'un point. Le calcul d'invariants différentiels à l'aide de composantes du local jet permet d'obtenir l'invariance à la rotation. Pour estimer de manière stable les dérivées, on utilise la convolution de l'image avec des dérivées gaussiennes. Freeman et Adelson [28] utilisent ces invariants dans leurs filtres orientables (steerable filters), Gouet et al. [69] proposent un descripteur pour les images couleur.

Moments généralisés. Van Gool et al. [26] ont introduit les moments généralisés pour décrire la nature multi-spectrale des données d'image. Le moment d'ordre p+q et de degré n sur une région Ω est défini par : $M_{pq}^n = \int_\Omega [I(x,y)]^n x^p y^q dxdy$. Ces moments généralisés caractérisent la forme et la distribution des intensités dans la région. Ils ont été utilisés par Tuytelaars et Van Gool pour leurs détecteurs et descripteurs EBR et IBR.

3.4.1 SIFT

Le détecteur SIFT a été décrit brièvement au 3.2.2. Le descripteur qui lui est associé est formé en calculant le gradient en chaque pixel d'une fenêtre 16x16 autour du point d'intérêt détecté, utilisant le niveau de la pyramide de gaussiennes auquel la détection a eu lieu. Les amplitudes de gradient sont pondérées par une gaussienne de lissage (le cercle sur la figure 3.10) qui réduisent l'influence des gradients loin du centre. Dans chaque quadrant 4x4, on construit un histogramme des orientations du gradient en ajoutant la

valeur pondérée à l'une des 8 orientations. Les 128 valeurs résultantes forment la version brute du vecteur descripteur de SIFT. Pour réduire les effets du contraste, le vecteur est ensuite normalisé à la longueur unité. La figure 3.10 illustre la méthode dans le cas simplifié d'une fenêtre 8x8.

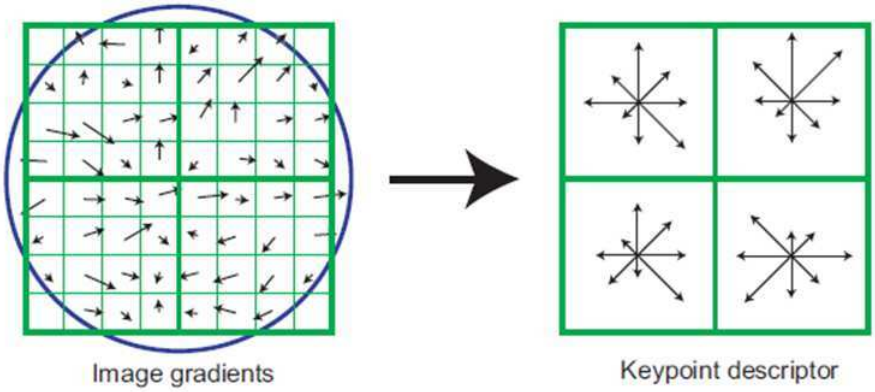


FIGURE 3.10: Descripteur SIFT : les éléments de la grille de gauche sont pondérés par une fenêtre gaussienne, puis agrégés dans les histogrammes d'orientations à droite. D'après [70].

3.4.2 PCA-SIFT

Ke et Sukthankar [44] proposent un descripteur inspiré par SIFT et d'un calcul plus simple, il calcule le gradient en x et en y sur une grille 39x39 orientée selon la direction dominante, puis procède à une réduction de la dimension du vecteur ainsi formé de 3042 à 20 en utilisant une Analyse en Composantes Principales (ACP, ou PCA en anglais). L'Analyse en Composantes Principales est une méthode statistique qui consiste à transformer des variables corrélées en nouvelles variables non corrélées entre elles, et qui forment les composantes principales, ou axes principaux d'inertie. Lorsque l'on veut compresser l'ensemble des N variables, les premiers axes constituent le meilleur choix du point de vue de l'inertie ou de la variance expliquée. La compression revient alors à ne conserver que les premiers vecteurs propres issus de la diagonalisation de la matrice de covariance. Selon les auteurs, l'utilisation de l'ACP permet de diminuer notablement le volume des calculs lors de la mise en correspondance d'autre part elle rend la méthode plus robuste à des transformations géométriques.

3.4.3 GLOH

Le descripteur *Gradient location-orientation histogram* de Mikolajczyk et Schmid est une extension du descripteur SIFT conçue pour en augmenter la robustesse et le carac-

tère discriminant. Le descripteur SIFT est calculé sur une grille en coordonnées polaires logarithmiques comportant 17 éléments locaux. Les orientations du gradient sont réparties en 16 orientations, construisant ainsi un histogramme à 272 valeurs. La taille du descripteur est alors réduite à 128 au moyen d'une ACP, comme dans la méthode PCA-SIFT.

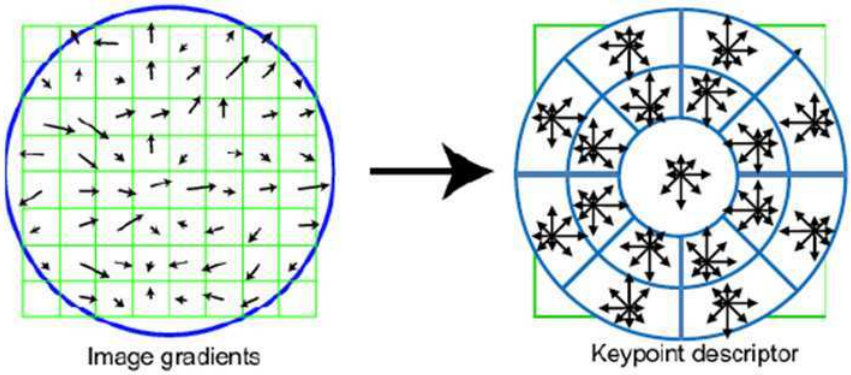


FIGURE 3.11: Le descripteur GLOH évalue les orientations du gradient sur 17 éléments en coordonnées log-polaires. D'après [72]

SURF

Autre variante de la méthode SIFT, SURF (Speed Up Robust Features) [4] vise à en améliorer les performances. Une des originalités de la méthode est d'utiliser des *images intégrales* pour accélérer les calculs, celles-ci sont obtenues en faisant la somme des intensités d'une image partielle. De même que SIFT, la méthode SURF comporte un détecteur et un descripteur de points d'intérêt. Plutôt que l'opérateur DoG utilisé dans SIFT, le détecteur défini dans SURF est construit sur un calcul approché de la matrice Hessienne. Quant au descripteur, SURF remplace le calcul du gradient par un filtrage en ondelettes de Haar au premier ordre. Dans une première étape, on calcule l'orientation des réponses des filtres, orientation selon laquelle est positionnée une région de forme carrée. Puis les réponses des filtres de Haar en x et en y, ainsi que leurs valeurs absolues, sont sommées sur des sous-régions pour former le descripteur (Figure (3.12)).

3.4.4 Shape context

Il s'agit d'un descripteur similaire à SIFT, mais basé sur les contours. Shape context [7] fournit un histogramme 3D des positions et des orientations de contours, ceux-ci sont extraits à l'aide du détecteur de Canny. Pour chacun des points de contour, on quantifie le *contexte de forme* selon 9 zones en coordonnées log-polaires, et l'orientation selon 4 directions, donnant lieu à un descripteur de dimension 36. La figure 3.13 illustre la

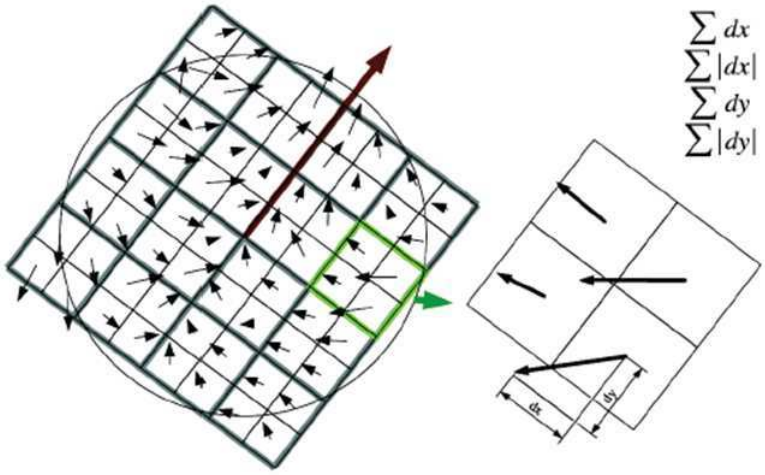


FIGURE 3.12: Le descripteur SURF calcule la somme des réponses des filtres ondelettes sur des sous-régions 4x4 d'une grille orientée. D'après [4]

méthode sur un exemple.

3.4.5 Les invariants différentiels couleur

Hilbert a montré en 1890 que tout invariant à une transformation affine pouvait s'exprimer comme un polynôme d'invariants irréductibles, ceux-ci s'exprimant comme des combinaisons des dérivées locales. Les invariants de Hilbert représentent l'ensemble de base des primitives qui permettent de décrire toutes les propriétés locales intrinsèques de l'image. Considérant l'invariance à la rotation, Gouet et al.[69] ainsi que Montesinos et al. [74] utilisent un vecteur formé de 8 invariants :

$$v(\sigma) = \begin{pmatrix} R \\ \|\nabla R\|^2 \\ V \\ \|\nabla V\|^2 \\ B \\ \|\nabla B\|^2 \\ \nabla R.\nabla V \\ \nabla R.\nabla B \end{pmatrix}$$

où R, V, B désignent les canaux respectivement rouge, vert et bleu, ∇représente l'opérateur gradient, et σ un paramètre de lissage par une fonction gaussienne. N'utilisant que les dérivées d'ordre 1, les invariants obtenus s'avèrent robustes et peu sensibles aux bruits.

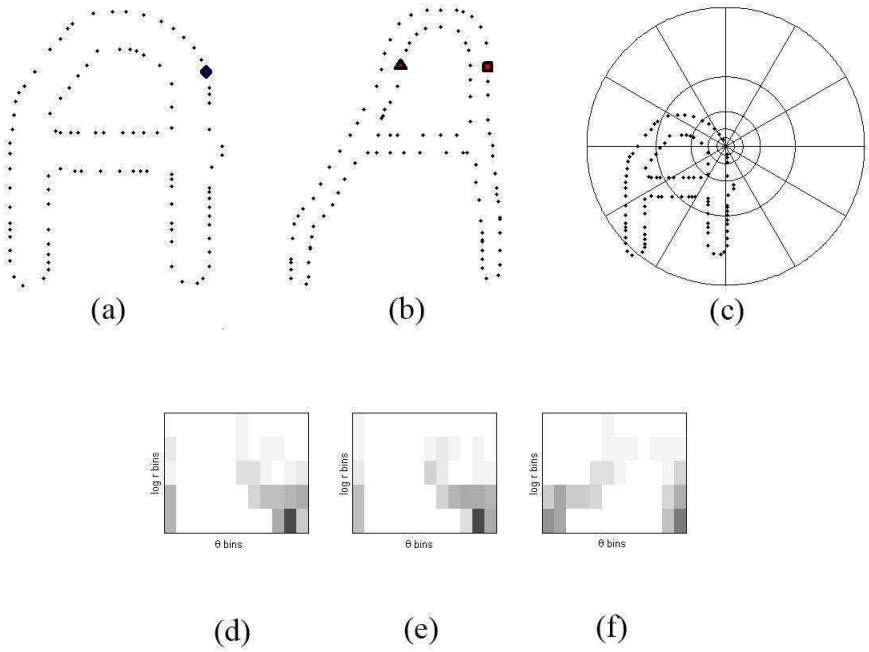


FIGURE 3.13: Shape context : les points de contours des formes (a) et (b) sont comparés en utilisant la grille log-polaire (c). Les *contextes de forme* (d) (e) (f) correspondent aux points repérés par un carré, un losange, un triangle. (d) et (e) ont des valeurs proches, (f) est très différent.

3.4.6 Les filtres orientables (steerable filters)

Les filtres orientables de Freeman et Adelson [28] permettent d'orienter les dérivées dans une direction particulière. Les orienter dans la direction du gradient les rend invariantes à la rotation.

Les filtres orientés sont souvent utiles pour l'extraction de primitives de bas niveau (contours, textures, mouvement, etc.). Freeman et Adelson [28] en proposent une construction élégante et efficace avec les filtres orientables, filtres d'orientation arbitraire qui sont calculés comme des combinaisons linéaires d'un ensemble réduit de filtres de base. En particulier, la famille de filtres suivante est orientable :

$$h(x,y) = \sum_{k=1}^{M}\sum_{i=1}^{k} \alpha_{k,i} \frac{\partial^{k-i}}{\partial x^{k-i}} \frac{\partial i}{\partial y^{i}} g(x,y) \tag{3.8}$$

Où $g(x,y)$ est une fonction isotropique arbitraire. La convolution d'une image I(x,y) avec une rotation quelconque du filtre h(x,y) peut ainsi s'exprimer comme une somme pondérée d'images filtrées :

$$I(x,y) * \left(R_\theta \begin{bmatrix} x \\ y \end{bmatrix} \right) = \sum_{k=1}^{M} \sum_{i=1}^{k} \alpha_{k,i} \frac{\partial^{k-i}}{\partial x^{k-i}} \frac{\partial i}{\partial y^i} g(x,y) \tag{3.9}$$

où R_θ est la matrice de rotation d'un angle θ.

La formulation de Freeman et Adelson est particulièrement intéressante lorsqu'elle est appliquée à une fonction $g(x,y)$ gaussienne, en raison de son caractère séparable. Dans ce cas, les filtres orientables sont synthétisés à l'aide d'un petit nombre de paires de filtres en quadrature. La figure 3.14 montre un exemple de filtres orientables G_1^θ : filtres gaussiens dérivés en x et de rotation θ.

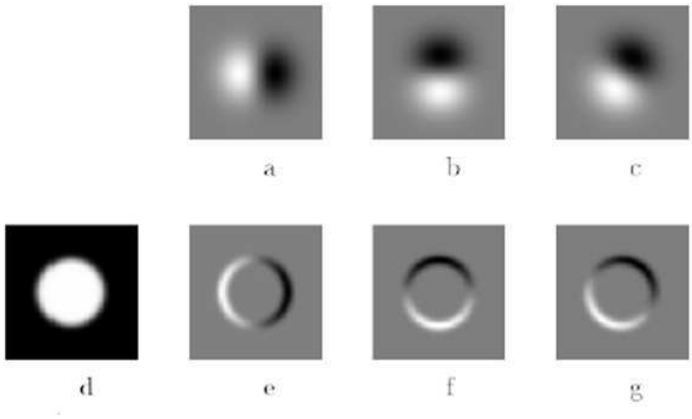


FIGURE 3.14: Filtres orientables. a)$G_1^{0^\circ}$ b)$G_1^{90^\circ}$c)$G_1^{60^\circ}$ d) image d'un disque (e)(f)(g) images de (d) aprés convolutuion avec (a)(b)(c). D'après[28].

Jacob et Unser [40] ont généralisé cette approche en optimisant les filtres orientables par rapport au critère de Canny pour détecter des contours. D'autres auteurs ont proposé à partir de ce cadre méthodologique des détecteurs sélectifs de points à caractéristiques orientées [84], [98]. Nous reviendrons au chapitre suivant sur ces filtres anisotropiques et nous détaillerons le filtre introduit par Perona [84]et repris par Knossow et al.[48].

3.4.7 DAISY

Cette méthode a été introduite récemment par Tola et al. [100][101] pour faire des appariements denses dans des images stéréo. De même que pour SIFT et GLOH, le descripteur est constitué d'un histogramme des orientations du gradient ; les cartes des directions sont pondérées par la norme du gradient, puis par des noyaux gaussiens de différentes tailles, représentés par des cercles sur la figure 3.15.

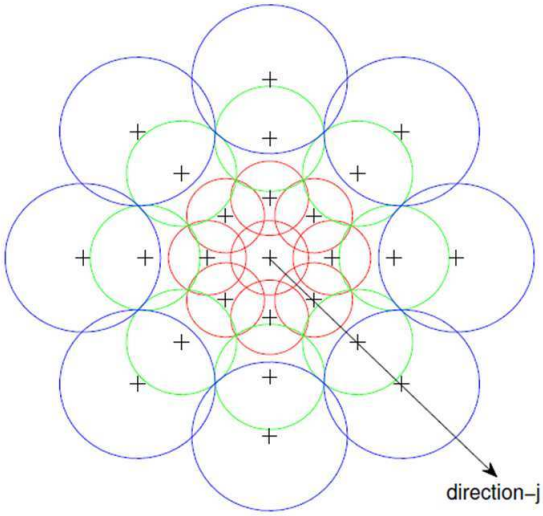


FIGURE 3.15: Le descripteur DAISY somme les orientations du gradient par convolution avec les noyaux gaussiens représentés ici. D'après [101].

La disposition des noyaux gaussiens sur une grille circulaire comme illustrée sur la figure 3.15 donne au descripteur DAISY l'apparence d'une fleur, d'où son nom. Notons que cette configuration s'appuie sur les travaux de Winder et Brown[105], qui ont analysé, optimisé et comparé différents descripteurs.

Enfin on notera que les calculs dans DAISY sont facilités par le caractère séparable des convolutions par les noyaux gaussiens.

3.4.8 La méthode Ferns

La méthode Ferns [83], [17] n'est pas à proprement parler une méthode de description locale ; il s'agit d'une méthode de mise en correspondance de zones d'intérêt dans des images, qui ne suit pas le schéma habituel détection de points d'intérêt, caractérisation et appariement. Nous la citons cependant comme une alternative à l'emploi de descripteurs locaux.

La méthode s'appuie pour cela sur des méthodes de classification statistiques qui comportent une phase d'apprentissage sur une large base de données, puis une phase de reconnaissance de la classe. Etant donnée une région située autour d'un point d'intérêt, on considère comme une seule classe l'ensemble de toutes les régions obtenues par un changement de point de vue. On construit pour cela un *fern*, ensemble de caractéristiques binaires définies comme la comparaison deux à deux des intensités des pixels d'une même région. Un classifieur bayésien consiste alors à maximiser la probabilité conditionnelle des caractéristiques sous une hypothèse d'indépendance. L'apprentissage des classes se fait sur des points d'intérêts d'un nombre réduit d'images, que l'on transforme ensuite

par une série de transformations affines, constituant ainsi une base de données qui admet tous les points de vue possibles. Les caractéristiques des ferns sont calculées en tirant aléatoirement les positions des pixels.

3.4.9 Conclusion

Nous avons fait dans ce chapitre un état de l'art qui, tout en n'étant pas complètement exhaustif, s'efforce de donner un aperçu des principales méthodes d'extraction de points d'intérêt dans les images. La littérature est en effet extrêmement abondante sur ce sujet, la recherche continuant à être très active, et nous nous sommes limités aux méthodes les plus utilisées ou les plus prometteuses. En matière de détection de points d'intérêt, les principales méthodes sont le détecteur de Harris-Stephens et ses dérivés (Harris-Laplace, Harris Affine), ainsi que le détecteur SIFT et ses variantes (SURF, ASIFT). Chaque méthode présente ses avantages, le tableau suivant en donne une indication qualitative. Bien souvent néanmoins, les performances des détecteurs dépendront du type d'application, du type d'images et de leur contenu. C'est ainsi que certaines méthodes tentent d'associer différents détecteurs ou de les faire fonctionner simultanément, afin de les comparer.

Feature Detector	*Corner*	*Blob*	*Region*	*Rotation invariant*	*Scale invariant*	*Affine invariant*	*Repeatability*	*Localization accuracy*	*Robustness*	*Efficiency*
Harris	X			X			+++	+++	+++	++
Hessian		X		X			++	++	++	+
SUSAN	X			X			++	++	++	+++
Harris-Laplace	X	(X)		X	X		+++	+++	++	+
Hessian-Laplace	(X)	X		X	X		+++	+++	+++	+
DoG	(X)	X		X	X		++	++	++	++
SURF	(X)	X		X	X		++	++	++	+++
Harris-Affine	X	(X)		X	X	X	+++	+++	++	++
Hessain-Affine	(X)	X		X	X	X	+++	+++	+++	++
Salient Regions	(X)	X		X	X	(X)	+	+	++	+
Edge-based	X			X	X	X	+++	+++	+	+
MSER			X	X	X	X	+++	+++	++	+++
Intensity-based			X	X	X	X	++	++	++	++
Superpixels			X	X	(X)	(X)	+	+	+	+

TABLE 3.1: Propriétés des principaux détecteurs de points d'intérêt. D'après [71]

Egalement objets de nombreux travaux de recherche, les descripteurs de points d'intérêt s'attachent à décrire l'environnement qui les entoure. Les descripteurs à base d'histogrammes d'orientations du gradient, comme dans la méthode SIFT, sont actuellement les plus populaires. Cependant, de nombreuses approches très différentes ont été proposées. En particulier, les filtres orientables ont été une source d'inspiration de nos travaux, nous présentons ainsi au chapitre suivant un nouveau descripteur utilisant le filtrage anisotrope pour la mise en correspondance d'images.

Chapitre 4

Un nouveau descripteur circulaire

Ce chapitre présente un nouveau descripteur de points et son utilisation pour la mise en correspondance. Un descripteur de points s'appuie sur des points d'intérêt, nous présenterons donc en premier lieu le type de primitive retenu et la méthode de sélection mise en œuvre. Dans un deuxième temps nous introduirons une nouvelle méthode de filtrage anisotrope et son utilisation en segmentation d'image permettant notamment de détecter des contours, des lignes de crêtes, des jonctions, etc. de manière robuste. Nous présenterons de nouveaux filtres dérivés de ces méthodes de segmentation afin d'étendre leur domaine d'application à la caractérisation de points d'intérêts, nous obtiendrons alors de nouveaux descripteurs basés sur la signature angulaire du signal image. Nous présenterons ensuite une méthode permettant de rendre invariants ces descripteurs face aux changements d'illumination présentés au paragraphe 2.2.5, enfin, nous présenterons les résultats obtenus.

4.1 Primitive "Point" retenue

Nous sommes dans un contexte de stéréo-vision ou mouvement et dans ces conditions, les structures présentes dans les images sont vues à des échelles proches. Dans ces conditions, l'utilisation d'un détecteur multi-échelle par rapport à un détecteur plus classique mono-échelle n'est donc, pas justifié, étant donné le coût calcul supplémentaire. Nous avons donc choisi d'utiliser un simple détecteur de "Harris", niveau de gris ou couleur. Le détecteur couleur aura notre préférence compte tenu de sa stabilité, en revanche un détecteur en niveau de gris autorisera des temps calculs inférieurs. Dans ce travail, l'accent est mis sur le descripteur, cependant nous avons mis en œuvre une méthode de sélection des points de "Harris" permettant d'assurer une forte répétabilité du détecteur.
Cette méthode de sélection des points permet d'une part d'augmenter la répétabilité du détecteur face aux changements de luminosité et d'autre part d'assurer un nombre de points indépendant du seuil de détection.

4.1.1 Sélection des points

Considérons une image couleur composée de trois cannaux $(R,\ G,\ B)$,

$$\vec{I}(x,\ y)\ =\ (R(x,\ y),\ V(x,\ y),\ B(x,\ y))^t$$

alors le tenseur multi-spectral g de Di-Zenzo [18] s'écrit :

$$g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \tag{4.1}$$

avec :

$$\begin{array}{rcl} g_{11} & = & {R_x}^2 + {V_x}^2 + {B_x}^2 \\ g_{22} & = & {R_y}^2 + {V_y}^2 + {B_y}^2 \\ g_{12} & = & R_x\,R_y + V_x\,V_y + B_x\,B_y \end{array}$$

Nous obtenons le tenseur de structure M par lissage gaussien des composantes de g :

$$M = \begin{pmatrix} G_\sigma(g_{11}) & G_\sigma(g_{12}) \\ G_\sigma(g_{21}) & G_\sigma(g_{22}) \end{pmatrix} \tag{4.2}$$

Où G_σ représente la convolution avec une gaussienne d'écart-type σ.

L'opérateur de Harris (en explicitant les coordonnées image) est alors obtenu par :

$$Har(x,\ y) = Det\,(M(x,\ y)) - 0.04 \cdot Trace^2\,(M(x,\ y)) \tag{4.3}$$

Dans [32], la sélection des points de Harris était effectuée par l'extraction des maxima locaux de l'image $Har(x,\ y)$ dans des fenêtres circulaires (7×7, 9×9, 11×11, etc.) puis seuillage.

Cette approche relativement simple et à première vue efficace, pose cependant plusieurs problèmes :

1. D'abord, le nombre de points obtenus par cette méthode est difficile à contrôler.
2. Ensuite, dans les cas qui nous intéressent (par exemple en stéréo à large base), la valeur obtenue pour la mesure de Harris va dépendre de la courbure locale de la fonction image, qui dépendra elle même du point de vue. Si l'on considère deux points d'intérêt proches, l'ordre des réponses obtenues par l'opérateur de Harris peut facilement s'inverser dans les deux images. L'un des deux points sera donc éliminé par la maximisation locale, dans la mesure où la distance entre ces deux points est inférieure au rayon du cercle de maximisation. Ce phénomène induit donc une diminition de répétabilité du détecteur en ne détectant pas le même point dans les deux images.
3. Enfin, le seuillage ne garantit pas une répartition optimale des points détectés. Les points extraits peuvent par exemple se trouver tous dans une même zone de l'image, ce qui posera des problèmes par exemple, lors de l'estimation de la matrice fondamentale.

Pour Cela nous avons mis en œuvre une méthode d'éxtraction basée sur un technique de "buckets" [86]. Un bucket est simplement défini par une zone rectangulaire dans une image, d'un point de vue informatique, c'est une structure de données qui pourra contenir des primitives. Ici les buckets forment un pavage de l'image sans recouvrement. Il s'agit de simplement de diviser l'image en $N_x \times N_y$ buckets (figure 4.1).

Nous voulons obtenir si possible, un nombre fixé de N points au total dans l'image, chaque bucket devra donc contenir au maximum $N/(N_x \times N_y)$ points (évidement un bucket d'intensité lumineuse contante ne contiendra aucun point). Dans ce cas tous les maximas locaux dans une fenètre 3×3 sont considéres bucket par bucket, sont uniquement conservés les $N/(N_x \times N_y)$ point les plus significatifs d'un point de vue de la mesure de Harris.

La figure 4.2 compare les résultats obtenus avec une méthode de maximisation locale et seuillage et la méthode par buckets. Il est difficile de contrôler le nombre de points obtenus avec la méthode par maximisation locale et seuillage. La méthode par buckets donne une répartition plus uniforme des points.

4.2 Des demi-filtres directionnels pour la segmentation d'image

4.2.1 Filtres Orientables

Nous allons nous intéresser à de nouvelles méthodes de filtrage anisotropes utilisant des filtres orientables. Les filtres orientables ou "steerable filters" ont été introduits par

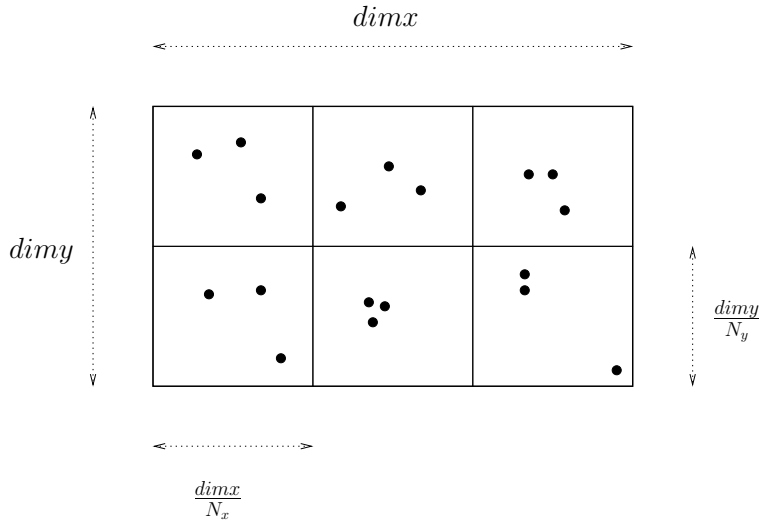


FIGURE 4.1: Sélection des points : Nous cherchons à obtenir au maximum 18 points dans l'image équitablement répartis dans 3 buckets en X et 2 buckets en Y, soit 3 points par bucket.

Freeman et Adelson [29]. Plus tard Jacob et Unser [40] ont généralisé ce type de filtres en introduisant le critère Canny [11] pour la détection de contours. Perona introduit des gaussiennes anisotropes dans [84], Geusebroek et al. [30], implémentent ce filtre de manière récursive. Plus récemment Knossow et al. [48] ont utilisé les gaussiennes anisotropes pour la détection de personnes dans des images couleur. Les auteurs montrent que ce type de filtrage détecte mieux les contours des personnes que la méthode de Canny [12] standard.

Même si cette dernière méthode est plutôt adaptée à la recherche de contours rectilignes, elle introduit cependant une nouvelle famille de filtres gaussiens anisotropes en segmentation d'images. Elle est le point de départ de travaux récents en détection de contours utilisant des demi-filtres gaussiens [85, 64, 63, 62, 61, 76]. Le nouveau descripteur introduit dans cette thèse est dérivé de ces travaux.

Avant d'introduire le nouveau descripteur proposé ici, nous décrirons plus en détails les gaussiennes anisotropes, puis nous aborderons ensuite les travaux de segmentation par demi-filtres. Les implémentations de ces méthodes seront abordées, ce qui nous permettra de comparer une implémentation spécifique pour la segmentation d'image avec une implémentation "éparse" pour des descripteur utilisant ce type de filtre.

4.2.2 Filtres gaussiens anisotropes pour la segmentation d'images

Nous présentons à l'équation 4.4 un des filtres de Perona et Geusebroeck et al., utilisés par Knossow : le filtre de dérivation selon l'axe Y. La figure 4.3 (a) présente le filtre de


a)


b)


c)

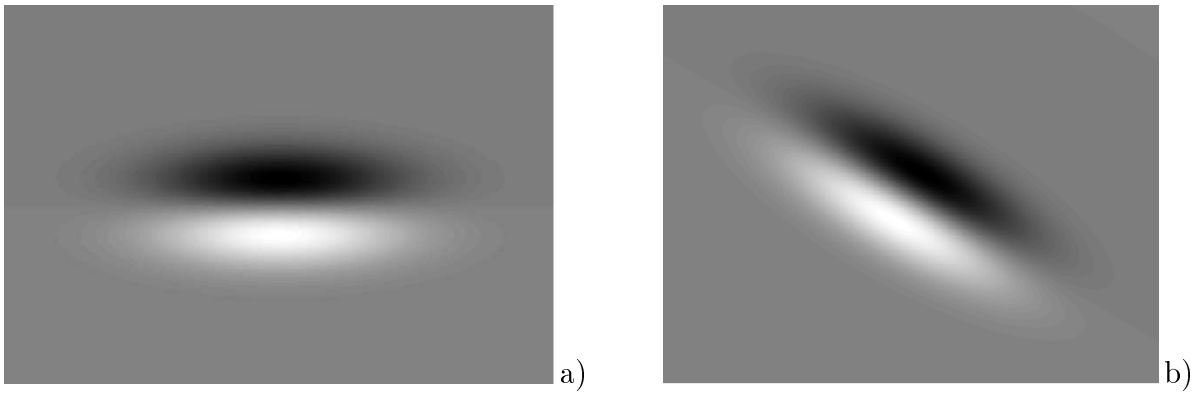
FIGURE 4.2: Sélection des points : a) image originale, b) Points de Harris couleur, maximima locaux dans une fenêtre circulaire de rayon 9 pixels, seuil 0.01 (461 points), c) Méthode des buckets 3 buckets en X 2 buckets en Y, 462 points, la répartition des points est plus uniforme, des points proches sont autorisés.

l'équation 4.4, la figure 4.3 (b) présente un filtre orienté à 30° de celui-ci.

$$G_{\sigma_x,\sigma_y}(x,\ y)\ =\ -C\cdot y\cdot e^{\frac{-x^2}{2\sigma_x^2}}\cdot e^{\frac{-y^2}{2\sigma_y^2}} \tag{4.4}$$

Où :

C	est un coefficient de normalisation
σ_x	est un l'ecart-type de la gaussienne suivant l'axe X
σ_y	est un l'ecart-type de la gaussienne suivant l'axe Y



a) b)

FIGURE 4.3: Filtres gaussiens orientés. a) Filtre de dérivation selon l'axe Y ($\sigma_x = 5,\ \sigma_y = 1$) . b) Filtre de dérivation orienté à 30° de l'axe Y.

Ce filtre en détection de contours, donne des résultats intéressants pour des contours rectilignes, en revanche, lorsque l'on s'approche des coins, la qualité de la détection baisse fortement. La figure 4.4 présente l'application des filtres gaussiens anisotropes sur le contour d'un objet. Le cas 1 représente le filtre appliqué à une portion linéaire du contour d'un objet : la réponse du filtre est forte. Le cas 2 représente le filtre appliqué en un coin : seule une partie du filtre prend en compte l'information "contour". Par conséquent, la réponse du filtre diminue fortement, le résultat sera donc d'autant plus influencé par le bruit.

Dans la suite de cet exposé, nous nous attarderons sur la mise en œuvre de ces filtres gaussiens anisotropes, puis nous discuterons des implémentations possibles pour ces filtres.

4.2.3 Mise en œuvre des filtres gaussiens anisotropes pour la segmentation d'images

Cette dernière méthode de détection de contours peut être vue comme une généralisation de la détection de contours couleur [18], dans laquelle les images convoluées avec les filtres de lissage à diverses orientations jouent le rôle des différents canaux. Considérons donc une image convoluée avec une banque de filtres de lissage gaussiens anisotropes :

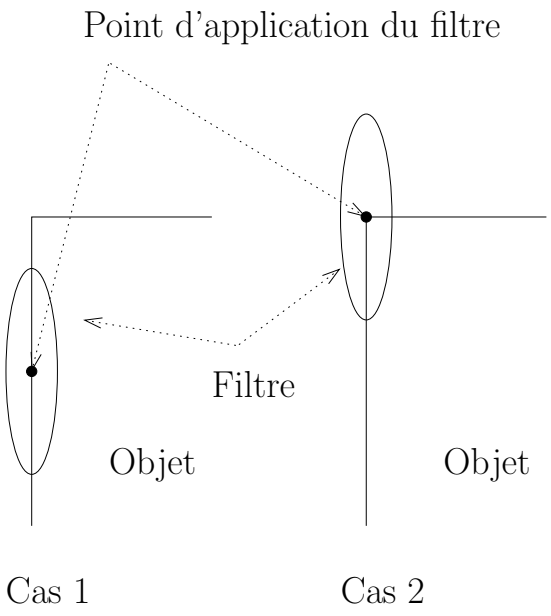


FIGURE 4.4: Application des filtres gaussiens orientés sur le contour d'un objet.

$$G_\theta(x,y) = C\, e^{-\begin{pmatrix} x & y \end{pmatrix} P_\theta^{-1} \begin{pmatrix} \frac{1}{2\sigma_\eta^2} & 0 \\ 0 & \frac{1}{2\sigma_\xi^2} \end{pmatrix} P_\theta \begin{pmatrix} x \\ y \end{pmatrix}}$$

Où :

σ_ξ et σ_η	représentent les deux écarts-type de la gaussienne anisotrope dans les deux directions principales de la gaussienne
ξ et η	représentent respectivement les directions du plus fort et du plus faible lissage
P_θ et P_θ^{-1}	représentent respectivement une matrice de rotation (dans le plan image) et son inverse

La figure 4.5 présente une gaussienne anisotrope ($G_{90}(x,y)$) dans laquelle $\sigma_\xi = 10$ et $\sigma_\eta = 2$ et :

$$P_{90} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \text{(matrice de rotation pour } \theta = 90°\text{)}.$$

et pour une orientation quelconque :

$$P_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

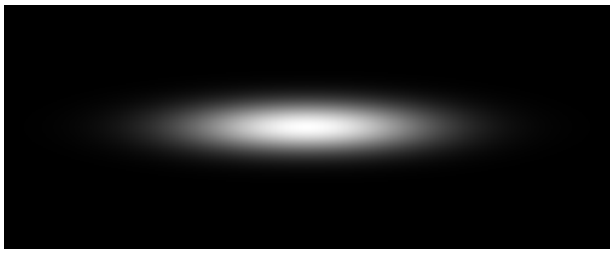


FIGURE 4.5: Filtre de lissage gaussien orienté selon l'axe des X ($\sigma_\xi = 5 \times \sigma_\eta$).

Nous présentons ici une mise en œuvre que nous avons effectuée dans notre laboratoire, qui passe par le calcul d'un tenseur d'órientation. La mise en œuvre est différente de celle de Geusebroeck et al. dans la mesure où pour nous, c'est l'image qui tourne et non le filtre. Nous avons alors un filtre de lissage d'orientation fixe, qui s'implémente récursivement de manière classique [87]. Considérons N filtres de lissage d'orientations croissantes $\theta_i \in [0°, 180°[$ par pas de $\Delta\theta$ (par exemple : $\Delta\theta = 2°,\ 5°,\ 10°, ...$), ces filtres forment une banque de filtres d'orientation. Par convolution de l'image avec cette banque de filtres, nour obtenons N images lissées. Chaque Image lissée est alors dérivée en X et en Y, ce qui nous donne $2 \times N$ images ($I_{\theta X}$ et $I_{\theta Y}$).

Nous pouvons alors former un tenseur d'orientation :

$$T = \begin{pmatrix} \sum_\theta I_{\theta X}^2 & \sum_\theta I_{\theta X}\, I_{\theta Y} \\ \sum_\theta I_{\theta X}\, I_{\theta Y} & \sum_\theta I_{\theta Y}^2 \end{pmatrix}$$

La valeur propre maximale de T nous donne le carré du gradient anisotrope, le vecteur propre associé à la valeur propre maximale, nous donne l'orientation du gradient. La généralisation à la couleur de cette méthode est immédiate, on considère alors simplement le tenseur d'orientation couleur comme la somme de trois tenseurs en niveau de gris ($T_R,\ T_V,\ T_B$) obtenus respectivement à partir des images $R,\ V,\ B$.

La figure 4.6 compare les résultats obtenus avec le filtre gaussien isotrope et ceux obtenus par cette méthode sur une image synthétique bruitée. Volontairement, nous avons considérés des seuils de détection identiques pour les deux types de filtres. Dans le cas du filtre gaussien isotrope, le résultat obtenu est très bruité, si l'on augmente les seuils nous perdons le contour de la forme géométrique du bas de l'image son contour est connecté au bruit. En revanche le filtre gaussien anisotrope présente une très bonne

immunité au bruit, cependant comme nous l'avons déja noté ce filtre conserve mal les coins des objets.

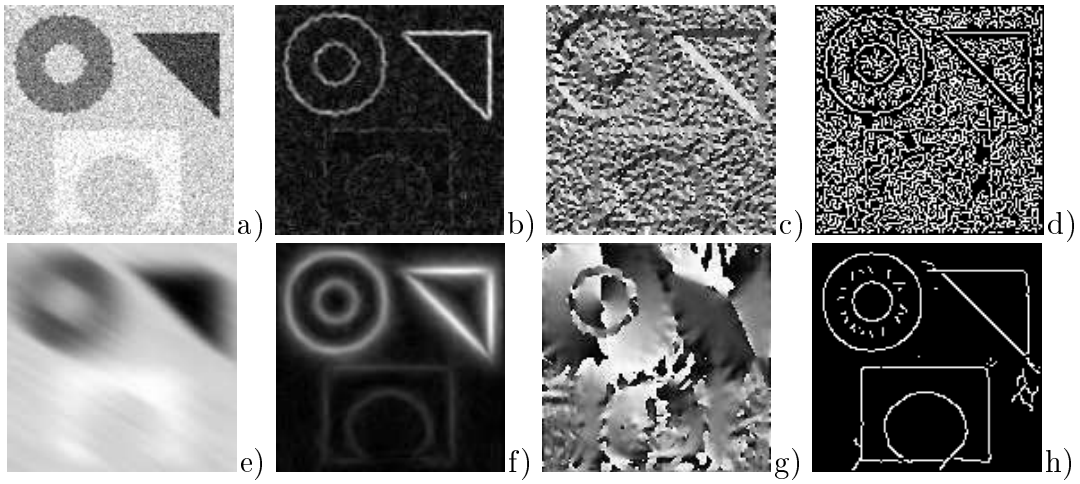


FIGURE 4.6: Résultats filtres gaussiens anisotropes ($\sigma_\eta = 1$, $\sigma_\xi = 10$, $\Delta_\theta = 5°$). a) image originale. b) Gradient gaussien isotrope ($\sigma = 1$). c) Angle du gradient isotrope ($\sigma = 1$). d) Détection de contours isotrope, seuillage par hystérésis, seuil haut ($s_h = 0.1$), seuil bas ($s_b = 0.01$). e) lissage anisotrope ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\theta = 120°$. f) Gradient gaussien anisotrope. g) Angle du gradient anisotrope. h) Détection de contours anisotrope, seuillage par hystérésis, seuil haut ($s_h = 0.1$), seuil bas ($s_b = 0.01$).

4.2.3.1 Implémentation efficace des filtres gaussiens anisotropes

Pour implémenter ce type de filtre, nous avons le choix entre plusieurs méthodes :

1. Convolution directe :
   Sachant qu'une gaussienne anisotrope n'est en général pas séparable sauf pour des orientations différentes des axes de l'image (0° / 90°), que au moins un des deux écarts-types de la gaussienne ne sera pas petit, l'implémentation par masque de convolution peut devenir complexe. En effet si l'on respecte la règle des $3\times\sigma$ pour la discrétisation du filtre, pour un sigma de 10, on peut facilement obtenir un masque de convolution de taille 61×61, ce qui coûterait 3721 opérations élémentaires par pixel, pour un θ donné.
2. Rotation inverse de l'image, convolution séparable, et rotation du résultat de filtrage :
   Cette fois, le filtre de convolution est séparable. A une orientation donnée, par exemple pour un filtre de $\sigma_\xi = 10$ et $\sigma_\eta = 1$, nous obtenons 61+7 opérations élémentaires par pixel auquelles il faut ajouter le coût des deux rotations (matrices 2×2) soit 8 opérations.

3. Convolution dans l'espace de Fourier :
   L'utilisation de la transformée de Fourier ou FFT ne se justifie que pour des valeurs de σ très grandes.
4. Equations de récurences :
   Deriche [87] a proposé une implémentation récursive du filtre gaussien à l'ordre 4. Dans la mesure où l'on peut se ramener au cas séparable par rotation inverse puis rotation du résultat après filtrage, nous obtenons une implémentation en 18 opérations élémentaires par pixel.
5. Equations aux Dérivées Partielles :
   La gaussienne étant la solution de l'équation de la chaleur, il est possible d'obtenir une image lissée par une gaussienne anisotrope en itérant des petits filtres de lissages, ce qui conduirait à une implémentation plus efficace que la convolution directe. Cependant la solution par équations de récurences reste la plus efficace.
6. Implémentation de Geusebroeck et al., qui approxime à la fois, le filtrage et la rotation par un filtre récursif. Cette implémentation est la plus efficace, cependant, elle introduit une approximation supplémentaire par rapport à la gaussienne récursive.

Une implémentation efficace de ce type de filtre passe donc par rotation inverse de l'image, filtrage récursif séparable, puis rotation du résultat du filtrage, ou par la méthode de Geusebroeck et al.

4.2.4 Des demi-filtres orientables pour la segmentation d'images

Afin de remédier aux problèmes rencontrés avec le filtre gaussien anisotrope décrit au paragraphe 4.2.1, une nouvelle méthode de filtrage et de segmentation a vu le jour [85, 64, 63, 62, 61, 76] pour la détection de contours, jonctions, de lignes de crêtes, la suppression de texture et la restauration d'images. L'idée ici est de couper une gaussienne directionnelle en deux parties comme le présente les figures 4.7 et 4.8 puis d'appliquer le filtre ainsi obtenu à plusieurs orientations sur l'image. Nous présentons le filtre de lissage et dérivation utilisé pour la détection de contours.

Cette fois le filtre de lissage n'étant plus symétrique selon la direction de son élongation maximale (direction du plus grand écart-type ξ) cf. equation 4.5, les orientations des filtres prennent des valeurs entre 0° et 360° au lieu de 0° à 180°. Cette non symétrie du filtre de lissage rend cette fois difficile le calcul d'un gradient via un tenseur d'orientation. Par conséquent, un filtre de dérivation est directement utilisé dans la direction du plus faible écart-type (direction η). Nous avons donc un demi-filtre de lissage dans la direction ξ et un filtre de dérivation dans la direction perpendiculaire η. Il est alors possible d'estimer un gradient anisotrope, simplement par la différence des réponses directionnelles maximale et minimale au filtre.

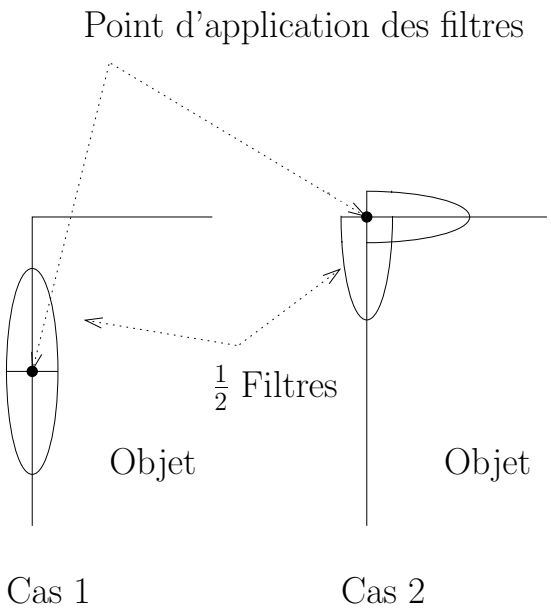


FIGURE 4.7: Application des demi-filtres gaussiens orientés sur le contour d'un objet.

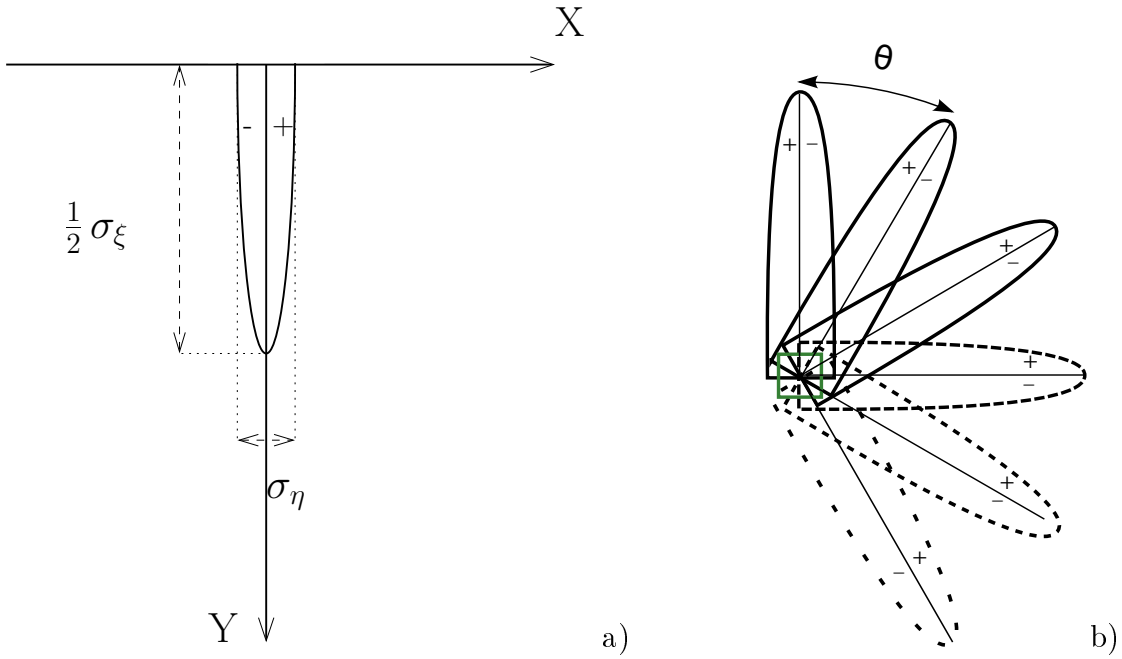


FIGURE 4.8: Demi-Filtres. a) Opérateur de détection de contour : Filtre de dérivation en X, demi-filtre de lissage en Y. b) Filtres tournés de θ degrés.

Mathématiquement, ce filtre est décrit par l'équation 4.5 explicitant la rotation :

$$G_\theta^{(1)}(x,y) = C \cdot (x\cos(\theta) - y\sin(\theta)) \cdot \; H_y\left(P_\theta \cdot \begin{pmatrix} x \\ y \end{pmatrix}\right) \cdot e^{-\begin{pmatrix} x & y \end{pmatrix} P_\theta^{-1} \begin{pmatrix} \frac{1}{2\sigma_\eta^2} & 0 \\ 0 & \frac{1}{2\sigma_\xi^2} \end{pmatrix} P_\theta \begin{pmatrix} x \\ y \end{pmatrix}} \tag{4.5}$$

Où :

σ_ξ et σ_η	représentent les deux écarts-type de la gaussienne anisotrope dans les deux directions principales de la gaussienne σ_ξ est pris dans la direction θ
P_θ et P_θ^{-1}	représentent respectivement une matrice de rotation (dans le plan image) et la matrice inverse
$H_y(x,\ y)$	représente une fonction de *Heaviside* selon l'axe Y

$$H_y(x,\ y) \;=\; \begin{cases} 0 & \text{si } y \;<\; 0 \\ \frac{1}{2} & \text{si } y \;=\; 0 \\ 1 & \text{si } y \;>\; 0 \end{cases}$$

C	est un coefficient de mormalisation calculé afin d'obtenir des dérivées exactes pour des polynomes

Ou encore exprimé plus simplement dans un repère lié au filtre :

$$G_\theta(x,y) = C \,.\, X \,.\, H(Y) \;.\, e^{-\frac{X^2}{2\sigma_\eta^2} - \frac{Y^2}{2\sigma_\xi^2}} \tag{4.6}$$

Avec :

$$\begin{pmatrix} X \\ Y \end{pmatrix} = P_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

$H(Y)$ représente la fonction de Heaviside

$$H(Y) = \begin{cases} 0 & \text{si } Y < 0 \\ \frac{1}{2} & \text{si } Y = 0 \\ 1 & \text{si } Y > 0 \end{cases}$$

Ces filtres sont comme au paragraphe précédent implémentés par rotation inverse de l'image, convolution par équations de récurence puis rotation du résultat (il n'y a qu'un seul filtre : seule l'image "tourne"). Nous avons ici une implémentation récursive des filtres qui assure une grande efficacité. Concernant le demi filtre de lissage, seule l'équation de récurence positive est conservée.

Nous verrons que la réponse à ces filtres tournants constitue une signature directionnelle du signal image autour de chaque pixel. Nous avons montré dans [85] que cette signature peut être interprétée en terme de gradient anisotrope. Mais aussi dans [64, 63, 62, 61, 76], ce type de signature a permis de caractériser des lignes de crêtes, des jonctions, des textures, etc.

Nous présentons à la figure 4.9 les résultats de détection de contours obtenus par cette méthode sur la mème image de test utilisée pour évaluer le filtre gaussien anisotrope, ces résultats sont comparés avec ceux obtenus pour ce filtre. Même si les résultats obtenus sont plus bruités que les résultats obtenus avec les filtres entiers, ces demi-filtres présentent un bon comportement vis à vis du bruit. En revanche les structures sont bien mieux conservées qu'avec les filtres entiers.

Nous présentons à la figure 4.10 la signature obtenue en certains points caractéristiques de l'image. Nous présentons en abscisse l'orientation du filtre (en degrés par rapport à l'axe Y de l'image) et en ordonnée, la réponse au filtre. Les points 0, 1 et 4 sont localisés sur des coins, la signature présente deux pics (un pic positif et un pic négatif), la différence entre les abscisses des deux pics nous donne l'angle du coin. La différence de hauteur des pics nous donne la norme gradient. Le point 3 est localisé entre deux structures, la courbe obtenue présente une influence mutuelle des deux structures. Enfin le point 2 est situé dans le bruit, la réponse obtenue est à peu près "plate". Malgré le bruit extrèmement important dans cette image, les réponses obtenues sont sans ambiguité. Plus de détails sont disponibles dans [76], [65], [60].

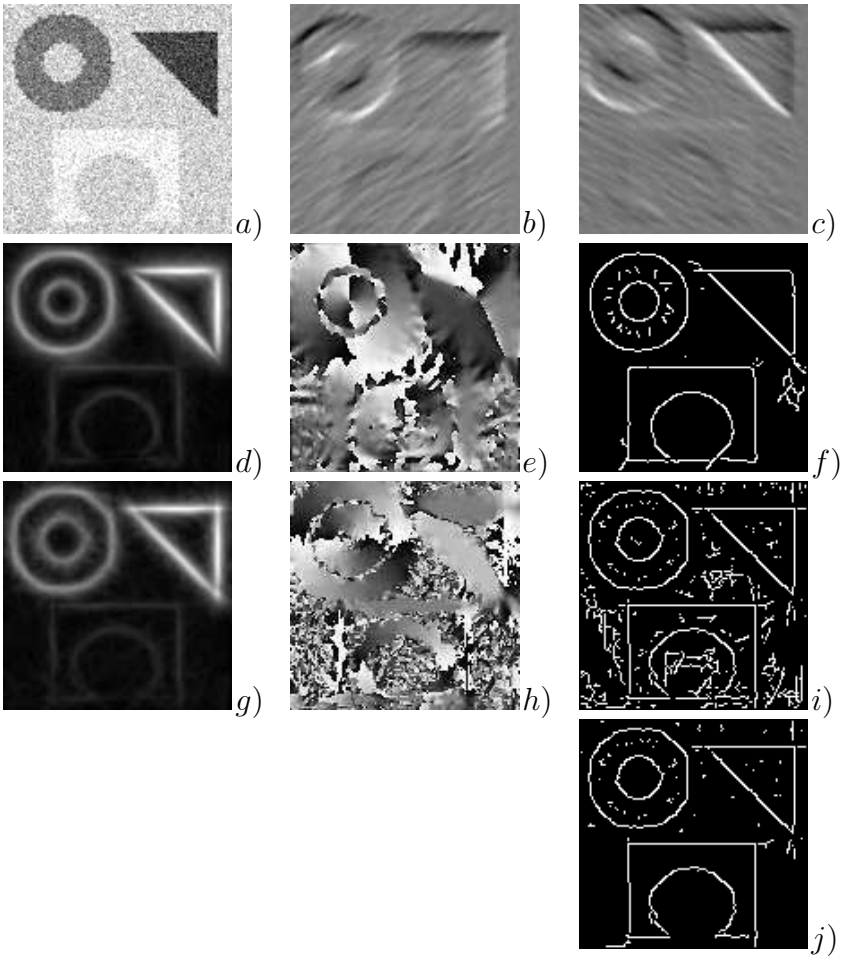
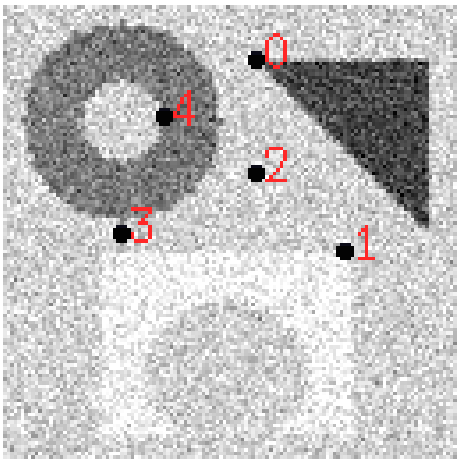
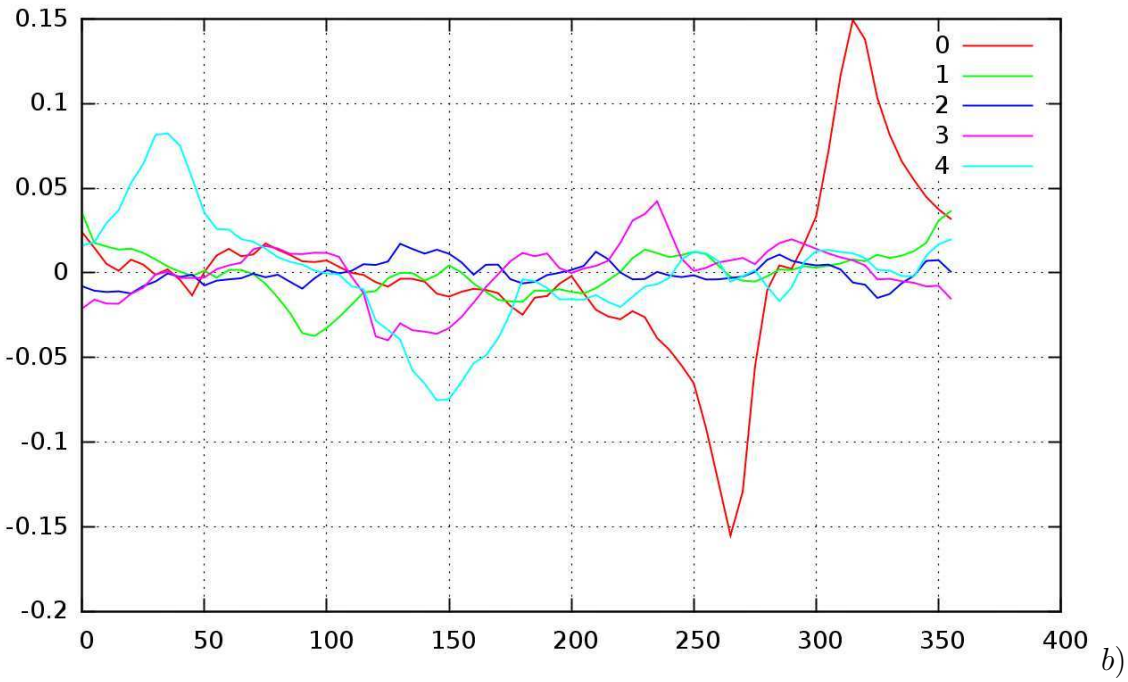


FIGURE 4.9: Résultats : demi-filtres ($\sigma_\eta = 1,\ \sigma_\xi = 10,\ \Delta_\theta = 5°$). a) image originale. b) Filtrage demi-lissage dans la direction $\xi = 55°$ et dérivation dans la direction η. c) Filtrage demi-lissage dans la direction $\xi = 120°$ et dérivation dans la direction η. d) Gradient gaussien anisotrope. e) Angle du gradient anisotrope. f) Détection de contours anisotrope, seuillage par hystérésis, seuil haut ($s_h = 0.1$), seuil bas ($s_b = 0.01$). g) Gradient gaussien anisotrope (demi-filtres). h) Angle du gradient anisotrope (demi-filtres). i) Détection de contours anisotrope (demi-filtres), seuillage par hystérésis, seuil haut ($s_h = 0.1$), seuil bas ($s_b = 0.01$). j) Détection de contours anisotrope (demi-filtres), seuillage par hystérésis, seuil haut ($s_h = 0.15$), seuil bas ($s_b = 0.01$) en augmentant légèrement le seuil haut, les objets sont extraits correctement.



a)



b)

FIGURE 4.10: Signatures. a) image originale. b) Signatures obtenues avec le filtre de détection de contours (equation 4.10) aux points marqués sur l'image originale ($\sigma_\eta = 1,\ \sigma_\xi = 10,\ \Delta_\theta = 10°$).

4.3 Demi-filtres directionnels pour la caractérisation de points d'intérêt

Dans le paragraphe précédent, le filtrage était utilisé pour la segmentation d'images. Dans ce cas la convolution devait être effectuée de manière dense (pour tous les pixels), d'où les implémentations récursives.

Cette fois nous nous intéressons principalement à la caractérisation de points d'intérêt, qui sont par nature épars. Dans ces conditions, un calcul pour tous les pixels n'est plus nécessaire, et donc la convolution directe devient plus efficace. Maintenant seul le filtre subit une rotation et non plus l'image : tous les filtres sont calculés une seule fois au début du processus de caractérisation des points. Nous définissons des familles de filtres de lissage ou de dérivation d'ordre 1 voire 2. Pour la dérivation d'ordre deux, il est possible d'appliquer une méthode dérivée des DOG (Difference Of Gaussians) en utilisant la différence de deux filtres de lissage (avec des σ_η différents). Nous présentons les expressions des filtres de caractérisation, ces expressions sont très similaires à celles des filtres utilisés pour la segmentation :

$$G_\theta^{(0)}(x,y) = C_\theta^{(0)} \cdot H(Y) \cdot e^{-\frac{X^2}{2\sigma_\eta^2}-\frac{Y^2}{2\sigma_\xi^2}} \tag{4.7}$$

pour les filtres de lissage et :

$$G_\theta^{(1)}(x,y) = C_\theta^{(1)} .X . H(Y) . e^{-\frac{X^2}{2\sigma_\eta^2}-\frac{Y^2}{2\sigma_\xi^2}} \tag{4.8}$$

Pour les filtres de dérivation d'ordre 1.

Dans le cas de la détection de contours, un seul filtre est implémenté donc nous avons un seul coefficient de normalisation à calculer pour les filtres d'ordre 0 et 1 [15]. Mais évidement ce coefficient dépend de l'orientation du filtre, (dans [75], les coefficients de normalisation dépendaient de la précision sub-pixélique) donc cette dépendance apparaît explicitement ici (les coefficients de normalisation sont indicés par θ.

Malheureusement, nos premières expérimentations ont montré l'apparition d'artéfacts dûs à la discrétisation des filtres ou des rotations de l'image dès que le pas augulaire entre les différents filtres $\Delta\theta$ devient inférieur à 10°. En fait, dans les deux cas que nous considérions le filtrage récursif ou le filtrage par convolution nous obtenons ces artéfacts génants.

La figure 4.11 présente les résultats obtenus sur une image synthétique avec les filtres de dérivation. Nous avons séléctionné ici deux points de Harris, un point correspondant à un coin et un point correspondant à une jonction. La courbe verte de la figure 4.11.b présente 2 pics principaux pour le coin, la courbe rouge de la figure 4.11.b présente 3 pics principaux pour la jonction. Ici le pas angulaire du filtrage est fixé à 2°, de nombreux maxima et minima secondaires dûs à la discrétisation image apparaissent. Il aurait été préférable d'obtenir des caractérisations plus lisses, cependant le paragraphe 4.3.1 introduit de nouveaux filtres anti-aliasés corrigeant ce problème.

Maintenant, si nous considérons les filtres obtenus après dicrètisation (figure 4.12) nous pouvons directement visualiser l'effet de l'aliasing dû à la fonction de Heaviside sur le filtre. La figure 4.12.a présente un filtre de lissage initial sans rotation, la direction principale du filtre sans rotation est toujours orientée selon l'axe Y. Dès que l'on applique une rotation, des effets d'aliasing apparaîssent, nous présentons les filtres obtenus avec des rotations de 44° et 89° respectivement aux figures 4.12.b et 4.12.c.

4.3.1 Filtres directionnels anti-aliasés

Afin de remédier à ces problèmes, nous avons modifié les filtres en échangeant la fonction de Heaviside avec une fonction sigmoïde. Nous obtenons alors de nouveaux filtres "anti-aliasés" d'équation :

$$G_\theta^{(0)}(x,y) = C_\theta^{(0)} . S(Y) \; . e^{-\frac{X^2}{2\sigma_\eta^2} - \frac{Y^2}{2\sigma_\xi^2}} \tag{4.9}$$

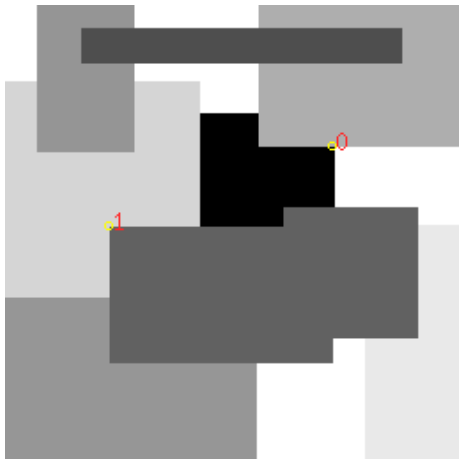
pour les filtres de lissage et :

$$G_\theta^{(1)}(x,y) = C_\theta^{(1)} . X \, . S(Y) \; . e^{-\frac{X^2}{2\sigma_\eta^2} - \frac{Y^2}{2\sigma_\xi^2}} \tag{4.10}$$

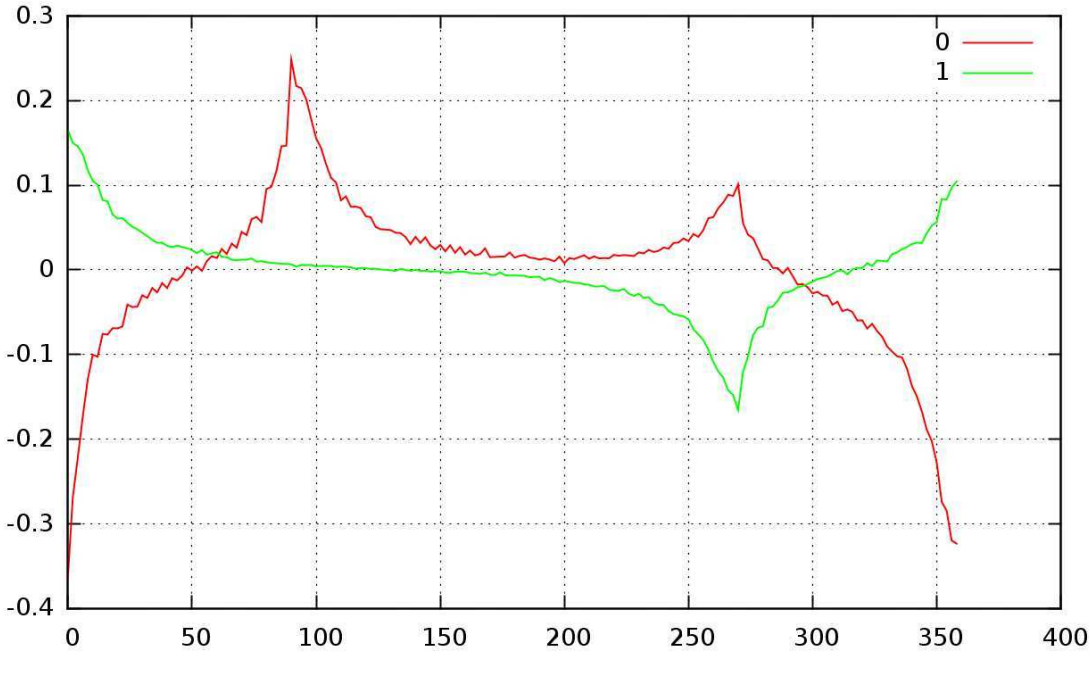
Pour les filtres de dérivation.
Avec :

$$S(Y) \; = \; \frac{1}{1+\exp(-\alpha Y+\beta)}$$

α contrôle la pente de la sigmoïde
β contrôle le centrage de la courbe



a)



b)

FIGURE 4.11: Exemples de signatures sur une image synthétique. a) image synthétique originale. b) Signatures obtenues aux points marqués sur l'image originale ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$).

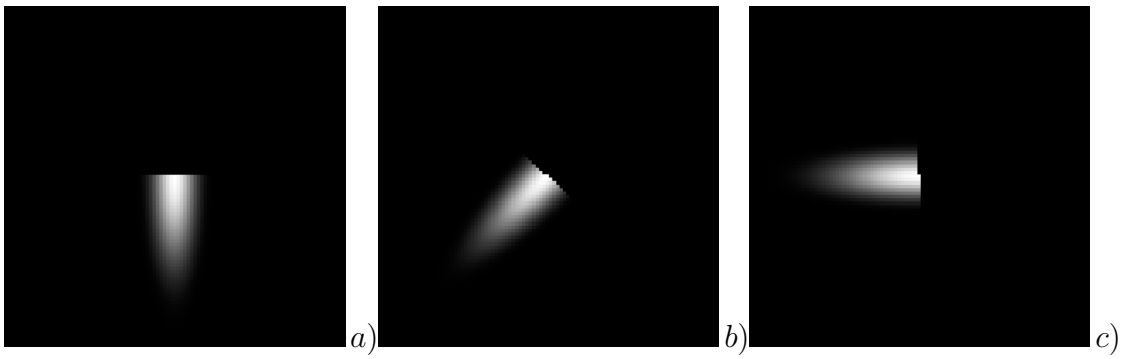


FIGURE 4.12: Discrétisation des demi-filtres. a) Filtre $\theta = 0°$. b) Filtre $\theta = 44°$. c) Filtre $\theta = 89°$. ($\Delta\theta = 2°$).

La figure 4.13 présente la courbe de la fonction sigmoïde pour $\alpha = 0.9$ et $\beta = 0$, enfin la figure 4.14 présente le profil du filtre de lissage anti-aliasé selon l'axe Y pour les paramètres $\alpha = 0.9$, $\beta = 0$ et $\sigma_\xi = 10$ (C_0 est fixé à 1). Dans la suite de nos expérimentations, les paramètres de la sigmoïde ont été fixés à $\alpha = 0.9$ et $\beta = 0$.

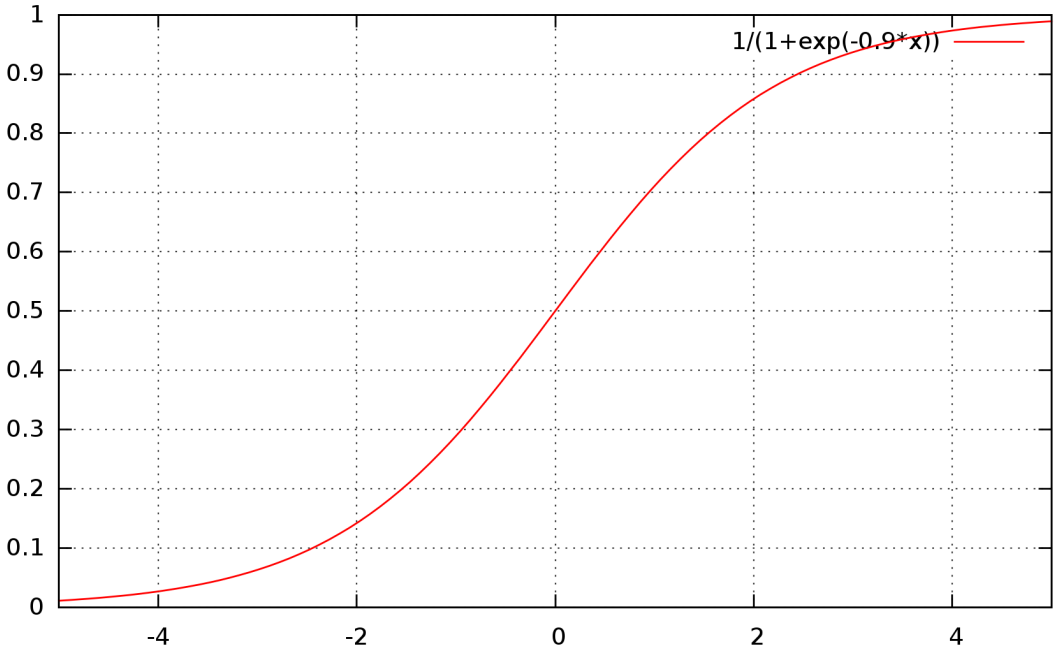


FIGURE 4.13: Courbe sigmoïde obtenue avec $\alpha = 0.9$ et $\beta = 0$.

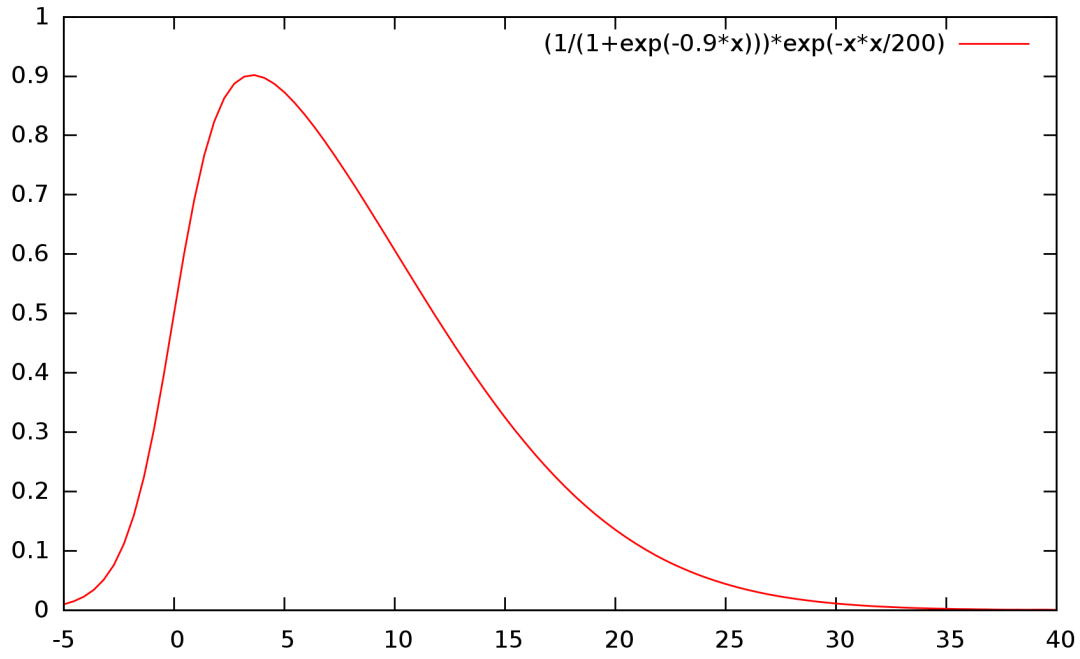


FIGURE 4.14: Profil du filtre de lissage anti-aliasé selon l'axe Y pour les paramètres $\alpha = 0.9$, $\beta = 0$ $\sigma_\eta = 1$ et $\sigma_\xi = 10$ (C_0 est fixé à 1).

4.3.2 Implémentation efficace

Selon les paramètres des filtres (Δ_θ, σ_ξ, σ_η, ordre de dérivation), une banque de filtres de convolution est calculée. Etant donné la forme allongée de ce filtre, la plupart des valeurs des masques de convolution sont nulles ou proches de zéro. Afin d'éviter des opérations inutiles : multiplications par zéro et sommation pendant le calcul des convolutions, un pré-processing sur les filtres est effectué pour sélectionner des listes de coordonnées-valeurs des filtres ayant une valeur significative. Cette opération nous permet de réduire de manière importante la complexité de la convolution. Sachant que le nombre de points d'intérêt maximal est donné en paramètre à notre méthode de sélection des points (sélection réalisé par découpage en buckets de l'image), nous avons en général peu de points d'intérêt à considérer (souvent de l'ordre de 200 ou 300 points), alors notre étape de caractérisation des points est extrêmement rapide.

4.3.3 Un descripteur invariant aux changements locaux affines de luminosité

Considérons le modèle affine de changement de luminosité à six paramètres décrit par l'équation 2.6 (section 2.2.5), il vient immédiatement que les filtres de dérivation présentés précédement (cf. figure 4.15)(d,e,f) éliminent le vecteur de translation du modèle $(\beta_R, \beta_V, \beta_B)^t$.

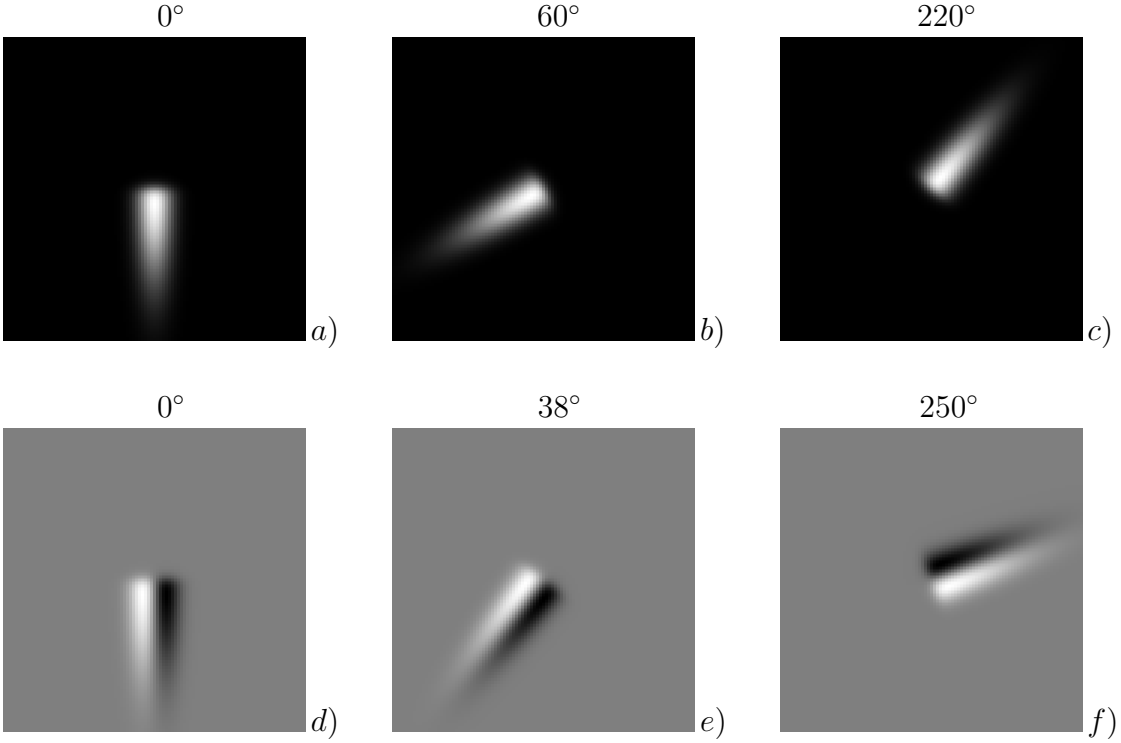


FIGURE 4.15: Exemples de filtres anti-aliasés $\sigma_\xi = 20$, $\sigma_\eta = 4$, $\alpha = 0.9$, $\beta = 0$: a) Filtre de lissage anti-aliasé $\theta = 0°$. b) Filtre de lissage anti-aliasé $\theta = 60°$. c) Filtre de lissage anti-aliasé $\theta = 220°$. d) Filtre de dérivation d'ordre 1 anti-aliasé $\theta = 0°$. e) Filtre de dérivation d'ordre 1 anti-aliasé $\theta = 38°$. d) Filtre de dérivation d'ordre 1 anti-aliasé $\theta = 250°$.

A ce stade, la méthode la plus simple permettant d'éliminer les paramètres $\alpha_R, \alpha_V, \alpha_B$ de la matrice de transformation (equation 2.6) est de normaliser chacune des trois signatures obtenues sur les canaux R, V, B, avec le maximum obtenu sur chaque signature.

La figure 4.16 présente les résultats obtenus avec les filtres de dérivation anti-aliasés, les signatures sont normalisées. Les courbes obtenues sont parfaitement lisses en comparaison avec celles de la figure 4.11.

La figure 4.17 présente les résultats obtenus sur deux images stéréoscopiques réelles (images déjà utilisées pour illustrer la géométrie épipolaire figure 2.7). Nous avons sélectionné deux points de Harris dans la première image (fig. 4.17.a) et deux points de Harris dans la deuxième (fig. 4.17.b) ces points se correspondent deux à deux. Les figures 4.17.c et 4.17.d présentent les signatures obtenues dans les deux images respectivement pour les points 0 et 1.

La figure 4.18 présente les résultats obtenus sur deux images stéréoscopiques réelles, l'images gauche est celle de la figure 4.17, en revanche l'image droite est différente, la base stéréoscopique est élargie, donc le changement de point de vue est important. Nous avons séléctionné deux points de Harris dans la première image (fig. 4.17.a) et deux points de Harris dans la deuxième (fig. 4.17.b) ces points se correspondent deux à deux. Les figures 4.18.c et 4.18.d présentent les signatures obtenues dans les deux images respectivement pour les points 0 et 1.

Nous allons voir au chapitre suivant que ces signatures vont permettre de définir une nouvelle méthode de mise en correspondance robuste et efficace, et ce même dans des cas où la transformation géométrique entre les deux images est importante.

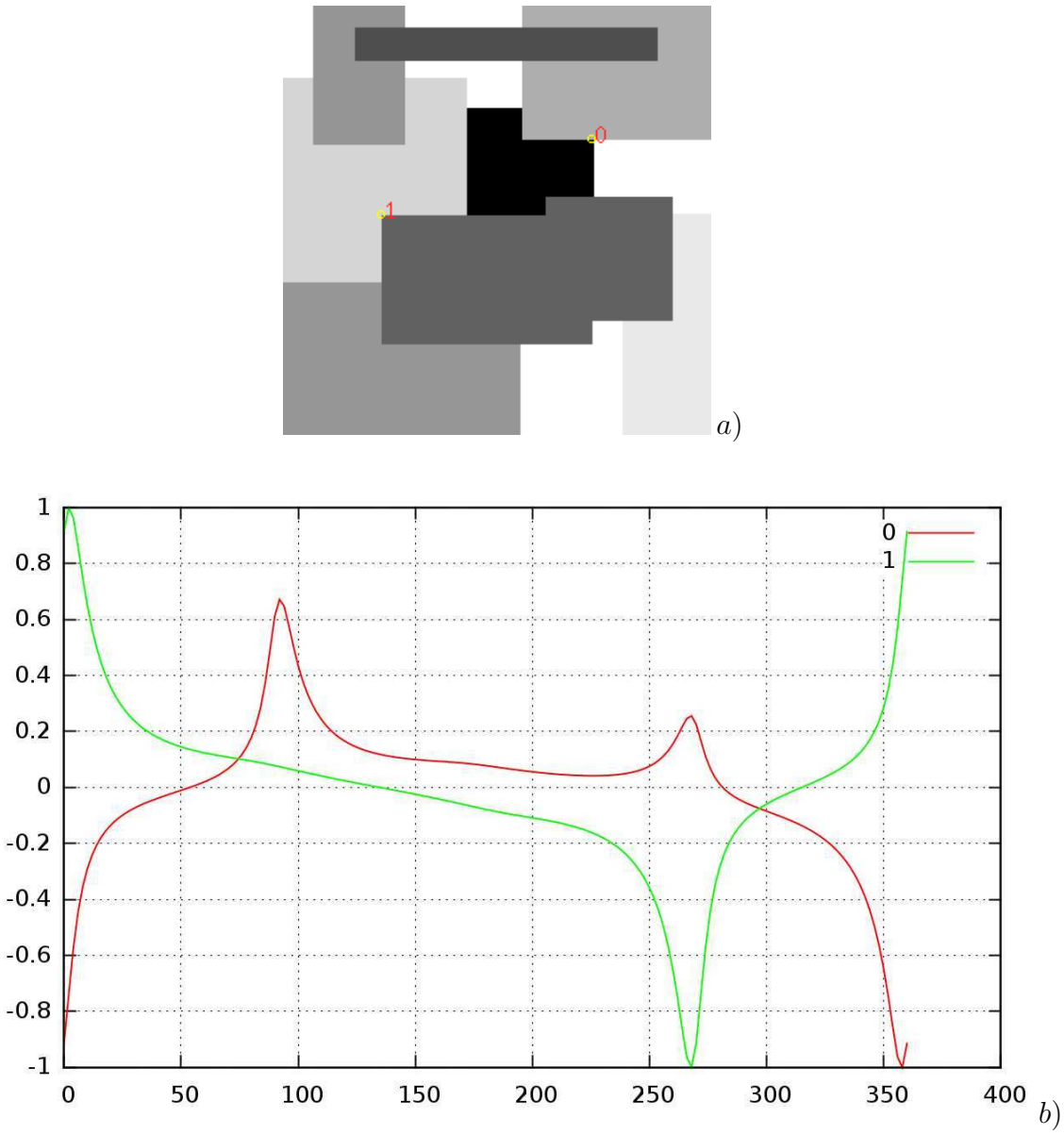
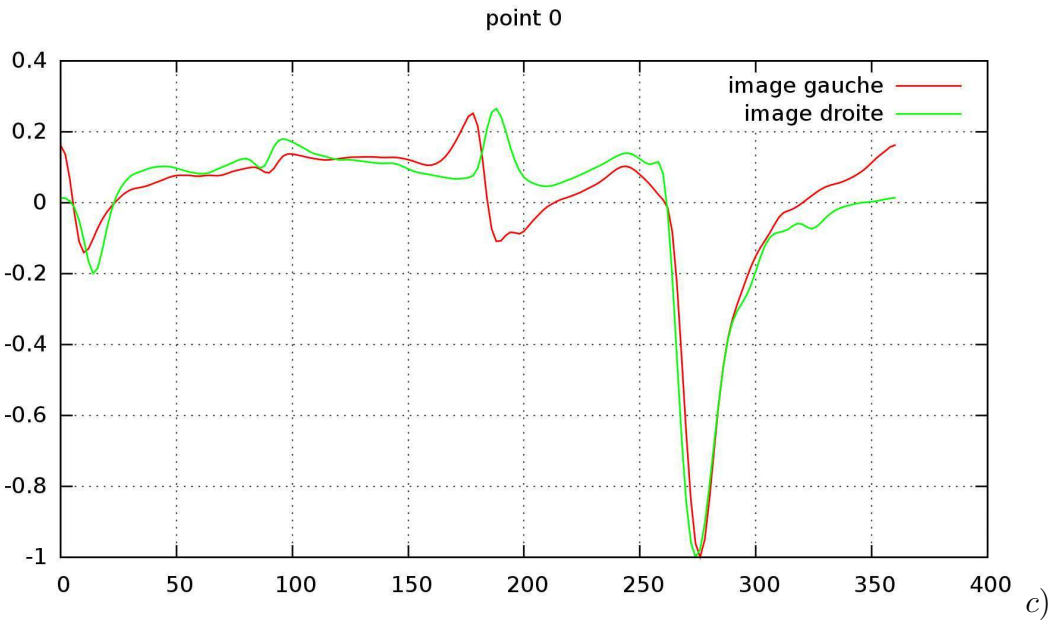


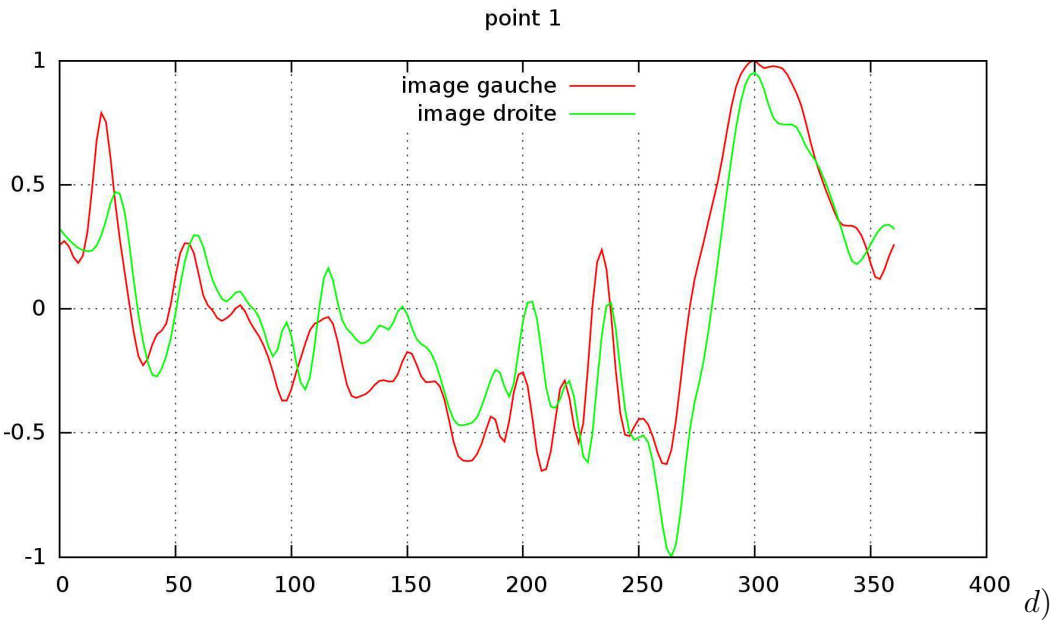
FIGURE 4.16: Signatures anti-aliasées pour une image synthétique. a) image synthétique originale. b) Signatures normalisées obtenues aux points marqués sur l'image originale avec les filtres dérivation anti-aliasés ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$).


a)


b)



c)

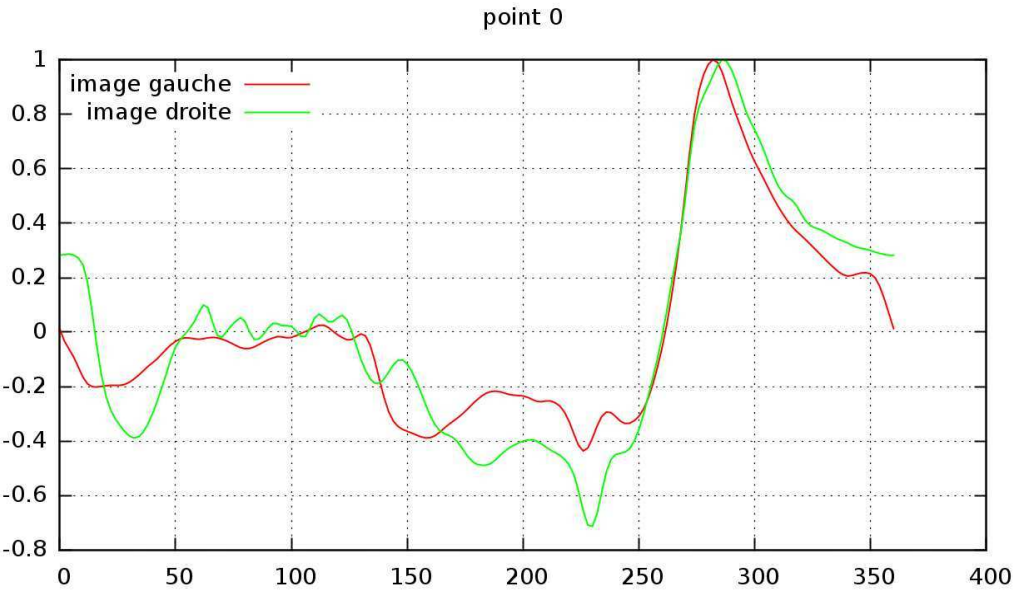


d)

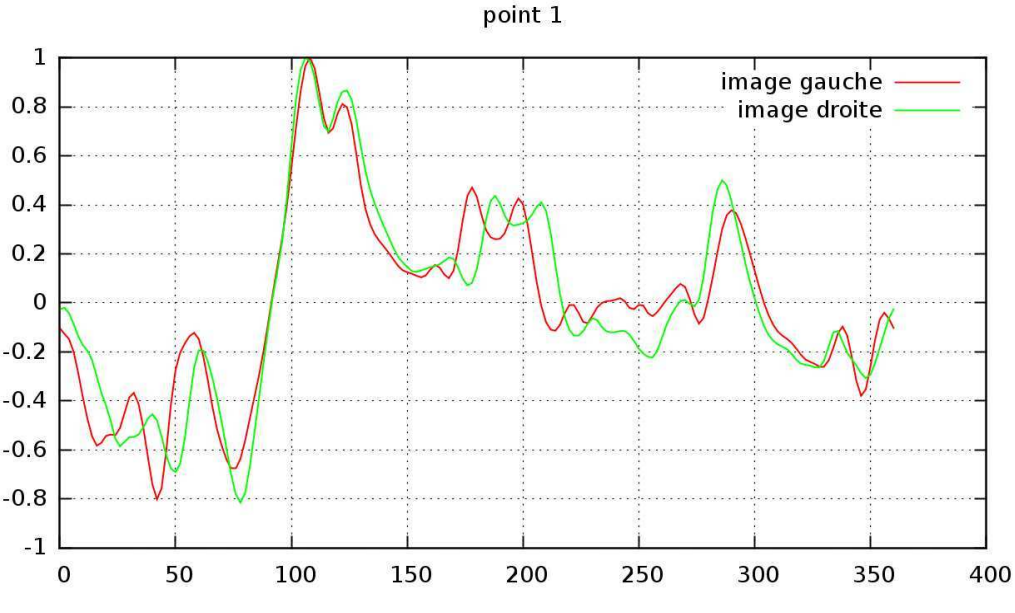
FIGURE 4.17: Signatures anti-aliasées pour dans un cas réel. a) image gauche. b) image droite. Deux points de Harris sélectionnés. c) Signatures obtenues pour le point 0 ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$). d) Signatures obtenues pour le point 1 ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$).


a)


b)



c)



d)

FIGURE 4.18: Signatures pour deux points en correspondance. a) image gauche. b) image droite. Deux points de Harris sélectionnés. c) Signatures obtenues pour le point 0 ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$). d) Signatures obtenues pour le point 1 ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$).

Chapitre 5

Une nouvelle méthode de mise en correspondance

Nous avons développé au chapitre précédent une nouvelle technique de description des points d'intérêt basée sur un filtrage dérivatif anisotrope. Nous obtenons en chaque point d'intérêt, une signature $s(\theta)$ qui décrit le signal image autour de ce point. Les signatures sont d'une part extrèmement robustes au bruit, et d'autre part elles peuvent être obtenues avec une discrétisation des orientations $\Delta\theta$ arbitraire. Nous sommes donc en mesure de caractériser très finement le signal image autour des points d'intérêt. Les expérimentations ont montré (cf. figure 4.17 et figure 4.18) que les signatures entre deux points identiques vus dans deux images différentes sont remarquablement similaires, cependant les courbes obtenues sont localement déformées et décalées. Nous proposons dans ce chapitre une nouvelle méthode affine de mise en correspondance de points d'intérêt utilisant ces signatures $s(\theta)$.

Disposant de points d'intérêt et de descripteurs, un processus de mise en correspondance s'appuie sur un calcul de scores de mise en correspondance ou de distance entre les descriptions. Nous avons vu que les descripteurs présentés au chapitre précédent, ne possèdent ni invariance euclidienne, ni invariance affine. Si nous voulons être capable de travailler avec des changements de point de vue importants, par exemple en stéréo-vision à grande base (cf. figure 4.18), ou en vidéo embarquée, alors c'est la méthode de mise en correspondance elle même, qui doit prendre en compte les transformations affines ou enclidiennes entre les images et donc entre les descripteurs.

La méthode proposée dans cette thèse se décompose en plusieurs étapes :

1. Une mise en correspondance croisée, basée sur un calcul de distance euclidienne puis affine entre les descripteurs, est réalisée.
2. Une technique de vote est utilisée afin d'éliminer les faux appariements.

Si nous comparons cette méthode avec SIFT (Lowe [56]), nous pouvons observer des différences importantes :

- SIFT est plutôt une méthode de mise en correspondance invariante à l'échelle qu'une méthode affine, SIFT n'est pas très robuste face aux changements de point de vue.
- Notre méthode n'est pas adaptée aux changements importants d'échelle, nous utilisons des points de Harris calculés avec une seule échelle. En revanche les transformations affines sont prises en compte directement.

5.1 Distance euclidienne entre descripteurs

Nos descripteurs sont des fonctions circulaires, une rotation dans le sens horaire (dans un repère image où l'axe Y est orienté vers le bas) dans le plan image va décaler la courbe vers la droite. Evidement dans ces conditions, si nous calculons la distance euclidienne entre les courbes, cette distance est affectée par la rotation. Il est donc nécessaire d'effectuer un décalage inverse d'une des deux signatures si nous voulons obtenir un résultat exploitable.

$$d(s_1,\ s_2)\ =\ \min_{\theta'}\ \sum_{\theta}\left(s_1\left(\theta\right)-s_2\left(\theta-\theta'\right)\right)^2 \tag{5.1}$$

Cette distance est liée aux maximun de corrélation entre les courbes. Ici les courbes à comparer sont réelles, la corrélation s'écrit donc :

$$c\left(\theta\right)\ =\ s_1\left(\theta\right)\ *\ s_2\left(-\theta\right) \tag{5.2}$$

En développant l'expression à minimiser dans l'équation 5.1, il vient :

$$\sum_{\theta}\left(s_1\left(\theta\right)-s_2\left(\theta-\theta'\right)\right)^2\ =\ \sum_{\theta}s_1^2\left(\theta\right)+\sum_{\theta}s_2^2\left(\theta-\theta'\right)-2\sum_{\theta}s_1\left(\theta\right)s_2\left(\theta-\theta'\right)$$

Nous avons affaire à des fonctions circulaires, les deux premières sommes sont donc constantes, nous reconnaissons la corrélation dans la troisième somme. Minimiser la distance euclidienne entre les courbes, revient donc à trouver le maximum de corrélation.

Nous avons donc plusieurs possibilités pour calculer une distance euclidienne :

1. Estimation de l'angle de rotation en considérant la différence d'angles entre les pics les plus forts des deux signatures, puis calcul de distance en tenant compte de cette différence d'angle.
   Le recalage entre les signatures est local.
2. Estimation de l'angle de rotation entre les points d'intérêt en utilisant le gradient image, puis calcul de distance en tenant compte de cette différence d'angle. Le recalage est donné par le signal image. Cette méthode nécessite le calcul de l'orientation du gradient, qui peut être estimée directement sur les courbes comme

dans [85], ou encore avec un calcul de gradient plus classique.
Même si cette méthode est intéressante, elle ne permet pas directement d'obtenir la distance entre nos signatures. Nous voulons obtenir le minimun global de distance, il faut donc encore rechercher le minimum global sur un sous ensemble réduit de décalages.

3. Effectuer tous les décalages θ', et calculer la distance minimale entre entre les signatures, nous obtenons un minimum global, en revanche cette méthode reste très coûteuse en temps calcul.
4. Effectuer une corrélation entre les signatures, le maximum de corrélation, nous permet alors d'obtenir le décalage θ' correspondant à la distance cherchée. Cette corrélation peut être obtenue avec un faible coût calcul dans l'espace de Fourier, en utilisant une méthode de FFT [41].

Ici nous avons donc choisi d'utiliser, la méthode de corrélation par FFT, qui nous donne une distance minimale directe. Nous présentons deux résultats obtenus sur deux images extraites d'une séquence couleur sous-marine réelle, d'un buste immergé. Les conditions d'acquisition difficiles (18 mètres de fond et beaucoup de particules en suspension) font que le canal rouge contient l'essentiel de l'information. Nous avons donc une paire stéréoscopique constituée des deux plans rouges extraits des images initiales dont le contraste est extrêmement faible.

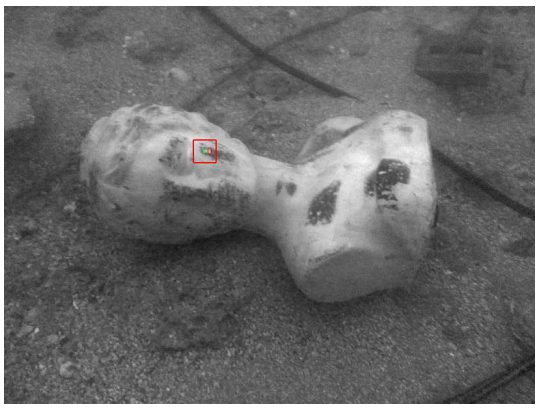
Nous avons sélectionné deux points de Harris sur le buste, nous présentons les résultats du recalage par corrélation dans l'éspace de Fourier sur deux figures : les figures 5.1 et 5.2 La figure 5.1 présente les résultat de recalage par corrélation dans l'espace de Fourier sur le point 0, les figures 5.1.a) et 5.1.b) présente les images et le point d'intérêt. La figure 5.1.c) présentent les deux courbes initiales. La figure 5.1.d) présentent les deux courbes après recalage.

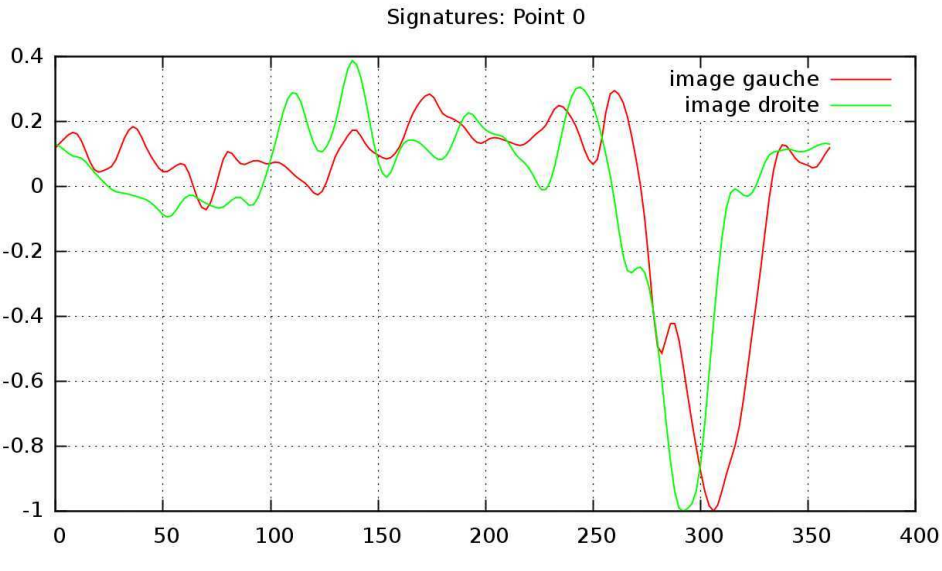
Dans cet exemple, le buste apparaît dans les deux images avec un changement de point de vue important. Le point 0 présente des distorsions projectives importantes ce qui se traduit par un décalage important des pics des signatures (figure 5.1.c). Après la mise en correspondance euclidienne par corrélation (figure 5.1.d) les pics principaux se correspondent mais ce n'est pas le cas pour les pics secondaires. Les distorsions de nature projective ne sont pas corrigées par la mise en correspondance euclidienne.

Dans le cas du deuxième point : point 1, cet effet est moins marqué, mais nous pouvons tout de même remarquer de légers décalages des pics secondaires après recalage.

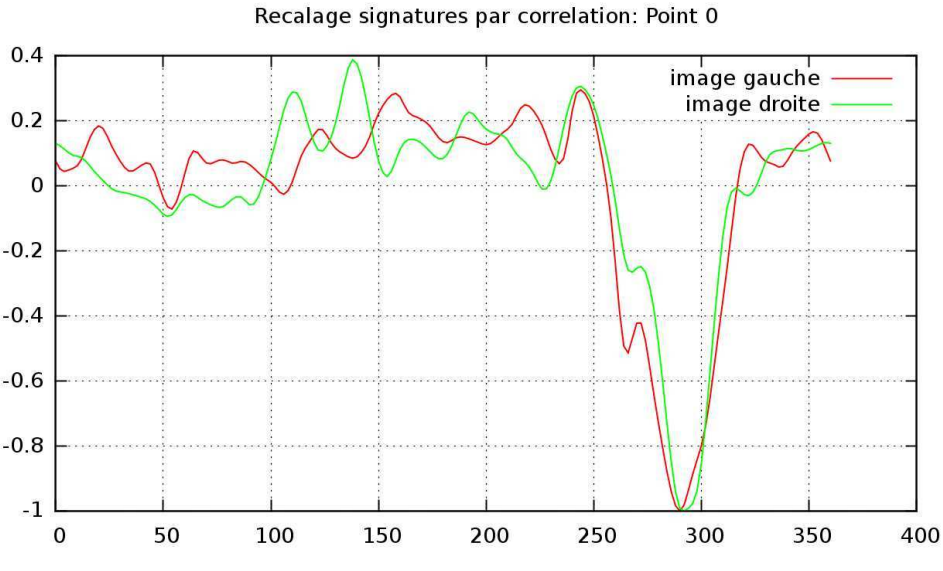
Cet exemple montre clairement que dès que le changement de point de vue devient important, la mise en correspondance euclidienne ne donne plus de résultats exploitables, il devient alors nécessaire de déformer les signatures au cours du processus de mise en correspondance.


a)


b)



c)

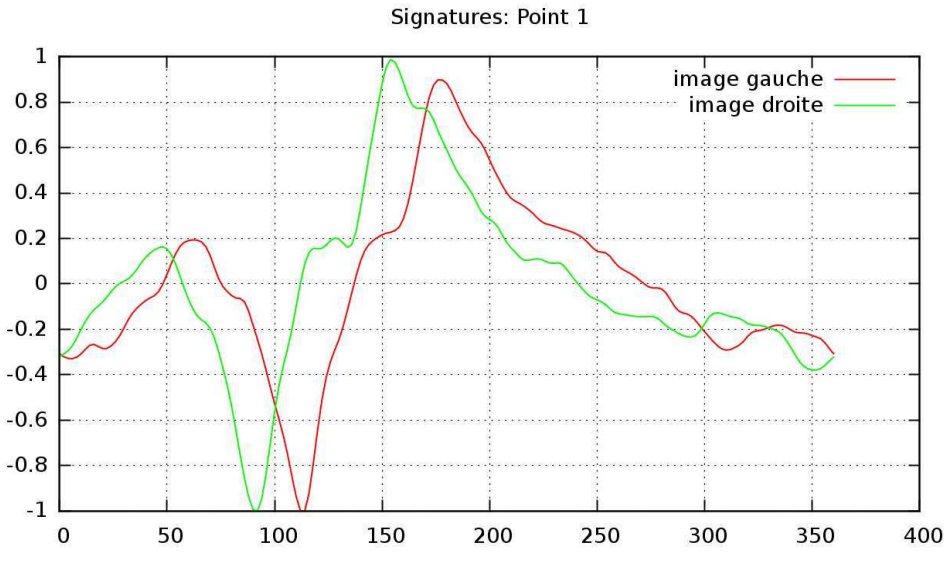


d)

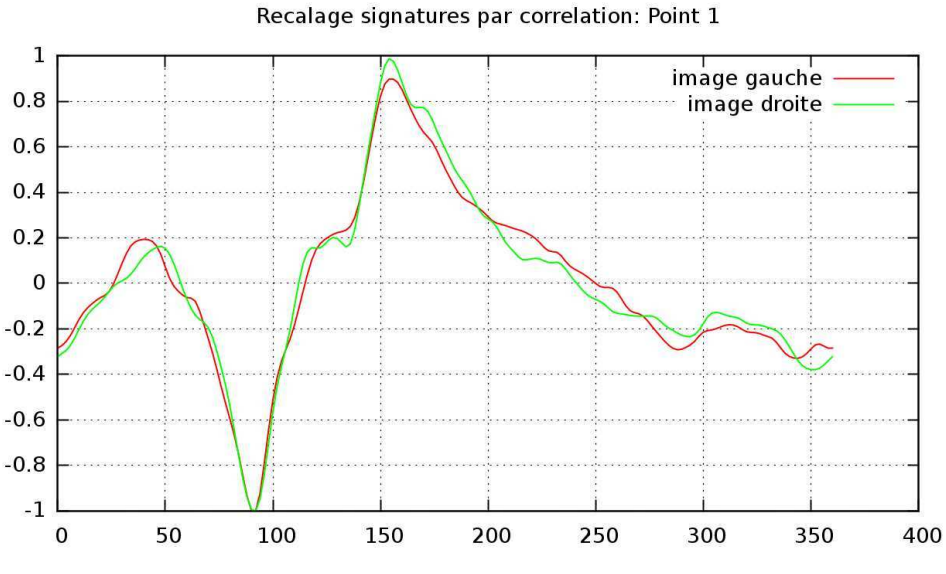
FIGURE 5.1: Recalage des signatures par corrélation (point 0). Un point de Harris sélectionné ($\sigma = 1$) : point 0. a) image gauche, b) image droite. c) Signatures initiales obtenues pour le point 0 ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$). d) Recalage par corrélation dans l'espace de Fourier.

a)

b)



c)



d)

FIGURE 5.2: Recalage des signatures par corrélation (point 1). Un deuxième point de Harris sélectionné ($\sigma = 1$) : point 1. a) image gauche, b) image droite. c) Signatures initiales obtenues pour le point 1 ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$). d) Recalage par corrélation dans l'espace de Fourier.

5.2 Distance "affine" entre descripteurs

Nous avons remarqué au chapitre 4 que les pics des signatures décrivant les points d'intérêt et plus particulièrement les coins correspondent à la valeur du gradient selon l'orientation du contour qui "traverse" ce coin. Sachant que les angles ne sont pas préservés par transformation projective ou affine (nous avons examiné deux exemples à la section précédente), les signatures vont présenter des déformations : dilatation pour certaines orientations et contraction pour d'autres, si nous voulons mettre en œuvre une distance entre signatures invariante affine, nous allons devoir nous affranchir des éventuelles déformations locales. Ceci est valable bien entendu, indépendamment du bruit, des problèmes d'illumination, des éventuels problèmes de discrétisation des images, etc.

Nous allons donc définir ici, une stratégie nous permettant de calculer des distances correctes entre signatures malgré la transformation due au changement de point de vue.

5.2.1 Recalage de signaux

Le recalage de signaux 1D est un problème largement étudié en traitement de signal ou des images [106], [8], [80]. Formellement, le problème du recalage de deux signaux s_1 et s_2, nous considérons uniquement ici des fonctions réelles, consiste à déterminer une transformation a telle que :

- $s_1 \circ a \simeq s_2$: la transformation a doit être telle que la distance entre $s_1 \circ a$ et s_2 soit la plus petite possible, compte tenu de la fonction distance choisie (l'opérateur $\circ$ correspond à l'opérateur mathématique de combinaison des applications).
- a doit être régulière. Dans notre cas nous aurons à considérer des transformation affines, mais dans le cas général il est souvent admis que a doit être inversible et dérivable.

On trouve le plus souvent deux grandes classes de méthodes :

- Des méthodes de recalages par "Landmarks". On entend par Landmarks, un ensemble de points caractéristiques du signal ou de l'image. Pour une courbe, il peut s'agir de points caractéristiques comme par exemple des maxima locaux ou des points d'inflexion. Pour une image, il peut même s'agir de points d'intérêt. Par exemple dans des applications basées sur du morphing de visage, les landmarks sont souvent les coins des yeux, de la bouche, etc. Pour des courbes, les landmarks sont souvent obtenus par une analyse multi-échelle du signal, par exemple dans [8] l'auteur utilise des transformées en ondelettes.
- Des méthodes globales, qui cherchent à déterminer une transformation régulière au sens d'une certaine distance. Dans cette catégorie, on trouve de nombreuses méthodes variationnelles. Dans le cas du recalage d'images, il existe de nombreuses applications médicales visant à recaler des données issues de plusieurs capteurs différents, etc.

Les méthodes développées ici sont souvent basées sur la résolution numérique d'une Equation aux Dérivées Partielles (EDP) obtenue par minimisation d'une énergie, energie dans laquelle apparaîssent des termes de distance et de régularisation.

5.2.2 Recalage des signatures par DTW

Nous nous intéréssons maintenant à l'estimation de la distance entre signatures. Dans la mesure où deux signatures qui correspondent au même point sont relativement proches (après recalage par corrélation), il semble qu'une méthode globale soit plus adaptée à notre problème. Compte tenu du nombre de distances que nous allons devoir calculer, nous avons besoin d'une méthode de recalage rapide et non itérative (la plupart des méthodes par EDP sont itératives).

Parmi le méthodes globales, il existe une technique utilisant une méthode d'optimisation par programmation dynamique [5], [6] : "Dynamic Time Warping" (DTW) [47], [45], [46], répondant parfaitement à notre problème. Cette méthode à été largement utilisée, notamment dans des probèmes de reconnaissance vocale, pour déformer localement des signaux temporels afin de les comparer.

La figure 5.3 présente notre problème de calcul de distance de signatures par recalage.

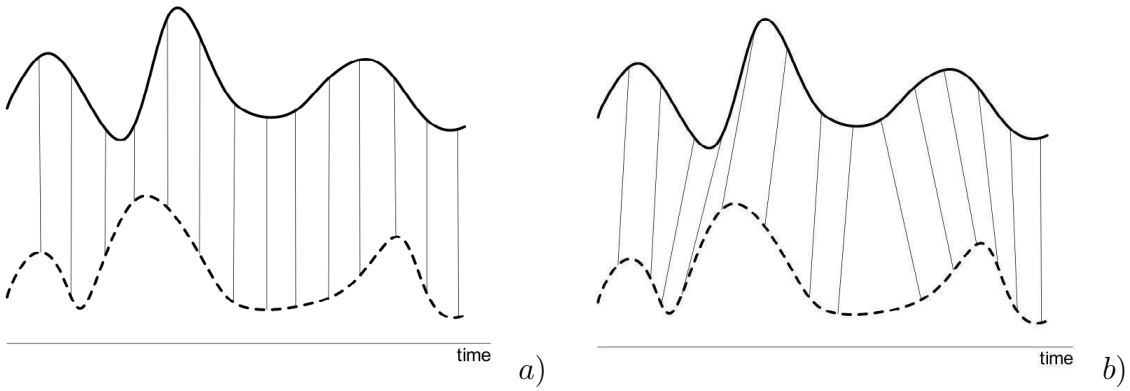


FIGURE 5.3: Calcul de distance entre signatures par recalage. a) calcul de distance point à point, la déformation entraine une fausse estimation de la "ressemblance" entre les courbes. b) Si nous connaissons une fonction de transformation a, alors nous pouvons estimer la "ressemblance" entre les deux courbes de manière plus réaliste.

Nous présentons ici la méthode DTW. Cette méthode est basée sur un algorithme de programmation dynamique. Il s'agit de construire une matrice de coût minimal des décalages possibles entre les courbes, le chemin minimal est alors trouvé par un parcours inverse dans cette matrice. La figure 5.4 présente le schéma de principe de cette méthode

DTW, soit la recherche d'un chemin dans cette matrice.

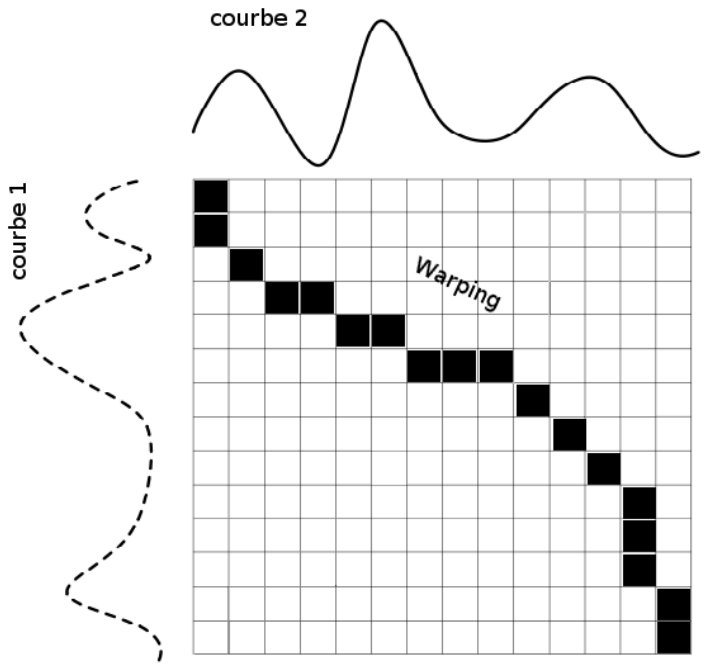


FIGURE 5.4: Transformation des courbes. La fonction de transformation de la courbe s_1 en la courbe s_2 est donnée par le parcours inverse des minima dans la matrice $D[i][j]$ en partant du coin bas-droite vers le coin haut-gauche.

Soit s_1 et s_2 les deux courbes, déjà recalées approximativement par corrélation. Considérons la matrice $d[i][j]$ du carré des différences des courbes point à point.

$$d[i][j] = \big(s_1[i] - s_2[j]\big)^2\big) = \begin{pmatrix} (s_1[1] - s_2[1])^2 & (s_1[1] - s_2[2])^2 & \cdots & (s_1[1] - s_2[n])^2 \\ (s_1[2] - s_2[1])^2 & (s_1[2] - s_2[2])^2 & \cdots & (s_1[2] - s_2[n])^2 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ (s_1[n] - s_2[1])^2 & \cdots & \cdots & (s_1[n] - s_2[n])^2 \end{pmatrix}$$

Puis la matrice $D[i][j]$ représentant la somme des carrés des distances le long d'un chemin de coût minimal, cette matrice est construite de la manière suivante :

- $D[1][1] = d[1][1]$
- $D[1][j] = d[1][j] + D[1][j-1]$
- $D[i][1] = d[i][1] + D[i-1][1]$

et

- $D[i][j] = d[i][j] + \min\{D[i-1][j-1], D[i-1][j], D[i][j-1]\}$

Enfin, la fonction de transformation de la signature s_1 en s_2 est donnée par un chemin dans la matrice $D[i][j]$. Ce chemin est obtenu en parcourant cette matrice du coin bas-droite vers le coin haut-gauche et en choisissant le minimum à chaque étape.

5.2.3 Recalage contraint

Afin d'éviter des transformations non compatibles avec les transformations affines que nous cherchons à prendre en compte, nous voulons contraindre la DTW à donner des chemins solutions proches de la diagonale de la matrice $D[i][j]$.

Pour cela, nous ajoutons simplement un terme de pénalisation à la matrice $d[i][j]$, qui vaut zéro sur la diagonale de la matrice et qui augmente lorsque l'on s'en éloigne. Ce terme de pénalisation est donné par une fonction $C(x)$ où x est la distance à la diagonale de la matrice prise dans la direction orthogonale. Nous avons besoin ici d'une fonction qui reste proche de zéro dans une certaine bande autour de la diagonale, les fonctions polynomes de degré pair supérieurs à 2 possèdent cette propriété. Cette propriété est illustrée à la figure 5.6. Les expérimentations ont montré qu'une simple fonction polynôme du $6^{ième}$ degré (equation 5.3) donnait les résultats les plus intéressants (nous avons testé des fonctions du deuxième jusqu'au huitième degré) :

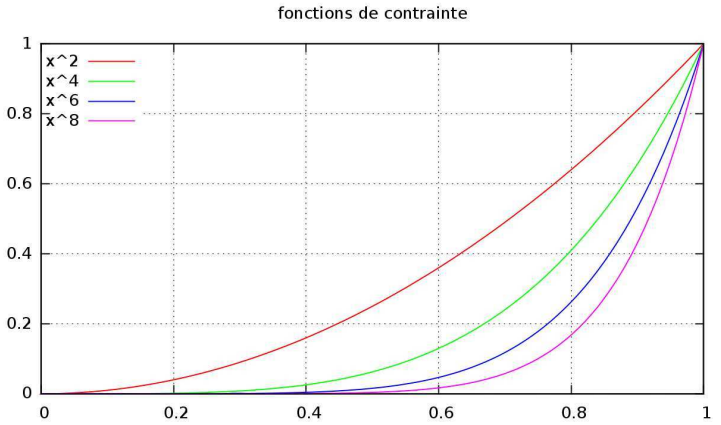
$$C : [0,1] \longrightarrow \Re, \qquad C(\ x\ ) \ = \ \epsilon \cdot x^6 \tag{5.3}$$

Où :

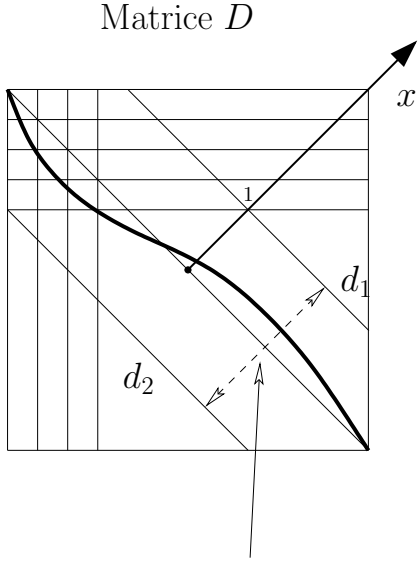
ϵ est un facteur de normalisation.

C'est la position des droites d_1 et d_2 par rapport à la diagonale qui fixe la valeur de ϵ, Dans nos expérimentations les droites d_1 et d_2 sont positionnées à $(\frac{1}{5})^{ième}$ de la demi-diagonale de la matrice D (de part et d'autre de la diagonale).

Nous présentons des résultats obtenus sur les deux points sélectionnés de la séquence "buste" (figures figure 5.2 et figure 5.1), nous présentons d'abord le point 1 (le cas le plus simple) puis le point 0.



a)



b)

FIGURE 5.5: Fonctions de pénalisation de la déformation. a) Fonctions de pénalisation de la déformation (polynômes du $2^{ième}$, $4^{ième}$, $6^{ième}$, $8^{ième}$ degré), les expérimentations ont montré qu'une fonction du $6^{ième}$ degré donnait de bons résultats. b) La fonction définie à l'équation 5.3 va définir une bande autour de la diagonale (droites d_1 et d_2) dans laquelle la déformation va être autorisée.

- Les figures 5.6 et 5.7 présentent les résultats de "warping" obtenus au point 1, Pour ce point, les signatures initiales étaient décalées (rotation) mais tout de même assez proches, et la mise en correspondance dans ce cas ne présente aucune difficulté pour notre méthode.
- Les résultats obtenus au point 0 sont présentés aux figures 5.8 et 5.9. Visuellement ces signatures sont beaucoup plus déformées que les précédentes. Néanmoins l'algorithme DTW contraint donne des résultats très intéressants.

Evidemment, un "warping" de courbes par DTW ne garantit pas que la transformation d'une courbe à une autre soit une transformation affine. Cependant, lorsque deux courbes sont très dissemblables, les contraintes que nous imposons via la fonction de contrainte utilisée éliminent des chemins irréalistes (loin de la diagonale) dans la matrice D, et conduisent à une distance entre courbes forte. Lorsque des courbes sont proches, la DTW trouve une "bonne" solution avec une erreur faible.

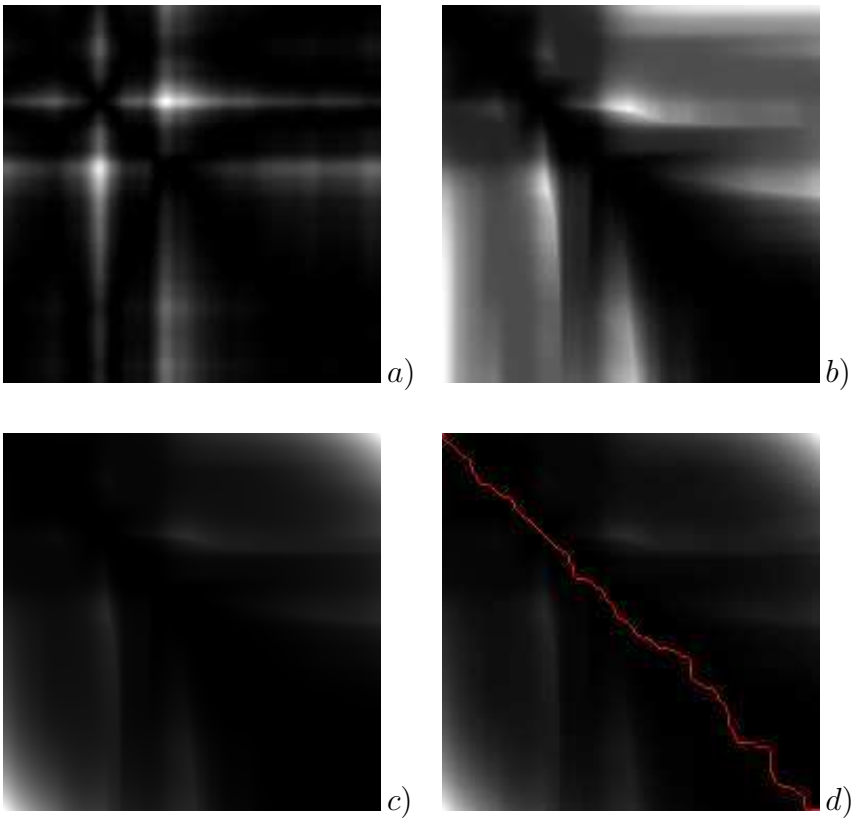


FIGURE 5.6: Résultat de "warping" obtenu sur la séquence "buste", point 1 (cas le plus simple). a) Image de coût de déformation : matrice $d[i][j]$. b) Matrice $D[i][j]$. c) Matrice $D[i][j]$ contrainte. d) Chemin obtenu avec contraintes.

Nous allons maintenant examiner le cas de points ne se correspondant pas. Nous avons sélectionné manuellement deux faux appariements dans la séquence "buste", ces points ainsi que les signatures initiales obtenues sont présentés à la figure 5.10. Les résultats de recalage par corrélation et de "warping" sont présentés sur les figures 5.11 pour le premier point et 5.12 pour le deuxième. Ces résultats montrent clairement que c'est notre fonction de contrainte qui empêche une déformation trop importante. En effet si nous supprimons cette contrainte, les courbes après "warping" deviennent trop similaires, la distance entre signatures est du même ordre que dans le cas de points en correspondance, et la méthode n'est plus discriminante. La figure 5.13 présente les résultats de "warping" avec suppression de la contrainte.

En résumé, pour qu'une mise en correspondance automatique donne de bons résultats, le "warping" est nécessaire pour éliminer les éventuelles transformations affines dues aux changements de point de vue (au bruit, etc.) mais doit rester dans des limites fixées par les droites d_1 et d_2 imposées par notre fonction de contrainte (cf. figure 5.6). Enfin nous présentons les résultats synthétiques de calcul de distance entre les signatures sur le tableau 5.1 pour les couples de points en correspondance (figure 4.17 points 0 et 1, figure 4.18 points 0 et 1, figure 5.1 points 0 et 1) et sur le tableau 5.2 nous comparons les distances calculées avec "warping" contraint et non contraint pour les deux faux appariements présentés ci-dessus (figure 5.10).

	fig. 4.17 point 0	fig. 4.17 point 1	fig. 4.18 point 0	fig. 4.18 point 1	fig. 5.1 point 0	fig. 5.1 point 1
initial	1.405465	8.117418	3.946277	9.610162	9.610162	21.414492
corrélation	1.405465	4.982680	3.604830	2.869964	2.869964	0.680834
"warping"	0.275893	1.160626	0.710255	0.215929	0.176799	0.085102

TABLE 5.1: Exemples de calcul de distances entre signatures. Calcul de distances obtenues sur les appariement présentés aux figures : 4.17, 4.18 et 5.1 ("warping" contraint).

fig. 5.10	couple 0 avec contrainte	couple 0 sans contrainte	couple 1 avec contrainte	couple 1 sans contrainte
initial	98.771941	98.771941	34.476251	34.476251
corrélation	17.120829	17.120829	29.769452	29.769452
"warping"	7.675630	0.716891	5.927044	3.944674

TABLE 5.2: Comparaison des distances contraintes et non-contraintes. Comparaison des calcul de distances obtenues sur les faux appariements de la séquence "buste" présenté à la figure 5.10 : "warping" contraint et "warping" non contraint.

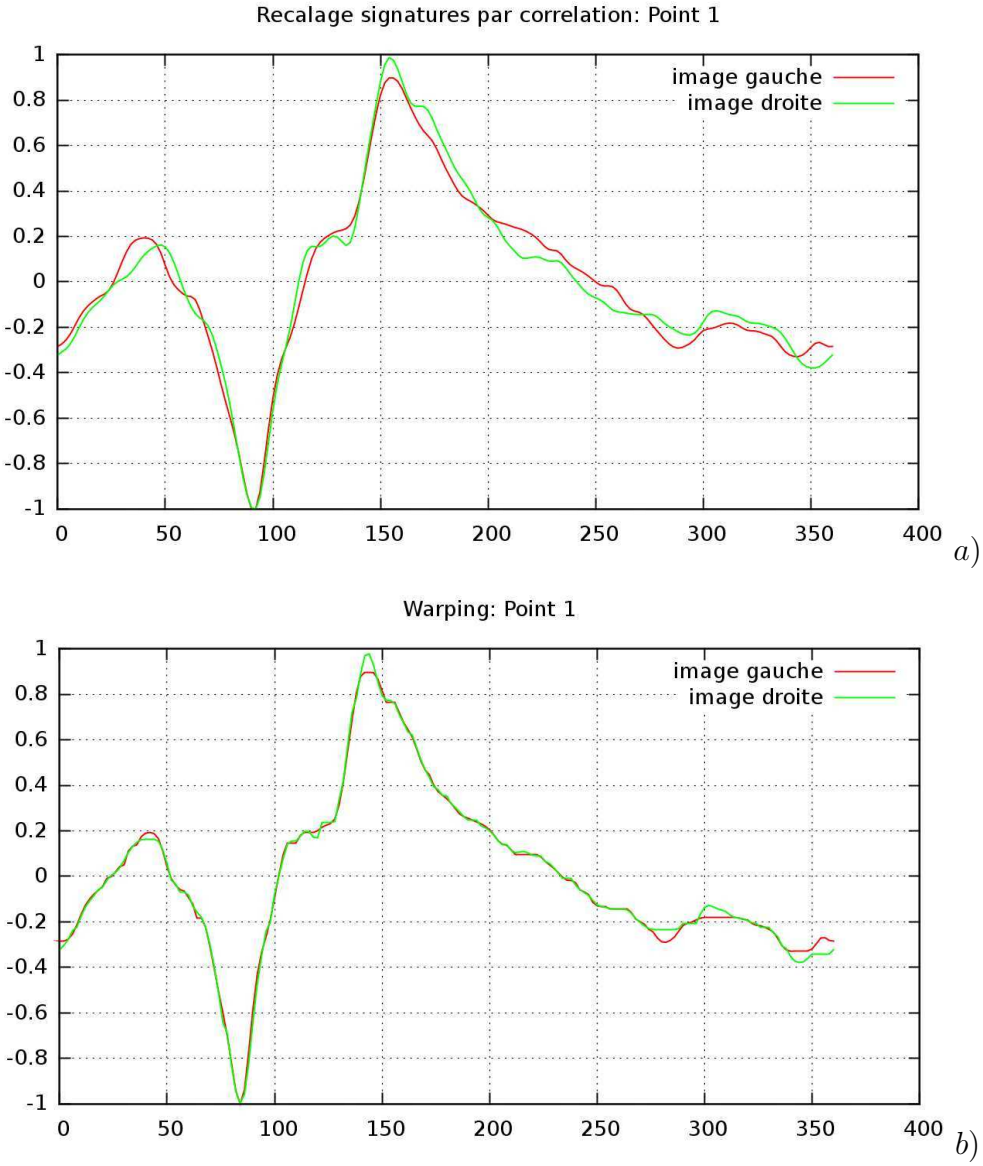


FIGURE 5.7: Résultat de "warping" obtenu sur la séquence "buste", point 1 (cas le plus simple). a) Signatures après recalage par corrélation. b) Mise en correspondance par DTW contrainte.

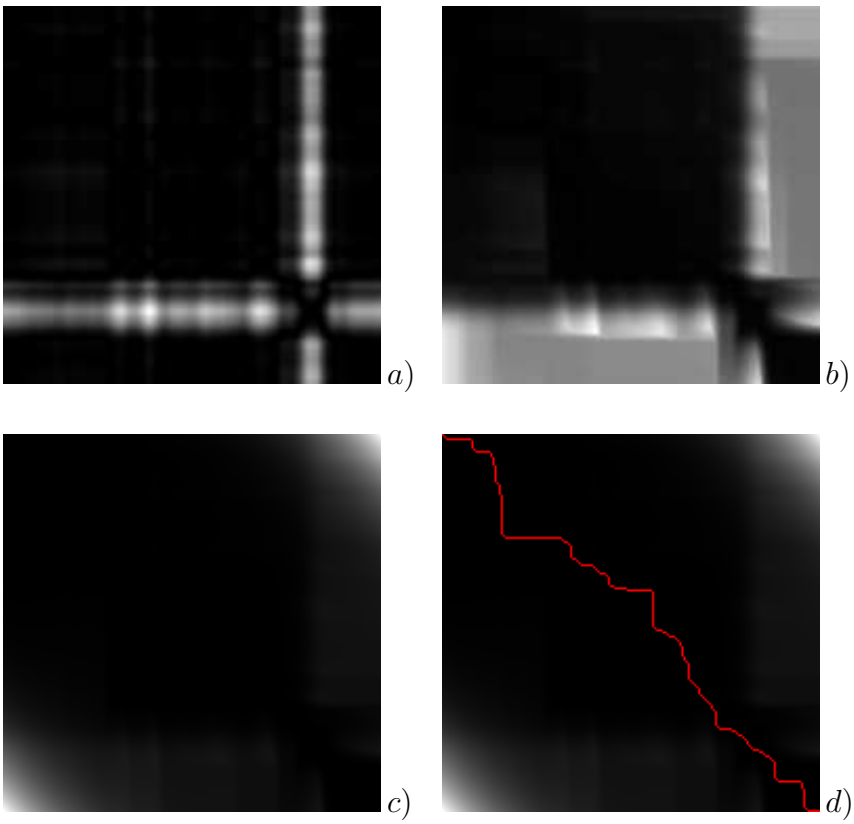


FIGURE 5.8: Résultat de "warping" obtenu sur la séquence "buste", point 0 (cas plus complexe). a) Image de coût de déformation : matrice $d[i][j]$. b) Matrice $D[i][j]$. c) Matrice $D[i][j]$ contrainte. d) Chemin obtenu avec contraintes.

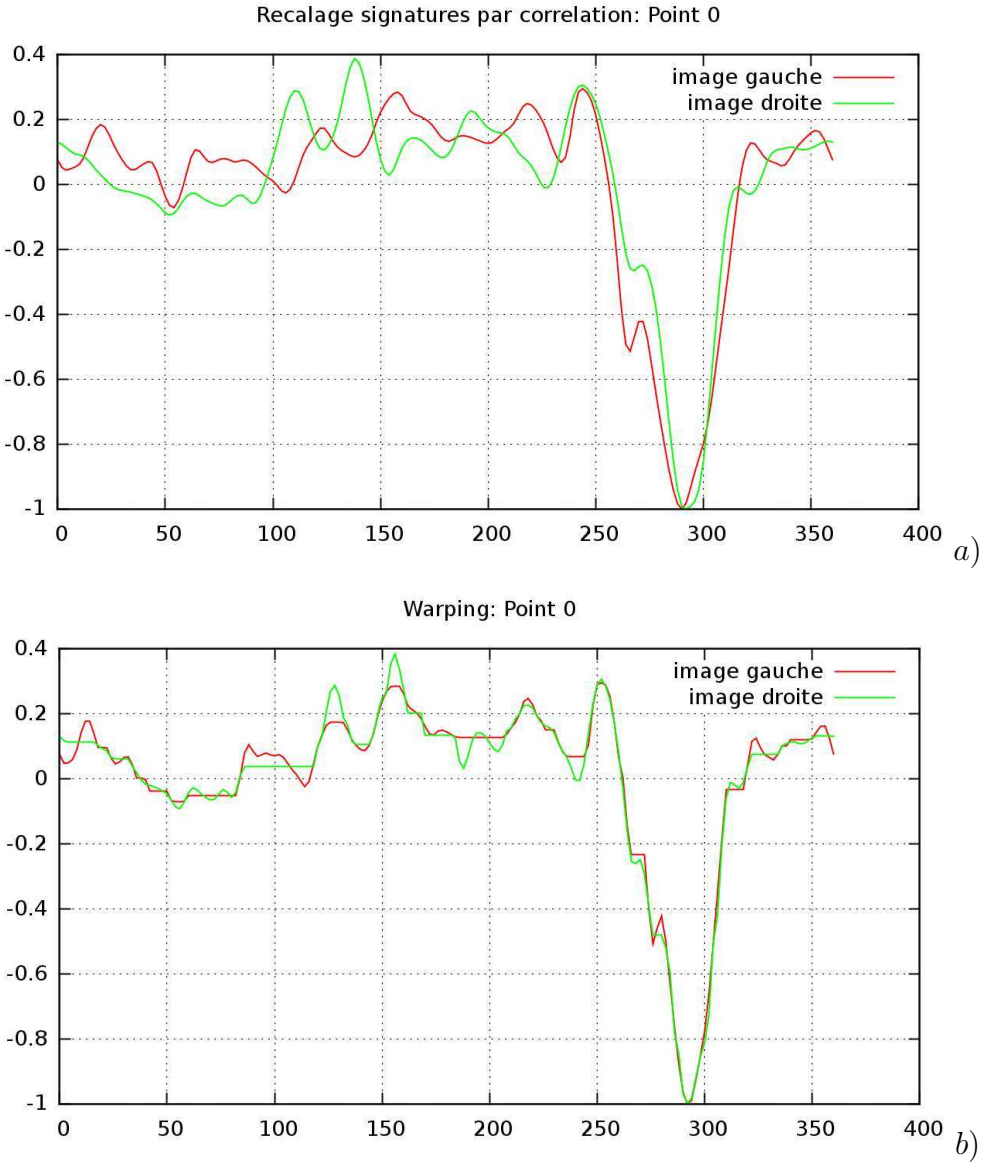
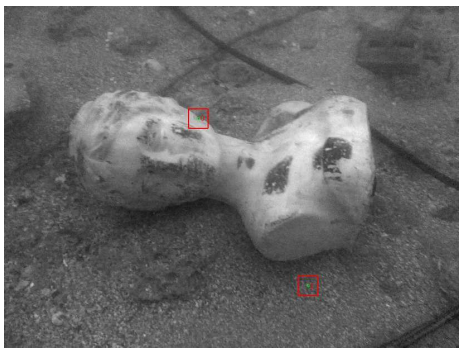
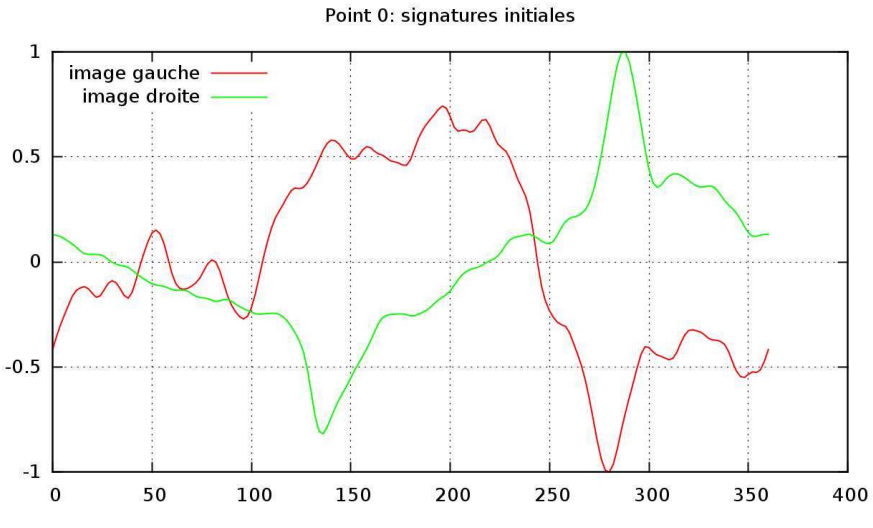


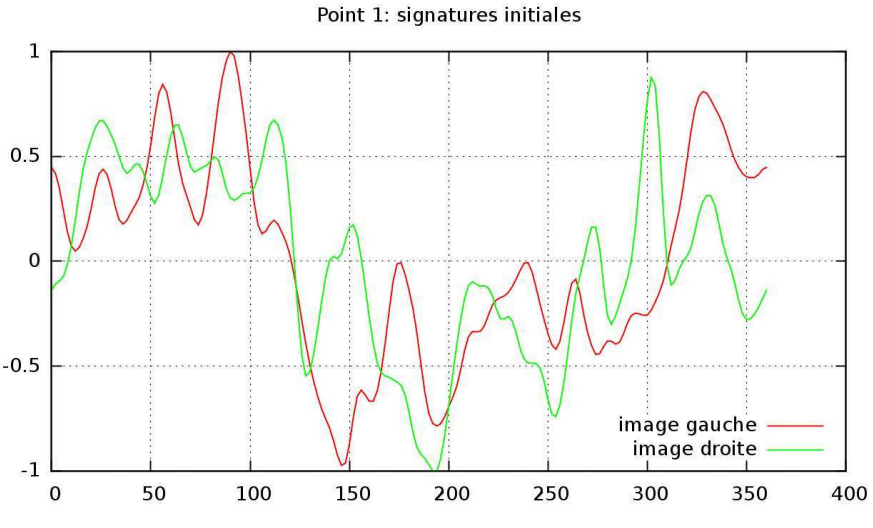
FIGURE 5.9: Résultat de "warping" obtenu sur la séquence "buste", point 0 (cas plus complexe). a) Signatures après recalage par corrélation. b) Mise en correspondance par DTW contrainte.


a)


b)



c)



d)

FIGURE 5.10: Recalage de signatures dans le cas de faux appariements. Deux faux appariements sont séléctionnés manuellement (points de Harris). le premier appariement comporte un point sur une structure géométrique et un point dans la texture. Le deuxième appariement comporte deux points électionné dans la texture : a) image gauche, b) image droite. c) Signatures initiales point 0 ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$). d) Signatures initiales point 1 ($\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta\theta = 2°$).

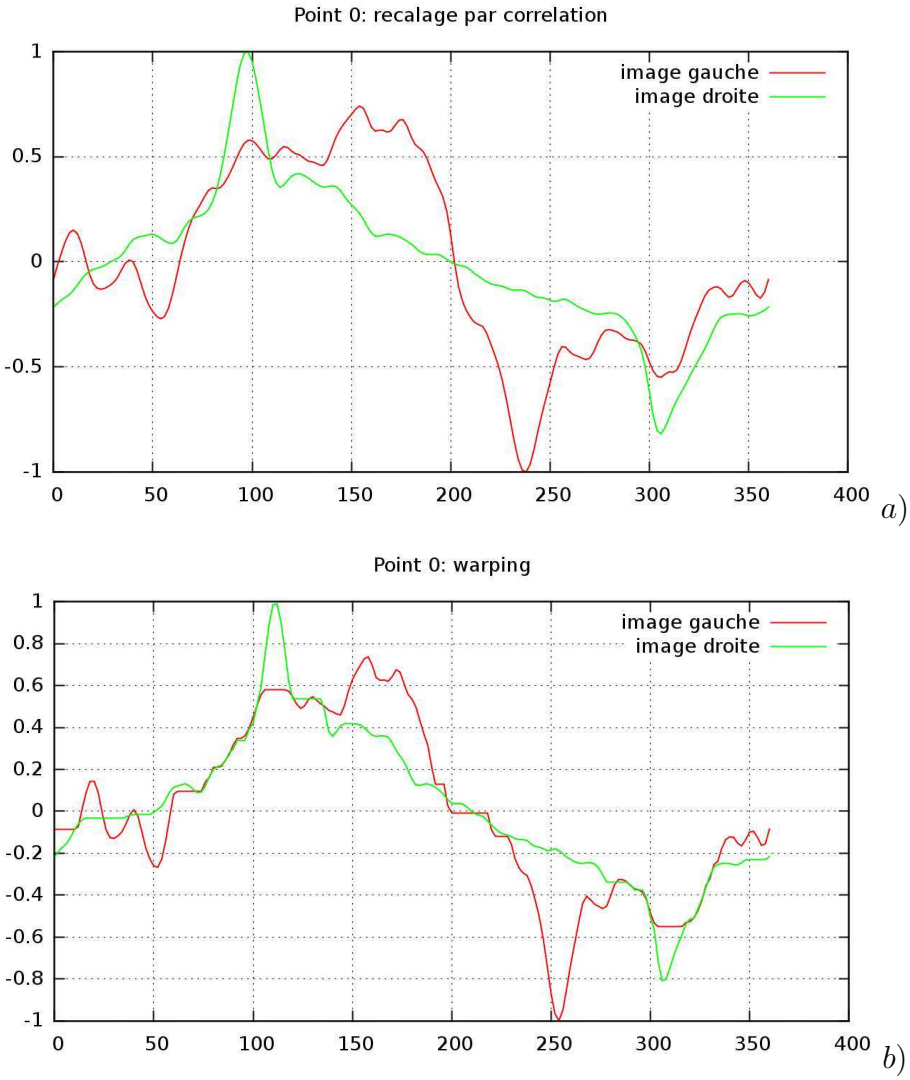


FIGURE 5.11: Résultats de recalage pour le point 0. a) Recalage par corrélation des signatures. b) "Warping" avec contrainte.

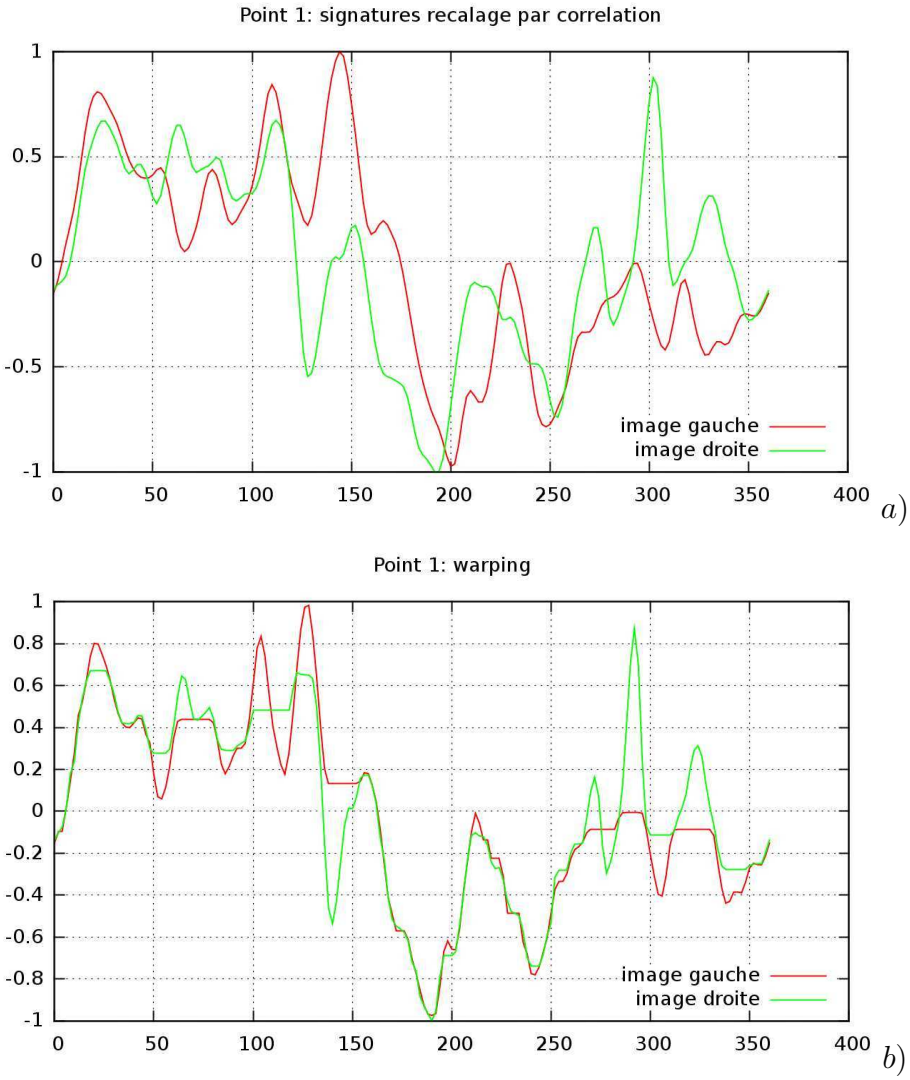


FIGURE 5.12: Résultats de recalage pour le point 1. a) Recalage par corrélation des signatures. b) "Warping" avec contrainte.

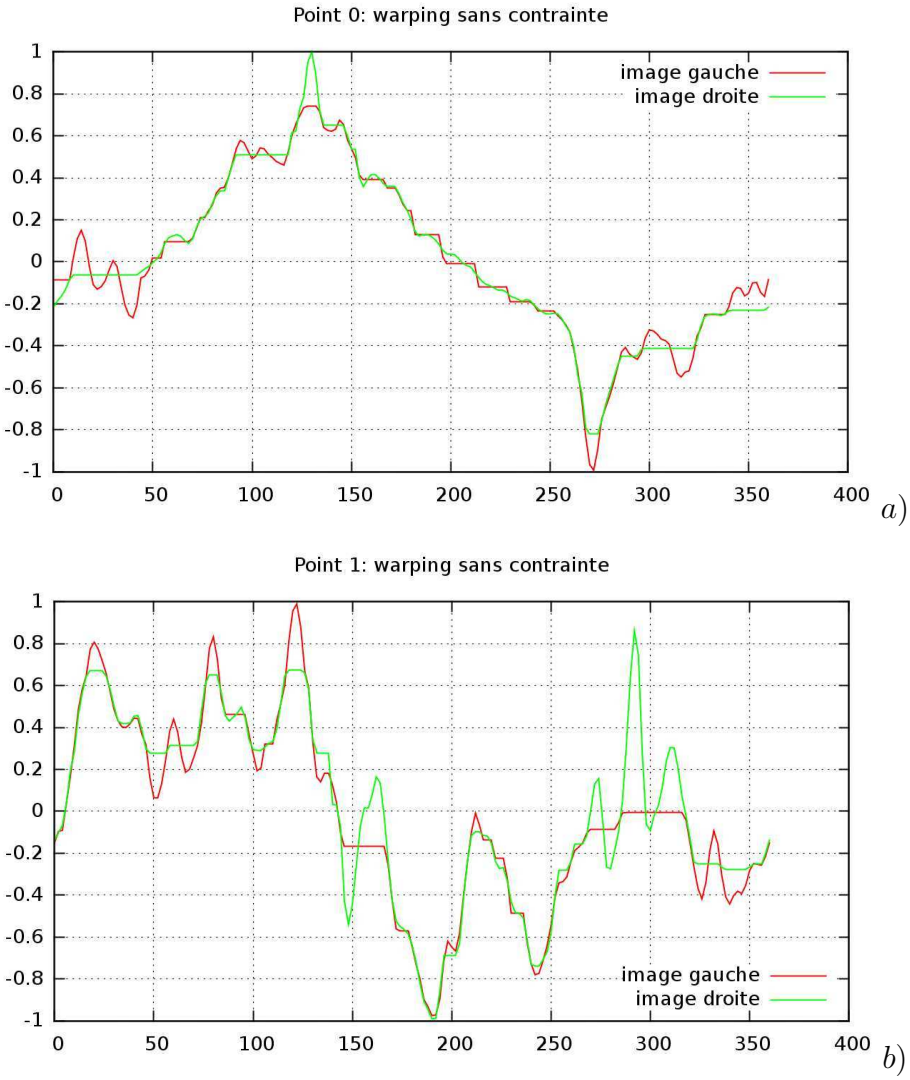


FIGURE 5.13: Résultats : "warping" sans contrainte. a) "Warping" point 0. b) "Warping" point 1.

5.3 Mise en correspondance

Au cours de ce chapitre, nous avons décrit point par point les différentes parties de notre méthode de mise en correspondance. Nous allons ici effectuer un récapitulatif et présenter la méthode dans sa globalité.

La méthode s'effectue en trois étapes :

1. Extraction des points et calcul des descripteurs.
   Cette première étape est décrite par l'algorithme 2.
   Nous calculons les signatures et leur transformée de fourier, qui serviront pour la corrélation dans l'espace de Fourier.

2. Mise en Correspondance.
   Nous effectuons ici une mise en correspondance croisée, l'algorithme 3 présente la mise en correspondance dans le sens image 1 vers image 2.
   Puis nous effectuons une mise en correspondance image 2 vers image 1, suivi d'une étape de filtrage des appariements.

3. Seuillage des appariements.
   Utilisation d'une méthode statistique basée un vote de vecteurs de déplacement, cette méthode est décrite en détails à la section 5.3.1.

Algorithme 2 *Extraction des points et calcul des descripteurs*

pour *chaque image* **faire**
- *extraire les points d'intérêt par buckets*

fin pour
pour *chaque point i provenant de la* $1^{ière}$ *image* **faire**
- *calcul_descripteur* $\rightarrow$ $s_1[i]$
- $FFT(\ s_1[i]\ ) \rightarrow \hat{s}_1[i]$

fin pour
pour *chaque point i provenant de la* $2^{ième}$ *image* **faire**
- *calcul_descripteur* $\rightarrow$ $s_2[i]$
- $FFT(\ s_2[i]\ ) \rightarrow \hat{s}_2[i]$

fin pour

Etant donné que deux points appartenant à des structures différentes dans l'image peuvent avoir des signatures proches, c'est le cas par exemple des structures répétitives dans des scènes d'intérieur (fenêtres, stores, étagères, etc.), alors la méthode de mise en correspondance développée ci-dessus peut être mise en défaut. Nous avons donc mis en

Algorithme 3 *Mise en correspondance affine*

pour *chaque point i provenant de la $1^{ière}$ image* **faire**
  pour *chaque point j provenant de la $2^{ième}$ image* **faire**
    - $maximum\_corrélation(\ \hat{s}_1[i],\ \hat{s}_2[j]\ ) \rightarrow phase$
    - $rotation(s_1[i],\ phase) \rightarrow s_{1,rot}[i]$
    - $CDTW(s_{1,rot}[i],\ s_2[j]) \rightarrow score$
    - $corresp[i] = j$ *si score est meilleur que le précédent*
  fin pour
fin pour

œuvre une méthode de vote qui permet (dans une certaine mesure) d'éliminer les faux appariements dus à l'ambigüité des scènes considérées, cette méthode est décrite à la section suivante.

5.3.1 Une méthode de vote pour éliminer les faux appariements

Cette technique est basée sur le principe de mouvement d'ensemble de points. En effet un ensemble de points proches dans une scène 3D projetés dans plusieurs images avec des points de vue différents vont rester proches en 2D. Disposant d'un ensemble d'appariements, nous voulons accumuler les vecteurs déplacement afin de pouvoir éliminer les couples isolés dans l'espace des déplacements (cf. figure 5.14). L'espace des déplacements est caractérisé par l'ensemble des vecteurs translation possibles d'une image I_1 vers une image I_2. La technique de vote que nous mettons en œuvre est très proche d'une transformation de Hough [38], [20].

Dans le cas de simples translations d'un objet dans l'image, nous obtenons une accumulation exacte des vecteurs translation, cependant les transformations que nous avons à prendre en compte sont beaucoup plus complexes, nous avons vu que ces transformations sont dans le cas général des transformations projectives. il est alors nécessaire d'introduire une notion d'incertitude dans les votes, nous accumulons donc une densité gaussienne pour chaque vecteur (en pratique un écart-type de 10 donne de bons résultats).

La méthode implémentée procède comme suit :

La figure 5.15 présente l'accumulateur obtenu dans un cas réel, à partir du couple stéréoscopique de la figure 2.7. La section suivante est dédiée aux résultats obtenus avec la méthode dans sa globalité.

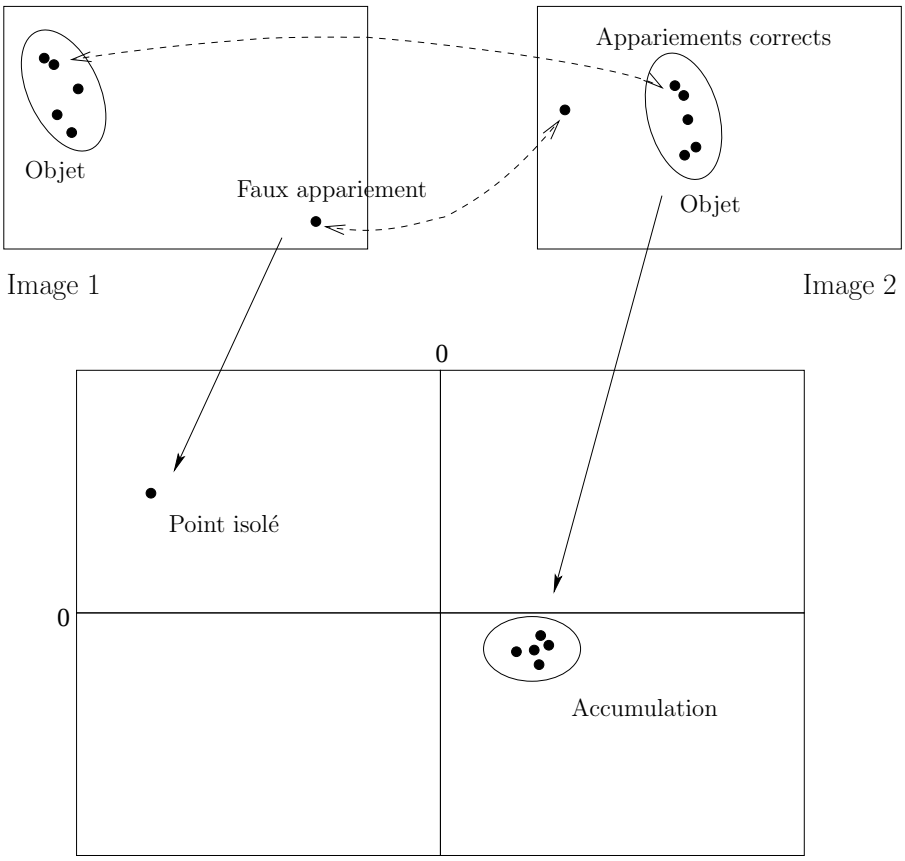


FIGURE 5.14: Espace des translations. Chaque couple de points appariés donne un point dans l'espace des déplacements.

Algorithme 4 *Seuillage des appariements*

pour *pour chaque couple de correspondants* **faire**
- *incrémenter l'accumulateur avec une densité gaussienne* ($\sigma = 10$).

fin pour
pour *pour chaque couple de correspondants* **faire**
si *la valeur de l'accumulateur dépasse un seuil (nombre de vecteurs déplacement)*
alors
- *garder ce couple.*

sinon
- *éliminer ce couple.*

fin si
fin pour


a)


b)

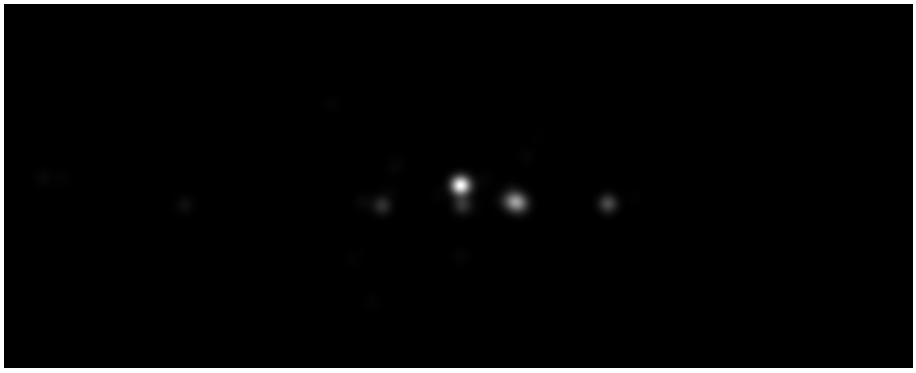

c)

FIGURE 5.15: Résultats d'accumulation. les deux pics pricipaux correspondent au personnage et au fond.

5.4 Résultats

5.4.1 Résultats Stéréo grande base

Nous présentons tout d'abord les resultats obtenus sur trois couples stéréoscopiques, deux d'entre eux ont servi à illustrer ces deux précédents chapitres (Gwenaelle, Buste).

Nous sommes dans un cadre de stéréo-vision non-calibrée, nous avons mis en œuvre notre méthode en trois étapes :

1. Première étape de mise en correspondance (500 points de Harris) sans contraintes géométriques.

2. Estimation de la matrice fondamentale en utilisant les résultats obtenus à l'étape 1.

3. Deuxième étape de mise en correspondance (plus de 10000 points) en utilisant la contrainte épipolaire.

5.4.1.1 Séquence Gwenaelle

Nous présentons sur la figure 5.16 les résultats complets sur le couple stéréoscopique à grande base déjà présenté à la figure 4.18, nous obtenons 632 appariements.

Nous présentons à la figure 5.17 la reconstruction $3D$ obtenue à partir de ces appariements, nous permet de mieux appréhender les erreurs d'appariement, cependant la base stéréoscopique étant très importante, la plupart des motifs du fond sont occultés par le personnage au premier plan, nous aurons donc, pour ce couple stéréoscopique, une mauvaise reconstruction du fond.

5.4.1.2 Séquence Eric

Nous présentons sur la figure 5.18 un deuxième couple stéréoscopique à grande base pour la reconstruction 3D de visages, nous obtenons 1267 appariements avec un taux d'erreur de 1% à 2%.

Nous présentons à la figure 5.19 la reconstruction $3D$ obtenue à partir de ces appariements. Nous avons affaire ici à des zones homogènes importantes, notre méthode de sélection essaie de trouver 10000 points de "Harris" dans chacune des images (par buckets), et ceci, même si le nombre total de points d'intérêt dans ces images est inférieur. Nous obtenons quelques faux appariements dans ces zones, que nous supprimons a posteriori manuellement (il aurait évidement été possible d'introduire un seuil pour éliminer les points non significatifs) nous obtenons alors un taux de faux appariements inférieur à 1%.

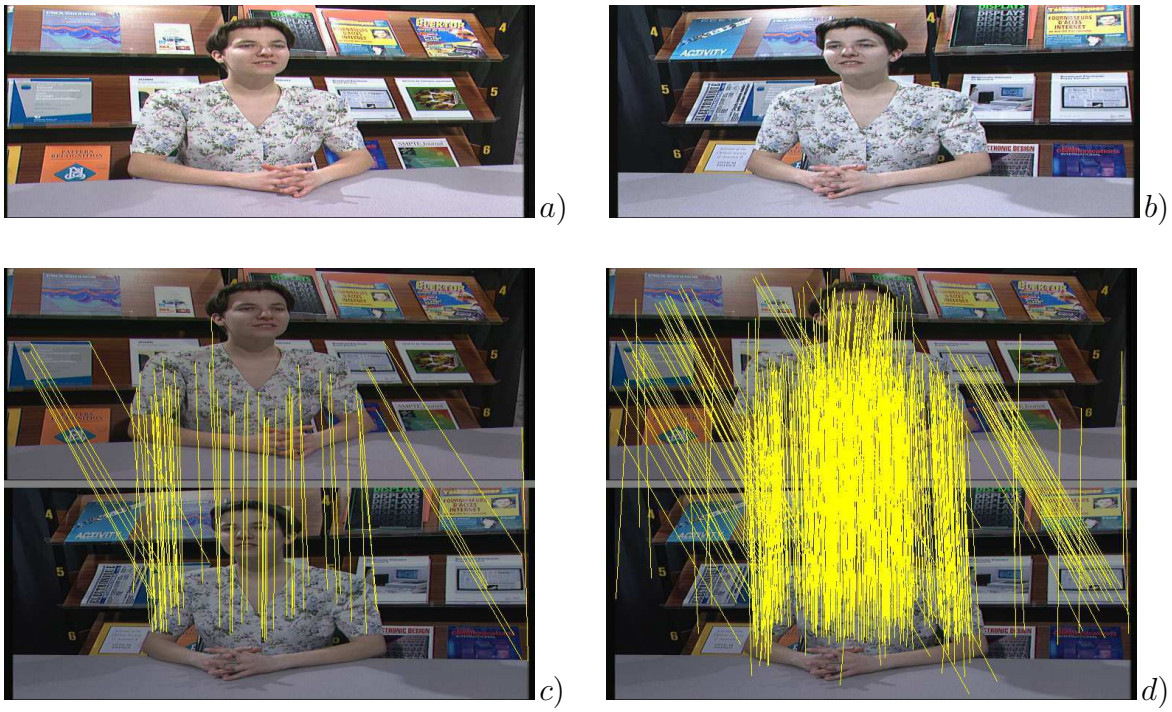


FIGURE 5.16: Mise en correspondance (Gwenaelle). images initiales : a) et b). c) Resultat de mise en correspondance en séléctionnant 500 points de Harris initiaux, $\sigma_x i = 10$, $\sigma_\eta = 1$, $\Delta\theta = 15°$, nous obtenons 64 appariements (1 faux). d) Resultat de mise en correspondance en séléctionnant 10000 points de Harris initiaux, utilisation de la contrainte épipolaire, nous obtenons 632 appariements (environ 20 faux appariements soit un taux de faux appariements inférieur à 5%).

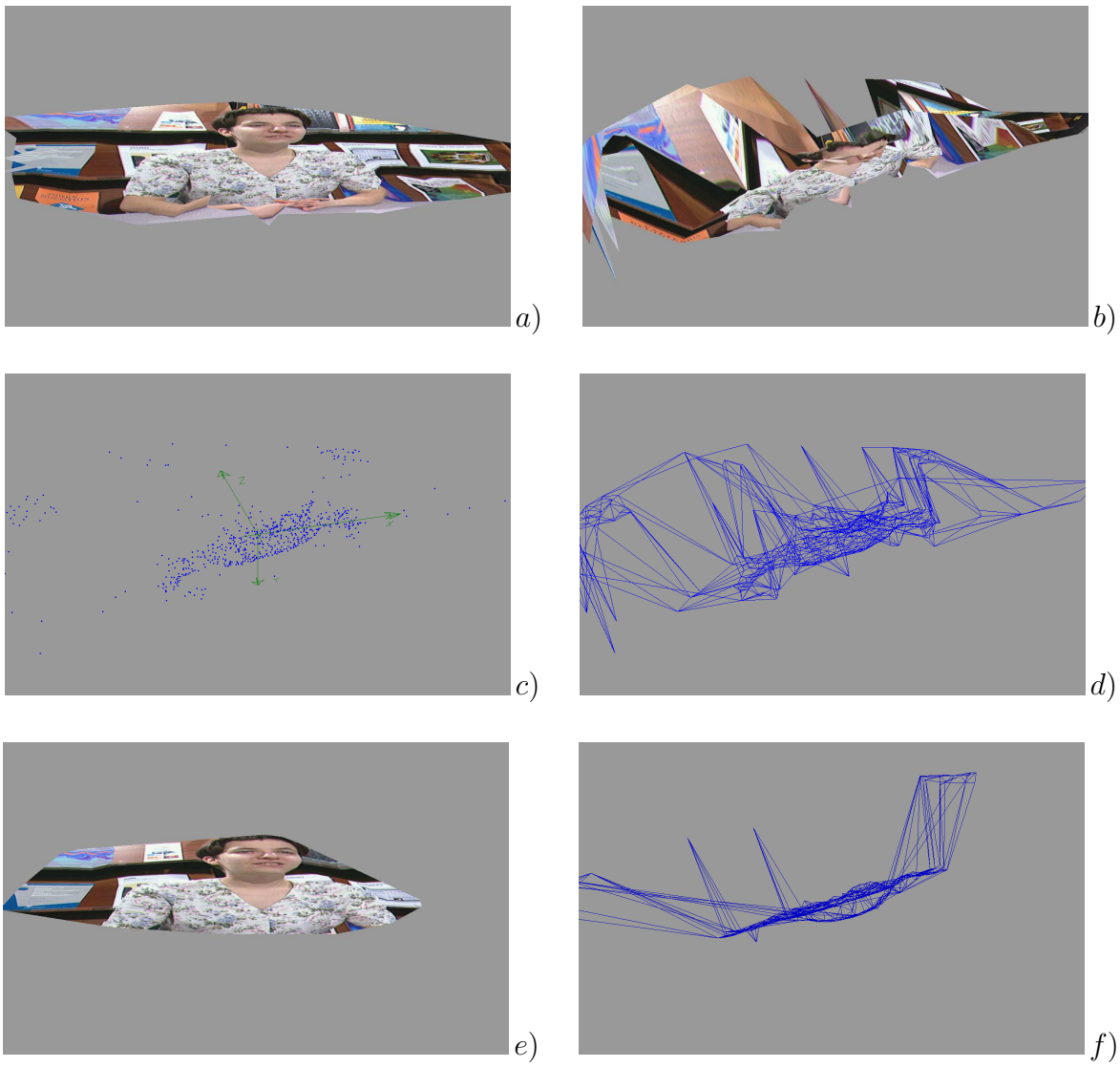


FIGURE 5.17: Reconstruction 3D (Gwenaelle). Reconstruction brute : Différentes vues 3D, triangles texturés, points, triangles filaire.

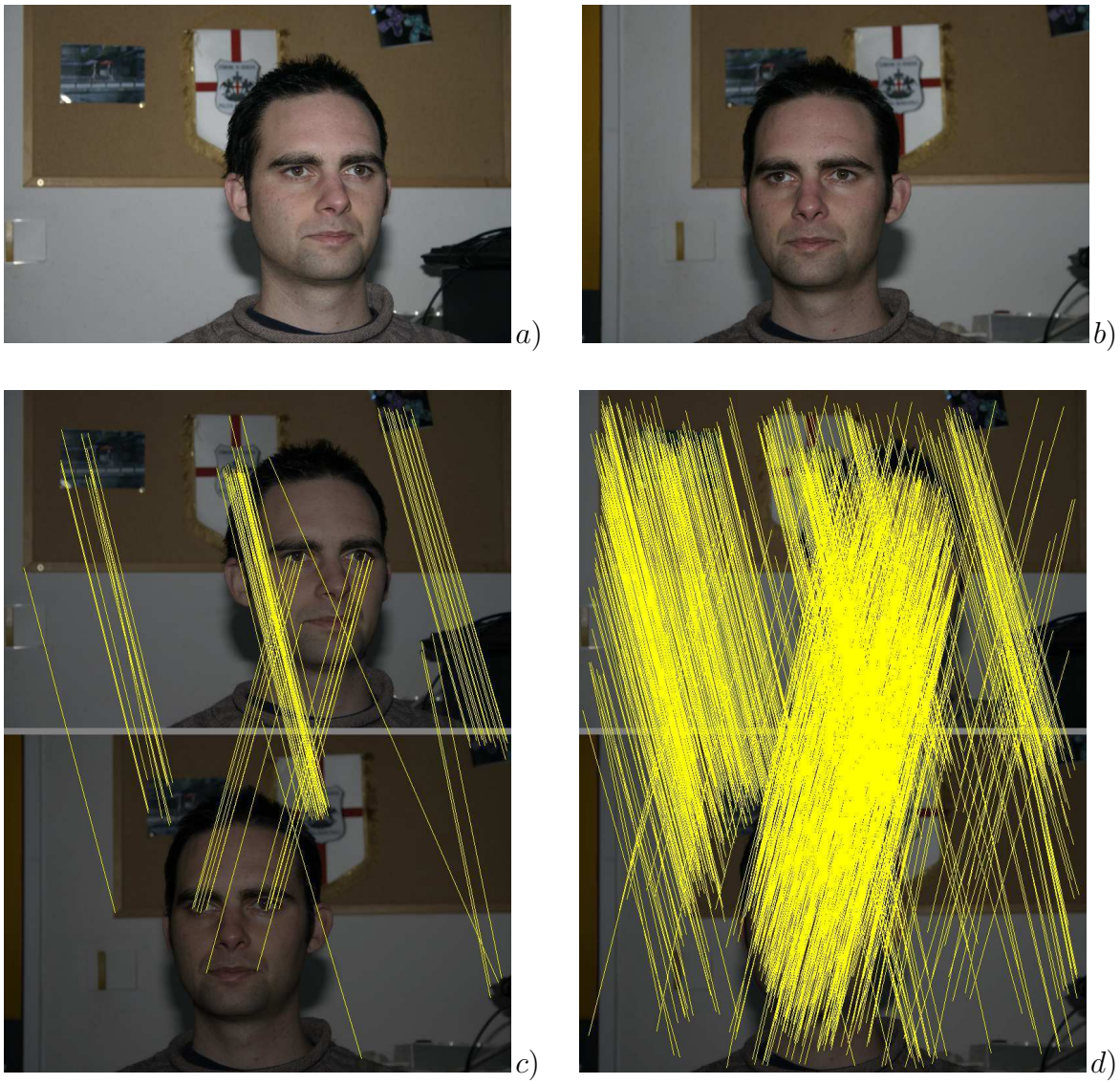


FIGURE 5.18: Mise en correspondance (Eric). images initiales : a) et b). c) Resultat de mise en correspondance en séléctionnant 500 points de Harris initiaux, $\sigma_x i = 10$, $\sigma_\eta = 1$, $\Delta\theta = 15°$, nous obtenons 77 appariements (2 sont faux). d) Resultat de mise en correspondance en séléctionnant 10000 points de Harris initiaux, utilisation de la contrainte épipolaire, nous obtenons 1267 appariements.


a)


b)

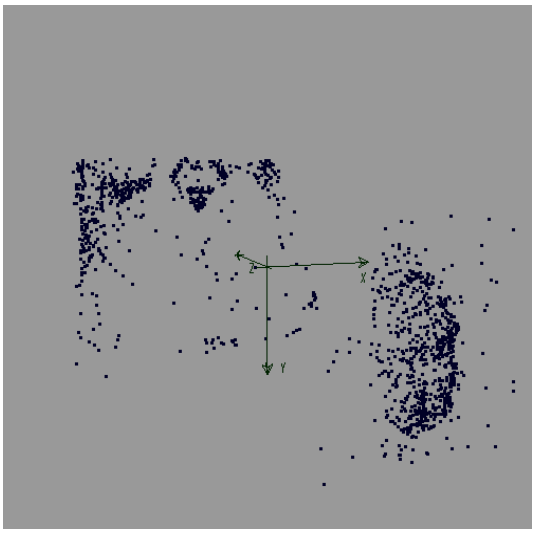

c)

FIGURE 5.19: Reconstruction 3D (Eric) à partir des images : 5.18.a et 5.18.b. a) et b) Deux vues texturées (on peut voir les quelques erreurs dans les zones homogènes. c) Points reconstruits 3D.

5.4.1.3 Séquence Buste

Nous avons déja présenté ce couple stéréoscopique à la figure 5.1, dans cet exemple nous utilisions uniquement le plan rouge, nous allons maintenant utiliser la couleur. Etant donné la mauvaise qualité des images initiales, nous allons effectuer un pré-traitement afin de faire ressortir le maximum d'information.

Dans un premier temps, les images sont normalisées indépendamment canal par canal afin déliminer la prépondérence du vert, puis quelques itérations d'équation de la chaleur inverse sont appliquées afin d'obtenir des images plus nettes [16]. La figure 5.20 présente cette étape (5.20.a et 5.20.b montrent les images originales, les résultats de restauration sont montrés en 5.20.c et 5.20.d). Même si les résultats obtenus ont l'air plutôt corrects, les détails du fond 5.20.e) et 5.20.f) illustrent la difficulté du problème : au niveau des petites structures (pierres, gravillons, sable) les images ne se ressemblent pas : d'une part la transformation projective est très importante et d'autre part le flou et les particules en suspension introduisent une modification du signal image extrêmement importante.

Comme pour les exemples traités précédemment, la mise en correspondance est réalisée en deux étapes :

1. Première étape de mise en correspondance sans contraintes géométriques. Nous séléctionnons 1000 points de Harris car les images sont grandes et contiennent beaucoup de détails. Nous obtenons 45 appariements dont 2 faux. Cette étape est utilisée pour estimer la matrice fondamentale en ajoutant manuellement quelques points sur le fond.
2. Deuxième étape de mise en correspondance en utilisant la contrainte épipolaire. Nous sélectionnons cette fois 50000 points. Nous obtenons 944 appariements : soit 242 sur le fond et 702 sur le buste. En grande majorité les appariements sur le fond sont faux. Les petits détails dans la texture sable du fond ne sont pas préservés d'une image à l'autre, ceci est dû en grande partie à la mauvaise qualité des images et aux nombreuses particules en suspension qui modifient le signal image. En conséquence les points de Harris obtenus dans le sable ne sont pas répétables, il ne peuvent donc pas être mis en correspondance.

   Considérons le buste : nous obtenons des appariements à l'intérieur de l'objet et des appariements localisés sur les contours. Compte tenu de la forme arrondie de l'objet (objet non géométrique) et du changement de point de vue important les appariements trouvés sur les contours du buste sont certainement faux, nous éliminons donc ces points (manuellement). Nous obtenons alors 682 appariements, en grande majorité corrects.

Afin de visualiser les résultats obtenus, nous ajoutons à nouveau manuellement quelques appariements sur le fond, puis nous effectuons la reconstruction 3D. Ces résultats sont présentés à la figure 5.21 pour les deux premières étapes de mise en correspondance puis

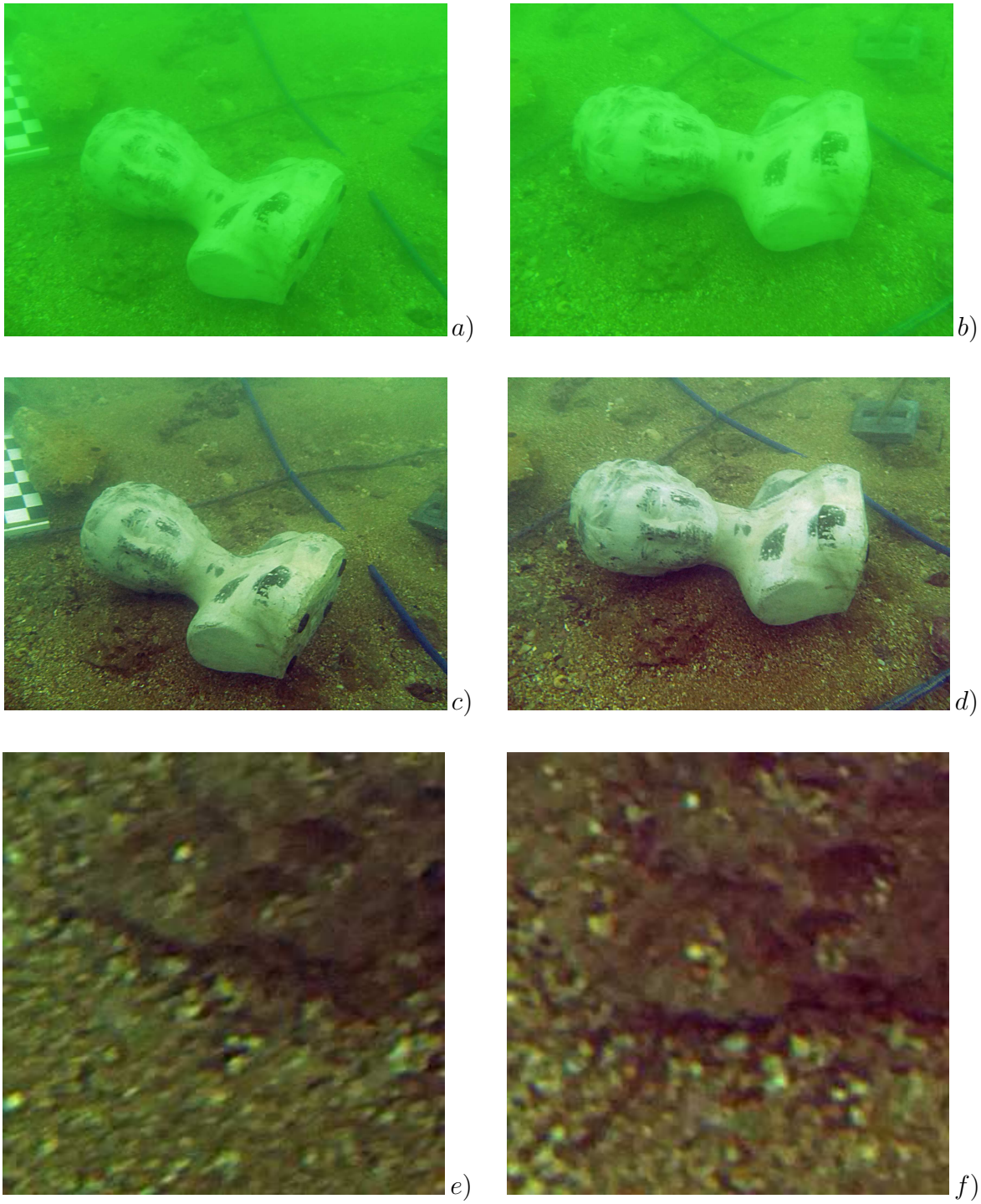
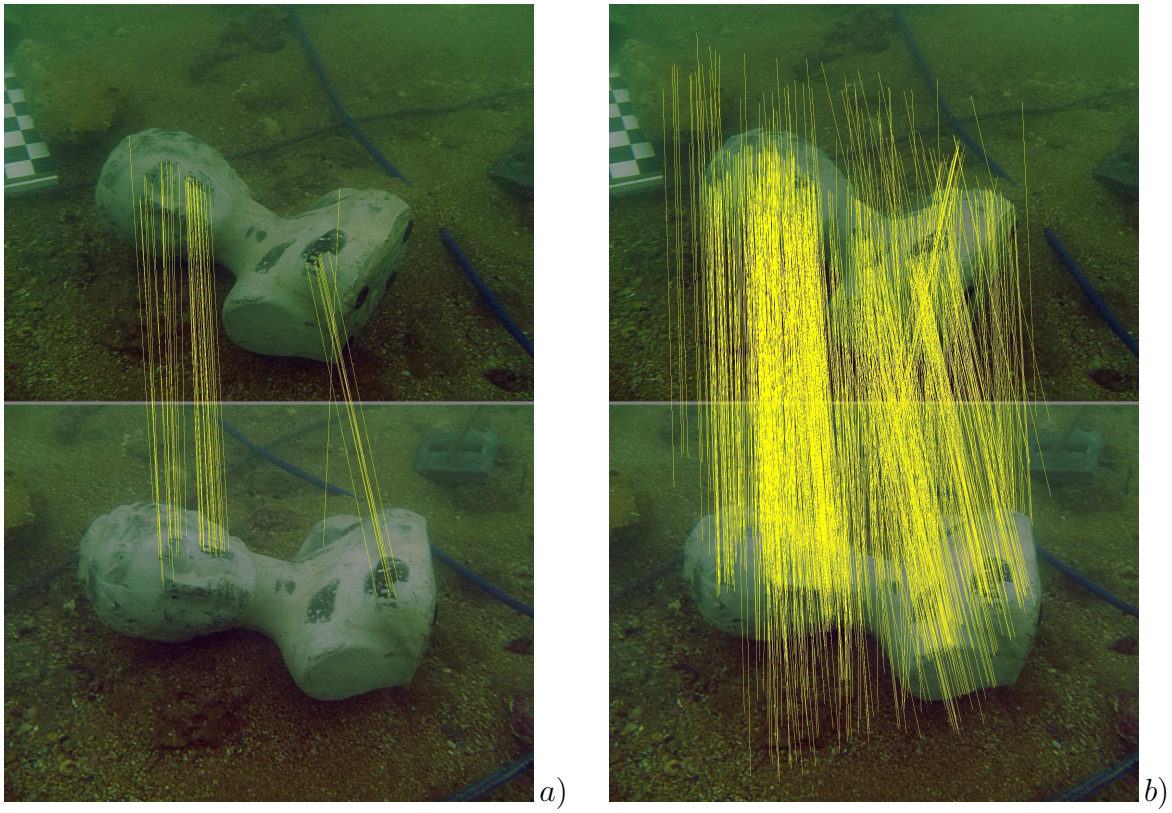


FIGURE 5.20: Restauration des images (buste). Correction colorimétrique et augmentation du contraste. a) et b) images droite et gauche originales. c) et d) images restaurées. e) et f) Détails du fond : pierre et sable devant la tête de la statue, la déformation du signal image rend difficile une détection répétable de points d'intérêt.

à la figure 5.22 pour la reconstruction 3D finale. Le buste se détache du fond, les détails du visage : nez, bouche, yeux sont parfaitement rendus en 3D.



a) *b)*

FIGURE 5.21: Mise en correspondance (buste). a) Mise en correspondance de 1000 points de Harris sélectionnés, nous obtenons 45 appariements (2 sont faux) ; ce résultat permet d'estimer la matrice fondamentale. b) Mise en correspondance de 50000 points de Harris, nous obtenons 944 appariements, tous les appariements sur le fond sont faux, les appariements sur le buste sont en grande majorité corrects (il reste en environ 2% de faux appariements).

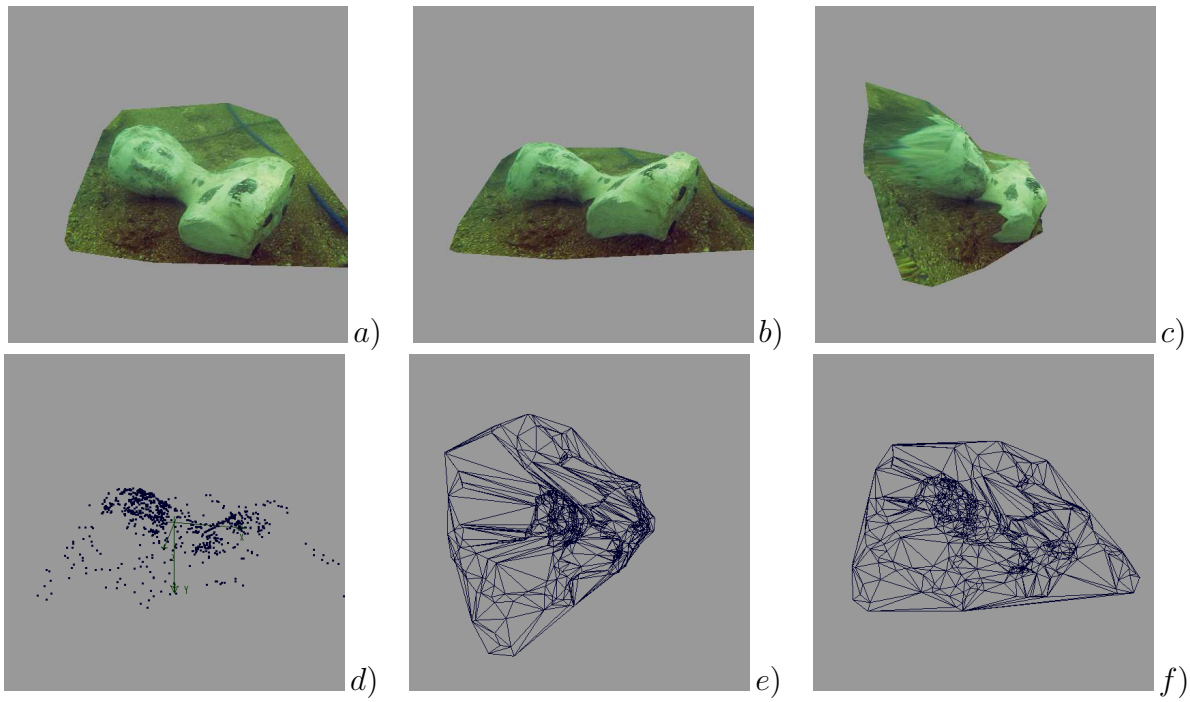


FIGURE 5.22: Reconstruction 3D (buste). a), b) et c) Différentes vues de la reconstruction finale sous plusieurs orientations. d) Points reconstruits, e) et f) Facettes obtenues par triangulation de Delaunay de 4).

5.4.2 Résultats de mise en correspondance dans une séquence vidéo

Dans cette nouvelle expérimentation, nous allons nous intéresser à la stabilisation d'images dans des vidéos, et plus particulièrement à la mise en correspondance de points d'intérêt au cours d'une séquence d'images. Nous nous intéressons ici au mouvement dans le plan image. D'un point de vue de vue de l'analyse du mouvement dans le plan image, de nombreux modèles existent : Par exemple dans [23] les auteurs utilisent un modèle affine associé à une profondeur donnée. Il n'est évidement pas question ici de segmenter finement le champ de mouvement [22] mais simplement d'obtenir une estimation grossière du mouvement global dans le plan image. Nous nous contenterons donc d'un simple modèle de translation global sans aucune considération de mouvement associé à la distance des objets à la caméra. Nous cherchons simplement à évaluer la méthode que nous avons développé dans un contexte de mouvement.

Nous appliquons notre méthode de mise en correspondance sur chaque couple d'images successives d'une séquence vidéo. Nous estimons la translation globale dans le plan image par la position du maximum de l'accumulateur d'appariements (cf. figure 5.15). Cette liste de coordonnées est lissée d'un point de vue temporel, puis intégrée de manière à obtenir la trajectoire dans le plan 2D. Le filtrage utilisé est réalisé à l'aide d'un filtre médian (taille 11) et d'un filtre gaussien d'écart-type $\sigma = 10$.

Nous présentons les résultats obtenus sur une séquence vidéo acquise à l'aide d'un simple appareil photo numérique tenu à la main et se déplacant sur un parking entre des véhicules à l'arrêt (la séquence comporte 255 images). La figure 5.23 présente un échantillonnage temporel de cette vidéo toutes les secondes, soit 11 images, la figure 5.24 présente quant à elle les résultats d'estimation du mouvement dans le plan image.

Les paramètres utilisés dans cette expérimentation sont :

- 300 points de Harris sont sélectionnés parmi 6 buckets (3 buckets en X et 2 buckets en Y).
- $\Delta\theta = 10°$, $\sigma_\xi = 20$, $\sigma_\eta = 2$.

D'un point de vue du temps calcul, la méthode de mise en correspondance prend en moyenne deux secondes par couple d'images sur une machine à processeur Intel cadencé à 2.4Ghz. Ici, aucun développement spécifique à la vidéo n'a été réalisé et un script est utilisé pour lancer les différentes commandes sur la séquence. Un gain substantiel de temps calcul pourrait être réalisé avec une implémentation spécifique.

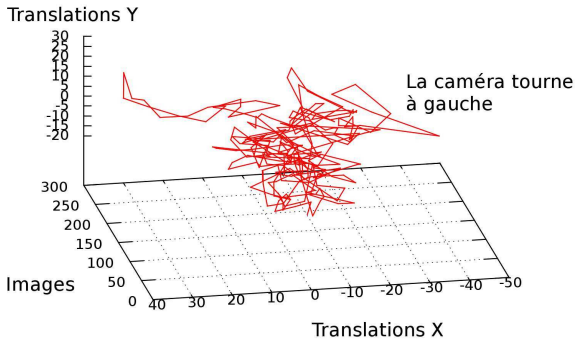
$t = 0s$ $t = 1s$ $t = 2s$

$t = 3s$ $t = 4s$ $t = 5s$

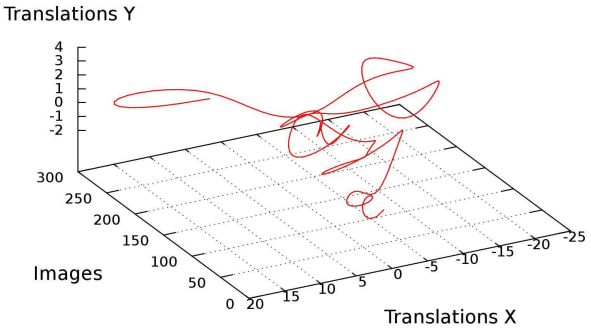
$t = 6s$ $t = 7s$ $t = 8s$

$t = 9s$ $t = 10s$

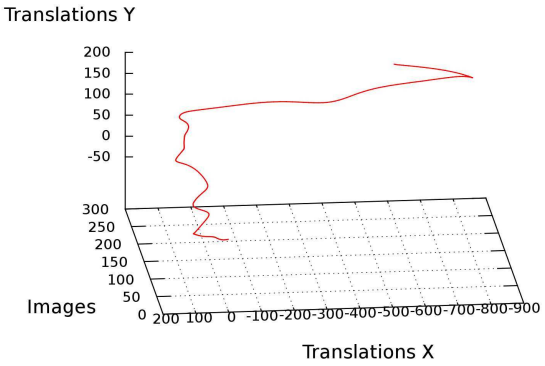
FIGURE 5.23: Images extraites de la séquence vidéo “Parking”.



a)



b)



c)

FIGURE 5.24: Mouvement dans le plan image obtenu sur la séquence vidéo "Parking". a) Mouvement estimé initial. b) Mouvement lissé. c) Intégration du mouvement lissé.

5.4.3 Comparaison avec SIFT

Dans cette section, nous comparons notre méthode avec la méthode de référence : SIFT sur trois couples d'images :

- Les images extraites de la séquence 'Buste', images déja présentées à la figure 5.1.
- Deux images extraites de la séquence 'Graffiti' déja présentée au paragraphe 2.13.
- Enfin, deux images d'un clavier, utilisées par Dro Désiré Sidibé dans [97]. Ce type d'images présentant de nombreuses ambiguités, pose en général d'importants problèmes aux méthodes de mise en correspondance.

5.4.3.1 Séquence 'Buste'

Nous avons comparé les résultats obtenus sur le buste immergé (en niveau de gris : images de la figure 5.1) par SIFT et par notre méthode. SIFT ne trouve aucun correspondant, nous obtenons 30 appariements justes avec comme paramètres $\sigma_\xi = 10$, $\sigma_\eta = 1$, $\Delta_\theta = 5$, $\sigma_{harris} = 1$, $\sigma_{vote} = 20$ avec 300 points de Harris initiaux (figure figure 5.25).

Evidement, le nombre d'appariements (ainsi que le nombre de faux appariements) dépend des seuils de détection et des paramètres utilisés pour la mise en correspondance. sur ce type d'images nous obtenons en général un taux d'erreur inférieur à 5%. Les paramètres que nous utilisons ici sont plutôt des "paramètres standard" qui donnent en général de bons résultats.

5.4.3.2 Séquence 'Graffiti'

SIFT travaille uniquement en niveau de gris, dans un premier temps, nous comparons donc les résultats obtenus par SIFT avec nos résultats en niveau de gris (figure 5.26). Dans un deuxième temps nous montrons l'apport de la couleur dans notre processus de mise en correspondance, cette fois en utilisant les deux images couleur de la séquence (figure 5.27). Pour notre méthode les paramètres utilisés étaient $\sigma_\eta = 1$, $\sigma_\xi = 40$, $\Delta_\theta = 2$, $\sigma_{harris} = 1$, $\sigma_{vote} = 20$ avec 2000 points de Harris initiaux. Nous avons ici un paramètre $\sigma_\xi = 40$ relativement important, il permet d'utiliser un voisinage important dans les descripteurs et la mise en correspondance.

5.4.3.3 Séquence 'Claviers'

Sur cette séquence présentant une ambiguïté importante, les points d'intérêt au niveau des touches du clavier ont tous une description à peu près identique, ce qui rend la mise en correspondance très difficile. Ici encore, nous utilisons un filtre très allongé avec un paramètre σ_ξ grand qui va nous permettre de discriminer les points grâce au voisinage. Nous avons donc comme pour la séquence 'Graffiti' les paramètres : $\sigma_\eta = 1$, $\sigma_\xi = 40$, $\Delta_\theta = 2$, $\sigma_{harris} = 1$, $\sigma_{vote} = 20$, avec 300 points de Harris initiaux. Nous présentons sur

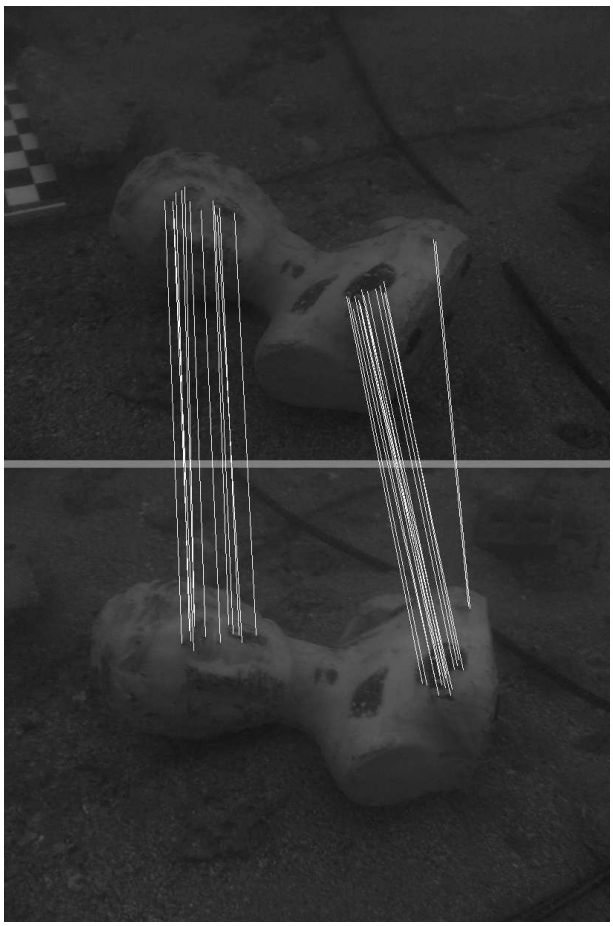


FIGURE 5.25: *[Comparaison : 'Buste'. Séquence 'Buste' : (SIFT ne trouve aucun appariement), notre méthode trouve 30 appariements (tous justes)*

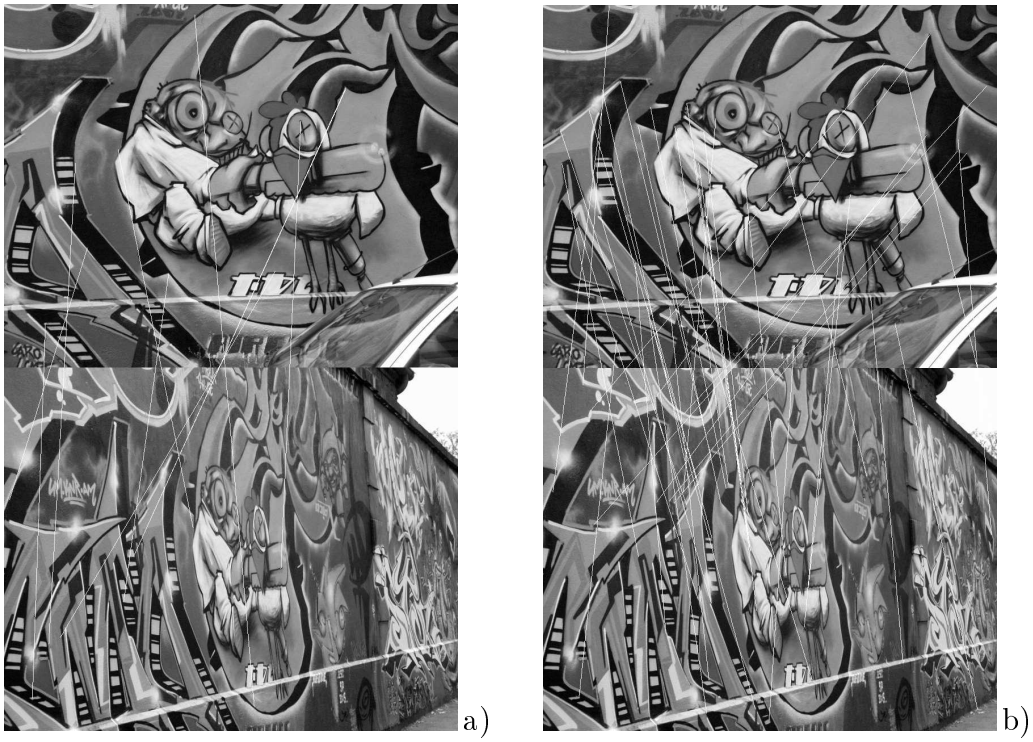

a) b)

FIGURE 5.26: *Comparaison : 'Graffiti'. Séquence 'Graffiti' : image 1 et 5. a) résultat obtenu avec SIFT. b) résultat obtenu par notre méthode (en niveau de gris).*


a)


b)

FIGURE 5.27: *'Graffiti' : mise en correspondance couleur. a) Séquence 'Graffiti' : image 1 et 5. a) Notre méthode (niveau de gris) : sélection des 10 meilleurs appariements). b) Notre méthode (couleur) : sélection des 30 meilleurs appariements.*

la figure 5.28 les résultats obtenus par SIFT et ceux obtenus par notre méthode, ces résultats sont résumés par le tableau suivant :

	appariements	appariements corrects	appariements faux
notre méthode	39	33	6
SIFT	33	14	19

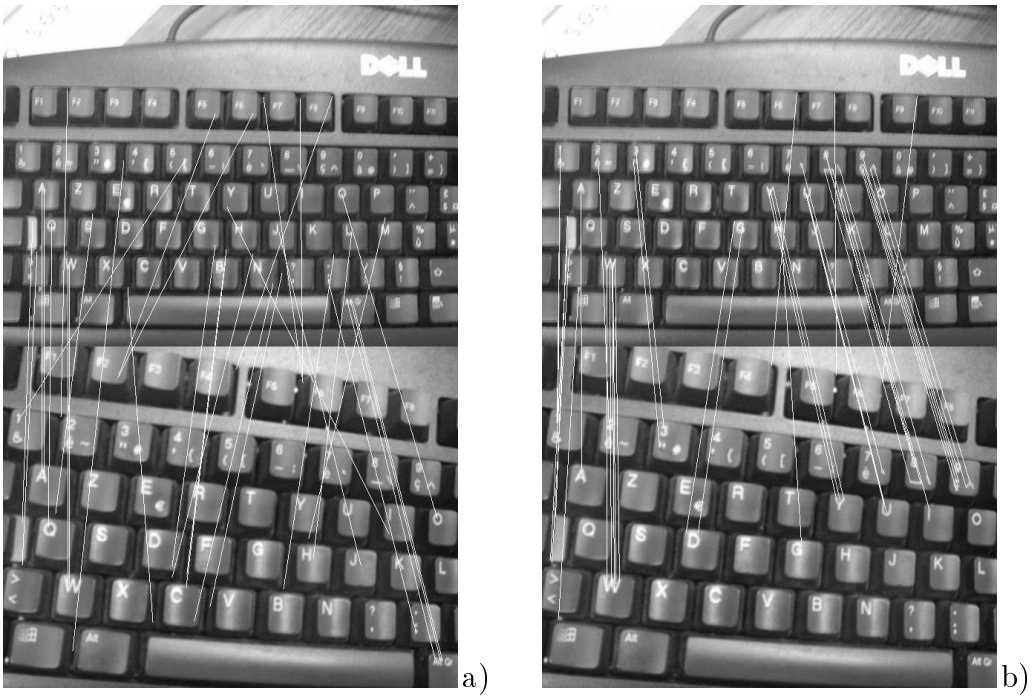


FIGURE 5.28: *a) Séquence 'claviers' : a) résultat obtenu avec SIFT, b) résultat obtenu avec notre méthode.*

5.4.4 Conclusion

Dans tous les exemples que nous avons testés (plus d'une centaine de couples d'images), la méthode décrite dans cette thèse a donné des résultats très intéressants, rivalisant souvent avec la méthode de référence : SIFT. Compte tenu des applications visées nécessitant la mise en correspondance de points précis (points de Harris précis), il n'était pas au départ prévu d'introduire une composante multi-échelle dans la détection et la caractérisation des points.

Les résultats présentés ici, notamment aux paragraphes 5.4.3.2 et 5.4.3.3, montrent que la prise en compte d'un aspect multi-échelle pourrait d'une part améliorer les résultats et d'autre part permettre de s'affranchir des paramètres des filtres gaussiens σ_η, σ_ξ, σ_{harris}.

Evidemment, comme cette méthode n'est pas multi-échelle, la comparaison avec d'autres méthodes comme SIFT est souvent difficile (on ne compare pas des méthodes de même nature). Cependant pour donner tout de même des éléments de comparaison.

Caractéristiques de la méthode SIFT :

- SIFT utilise des points d'intérêt multi-échelle,
- SIFT utilise une description statistique des points d'intérêt par des histogrammes.
- SIFT n'est pas une méthode invariante affine, mais invariante à l'échelle.

Concernant la méthode que nous venons de présenter :

- Nous utilisons un détecteur mono-échelle et précis pour des applications nécessitant de la précision (stéréo, mouvement).
- Nous utilisons une description précise de l'environnement du point, à l'aide de filtres anisotropes robustes vis-à-vis du bruit.
- Notre méthode est relativement robuste face aux transformations affine, mais beaucoup moins face aux changements d'échelle.

En conclusion, nous avons une nouvelle méthode de mise en correspondance, qui donne déjà de bons résultats dans de nombreux cas où l'échelle n'intervient pas, c'est bien sûr le cas des applications que nous avons étudiées. Cette méthode pourrait donc facilement être étendue afin de tenir compte d'un facteur d'échelle, d'une part en utilisant un détecteur multi-échelle et d'autre part en adaptant l'échelle de la description.

Chapitre 6

Conclusion

Dans cette thèse nous avons abordé le problème de la mise en correspondance d'images dans un cadre stéréoscopique à grande base ou de stabilisation de mouvement. Nous nous posons le problème de scènes complexes comme des scènes d'extérieur dans lesquelles les méthodes existantes comme par exemple SIFT sont inefficaces.

Dans les deux premiers chapitres, nous effectuons une revue des problèmes auxquels nous sommes confrontés, et nous introduisons les principales contraintes géométriques utilisées pour les applications visées. Nous effectuons un état de l'art des primitives et des descripteurs couramment utilisés, nous développons principalement les approches par points d'intérêt.

Les deux chapitres suivants sont consacrés à une nouvelle méthode de mise en correspondance. Le chapitre trois présente un nouveau descripteur de points qui s'appuie sur une technique de filtrage anisotrope, extrêmement robuste et fortement discriminante. Ces descripteurs caractérisent des points d'intérêt de type Harris ou Harris couleur mono-échelle. Ils sont capables de scanner l'environnement des points d'intérêt de manière très fine et ils fournissent une signature mono-dimensionnelle de l'environnement d'un point. Le chapitre quatre, quant à lui propose une nouvelle méthode d'appariement, invariante aux transformations euclidiennes et affines. Cette méthode repose sur la corrélation et le warping des signatures qui permet d'obtenir une mesure de distance. Cette distance est ensuite utilisée par un algorithme de mise en correspondance croisée.

Les résultats obtenus notamment sur des images difficiles, comme par exemple des images sous marines bruitées sont très convaincants et laissent envisager des prolongements prometteurs.

Limites de la méthode

Dans le cadre des applications visées au départ, la notion d'échelle ne présente pas une importance capitale, en revanche dès lors que nous nous intéressons à d'autres applications comme par exemple l'indexation d'image ou la reconnaissance d'objets, la méthode

proposée est moins performante que les méthodes classiques. D'une part les points de Harris mono-échelle ne sont plus suffisamment répétables pour obtenir une mise en correspondance correcte et d'autre part les descripteurs ne caractérisent plus la bonne zone d'intérêt.

Une deuxième limitation provient du temps d'exécution qui n'est à l'heure actuelle pas compatible avec la vidéo temps réel. A titre d'exemple,la recherche de correspondants parmi 300 points pris dans une image et 300 points dans une autre image, demande un temps calcul inférieur à la seconde, dans des conditions standard de taille d'image et de puissance calcul de la machine (image de taille $\leq 1000 \times 1000$ et CPU cadencé à 2.4GHz), et ce sans aucune contrainte géométrique de type matrice fondamentale. L'analyse montre que le temps calcul est important au niveau de la méthode de warping, dont la complexité est en $O(N^2)$ par rapport à la taille du descripteur.

Perspectives

Nos venons d'évoquer les principales limitations de notre méthode, nous introduisons à présent des propositions d'amélioration et de généralisation.

Tout d'abord l'extension au cas multi-échelle ne nous paraît pas présenter de difficulté particulière : l'utilisation par exemple d'un détecteur de Harris multi-échelle donnerait une indication sur l'échelle avec laquelle un point d'intérêt est détecté, qui pourrait être utilisée pour paramétrer les noyaux semi-gaussiens de nos descripteurs.

Au niveau du temps calcul, nous avons plusieurs pistes à envisager :

- Améliorer l'algorithme de warping, en limitant le calcul dans la zone de déformation autorisée par les contraintes géométriques, ce qui permettrait d'obtenir une complexité en $O(N)$ par rapport à la taille du descripteur.
- Introduire une mise en correspondance utilisant un arbre de décision, en effectuant le warping uniquement sur des descripteurs potentiellement ressemblants (seuil au niveau de la corrélation des descripteurs).
- Introduire une précision variable en ajoutant une dimension d'échelle au pas angulaire des signatures. Un pas grossier permettrait de repérer très rapidement les appariements possibles.
- Enfin, une implémentation utilisant des processeurs multi-coeurs dans notre laboratoire, nous permet déjà d'obtenir des résultats de détection des points d'intérêts en un temps calcul compatible avec la cadence vidéo. Par ailleurs, il est encore envisageable de paralléliser le code ou certaines parties du code sur un processeur de type GPU, ce qui devrait encore améliorer les performances.

Bibliographie

[1] E. Alvernhe. *Distance au minimum local pour le problème de la stéréo-vision.* PhD thesis, Université de Montpellier II, Sciences et Techniques du Languedoc, Septembre 2006.

[2] N. Ayache and F. Lustman. Trinocular stereo for robotics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1) :73–85, January 1991.

[3] E. Ballet, J. Rousseau, D. Gibon, J. Cussac, C. Vasseur, and X. Marchandise. Mise en correspondance d'images scintigraphiques et échographiques de la thyroïde par une méthode de vision stéréoscopique. *Médecine nucléaire*, 21(7) :409–414, 1997.

[4] H. Bay, T. Tuytelaars, and L. Van Gool. Surf : Speeded up robust features. *Computer Vision–ECCV 2006*, pages 404–417, 2006.

[5] R. Bellman, I. Glicksberg, and O. Gross. The theory of dynamic programming as applied to a smoothing problem. 2(2) :82–88, June 1954.

[6] R. E. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, 1957.

[7] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4) :509–522, 2002.

[8] J. Bigot. *Recalage de signaux et analyse de variance fonctionnelle par ondelettes. Applications au domaine biomédical.* PhD thesis, Université Joseph Fourier, Grenoble, France, 2003.

[9] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 510–517. IEEE, 2005.

[10] H. C. and S. M. A combined corner and edge detector. *4th alvey Vision Conf*, pages 147–151, 1988.

[11] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6) :679–698, 1986.

[12] J. F. Canny. A variational approach to edge detection. In *AAAI-83*, pages 54–58, 1983.

[13] H. Chabbi. *Construction de facettes 3D par stéréovision intégrant des principes de géométrie projective.* PhD thesis, Institut National Polytechnique de Lorraine, Février 1993.

[14] D. Coquin, P. Bolon, and A. Onea. *Objective metric for colour image comparison*, volume 1, pages 119–122. 2000.

[15] R. Deriche. Fast algorithms for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(12) :78–88, January 1990.

[16] R. Deriche, Dr. and O. Faugeras. Les EDP en traitement des images et vision par ordinateur. Technical Report RR-2697, INRIA, Nov. 1995.

[17] E. Derner, T. Svoboda, and K. Zimmermann. Random ferns for keypoint recognition, image matching and tracking – implementation and experiments. Research Report CTU–CMP–2010–16, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, October 2010.

[18] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33(1) :116–125, 1986.

[19] M. Donoser and H. Bischof. Efficient maximally stable extremal region (mser) tracking. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 553–560. Ieee, 2006.

[20] R. Duda and P. Hart. Use of the hough transformation to detect lines and curves in pictures. Technical Report 36, AI Center, SRI International, 333 Ravenswood Ave, Menlo Park, CA 94025, Apr 1971. SRI Project 8259 Comm. ACM, Vol 15, No. 1.

[21] R. Duda and P. Hart. Use of the hough transform to detect lines and curves in pictures. *Commun. ACM*, 15(1) :11–15, January 1972.

[22] R. Dupont. *Suivi des Parties Cachées dans une Séquence Vidéo et Autres Problèmes Soulevés par la Reconstruction Tridimensionnelle d'un Environnement Urbain.* PhD thesis, Ecole Nationale des Ponts et Chaussées., 2006.

[23] R. Dupont, N. Paragios, R. Keriven, and P. Fuchs. *Extraction de Couches de Même Mouvement Via des Techniques Combinatoires.* 2006.

[24] O. D. Faugeras and M. Berthod. Improving consistency and reducing ambiguity in stochastic labeling : An optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(4) :412–424, July 1981.

[25] G. Finlayson. *Coefficient Color Constancy.* PhD thesis, PhD thesis, School of Computer Science, Simon Fraser University,Vancouver, Canada, April 1995.

[26] G. Finlayson, M. Drew, and B. Funt. Color constancy : generalized diagonal transforms suffice. *JOSA A*, 11(11) :3011–3019, 1994.

[27] D. Forsyth. A novel approach to colour constancy. *Proc. 2'nd International Conference on Computer Vision*, pages 9–18, 1988.

[28] W. Freeman, E. Adelson, M. I. of Technology. Media Laboratory. Vision, and M. Group. The design and use of steerable filters. *IEEE Transactions on Pattern analysis and machine intelligence*, 13(9) :891–906, 1991.

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

[30] J. M. Geusebroek, A. W. M. Smeulders, and J. van de Weijer. Fast anisotropic gauss filtering. *IEEE Transactions on Image Processing*, 12(8) :938–943, 2003.

[31] J. Gibson. The perception of the visual world. 1950.

[32] V. Gouet. *Mise en Correspondance d'Images en Couleur, Application à la Synthèse de Vues Intermédiaires.* PhD thesis, Université de Montpellier II, Sciences et Techniques du Languedoc, Octobre 2000.

[33] V. Gouet, P. Montesinos, and D. Pelé. A fast matching method for color uncalibrated images using differential invariants. In *British Machine Vision Conference, Southampton, UK*, 1998.

[34] H. Greenspan, J. Goldberger, and L. Ridel. A continuous probabilistic framework for image matching, 2001.

[35] P. Gros, G. Mclean, R. Delon, R. Mohr, C. Schmid, and G. Mistler. Utilisation de la couleur pour l'appariement et l'indexation d'images. Technical Report 3269, INRIA, 1997.

[36] H. Hirschmüller, P. R. Innocent, and J. Garibaldi. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, 47 :229–246, 2002. 10.1023/A :1014554110407.

[37] B. Horn and B. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3) :185–203, 1981.

[38] P. Hough. Method and means for recognizing complex patterns. 1962.

[39] R. Hummel and S. Zucker. On the foundations of relaxation labeling processes. 5(3) :267–287, May 1983.

[40] M. Jacob and M. Unser. Design of steerable filters for feature detection using Canny-like criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8) :1007–1019, August 2004.

[41] S. G. Johnson and M. Frigo. Implementing FFTs in practice. In C. S. Burrus, editor, *Fast Fourier Transforms*, chapter 11. Connexions, Rice University, Houston TX, September 2008.

[42] T. Kadir and M. Brady. Scale saliency : A novel approach to salient feature and scale selection. In *Visual Information Engineering, 2003. VIE 2003. International Conference on*, pages 25–28. IET, 2003.

[43] T. Kanade, P. Rander, S. Vedula, and H. Saito. Virtualized reality : Digitizing a 3d time-varying event as is and in real time. In H. T. Yuichi Ohta, editor, *Mixed Reality, Merging Real and Virtual Worlds*, pages 41–57. Springer-Verlag, 1999.

[44] Y. Ke and R. Sukthankar. PCA-SIFT : A more distinctive representation for local image descriptors. *Computers, IEEE Transactions on*, 2004.

[45] E. Keogh and M. Pazzani. Scaling up dynamic time warping for datamining applications. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 285–289. ACM, 2000.

[46] E. Keogh and C. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3) :358–386, 2005.

[47] E. J. Keogh and M. J. Pazzani. Derivative dynamic time warping. In *In First SIAM International Conference on Data Mining (SDM 2001*, 2001.

[48] D. Knossow, J. V. D. Weijer, R. Horaud, and R. Ronfard. Articulated-body tracking through anisotropic edge detection. In R. Vidal, A. Heyden, and Y. Ma, editors, *ECCV 2006 Workshop on Dynamical Vision, May, 2006*, volume 4358 of *Lecture Notes in Computer Science*, pages 86–99, Graz, Autriche, Mar. 2007. Springer.

[49] F. Lauze, P. Kornprobst, C. Lenglet, R. Deriche, and M. Nielsen. About some optical flow methods from structure tensors : review and contribution. In *RFIA 2004, Actes du 14e Congres Francophone AFRIF-AFIA*, volume 1, pages 283–292, 2004.

[50] S. Z. Li, J. Kittler, and M. Petrou. Matching and recognition of road networks from aerial images. pages 857–861. European Conference on Computer Vision (ECCV), 1992.

[51] P. Limozin-Long. *Vision stéréoscopique appliquée à la robotique.* PhD thesis, Universite de NICE, Octobre 1986.

[52] P. Limozin-Long. *Vision stéroscopique appliquée à la robotique*, chapter 4- Relaxation, pages 27–60. Octobre 1986. extrait de These de Doctorat, Universite de NICE.

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

[54] D. Lowe. Object recognition from local scale-invariant features. In *iccv*, page 1150. Published by the IEEE Computer Society, 1999.

[55] D. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2) :91–110, 2004.

[56] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2) :91–110, 2004.

[57] B. Lukas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Image Understanding Workshop*, 1981.

[58] Q. Luong. La couleur en vision par ordinateur : Une revue. *Traitement du Signal*, 8(1), 1991.

[59] Q. Luong. *Matrice fondamentale et calibration visuelle sur l'environnement, vers une plus grande autonomie des systèmes robotiques.* PhD thesis, Université de Paris-Sud centre d'Orsay, France, 1992.

[60] B. Magnier. *Détection de Contours et Diffusion Anisotropique dans les Images.* PhD thesis, Université de Montpellier II, 12 Decembre 2011.

[61] B. Magnier, D. Diep, and P. Montesinos. Perceptual crest line extraction. In *2011 IEEE IVMSP, Image, Video, and Multidimensional Signal Processing*, Ithaca, New York, June 16-17 2011.

[62] B. Magnier, P. Montesinos, and D. Diep. Ridge and valley junctions extraction. In *IPCV'11, The 2011 International Conference on Image Processing, Computer Vision, and Pattern Recognition*, Las Vegas, Nevada, USA, July 18-21 2011.

[63] B. Magnier, P. Montesinos, and D. Diep. Texture removal by pixel classification using a rotating filter. In *The 36th International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Prague, Czech Republic, May 22-27 2011.

[64] B. Magnier, P. Montesinos, and D. Diep. Texture removal in color images by anisotropic diffusion. In *VISAPP 2011, International Conference on Computer Vision Theory and Applications*, Algarve, Portugal, March 05-07 2011.

[65] B. Magnier, P. Montesinos, and D. Diep. Fast anisotropic edge detection using gamma correction in color images. In *7th IEEE International Symposium on Image and Signal Processing and Analysis (ISPA 2011), Dubrovnik, Croatia*, pages 212–217, September 4-6, 2011.

[66] T. Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A*, 3(10) :1673–1683, 1986.

[67] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Proc. of British Machine Vision Conference*, pages 384–396, 2002.

[68] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10) :761–767, 2004.

[69] *Gouet V., Montesinos P., Deriche R. and Pelé D.*. Evaluation de détecteurs de points d'intêret pour la couleur. *12ième Congré Francophone AFRIF-AFIA, Reconnaissance des Formes et Intelligense Artificielle*, 2 :257–266, Fèvrier 2000.

[70] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 525–531. IEEE, 2001.

[71] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1) :63–86, 2004.

[72] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, pages 1615–1630, 2005.

[73] R. Mohr, P. Gros, B. Lamiroy, S. Picard, and C. Schmid. Indexation et recherche d'images. In *16Â° Colloque sur le traitement du signal et des images, FRA, 1997.* GRETSI, Groupe Etudes du Traitement du Signal et des Images, 1997.

[74] P. Montesinos. Matching color uncalibrated images using differential invariants. *Image and Vision Computing*, 18(9) :367–376, 2000.

[75] P. Montesinos and S. Dattenny. Sub-pixel accuracy using recursive filtering. *Proceedings of The* 10^{th} *Scandinavian Conference on Image Analysis*, 1(10), 1997.

[76] P. Montesinos and B. Magnier. A new perceptual edge detector in color images. In *Advanced Concepts for Intelligent Vision Systems, ACIVS 2010*, Macquarie University, Sydney, Australia, Dec. 13-16 2010.

[77] H. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. *tech report CMURITR8003 Robotics Institute Carnegie Mellon University doctoral dissertation Stanford University*, 1980.

[78] J. Morel and G. Yu. Asift : A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2) :438–469, 2009.

[79] L. Najman and M. Couprie. Building the component tree in quasi-linear time. *Image Processing, IEEE Transactions on*, 15(11) :3531–3539, 2006.

[80] K. Nasrredine, A. Benzinou, and R. Fablet. Recalage de signaux et d'images : application au décryptage d'archives biologiques marines. *Traitement du signal*, 26(4) :255–268, 2009. 9462 9462.

[81] D. Nistér and H. Stewénius. Linear time maximally stable extremal regions. *Computer Vision–ECCV 2008*, pages 183–196, 2008.

[82] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming, 1985.

[83] M. Özuysal, P. Fua, and V. Lepetit. Fast keypoint recognition in ten lines of code. In *In Proc. IEEE Conference on Computing Vision and Pattern Recognition*, 2007.

[84] P. Perona. Deformable kernels for early vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(5) :488–499, 1995.

[85] P.Montesinos, B. Magnier, and J. L. Palomares. A new perceptually edge detector. In *Proceedings of the third International Workshop on Image Analysis : IWIA 2010*, pages 185–192. Ecole des Mines d'Alès, Presses des Mines, 2010.

[86] F. P. Preparata and M. I. Shamos. *Computational geometry : an introduction.* Springer-Verlag New York, Inc., New York, NY, USA, 1985.

[87] D. R. Recursively implementing the gaussian and its derivatives. In *Proceedings of the 2nd Singapore International Conference on Image Processing (ICIP92). A longer version is INRIA Research Report RR-1893*, pages 263–267, 1992.

[88] A. Rosenfeld. Relaxation methods in image processing and analysis. pages 181–185, 1978.

[89] A. Rosenfeld, R. Hummel, and S. Zucker. Scene labeling by relaxation operations. 6(6) :420–433, June 1976.

[90] Y. Rubner, L. Guibas, and C. Tomasi. The earth mover distance, multi-dimensional scaling, and color-based image retrieval. In *Proceedings of the ARPA Image Understanding Workshop*, pages 661–668, 1997.

[91] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover s distance as a metric for image retrieval. *International Journal of Computer Vision*, 40 :2000, 2000.

[92] B. Schiele and J. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1) :31–50, 2000.

[93] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris. Application à l'indexation d'une base d'objets.* PhD thesis, 1996.

[94] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of computer vision*, 37(2) :151–172, 2000.

[95] S. Seitz and C. Dyer. Toward image-based scene representation using view morphing. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 1, pages 84–89. IEEE, 1996.

[96] D. Sidibe, P. Montesinos, and S. Janaqi. Matching Local Invariant Features with Contextual Information : An Experimental Evaluation. *ELCVIA*, 7(1), 2008.

[97] D. D. Sidibé. *Une technique de relaxation pour la mise en correspondance d'images : Application à la reconnaissance d'objets et au suivi du visage.* PhD thesis, Université de Montpellier II, Decembre 2007.

[98] E. P. Simoncelli and H. Farid. Steerable wedge filters for local orientation analysis. *IEEE Trans. Image Processing*, 5 :1377–1382, 1996.

[99] S. Smith and J. Brady. Susan—a new approach to low level image processing. *International journal of computer vision*, 23(1) :45–78, 1997.

[100] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. In *In CVPR*. Citeseer, 2008.

[101] E. Tola, V. Lepetit, and P. Fua. Daisy : An efficient dense descriptor applied to wide-baseline stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(5) :815–830, 2010.

[102] T. Tuytelaars and L. J. V. Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1) :61–85, 2004.

[103] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors : a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3) :177–280, 2008.

[104] O. Veksler. Stereo correspondence by dynamic programming on a tree. In *CVPR (2)'05*, pages 384–390, 2005.

[105] S. Winder and M. Brown. Learning local image descriptors. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[106] L. Younes. Deformations, warping and object comparison – a tutorial. 2000.

[107] B. Zitova and J. Flusser. Image registration methods : a survey. *Image and vision computing*, 21(11) :977–1000, 2003.

[108] S. W. Zucker, A. Dobbins, and L. Iverson. Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, 1 :68–81, 1989.