Appariement d'images par invariants locaux de niveaux de gris. Application à l'indexation d'une base d'objets

Cordelia Schmid

THÈSE

présentée par

Cordelia SCHMID

pour obtenir le titre de DOCTEUR
de l'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE
(Arrêté ministériel du 30 mars 1992)
Spécialité INFORMATIQUE

APPARIEMENT D'IMAGES PAR INVARIANTS LOCAUX DE NIVEAUX DE GRIS

APPLICATION À L'INDEXATION D'UNE BASE D'OBJETS

Soutenue le 2 juillet 1996 devant la commission d'examen :

Président :	**Jan-Olof EKLUNDH**
Rapporteurs :	**Andrew ZISSERMAN**
	Jean PONCE
Examinateurs :	**Luc VAN GOOL**
	James L. CROWLEY
	Roger MOHR

Thèse préparée au sein du laboratoire GRAVIR - IMAG - INRIA
sous la direction de Roger MOHR

Remerciements

Je tiens tout d'abord à remercier Andrew Zisserman et Jean Ponce qui ont accepté de juger ce travail et d'en rédiger les rapports. Je remercie également Jan Olof Eklundh, Luc Van Gool et James L. Crowley pour l'intérêt qu'ils portent à ce travail en acceptant d'en être les examinateurs.

Je suis tout particulièrement reconnaissante à Roger Mohr de m'avoir accueillie dans son équipe. Toujours porteur de nouvelles idées, il a su donner à cette thèse les bonnes orientations. Je voudrais également souligner son encouragement lors de moments difficiles.

Je n'oublie pas les membres de l'équipe MOVI pour la bonne ambiance qui a permis un travail agréable. Je tiens à remercier tout particulièrement Jerôme Blanc comme relecteur patient de ce rapport. Parmi les stagiaires que j'ai encadrés, je souhaite nommer Marianne Hardt pour les discussions et pour le tableau de Sanja ainsi que Christian Bauckhage pour avoir effectué un travail remarquable et pris puis repris de nombreuses séquences d'images.

Cette thèse a débuté au LIFIA, j'ai donc une pensée pour les personnes que j'ai pu y rencontrer et notamment les membres de l'équipe PRIMA. Elle s'est terminée dans les locaux de l'INRIA que je remercie pour son support et son ambiance.

Merci à la communauté Européene pour avoir financé deux ans de ce travail dans le cadre du programme Capital Humain et Mobilité. Merci également à ISTAR pour avoir mis à ma disposition les images aériennes utilisées dans cette thèse.

Enfin, merci beaucoup à Philippe pour m'avoir supporté pendant ces six derniers mois et pour nos nombreuses discussions.

Table des matières

Chapitre 1

Introduction

Les travaux présentés dans ce rapport s'inscrivent dans le domaine de l'appariement, encore appelé mise en correspondance. Il s'agit d'un domaine fondamental et très vaste de la vision par ordinateur. Il recouvre des problèmes très variés allant de celui de l'appariement entre deux images à celui de la mise en correspondance d'une image avec un modèle CAO défini par des primitives géométriques. L'approche proposée dans cette thèse apporte une solution générique aux problèmes liés à l'appariement. Dans ce chapitre, nous présentons d'abord le contexte dans lequel nous nous plaçons. Ensuite, l'approche proposée dans cette thèse est expliquée et sa position par rapport aux méthodes existantes est discutée. Les contributions de ce travail puis un plan détaillé de ce document terminent ce chapitre.

1.1 Contexte

Les techniques utilisées pour résoudre les problèmes d'appariement sont très différentes. En effet, dans les approches existantes d'appariement entre une image et un modèle CAO, une recherche de ressemblance est effectuée entre quelques dizaines de primitives géométriques tridimensionnelles (segments de droites, ellipses, etc.) définies pour le modèle et des primitives extraites des images. En revanche, dans le cas de la recherche d'une image dans une base d'images, il faut mettre en correspondance plusieurs centaines de milliers de points.

Trouver une solution générale au problème de l'appariement a de très nombreuses applications comme par exemple :

- savoir quel point d'une image correspond à quel autre point d'une seconde image.
  Ceci est utile dans un contexte d'appariement stéréoscopique et permet de calculer la géométrie épipolaire existant entre ces deux images.
- retrouver une image dans une base d'images.
  La recherche dans une base d'images permet par exemple d'identifier un tableau volé ou de vérifier l'existence d'un copyright. Mais l'application la plus riche - et aussi la plus difficile à réaliser - est la documentation : trouver l'image qui illustre

tel événement politique ou scientifique par exemple. Cette aspect prend une dimension particulière avec les potentialités de consultation qu'offre maintenant le réseau Internet.
- savoir quel objet est contenu dans une scène et localiser ses parties.
  L'identification de l'objet puis sa localisation peut être utilisée pour des tâches d'asservissement visuel ou de navigation en robotique mobile. En créant une représentation d'une région à partir d'images aériennes, cette application permet de localiser la position d'un observateur.
- savoir quelle partie d'une image correspond à un élément d'un modèle CAO.
  Ceci permet par exemple de savoir quelle partie d'une image correspond à l'anse d'une tasse ou au pied d'un dinosaure.

De nombreuses solutions ont été proposées pour résoudre les différents problèmes liés à l'appariement. Elles ont donné lieu à des applications variées. Cependant, elles présentent de fortes limitations : elles ne permettent pas de retrouver un objet dans une grande base d'objets sous des conditions générales ; elles ne permettent pas non plus de mettre en correspondance deux images entre lesquelles il existe une forte rotation ou un changement de taille important. Enfin, les méthodes proposées sont fortement combinatoires et ne parviennent pas à traiter des données volumineuses ou complexes en un temps raisonnable ; elles nécessitent parfois jusqu'à plusieurs heures de calcul pour obtenir un résultat. L'objet de cette thèse a été de proposer une méthode innovante par rapport à ces méthodes et à leurs limitations.

1.2 Approche proposée

Parmi les applications potentielle de l'appariement, nous nous intéressons plus particulièrement dans ce travail à l'appariement entre deux images, à la recherche d'une image dans une base et à la localisation d'un objet ou d'une de ses parties dans une image. Pour ce faire, nous proposons une solution unifiée qui permet de tenir compte des spécificités de chacun de ces problèmes. Notre approche permet en outre de s'affranchir des limites des approches existantes. Elle permet d'obtenir de très bons résultats dans des conditions où les approches classiques ne fonctionnent plus.

Plus particulièrement, nous nous plaçons dans les conditions suivantes : mettre en correspondance des objets qui peuvent apparaître dans des scènes complexes différentes, et cela même s'ils sont partiellement visibles et s'ils sont observés de différents points de vues. La visibilité partielle comprend la présence d'occultations et le fait qu'une partie de l'image est seulement observée, par exemple une portion d'un tableau de maître. En outre, nous avons étendu la solution de la mise en correspondance au problème suivant : retrouver à partir d'une seule image l'image correspondant dans une volumineuse base d'images et ceci dans des délais raisonnables. Enfin, il est également possible de localiser des parties d'un objet dans l'image recherchée.

L'approche que nous détaillons dans la suite modélise les images à partir de ce qui est vu et ne repose sur aucune représentation abstraite. Cette modélisation repose sur une caractérisation particulière de l'image. Cette caractérisation est discriminante du fait qu'elle est basée sur les informations contenues dans le signal de niveaux de gris. De plus elle est locale et applicable dans un contexte d'appariement. Par ce biais, ce travail apporte quelques contributions au problème de la mise en correspondance. En outre, notre

approche est robuste, ce qui permet de traiter les incertitudes inhérentes à tout processus de vision par ordinateur.

Il existe dans la littérature d'autres méthodes basées sur les informations contenues dans le signal (cf. section 5.1). L'avantage de ces méthodes est qu'elles permettent de distinguer des objets de n'importe quelle classe sans faire d'hypothèse initiale. En effet, les approches basées sur des données géométriques ne permettent pas de traiter des objets compliqués (cf. figure 1.1). Toutefois les approches existantes basées sur le signal sont globales et ne sont donc pas robustes aux occultations ni à la présence d'arrières-plans complexes. En outre, elles ne sont invariantes à aucune transformation. Notre approche s'affranchit des limites de ces méthodes.

Notre méthode de mise en correspondance se compose de trois étapes. Celles-ci suivent le schéma classique de la vision par ordinateur : un traitement de bas niveau qui permet de traiter le signal et d'extraire des primitives, un calcul de grandeurs numériques à partir des primitives extraites et ensuite une interprétation des grandeurs obtenues. Dans notre travail, cette dernière étape consiste en l'identification et la localisation d'un objet. Il s'est avéré que chacun des choix lors de ces étapes est important ; c'est la combinaison de l'ensemble qui nous a permis d'obtenir un algorithme robuste. Revenons maintenant sur ces étapes.

Parmi les différentes possibilités de traitement bas niveau existantes, nous avons choisi d'extraire des points d'intérêt. Ils correspondent bien à nos objectifs : localité et richesse de l'information contenue dans le signal en ces points. En outre, les expériences menées par Zhang [Zha 95] et dans notre équipe [Cot 94] ont montré l'intérêt d'utiliser de tels points pour le calcul de la géométrie épipolaire.

En ce qui concerne l'étape suivante de quantification de l'information, plusieurs choix étaient possibles. On aurait par exemple pu choisir d'utiliser des grandeurs géométriques, par exemple des rapports de longueurs entre différents points d'intérêt. Toutefois de telles caractéristiques sont moins significatives que l'information photométrique que nous avons choisi d'utiliser. En effet, les grandeurs géométriques sont issues de primitives symboliques ce qui entraîne inévitablement une perte d'information. Le type d'information que nous avons retenu caractérise un point localement. Cette information est calculée aux points d'intérêt et stockée dans des vecteurs (cf. figure 1.1). Elle permet de caractériser localement le signal observé. Le fait qu'elle soit calculée aux points d'intérêt la rend très significative et particulièrement discriminante. La caractérisation utilisée dans ce travail est basée sur les travaux théoriques de Koenderink [Koe 87].

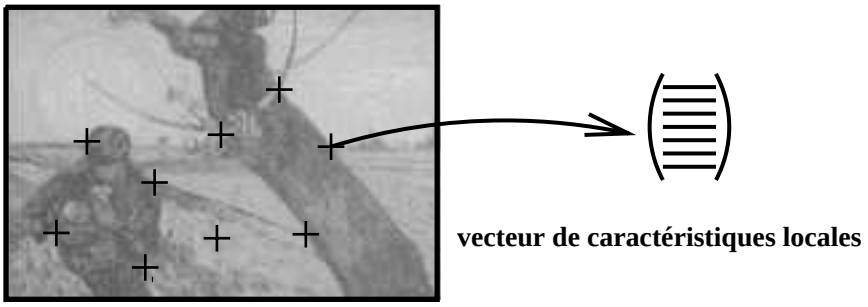

FIG. 1.1 – *Représentation d'une image.*

La troisième et dernière étape de la méthode proposée est la phase d'appariement

proprement dite. Elle consiste à retrouver les vecteurs les plus semblables entre images. L'ajout de contraintes semi-locales permet d'augmenter la robustesse de cette mise en correspondance. Dans le cas d'une mise en correspondance entre deux images, il suffit de rechercher les points les plus semblables. Dans le cas de l'appariement d'une image avec une base d'images, la multiplicité des correspondances ne permet plus d'avoir de réponse satisfaisante ; il faut faire émerger la réponse par une méthode de vote, méthode simple et statistiquement robuste. Enfin, le volume d'informations nécessite le développement d'un outil de recherche rapide par un mécanisme d'indexation.

Étant en mesure de retrouver une image dans une base d'image, il est ensuite possible de modéliser un objet 3D à partir d'une collection d'images. Ces images sont prises de points de vue différents et doivent être représentatives des différents aspects de l'objet. Nous utilisons donc ce qui est perçu pour modéliser un objet 3D. Ceci facilite la reconnaissance d'un objet 3D. D'autre part l'ajout de données symboliques 3D aux différents aspects de l'objet stockés dans la base permet ensuite la localisation de ces données tridimensionnelles dans une nouvelle image.

1.3 Contributions

La contribution principale de cette thèse est d'avoir développé une nouvelle méthode de mise en correspondance. Cette méthode est robuste, rapide et n'est pas restreinte à une classe particulière d'images ou d'objets observés. Les résultats présentés prouvent la robustesse de la méthode face aux transformations d'images importantes, aux occultations et en présence d'arrières-plans complexes.

Le succès de l'approche présentée s'explique d'une part par l'utilisation d'un algorithme statistiquement robuste et d'autre part par les choix effectués à chaque étape de notre algorithme. Par exemple quand nous avons observé que l'instabilité des points d'intérêt influence la stabilité de notre caractérisation, une évaluation de différents détecteurs de points d'intérêt a été effectuée. La répétabilité des points nous a permis de cerner le détecteur qui correspond le mieux aux besoins de notre méthode. D'autre part nous avons montré que les invariants différentiels peuvent être appliqués avec des tailles de fenêtre raisonnables. Ceci permet la mise en œuvre d'une approche multi-échelle. Il a été montré qu'une telle approche est rendue nécessaire par la difficulté d'utilisation des invariants à l'échelle. Pour une telle approche nous avons montré qu'un espacement de 20% entre des échelles consécutives est nécessaire. D'autre part la réalisation d'un algorithme d'indexation a permis une recherche rapide.

Une autre contribution de ce travail est d'avoir proposé une nouvelle méthode de modélisation d'objet 3D qui autorise non seulement l'identification d'objets, mais aussi la localisation d'information tridimensionnelle : nous utilisons les images pour modéliser les objets plutôt qu'une représentation abstraite trop éloignée de la réalité du signal et des performances des algorithmes de vision par ordinateur. Un objet 3D est alors modélisé à partir de plusieurs images. Ensuite on ajoute une information symbolique à chaque image de la base. Le tenseur trilinéaire qui lie les coordonnées des points entre différentes images permet alors de retrouver cette information symbolique dans une nouvelle image. Ceci peut directement servir à des tâches de positionnement d'outils en commande référencée vision.

1.4 Plan du mémoire

Ce rapport présente d'abord les trois étapes de notre approche, puis il présente deux applications de la méthode d'appariement développée : la recherche d'une image dans une base d'image et la modélisation d'objet 3D.

Le deuxième chapitre décrit donc l'étape de bas niveau : l'extraction des points d'intérêt. Nous présentons d'abord l'avantage des points d'intérêt par rapport à d'autres caractéristiques de bas niveau. Ensuite nous comparons différents détecteurs de points d'intérêt. Les critères de comparaison retenus sont la stabilité en présence du bruit de la caméra et la répétabilité en présence de différentes transformations. Cette répétabilité signifie que le point est retrouvé à la même position indépendamment de toute transformation de l'image. Une telle répétabilité influence de façon très importante la stabilité de la caractérisation, l'étape qui suit l'extraction de points.

La caractérisation locale du signal utilisée par la suite est présentée dans le troisième chapitre. Cette caractérisation est basée sur des combinaison de dérivées invariantes aux rotations image. L'utilisation de ces invariants différentiels dans un cadre multi-échelle permet d'obtenir des invariants aux similitudes image. De plus, ces invariants sont des quasi-invariants à une transformation perspective.

Le quatrième chapitre décrit l'étape de mise en correspondance. La méthode d'appariement proposée repose sur un calcul de distance entre deux vecteurs de caractéristiques. L'utilisation de la distance de Mahalanobis permet de tenir compte des incertitudes sur les vecteurs ainsi que de la corrélation éventuelle de leurs composantes. L'ajout de contraintes semi-locales de voisinage augmente la robustesse de la mise en correspondance. Ce chapitre permet d'évaluer la stabilité et l'invariance de la caractérisation retenue au chapitre précédent.

La mise en correspondance entre deux images mène directement à la recherche d'image qui est un problème de mise en correspondance entre l'image recherchée et les images stockées dans la base. Le cinquième chapitre aborde ce problème. La méthode proposée repose sur un algorithme de vote qui permet de gérer la ressemblance entre images de façon robuste. Toutefois un tel algorithme est fortement combinatoire. Nous introduisons donc un mécanisme d'indexation via une table de hachage multi-dimensionnelle. Ceci nous permet de retrouver une image dans une base contenant plus de mille images en moins de cinq secondes avec un taux de reconnaissance supérieur à 99%.

Le sixième chapitre étend la méthode de recherche d'image à des objets tridimensionnels et traite de la modélisation d'un objet 3D à partir de plusieurs images. Le problème est de déterminer le nombre de vues nécessaires pour modéliser un objet 3D. Ayant apporté un élément de réponse à ce problème, nous montrons que la modélisation retenue permet de reconnaître correctement des objets 3D à partir d'une image. Pour pouvoir obtenir une description symbolique de l'objet, des données symboliques sont ajoutées aux images de la base. Ces données peuvent alors être retrouvées pour une nouvelle image en utilisant la contrainte trilinéaire.

La conclusion présentée au chapitre 7 dégage les perspectives ouvertes par ce travail.

Chapitre 2

Détecteurs de points d'intérêt

Dans ce chapitre nous présentons l'étape initiale de notre algorithme d'appariement : l'extraction de points d'intérêt. Le choix des points d'intérêt comme primitives de bas-niveau est d'abord expliqué à la section 2.1. Ensuite un état de l'art des différents détecteurs existants est présenté à la section 2.2. La section 2.3 montre alors comment améliorer la qualité du détecteur de Harris. Afin de choisir un détecteur, il est nécessaire de les comparer. La section 2.4 présente la méthode d'évaluation utilisée dans ce travail. Le critère d'évaluation utilisé pour juger des résultats obtenus est la répétabilité. Une étude comparative pour ce critère est menée à la section 2.5 en présence de différentes transformations. Une approche multi-échelle est ensuite présentée à la section 2.6. Elle rend la détection plus robuste à un changement d'échelle.

2.1 Choix de points d'intérêt

Parmi les différents types de caractéristiques bas-niveau, nous avons choisi d'utiliser les points d'intérêt[1]. Un point d'intérêt correspond à un changement bidimensionnel du signal. Des exemples en sont les coins et les jonctions en T, mais aussi les endroits où la texture varie fortement. Ce choix repose sur le fait que le signal contient plus d'information en ces points qu'en des points correspondant à des changements unidimensionnels du signal (lignes de contours) ou à des régions homogènes.

L'utilité des points d'intérêt a été constatée par Brady [Bra 87] qui a remarqué qu'ils imposent plus de contraintes sur les processus visuels que les contours. Selon lui, ces points fournissent des endroits de calcul fiable. De même, Dreschler et Nagel [Dre 82] ont constaté que le flot optique peut être calculé uniquement aux endroits des points d'intérêt. On peut également citer le travail de Zhang [Zha 95]. Il a montré que l'utilisation de points d'intérêt pour le calcul de la géométrie épipolaire donne de bons résultats. Dans son travail, les points détectés sont appariés par corrélation, donc par une mesure du signal.

D'autre part, les points d'intérêt sont locaux. Leur calcul est effectué sur une fenêtre locale, au moins en ce qui concerne les méthodes basées sur le signal. En présence d'oc-

1. Points d'intérêt et coins sont souvent utilisés de manière équivalente dans la littérature. En fait, point d'intérêt est plus général que coin et ne comporte pas de connotation symbolique.

cultation, de telles méthodes sont donc robustes. Ceci est beaucoup moins vrai pour les algorithmes d'extraction de contours ou de régions, qui ont besoin d'une étape de chaînage ou de fusion, étape qui par expérience reste très fragile.

Les points d'intérêt ont également un caractère général. Leur extraction fonctionne aussi bien pour des objets simples que pour les objets complexes. Un exemple d'objet complexe est le semeur de "Van Gogh" (voir figure 1.1). Pour un tel exemple, l'extraction de contour est pratiquement impossible du fait de la texture contenue dans cette scène.

2.2 État de l'art

Les détecteurs de points d'intérêt peuvent être classés en trois catégories. La première contient les méthodes basées sur les contours, c'est-à-dire à partir de chaînes de contours les endroits avec une courbure maximale ou un point d'inflexion sont recherchés. La deuxième extrait le point d'intérêt directement à partir du signal de niveaux de gris et la dernière approxime les points recherchés avec un modèle théorique.

2.2.1 Méthodes basées sur les contours

Le principe des méthodes basées sur les contours est soit de rechercher les points de courbure maximale le long des chaînes de contour soit d'effectuer une approximation polygonale en vue d'en déduire des points particuliers (intersection, inflexion, ...). De telles méthodes existent depuis longtemps, nous détaillerons dans la suite quelques unes des plus récentes.

Asada et Brady [Asa 86] extraient des points d'intérêt pour des objets 2D à partir de courbes planes. Ils constatent que les courbes planes ont des caractéristiques significatives : les changements de courbure. Ces changements sont classés en plusieurs catégories : coin, terminaison, etc. Pour pouvoir les détecter d'une manière robuste, l'algorithme est intégré dans un cadre multi-échelle. Une approche similaire a été proposée par Mokhtarian et Mackworth [Mok 86]. Au lieu d'utiliser les changements de courbure d'une courbe plane, ils utilisent les points d'inflexion de celle-ci.

Medioni et Yasumoto [Med 87] approximent les contours avec des B-splines. Les points d'intérêt sont des maxima de courbure calculés à partir des coefficients de ces B-splines.

Horaud et al. [Hor 90] recherchent des groupements dans une image de contours pour établir une représentation intermédiaire. Cette représentation repose sur la structuration de segments extraits dans l'image. L'intersection de ces segments donne les points d'intérêt.

2.2.2 Méthodes basées sur le signal

Les méthodes basées sur le signal ne dépendent pas des contours ni d'un modèle théorique du signal. La mesure qui indique s'il y a un point d'intérêt à un endroit donné est calculée directement à partir du signal.

Beaudet [Bea 78] a proposé le premier détecteur de points d'intérêt. Cet opérateur utilise les dérivées deuxièmes du signal pour calculer une mesure "DET" :

$$DET = I_{xx}I_{yy} - I_{xy}^2$$

où $I(x, y)$ représente la surface d'intensité de l'image.

Cette mesure est invariante en rotation et liée à la courbure gaussienne du signal. Les points où cette mesure est maximale sont les points d'intérêt. Pour obtenir effectivement les points d'intérêt, la valeur absolue de cette mesure est seuillée. Il faut noter que cet opérateur détecte les points d'intérêt près des coins mais pas sur les coins, pour autant que la notion de coin existe dans le signal.

Moravec [Mor 79, Mor 81] a proposé un détecteur basé sur la fonction d'auto-corrélation du signal. Cette fonction mesure les différences entre une fenêtre du signal et ses quatre fenêtres voisines. En effet, le voisinage n'est considéré que de manière discrète et dans les directions parallèles aux lignes et colonnes de l'image. Lorsque le minimum de ces quatre différences est supérieur à un seuil, ceci indique la présence d'un point d'intérêt.

Kitchen et Rosenfeld [Kit 82] ont proposé un détecteur de points d'intérêt qui repose sur la courbure de courbes planes. Ils recherchent les maxima de courbure des isophotes du signal. Cependant, un isophote peut présenter une courbure importante du fait du bruit sans que cela corresponde à un point d'intérêt. Cela peut par exemple survenir sur une zone quasi-uniforme, d'autant plus que le calcul fait de la courbure est très approximatif. Kitchen et Rosenfeld proposent donc de multiplier la courbure par la magnitude de gradient de l'image. La mesure K qu'ils utilisent s'écrit de la manière suivante :

$$K = \frac{I_{xx}I_y^2 + I_{yy}I_x^2 - 2I_{xy}I_xI_y}{I_x^2 + I_y^2}$$

La magnitude du gradient est assez diffuse, aussi cet opérateur est très imprécis en localisation. Pour que les points d'intérêt ne soient pas trop épais, les maxima locaux de l'image de magnitude sont extraits avant d'effectuer la multiplication.

Dreschler et Nagel [Dre 82] ont constaté comme défaut à l'approche de Beaudet que la courbure gaussienne peut devenir grande sur des contours marqués, c'est-à-dire sur des contours pour lesquels les deux niveaux de gris définissant ce contour sont très différents. Ceci est dû au fait que la courbure gaussienne est le produit des deux courbures principales d'une surface, et sur un contour marqué une des deux courbures devient très importante. En utilisant un modèle théorique d'un coin, ils constatent qu'autour d'un coin la courbure gaussienne change de signe et qu'elle possède un maximum positif et un minimum négatif. Ils proposent donc de localiser un point d'intérêt sur la ligne joignant ce minimum et ce maximum, notamment à l'endroit où la pente du signal est maximale. A cet endroit la courbure s'annule et change de signe. Par la suite [Nag 83] et [Sha 84] ont montré que les approches de Nagel, Kitchen et Zuniga [Zun 83] sont équivalentes.

Harris [Har 88] a amélioré l'approche de Moravec en calculant une matrice liée à la fonction d'auto-corrélation qui prend en compte les valeurs des dérivées premières du signal sur une fenêtre. Ceci est une amélioration par rapport à Moravec, car la discrétisation utilisée pour calculer la fonction d'auto-corrélation, due au déplacement et aux directions choisies, n'est plus nécessaire. Il obtient donc la matrice suivante :

$$\exp^{-\frac{x^2+y^2}{2\sigma^2}} \otimes \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix}$$

Les valeurs propres de cette matrice sont les courbures principales de la fonction d'auto-corrélation. Si ces deux courbures sont grandes, ceci indique la présence d'un point d'intérêt. L'utilisation des courbures est plus précise que l'utilisation de la valeur minimale comme l'avait proposé Moravec. Toutefois, pour ne pas extraire les valeurs propres, Harris

utilise une mesure reposant sur le déterminant et la trace de la matrice. Cette mesure est supérieure à zéro dans le cas d'un coin. Noble [Nob 88] a montré que l'approche de Harris est optimale uniquement pour des coins en forme de "L". Cottier [Cot 94] a proposé une autre réalisation du détecteur de Harris. Pour améliorer la localisation des points détectés, il applique ce détecteur uniquement sur les contours de l'image et utilise successivement deux tailles de support différentes. Dans [Bau 96] nous avons proposé une amélioration par l'utilisation de dérivées précises, on reviendra sur ce point en section 2.3.

Förstner [För 87, För 94] propose une approche basée sur la statistique locale d'une image. Ceci lui permet d'estimer les paramètres de son algorithme de manière automatique. La première étape de son algorithme est d'estimer la variance du bruit. Il utilise ensuite cette estimation pour restaurer le signal. Puis, les pixels sont classés dans les catégories régions, contours et points d'intérêt. Pour ce faire, il utilise la fonction d'auto-corrélation de la même manière que Harris. Enfin, il classe les points d'intérêt en jonctions ou points isolés. De plus, il effectue une estimation sous-pixellique.

Heitger et Rosenthaler [Hei 92, Ros 92] ont proposé une autre approche inspirée des mécanismes neuro-biologiques. Leur approche consiste à convoluer l'image avec des filtres directionnels pairs et impairs. Ces filtres sont des fonctions sinusoïdales sur une enveloppe gaussienne de moyenne nulle. Ils ressemblent à des filtres de Gabor. Les résultats des filtres pairs et impairs d'une même direction permettent de calculer l'énergie locale de l'image dans cette direction. Cette énergie correspond aux caractéristiques 1D de l'image. Pour obtenir les caractéristiques 2D de l'image, leur approche consiste à calculer pour chaque direction les dérivées premières et deuxièmes de cette énergie. La mesure ainsi obtenue permet de détecter les caractéristiques 2D mais aussi les fausses réponses sur les caractéristiques 1D. Une méthode reposant sur la nature systématique des erreurs permet ensuite d'éliminer les réponses sur les caractéristiques 1D. Les réponses restantes sont seuillées.

Reisfeld et al. [Rei 95] ont proposé un opérateur motivé par des évidences psychophysiques. Cet opérateur est basé sur la notion de symétrie. Ils calculent une carte de symétrie qui contient pour chaque pixel une " magnitude de symétrie " et une orientation. Cette symétrie est calculée localement en regardant la magnitude et la direction des dérivées des points voisins. Cette carte de symétrie peut être appliquée à des tâches diverses, notamment à l'extraction de points d'intérêt. Les endroits avec une symétrie importante sont des points d'intérêt, des lignes de symétrie représentent les axes de symétrie.

2.2.3 Méthodes basées sur un modèle théorique du signal

En ce qui concerne les méthodes basées sur un modèle théorique du signal, le but est d'obtenir une précision sous-pixellique en approximant le signal par un modèle théorique. De telles méthodes ne sont utilisables que pour des types bien précis de points d'intérêt, par exemple des coins. La figure 2.1 montre un modèle théorique pour un coin avec un angle de 90 degrés. Une telle approche est inutilisable dans un contexte général de détection de points d'intérêt. Cercles, lignes etc. peuvent également être modélisés par une telle approche.

Rohr [Roh 90, Roh 92] modélise les jonctions de plusieurs lignes. Pour ce faire, il convolue un modèle binaire de jonction avec une gaussienne afin de modéliser le flou. Dans le cas d'un coin les paramètres du modèle sont l'angle définissant l'orientation de l'axe de symétrie, l'angle définissant l'ouverture du coin, les niveaux de gris, la position du point et

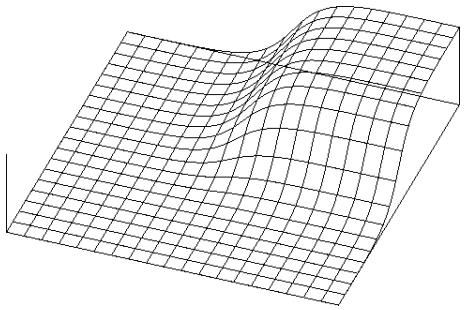

FIG. 2.1 – *Modèle théorique d'un coin.*

le flou. Ces paramètres sont ajustés pour que le signal théorique soit le plus proche possible du signal observé. Cette recherche repose sur une minimisation au sens des moindres carrés. Les coins obtenus par cette méthode sont très précis. Toutefois, la qualité de l'approximation repose sur une bonne estimation initiale de la position. Rohr utilise les segments extraits pour déterminer les paramètres initiaux du modèle : le type de coin, les angles initiaux ainsi que la position du coin.

Deriche et Blaszka [Der 93b] ont proposé une amélioration de la méthode de Rohr au niveau du temps de calcul en remplaçant la fonction gaussienne de lissage par une fonction exponentielle. Deriche propose, en outre, une solution pour l'initialisation des paramètres. Une fois que la position du coin a été donnée de façon interactive par l'utilisateur, l'ouverture du coin est estimée de façon automatique à partir du gradient sur les bords de la fenêtre. Le point donné par l'utilisateur est ensuite amélioré avec une méthode de descente de la variance des niveaux de gris. Deriche et al. ont montré la bonne précision en position des résultats en présence de bruit synthétique. Cette étude a aussi permis de voir que pour assurer la convergence de la minimisation, le support utilisé doit être assez grand. Ceci constitue un défaut de ces méthodes, car lorsque le signal réel comporte plusieurs signaux sur ce support, la convergence est fortement affectée.

Brand et Mohr [Bra 94] proposent une implémentation différente. Le modèle théorique est ajusté au signal en utilisant une transformation affine. L'importance de leurs travaux repose sur l'évaluation de la qualité de la localisation d'une telle approche sur des données réelles. Pour cette évaluation ils mettent en place plusieurs tests : un test d'alignement, une reconstruction 3D et un calcul de la géométrie épipolaire. Ces tests permettent de valider une précision de 0.1 pixel.

Deriche et Giraudon [Der 90, Gir 91, Der 93c] ont présenté une méthode différente pour améliorer la précision de la détection, tout en utilisant un modèle théorique pour un coin. À partir de ce modèle ils analysent d'une manière théorique le comportement de différents détecteurs. Ils constatent que la réponse de nombreux détecteurs varie suivant l'échelle. Toutefois, il existe une relation entre la position de la véritable caractéristique à détecter et les réponses dans l'espace d'échelle. Pour le détecteur de Beaudet, les réponses se situent, pour un coin donné, sur la bissectrice de l'angle définissant ce coin. Ceci, et le fait que le Laplacien du signal s'annule à l'endroit précis du coin indépendamment de l'échelle considérée, leur permet de proposer la méthode suivante : d'abord, un détecteur de Beaudet est appliqué à deux échelles différentes. Ensuite les points correspondant à un même coin sont recherchés dans les deux images. La droite reliant ces deux points définit la bissectrice de ce coin. La position exacte du coin est alors déterminée sur cette droite à

l'endroit le plus proche du point où le Laplacien s'annule.

2.3 Stabilisation du détecteur de Harris

Le calcul des dérivées est mal conditionné dans le sens où il manque de robustesse vis-à-vis du bruit dans les données d'entrée. Même un bruit faible peut perturber ce calcul de façon importante. Pour illustrer ce manque de robustesse, considérons les fonctions $f(x)$ et $\hat{f}(x) = f(x) + \varepsilon \sin(\omega x)$. Elles sont similaires, si ε est petit. Toutefois $f'(x)$ peut différer beaucoup de $\hat{f}'(x)$ pour un ω grand ($\varepsilon << \omega$). En conséquence un bruit de relativement haute fréquence dans une image peut modifier considérablement la première dérivée et donc a fortiori les dérivées d'ordre supérieur.

Préalablement à tout calcul de dérivation, il est donc nécessaire d'effectuer un lissage. Puisque la différenciation commute avec la convolution : $\partial_i(g * f) = g * \partial_i f = \partial_i g * f$, un tel lissage peut être obtenu soit en lissant l'image soit en lissant l'opérateur de dérivation. Un moyen simple pour stabiliser les calculs de dérivation est donc d'utiliser les dérivées d'une fonction de lissage. Le choix le plus communément fait pour une telle fonction est la gaussienne [Wit 83, Tor 86, Rom 94a, Flo 93, Lin 94]. La formule de la gaussienne $G(\vec{x}, \sigma)$ pour un $\vec{x}$ de dimension 2 est :

$$G(\vec{x}, \sigma) = \frac{1}{2\pi\sigma^2} exp(-\frac{\vec{x}^2}{2\sigma^2}) \tag{2.1}$$

La dérivée n-ième de cette fonction par rapport aux variables x_{i_k} $(k = 1 \ldots n)$ est la suivante :

$$G_{i_1 \ldots i_n}(\vec{x}, \sigma) = \frac{\partial^n}{\partial i_1 \ldots \partial i_n} G(\vec{x}, \sigma) \qquad n = 0 \ldots N \tag{2.2}$$

où i_k indique l'axe du système de référence et donc $i_k = 1, 2$ dans le cas d'un système de dimension 2.

Pour la nouvelle version du détecteur de Harris, nous utilisons de telles dérivées. Les convolutions avec les dérivées des gaussiennes intervenant dans le calcul sont implémentées de manière récursive ce qui permet des calculs rapides [Der 93a]. Pour effectuer ces calculs récursifs la gaussienne et ses dérivées sont approximées de façon très précise.

La figure 2.2 compare la version stabilisée avec une implémentation standard qui utilise des dérivées [-1 0 1]. Le critère de comparaison est le taux de répétabilité qui est introduit dans la section suivante. La figure 2.2 montre cette comparaison pour une rotation image (graphe de gauche) et pour un changement d'échelle (graphe de droite). La scène utilisée est "Van Gogh". On peut observer que la version stabilisée donne de meilleurs résultats. Dans le cas d'une rotation les résultats obtenus avec la version standard se détériorent pour un angle de 45 degrés. Ceci est dû au fait que cette implémentation standard du détecteur de Harris utilise des filtres trop discrets pour approximer de façon précise et isotrope les dérivées gaussiennes. La figure A.10 dans l'annexe A montre des résultats similaires pour la scène "Astérix".

2.4 Méthode d'évaluation

Précédemment, nous avons vu qu'il existe beaucoup de travaux sur les détecteurs de points d'intérêt. Toutefois la méthode d'évaluation la plus répandue a été de vérifier

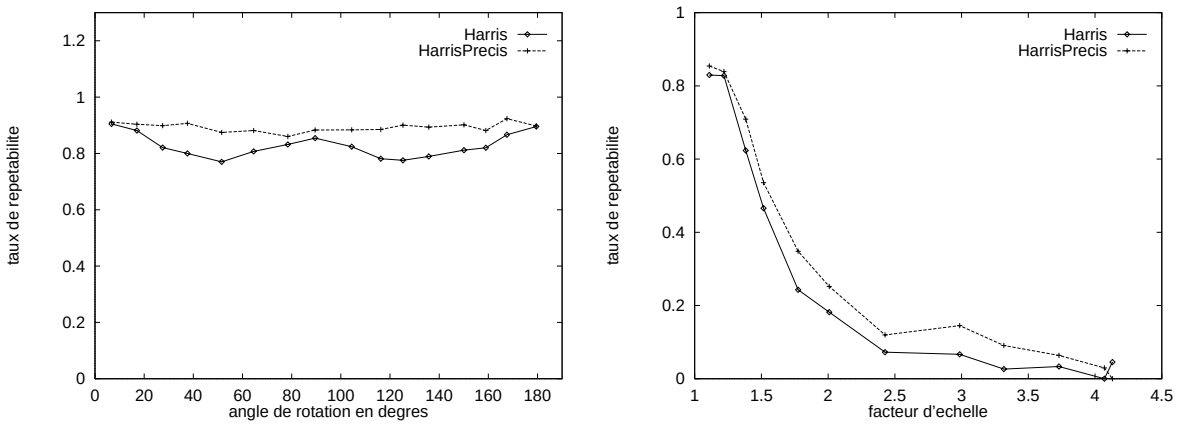

FIG. 2.2 – *Comparaison de Harris et HarrisPrécis. À gauche pour la séquence rotation image et à droite pour la séquence changement d'échelle. La scène utilisée est "Van Gogh" et $\varepsilon = 1.5$.*

visuellement la qualité des résultats. Ceci n'est pas systématique et risque en plus de donner des résultats subjectifs : on compare le résultat avec ce qu'on évalue comme étant un point d'intérêt et non pas avec ce qui est important pour l'étape qui suit l'extraction de ces points. Un exemple pour une évaluation systématique est le travail de Brand et Mohr [Bra 94] mentionné dans la section précédente.

Nous présentons maintenant différents critères d'évaluation et en retenons un particulier : la répétabilité. Une méthode pour mesurer ce critère est ensuite définie et mise en œuvre.

2.4.1 Critères d'évaluation

D'une manière générale un détecteur est caractérisé par sa répétabilité et sa localisation. La répétabilité signifie qu'un même point est détecté dans une suite d'images. La localisation, par contre, signifie qu'un point détecté dans l'image correspond de façon précise à un point 3D de la scène. Plusieurs travaux ont constaté [Wan 92], [Der 93c] qu'il existe une contradiction entre répétabilité et localisation. En effet, un lissage améliore la répétabilité, mais rend plus mauvaise la localisation, comme l'a constaté Canny [Can 86].

Selon l'application pour laquelle sert la détection, les critères d'évaluation diffèrent. Pour des applications qui ne nécessitent pas de connaître la position 3D, comme le calcul de la géométrie épipolaire, l'appariement ou la reconnaissance d'objet, le seul critère important est la répétabilité. Par contre pour la reconstruction 3D ou le calibrage, la localisation et la répétabilité sont toutes les deux nécessaires.

2.4.2 Définition de la répétabilité

Définition 2.1 *Répétabilité*
Soient I_1 et I_i deux images d'une même scène et M_1 et M_i les matrices de projection correspondantes. La détection des points image p_1 et p_i appartenant respectivement à I_1 et I_i est répétable si et seulement si $p_1 = M_1 P$ et $p_i = M_i P$.

La figure 2.3 illustre cette définition. Pour mesurer la répétabilité d'un détecteur, il faut établir une relation entre p_1 et p_i à partir des images. Dans le cas général, il n'existe

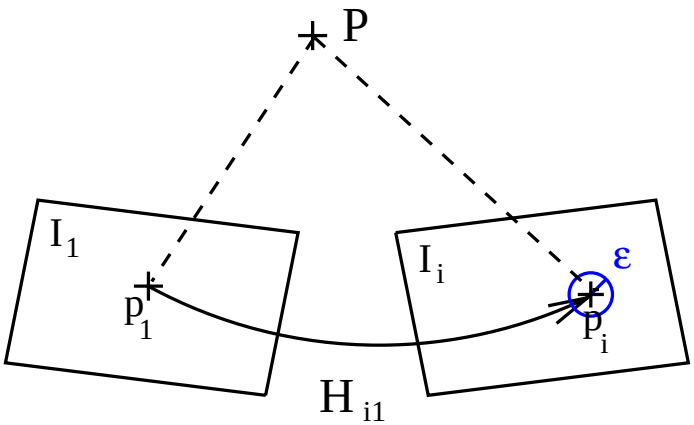

FIG. 2.3 – *Définition de la répétabilité.*

pas de telle relation. Toutefois, si I_1 et I_i sont des images d'une scène plane, cette relation est définie par une homographie :

$$p_i = H_{i1} p_1 \quad \text{où } H_{i1} = M_i M_1^{-1}$$

Dans le cas d'une répétabilité parfaite entre I_1 et I_i, on obtient pour les points détectés sur ces images $\{p\}_1$ et $\{p\}_i$:

$$H_{i1}\{p\}_1 = \{p\}_i$$

Dans des conditions réelles, uniquement un sous-ensemble de points est répété. En outre un point n'est souvent pas répétable de façon exacte, mais il est répétable dans un voisinage. Nous allons dans la suite introduire une mesure de répétabilité qui prend en compte ces deux faits.

2.4.3 Mesure de répétabilité

La mesure de répétabilité compare les ensembles de points $\{p\}_i$ et $\{p\}_1$. Il faut tout d'abord noter qu'un certain nombre de points ne peuvent pas être répétés, car ils correspondent à une partie de la scène qui n'est pas vue dans les deux images. Il faut donc tenir compte uniquement de la partie commune effective des images, c'est-à-dire de la partie de la scène vue dans les deux images. Soient $\{d\}_1$ et $\{d\}_i$ les sous ensembles de points détectés correspondant à la partie commune :

$$\{d\}_1 = \{p_1 \ / \ H_{i1} p_1 \in I_i\} \text{ et } \{d\}_i = \{p_i \ / \ H_{1i} p_i \in I_1\}$$

L'ensemble des points détectés à la même position dans l'image I_i est formellement :

$$D = \{(d_i, d_1) / d_i = H_{i1} d_1\}$$

Le support de l'image étant discret, cette intersection n'a pas de réalité physique. Pratiquement, l'ensemble des points détectés à la même position est déterminé par un seuil de distance ε (on parlera alors d'"ε-répétabilité") :

$$D(\varepsilon) = \{(d_i, d_1) / dist(d_i, H_{i1} d_1) < \varepsilon\}$$

Soient $n_i = |\{d\}_i|$ et $n_1 = |\{d\}_1|$ le nombre de points détectés dans les images respectives. Le taux de répétabilité $r_i(\varepsilon)$ pour l'image i est défini par :

$$r_i(\varepsilon) = \frac{|D(\varepsilon)|}{\min(n_i, n_1)}$$

2.4.4 Cadre d'évaluation

Pour nos expériences nous avons enregistré des séquences d'images correspondant aux différentes transformations à examiner. Pour ces séquences il faut estimer les homographies de façon précise et indépendante des points détectés. En effet, si l'homographie est estimée à partir de points détectés et que ceux-ci présentent un biais, l'homographie va être faussée par ce biais. Ceci favorise le détecteur avec lequel ont été détectés les points.

Nous avons donc besoin d'une détection indépendante et si possible sous-pixellique de points pour le calcul de l'homographie. Pour ce faire, deux images de la scène sont prises pour chaque position de la caméra : une image de la scène originale et une image de la scène sur laquelle sont projetées des cibles noires. La figure 2.4 montre une telle paire d'images pour la scène "Astérix".

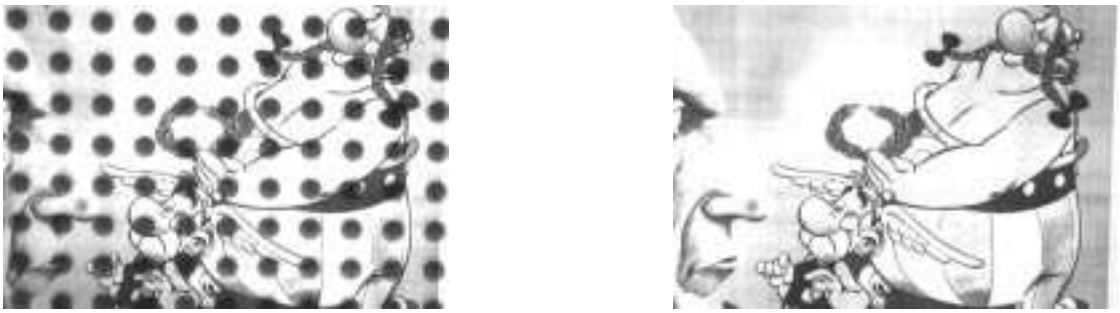

FIG. 2.4 – *Image "Astérix" avec et sans cibles.*

Le processus de projection est illustré par la figure 2.5. Des cibles noires sont projetées sur la scène par un rétroprojecteur. Pendant toute la prise de vue, la scène et le mécanisme de projection des cibles sont fixes. Seule la caméra bouge.

Dans la suite les images avec cibles sont utilisées pour le calcul de l'homographie et la détection des points d'intérêt est effectuée sur les images sans cibles. Pour le calcul de l'homographie on extrait d'abord les centres des cibles d'une manière précise en utilisant l'algorithme de Brand [Bra 95] qui repose sur une approximation du modèle théorique du signal. À partir de ces centres des cibles le calcul de l'homographie est effectué avec une méthode robuste de moindres carrés médians.

2.5 Étude comparative de répétabilité

À la section 2.2 nous avons vu qu'il existe une grande variété de détecteurs. Vu qu'il est impossible de les comparer tous, un sous-ensemble a été choisi. L'évaluation a été effectuée par rapport à différentes transformations, notamment une rotation image, un changement d'échelle, un changement de luminosité et un changement de point de vue. La stabilité au bruit de la caméra a également été testée. L'évaluation est effectuée sur deux

FIG. 2.5 – *Mécanisme de projection des cibles.*

scènes différentes, référencées dans la suite "Astérix" et "Van Gogh" (voir figure 2.4 et figure 2.6). On peut constater la nature différente de ces deux images : l'image "Astérix" contient surtout des contours et l'image "Van Gogh" contient beaucoup de texture. Avant de donner les résultats obtenus pour les différentes transformations, nous présentons un exemple de points détectés et nous spécifions les détecteurs considérés.

2.5.1 Exemples de détections et détecteurs considérés

La figure 2.6 montre les points d'intérêt détectés sur une même scène pour des images prises sous différentes transformations de la caméra. Entre l'image de gauche et l'image du milieu il y a une rotation image de 155 degrés. Et entre l'image de gauche et l'image de droite il y a un changement d'échelle de 1.4. Le détecteur utilisé pour cet exemple est le détecteur de Harris. On peut constater que la répétabilité obtenue est bonne. Même sur les zones texturées les points obtenus sont répétables.

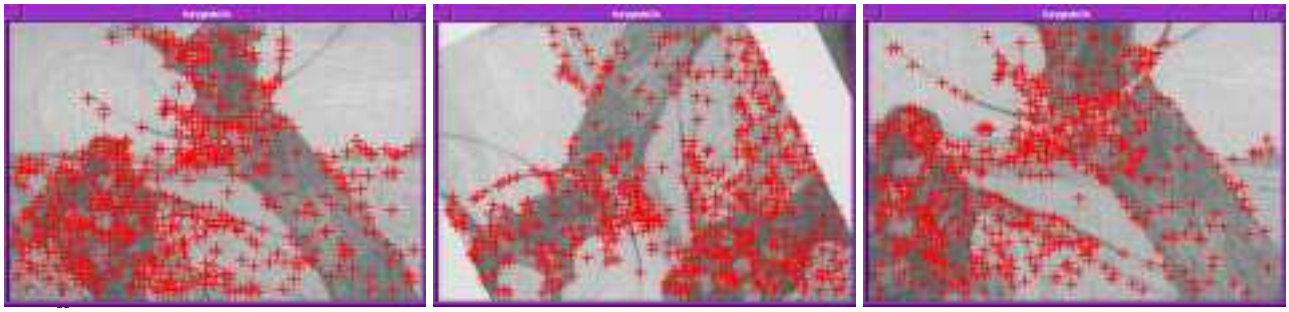

FIG. 2.6 – *Points d'intérêt détectés sur la même scène pour différentes transformations de la caméra. La rotation entre l'image de gauche et l'image du milieu est de 155 degrés. Le facteur d'échelle entre l'image de gauche et l'image de droite est de 1.4.*

Pour notre comparaison nous avons retenu les détecteurs suivants :

- **HarrisPrécis[Bau 96]**
- **Heitger[Hei 92]**
- **Förstner[För 94]**
- **Horaud[Hor 90]**
- **Cottier[Cot 94]**

Dans chaque cas nous avons utilisé le programme développé par les auteurs correspondants. Pour une description de ces détecteurs le lecteur pourra se reporter à la section 2.2 et à la section 2.3. Nous n'avons pas inclut les algorithmes de la dernière catégorie dans notre comparaison. En effet, elle nécessite des modèles typiquement obtenus par l'usage de cibles ce qui est trop restrictif pour le contexte générale dans lequel nous nous plaçons. Les sections suivantes étudient la répétabilité de chacun des détecteurs par rapport aux transformations considérées. L'évaluation pour une rotation image est présentée à la section 2.5.2, pour un changement d'échelle à la section 2.5.3, pour un changement de luminosité à la section 2.5.4 et pour un changement de point de vue à la section 2.5.5. La stabilité au bruit est évaluée à la section 2.5.6.

2.5.2 Rotation image

Pour obtenir une séquence de rotations image, nous avons tourné la caméra approximativement autour de son axe optique. Ceci est rendu possible par le mécanisme particulier de notre objectif qui permet une rotation autour de la bague de montage. La figure 2.6 montre deux images de la séquence rotation image entre lesquelles l'angle de rotation est de 155 degrés. La scène utilisée est "Van Gogh". La figure 2.7 montre les résultats obtenus pour cette séquence. Les angles de rotation varient entre 0 et 180 degrés. Pour le graphe de gauche l'erreur de localisation ε est de 0.5 pixel ce qui correspond à une précision du pixel. Le graphe de droite représente les résultats obtenus pour un erreur de localisation ε de 1.5 pixels. Ceci indique que le point détecté se trouve dans un des pixels voisins du point prédit.

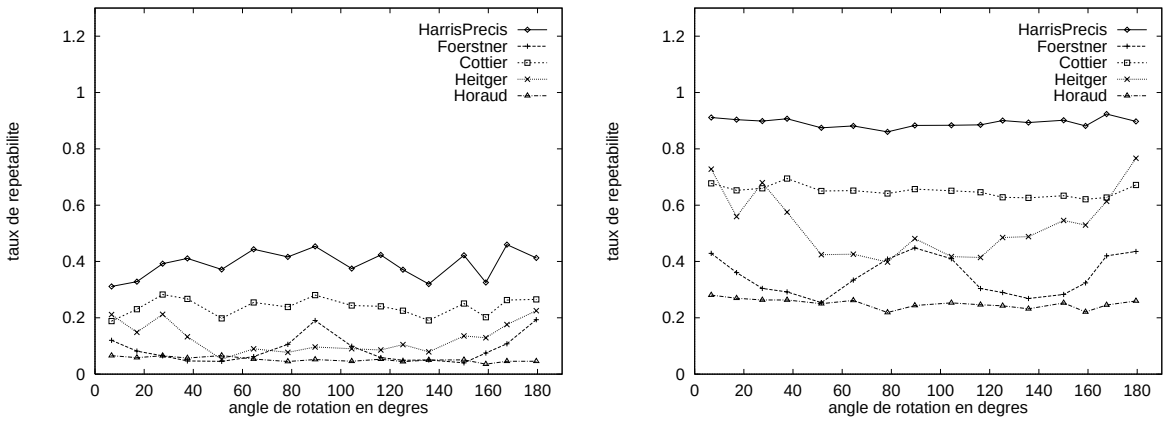

FIG. 2.7 – *Taux de répétabilité pour la séquence rotation image et la scène "Van Gogh". Pour le graphe de gauche ε = 0.5 et pour le graphe de droite ε = 1.5.*

Pour les deux erreurs de localisation le détecteur "HarrisPrécis" donne les meilleurs résultats. Pour une ε de 1.5 il obtient un taux de répétabilité de presque 100%. En ce qui concerne les autres détecteurs, on observe que le fait de calculer Harris uniquement sur les contours (Cottier) dégrade les résultats. Ceci est dû à l'instabilité supplémentaire de l'extraction des contours. On peut également voir que le détecteur de Heitger n'est pas du tout invariant à une rotation. Pour ce détecteur les résultats sont les plus mauvais pour une rotation de 90 degrés ce qui s'explique par l'utilisation des directions choisies pour les filtres directionnels. Le détecteur de Förstner donne de mauvais résultats pour des rotations de 45 degrés ce qui peut s'expliquer par l'utilisation de filtres anisotropes. La méthode basée sur l'intersection des segments (Horaud) donne les plus mauvais résultats. En effet, l'extraction des contours, l'extraction de segments ainsi que l'étape d'intersection des segments introduisent tous des erreurs qui se propagent. La figure A.2 dans l'annexe A montre des résultats similaires pour la scène "Astérix".

La figure 2.8 montre le taux de répétabilité en fonction de l'erreur de localisation ε pour un angle de rotation constant. L'erreur le localisation varie entre 0.5 pixel et 5 pixel. On peut observer que les résultats de tous les détecteurs s'améliorent si l'on tolère de plus grandes erreurs de localisation. Toutefois, le détecteur de HarrisPrécis est supérieur aux autres et augmente plus rapidement. On peut voir que pour ce détecteur on obtient de bons résultats pour une erreur de localisation de 1 pixel. La figure A.3 dans l'annexe A montre des résultats similaires pour la séquence "Astérix".

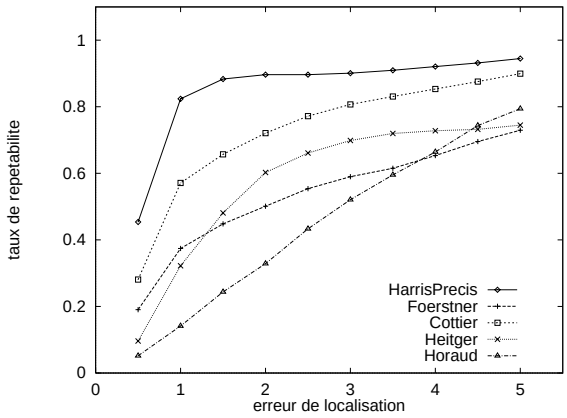

FIG. 2.8 – *Taux de répétabilité pour un angle de rotation de 89 degrés et la scène "Van Gogh".*

2.5.3 Changement d'échelle

Pour obtenir un changement d'échelle nous avons changé la longueur focale du zoom de la caméra. La figure 2.9 montre l'image de référence de la séquence zoom et la dernière image de la séquence. Le changement d'échelle entre ces deux images est de 4.1. Il a été déterminé par le rapport des focales utilisées. La figure 2.10 montre le taux de répétabilité des différents détecteurs pour un changement d'échelle. Le graphe de gauche représente les résultats obtenus pour un ε de 0.5, celui de droite les résultats obtenus pour un ε de 1.5 pixels. La figure A.5 de l'annexe A montre les résultats obtenus pour la séquence

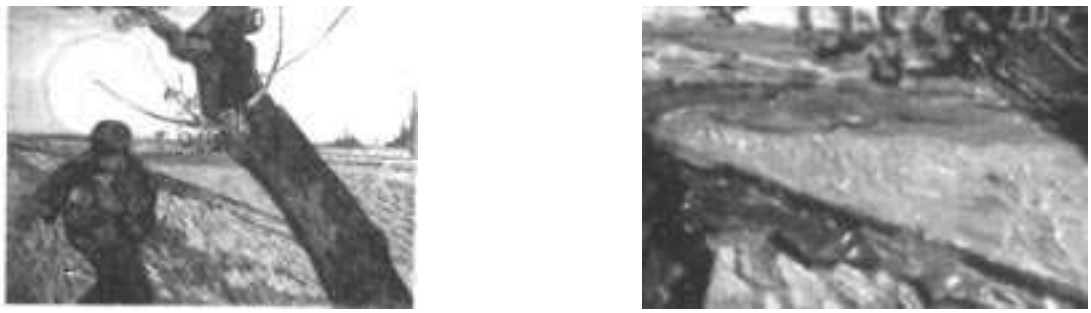

FIG. 2.9 – *À gauche l'image de référence pour la séquence changement d'échelle et à droite la dernière image de cette séquence. Le changement d'échelle entre les deux est de 4.1.*

"Astérix".

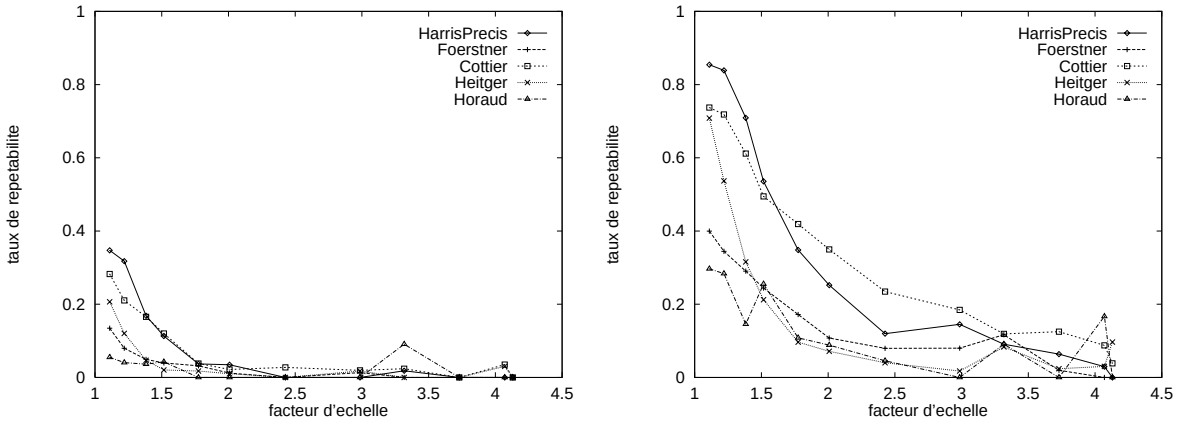

FIG. 2.10 – *Taux de répétabilité pour la séquence changement d'échelle et la scène "Van Gogh". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

La figure 2.10 montre que les détecteurs sont tous très sensibles à un changement d'échelle. Pour un ε de 0.5, la répétabilité est très mauvaise pour un facteur supérieur à 1.5. Pour un ε de 1.5, les résultats deviennent tous très mauvais au-dessus d'un facteur de 2. On peut également observer que le détecteur de HarrisPrécis et le détecteur de Cottier donnent les meilleurs résultats. Les autres donnent des résultats difficilement exploitables. Les résultats obtenus au-dessus d'un facteur d'environ 2.5 sont dus à des artéfacts. A une échelle plus grande, on trouve beaucoup plus de points du fait de la texture de la scène. Des points peuvent donc plus facilement se correspondre par hasard.

La figure 2.11 montre pour un changement d'échelle constant le taux de répétabilité en fonction de l'erreur de localisation ε. On peut observer que les résultats s'améliorent si l'on tolère de plus grandes erreurs de localisation. Toutefois, le taux de répétabilité des détecteurs de HarrisPrécis et de Cottier augmentent plus rapidement que les autres. La figure A.6 dans l'annexe A montre des résultats similaires pour la scène "Astérix".

2.5.4 Changement de luminosité

Dans la suite deux types de changement de luminosité sont examinés : un changement uniforme et un changement complexe. Dans le cas d'un changement uniforme, unique-

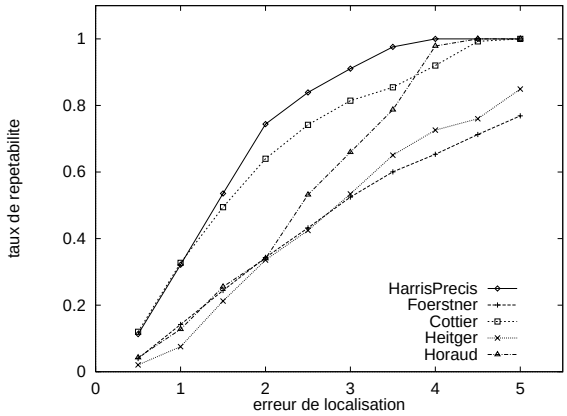

FIG. 2.11 – *Taux de répétabilité pour un changement d'échelle de 1.5 et la scène "Van Gogh".*

ment l'intensité de la luminosité varie et dans le cas d'un changement complexe la source lumineuse est déplacée.

Changement uniforme de luminosité

Un changement uniforme de luminosité est obtenu en changeant l'ouverture de la caméra. Pour mesurer ce changement nous avons introduit la notion de "niveau de gris relatif". Cette mesure est le rapport de la moyenne des niveaux de gris d'une image de la séquence par rapport à une image de référence. L'image de référence est une image au milieu de la séquence, c'est à dire une image qui n'est ni très sombre ni très claire. La figure 2.12 montre deux images de la séquence, une très sombre avec un niveau de gris relatif de 0.6 et une claire avec un niveau de gris relatif de 1.7.

FIG. 2.12 – *À gauche l'image de la séquence changement uniforme de luminosité avec un niveau de gris relatif de 0.6 et à droite l'image avec un niveau de gris relatif de 1.7.*

La figure 2.13 montre les résultats obtenus pour un changement uniforme de luminosité. Le graphe de gauche montre les résultats obtenus pour un ε de 0.5, le graphe de droite représente les résultats obtenus pour un ε de 1.5 pixels. Pour ces deux graphes, le taux de répétabilité diminue régulièrement en fonction du niveau de gris relatif.

Pour l'image de niveau de gris relatif de 1, le taux de répétabilité n'est pas de 100% du fait du bruit dans les images (deux images de niveau de gris relatif de 1 ont été prises pour la séquence : une de référence et une de test).

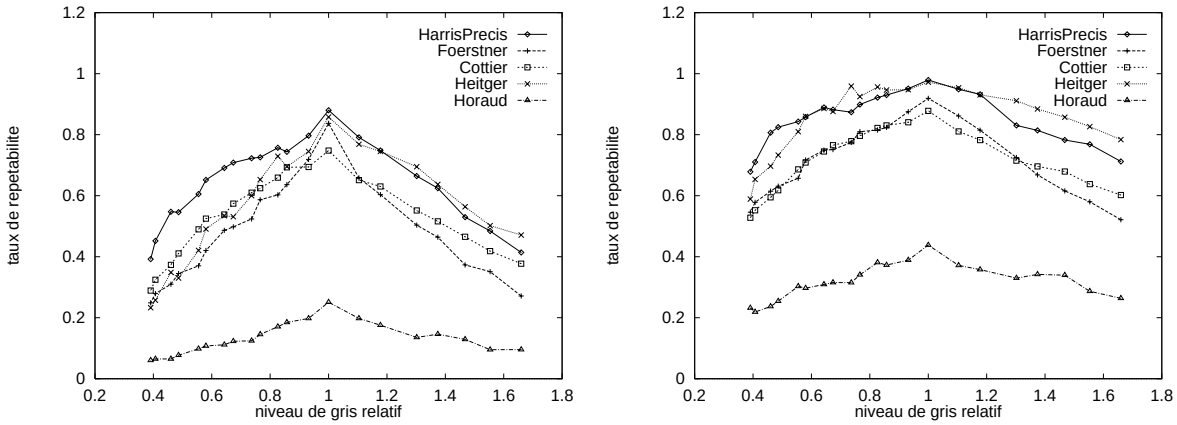

FIG. 2.13 – *Taux de répétabilité pour la séquence changement uniforme de luminosité et la scène "Van Gogh". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

Les détecteurs de HarrisPrécis et de Heitger donnent des résultats supérieurs aux autres détecteurs. Les résultats de ces deux détecteurs sont équivalents. On peut dire que HarrisPrécis est légèrement meilleur, surtout si on considère aussi les résultats obtenus pour la scène "Astérix" qui sont présentés à la figure A.2 de l'annexe A.

Les résultats obtenus pour "Astérix" sont meilleurs que ceux de "Van Gogh" ce qui s'explique surtout par les différents intervalles de luminosité choisis. L'image la plus sombre de la scène "Van Gogh" est beaucoup plus sombre que celle de la scène "Astérix".

Changement complexe de luminosité

Un changement non uniforme de luminosité est obtenu en bougeant la source lumineuse sur un cercle entre approximativement -45 degrés et 45 degrés. La figure 2.14 montre trois images de la séquence. L'image 0 est prise pour la position de la source lumineuse la plus à droite. Cette image sert comme image de référence pour nos tests. Pour l'image 6 l'éclairage vient de face. Une partie de cette image est saturée. En ce qui concerne l'image 11, la source lumineuse est la plus à droite.

FIG. 2.14 – *À gauche l'image 0 de la séquence changement complexe de luminosité, au milieu l'image 6 et à droite l'image 11. L'image 0 est prise pour la position de la source lumineuse la plus à droite. Pour l'image 6 l'éclairage vient de face. Et pour l'image 11, la source lumineuse est la plus à droite.*

La figure 2.15 montre les résultats obtenus pour un changement complexe de luminosité. Le graphe de gauche montre les résultats obtenus pour un ε de 0.5, le graphe de droite représente les résultats obtenus pour un ε de 1.5 pixels.

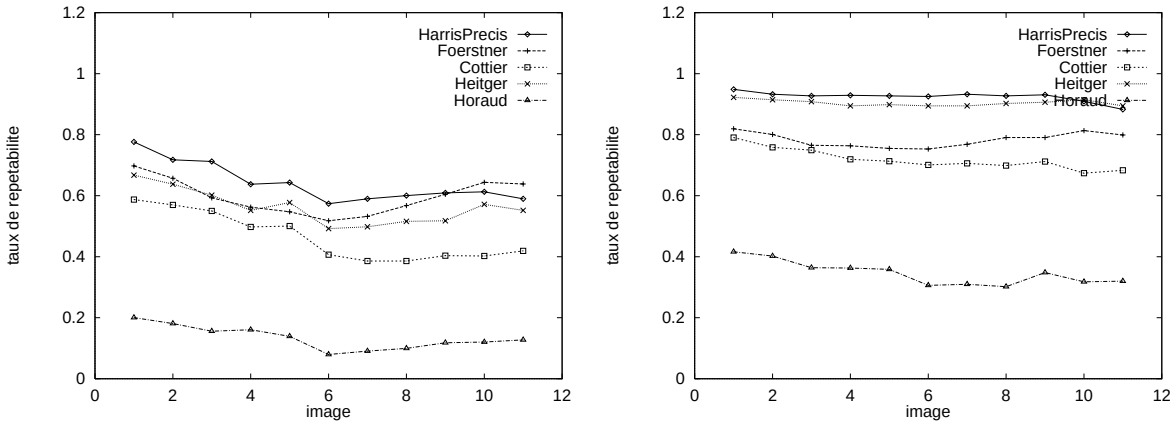

FIG. 2.15 – *Taux de répétabilité pour la séquence changement complexe de luminosité et la séquence "Van Gogh". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

Le détecteur de HarrisPrécis donne des résultats supérieurs aux autres détecteurs. Pour un ε de 0.5, on peut constater une dégradation pour l'image au milieu de la séquence. Ceci provient de la saturation importante de cette image due à un léger reflet. Par contre, pour un ε de 1.5 les résultats sont peu modifiés par un changement complexe de luminosité et globalement constants. Ceci s'explique par la localité des détecteurs évalués.

2.5.5 Changement de point de vue

Pour mesurer la répétabilité des détecteurs par rapport à un changement de point de vue, la caméra a été déplacée autour de la scène sur un cercle entre approximativement -50 et 50 degrés. Les différentes prises de vues sont à peu près équi-réparties. La figure 2.16 montre trois images de la séquence. L'image 0 est prise pour la position de la caméra la plus à droite. Pour l'image 7 la caméra est positionnée en face du tableau. Cette image sert comme image de référence pour nos tests. En ce qui concerne l'image 15, c'est la prise de vue la plus à gauche.

FIG. 2.16 – *À gauche l'image 0 de la séquence pour un changement complexe de luminosité, au milieu l'image 7 et à droite l'image 15. L'image 0 est prise pour la position de la caméra la plus à droite. Pour l'image 7 la caméra est positionnée en face du tableau. En ce qui concerne l'image 15, c'est la prise de vue la plus à gauche.*

La figure 2.17 montre les résultats obtenus pour un changement du point de vue. Le graphe de gauche montre les résultats obtenus pour un ε de 0.5, le graphe de droite représente les résultats obtenus pour un ε de 1.5 pixels.

Le détecteur de HarrisPrécis donne des résultats supérieurs aux autres détecteurs. On

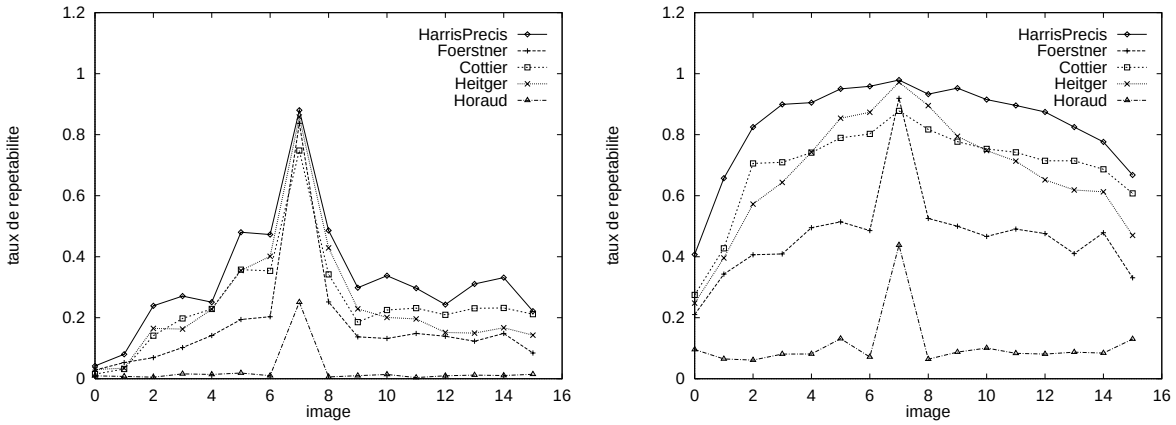

FIG. 2.17 – *Taux de répétabilité pour la séquence changement de point de vue et la scène "Van Gogh". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

peut constater que les résultats se dégradent rapidement pour un ε de 0.5. Par contre, la dégradation est beaucoup moins forte pour un ε de 1.5. Dans ce cas le taux de répétabilité est toujours supérieur à 60% à l'exception de l'image 0. La qualité des résultats obtenue avec cette expérience montre un bon comportement vis à vis de déformations perspectives de l'image.

2.5.6 Bruit de la caméra

Pour observer le comportement du détecteur en présence du bruit de la caméra, une scène statique a été prise plusieurs fois. Les résultats obtenus pour une telle expérience sont montrés sur la figure 2.18.

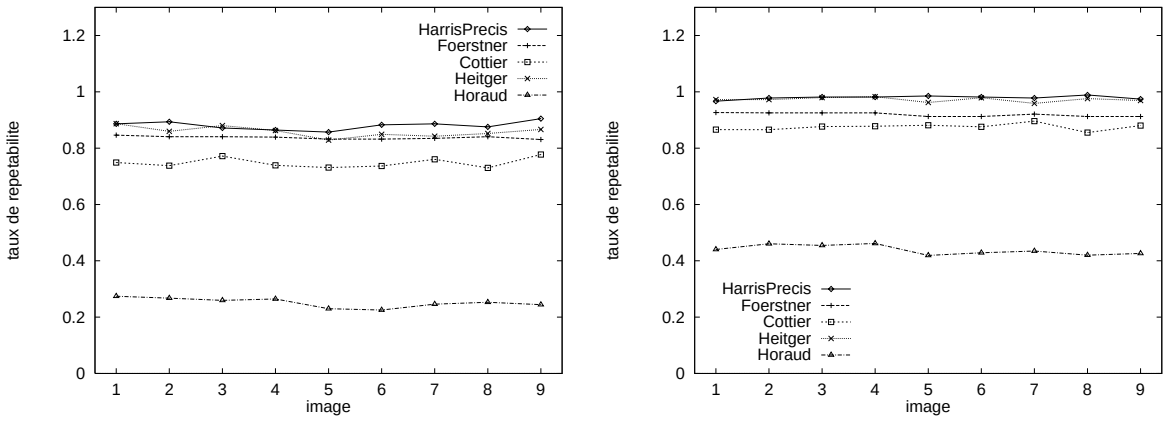

FIG. 2.18 – *Taux de répétabilité pour la séquence bruit de la caméra et la scène "Van Gogh". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

On peut voir que tous les détecteurs donnent de bons résultats à part celui de Horaud. Le détecteur de HarrisPrécis donne les meilleurs résultats. Ces résultats sont légèrement supérieurs à ceux obtenus avec le détecteur de Heitger. Pour un ε de 1.5 ces deux détecteurs obtiennent un taux avoisinant 100%. La figure A.9 dans l'annexe A montre des résultats similaires pour la séquence "Astérix".

2.6 Robustesse à l'échelle - une approche multi-échelle

Notre version du détecteur de Harris utilise des gaussiennes pour calculer les dérivées. Ceci rend plus stable les calculs et permet également de traiter l'image à plusieurs échelles. En effet, la taille σ de la gaussienne peut être adaptée à un changement d'échelle de l'image.

Dans ce qui suit, un changement d'échelle signifie un changement du facteur d'agrandissement d'une image, dénommé par α pour bien le distinguer de la taille σ de la gaussienne utilisée pour le calcul des dérivées. Dans le cas d'une image, nous devons aussi tenir compte du fait que pour le calcul de dérivées nous utilisons un masque gaussien. Il faut donc adapter la taille de ce masque au changement d'échelle. Ainsi, pour deux images I_1 et I_2, où I_2 est changée par un facteur d'échelle α nous avons:

$$\int_{-\infty}^{+\infty} I_1(\vec{x}) G_{i_1 \ldots i_n}(\vec{x}, \sigma) d\vec{x} = \alpha^n \int_{-\infty}^{+\infty} I_2(\vec{u}) G_{i_1 \ldots i_2}(\vec{u}, \sigma\alpha) d\vec{u} \tag{2.3}$$

où les G_{ii} sont les dérivées de la gaussienne comme définies dans l'équation 2.2.

L'équation 2.3 montre que si l'on connaît le changement d'échelle entre 2 images, alors il est possible d'adapter la taille de la gaussienne utilisée pour le calcul des dérivées. Ainsi, il est possible d'obtenir les mêmes points d'intérêt indépendamment du facteur α.

La figure 2.20 montre une telle adaptation pour un changement d'échelle de la séquence "Astérix" (voir figure A.12 dans l'annexe A pour la séquence "Van Gogh"). On peut voir que la courbe "HarrisPrécis adapté" donne des résultats nettement meilleurs. Toutefois au-dessus d'un facteur de 3 les résultats se dégradent. Ceci est dû au fait qu'au-dessus d'un tel facteur les changements du signal ne ressemblent plus aux changements théoriques modélisés.

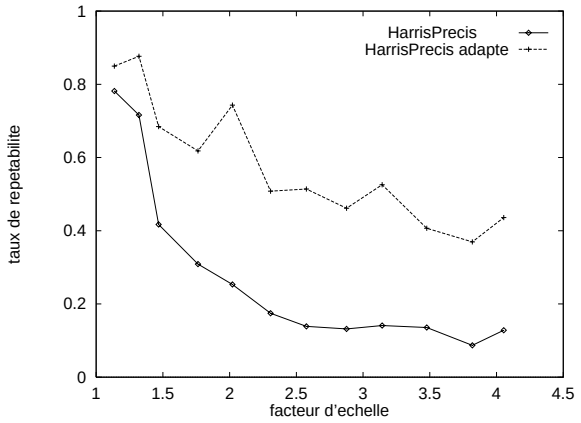

FIG. 2.19 – *Adaptation de HarrisPrécis à un changement d'échelle. La scène utilisée est "Astérix" et ε = 1.5.*

Toutefois, le changement d'échelle est souvent inconnu. À ce moment il est indispensable d'introduire un cadre multi-échelle, c'est à dire d'effectuer les calculs à des échelles pré-définies. Nous avons choisi un espacement de 0.5, c'est à dire les calculs sont effectués aux échelles 1.0, 1.5, 2.0 etc. En théorie il faudrait un espacement exponentiel suivant une suite géométrique. Toutefois vues les différences qui apparaissent dans les signaux avec

des échelles trop différentes, un espacement linéaire est un compromis raisonnable. Les résultats sont donnés sur la figure 2.20 (voir figure A.12 dans l'annexe A pour la séquence "Van Gogh"). Pour chaque détection, l'échelle pour laquelle ont été effectués les calculs est notée. Ces résultats montrent que si nous effectuons des calculs à un tel espacement, une détection robuste à un changement d'échelle est possible.

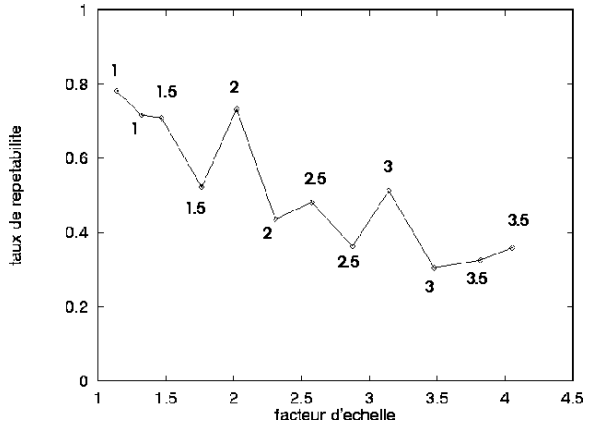

FIG. 2.20 – *Utilisation d'un cadre multi-échelle pour HarrisPrécis. La scène utilisée est "Astérix" et $\varepsilon = 1.5$.*

2.7 Conclusion

Dans ce chapitre nous avons mené une évaluation comparative de différents détecteurs. Le critère de cette évaluation est la répétabilité de la détection lorsque l'image est prise dans des conditions différentes : une rotation image, un changement d'échelle, un changement de luminosité, un changement du point de vue et le bruit du système de prise de vue. Dans tous ces cas le détecteur de HarrisPrécis donne des résultats meilleurs ou équivalents aux autres détecteurs. Ce chapitre a également montré qu'une implémentation stable des dérivées du signal permet d'améliorer de façon importante les résultats du détecteur de Harris standard. D'autre part, les expérimentations de ce chapitre ont montré que les détecteurs basés sur les contours sont moins stables, car leur performance est dépendante du succès ou de l'échec de l'extraction de contour.

Les résultats obtenus pour les différentes transformations peuvent être résumés comme suit. Dans le cas d'une rotation de l'image le détecteur de HarrisPrécis donne des résultats nettement supérieurs aux autres détecteurs. Ceci est dû au fait qu'il est basé sur des mesures invariantes à une telle transformation. Le détecteur de Heitger qui utilise un calcul dans plusieurs directions résiste moins bien aux rotations image. Ceci est confirmé par les observations de [Per 95] qui a constaté que le calcul pour plusieurs directions est moins robuste à une rotation image.

Dans le cas d'un changement d'échelle, les détecteurs de HarrisPrécis et de Cottier donnent les meilleurs résultats. En outre, ces détecteurs peuvent être facilement adaptés à un changement d'échelle. Nous avons montré qu'une telle adaptation est réalisable et permet d'obtenir de bons résultats jusqu'à un facteur d'échelle de 3.

Dans le cas d'un changement de luminosité et du bruit de la caméra, les détecteurs de HarrisPrécis et de Heitger donnent les meilleurs résultats. Enfin, dans le cas d'un

changement de point de vue le détecteur de HarrisPrécis donne des résultats supérieurs aux autres détecteurs.

En conclusion, L'évaluation comparative montre que le détecteur de HarrisPrécis est le plus répétable. Ce détecteur est donc utilisé par la suite pour notre algorithme d'appariement.

Chapitre 3

Caractérisation locale

Ce chapitre présente la deuxième étape de notre algorithme d'appariement : la caractérisation locale du signal autour d'un point. Dans le chapitre précédent nous avons vu comment détecter des points d'intérêt. Ces points ont été retenus parce qu'à priori ils sont à des endroits où le signal présente un contenu informatif important. Il s'agit maintenant de valider cette hypothèse en capturant cette information. Pour ce faire, il faut décrire localement la fonction signal au voisinage d'un point d'intérêt. La figure 3.1 représente cette fonction $I(x, y)$, encore appelée fonction d'intensité lumineuse.

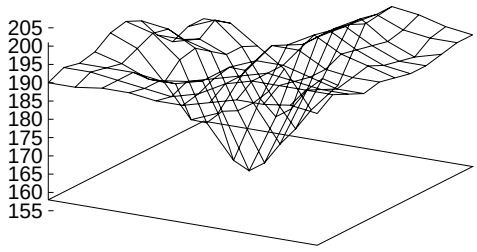

FIG. 3.1 – *Fonction d'intensité lumineuse autour d'un point d'intérêt*

Le but de la caractérisation présentée dans ce chapitre est de décrire cette fonction de la manière la plus précise et la plus complète possible. D'autre part, nous recherchons une description qui soit invariante aux transformations usuelles de l'image. Après avoir décrit à la section 3.1 les méthodes existantes permettant d'obtenir une caractérisation du signal, la section 3.2 présente une brève introduction aux invariants. Il est alors montré à la section 3.3 comment des mesures différentielles - la méthode de caractérisation retenue - peuvent être invariantes aux différentes transformations image considérées. Les expériences menées dans un contexte d'appariement à la section 4.3 du chapitre suivant montrent

l'invariance et la discriminance de cette caractérisation. Ceci est brièvement discuté en section 3.4.

3.1 Méthodes de caractérisation locale

De nombreuses méthodes de caractérisation locale sont possibles. Dans le cas d'images de niveaux de gris des exemples en sont les dérivées, les filtres de Gabor et plus généralement les ondelettes ainsi que les moments. Des caractéristiques basées sur la couleur sont également utilisables. Elles ne sont cependant pas détaillées par la suite, puisque nous nous plaçons dans un contexte d'images de niveaux de gris.

3.1.1 Dérivées

Une fonction peut être approximée localement par ses dérivées. Sachant calculer les dérivées d'une fonction en un point jusqu'à un ordre N, la série de Taylor décrit cette fonction localement jusqu'à cet ordre :

$$f(x_0+x, y_0+y) = f(x_0, y_0) + x\frac{\partial}{\partial x}f(x_0, y_0) + y\frac{\partial}{\partial y}f(x_0, y_0)\ldots + \sum_{p=1}^{N} x^p y^{N-p}\frac{\partial^N}{\partial x^p \partial y^{N-p}}f(x_0, y_0) + O(x^N, y^N)$$

De ce fait il est possible de décrire une image en un point en stockant dans un vecteur l'ensemble des dérivées en ce point. Un tel vecteur a été utilisé par Koenderink [Koe 87] qui l'a nommé jet local. Koenderink calcule en outre le jet local de manière stable en utilisant un filtre passe-bas : la gaussienne et ses dérivées (cf. section 2.6). La définition du jet local est la suivante :

Définition 3.1 *Jet local*
Soit I une image et σ un facteur d'échelle. Le jet local d'ordre N en un point $\vec{x}$, noté $J^N[I](\vec{x}, \sigma)$, est défini par

$$J^N[I](\vec{x}, \sigma) = \{L_{i_1 \ldots i_n}(\vec{x}, \sigma) \mid (\vec{x}, \sigma) \in I \times I\!R^+, n = 0, \ldots, N\}$$

où $L_{i_1 \ldots i_n}(\vec{x}, \sigma)$ est la dérivée n-ième de l'image par rapport aux variables i_k $(k = 1 \ldots n)$ obtenue par la convolution de I avec la différentielle $G_{i_1 \ldots i_n}(\vec{x}, \sigma)$ de la fonction gaussienne :

$$L_{i_1 \ldots i_n}(\vec{x}, \sigma) = G_{i_1 \ldots i_n}(\vec{x}, \sigma) * I(\vec{x})$$

$G(\vec{x}, \sigma)$ et $G_{i_1 \ldots i_n}(\vec{x}, \sigma)$ sont définis par les équations 2.1 et 2.2. L'opérateur $*$ représente l'opérateur de convolution.

Un jet local est calculé au voisinage d'un point et décrit la géométrie locale de ce voisinage. Pour un point donné le jet local est fonction d'un paramètre : la taille σ de la gaussienne. Ce paramètre permet de caractériser une fonction à plusieurs niveaux d'échelle ou il peut être adapté à l'échelle de l'image considérée.

Le jet local est basé sur la dérivation du signal. Toutefois, comme on ne connaît pas la fonction du signal de manière analytique, les dérivées doivent être estimées de façon numérique : elles sont calculées par convolution avec une gaussienne et ses dérivées. On peut donc interpréter le jet local comme la projection - la décomposition - du signal sur une base : la gaussienne et ses dérivées.

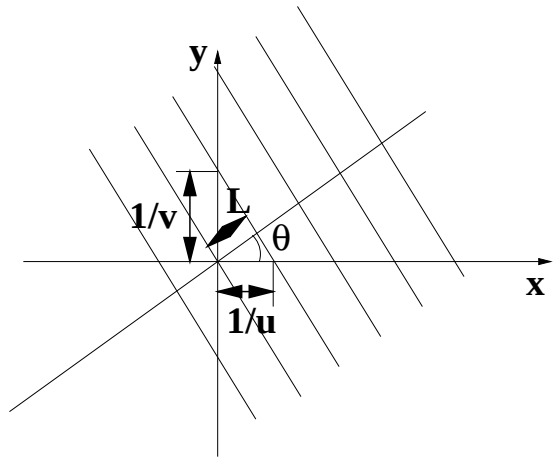

FIG. 3.2 – *La transformée de Fourier est directionelle.*

3.1.2 Descriptions fréquentielles

On peut également caractériser une fonction par une description fréquentielle. Un exemple de description fréquentielle globale est la transformée de Fourier, dont la formulation dans le cas d'un signal bidimensionnel est la suivante :

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

La transformée de Fourier est un cas spécial de décomposition d'une fonction dans une base infinie de fonctions complètes et orthogonales. Cette base de fonctions est constituée des fonctions sinusoïdales donc périodiques et infinies. D'autre part, cette transformée est directionelle : la transformée est calculée dans la direction définie par u et v (cf. figure 3.2). Il existe pour chaque paire de fréquences (u,v) une direction, une magnitude et une phase.

L'intérêt de la transformée de Fourier, ainsi que des autres méthodes de description fréquentielle, réside dans le fait que la phase est " normalisée ", c'est à dire indépendante de la luminosité des images ainsi que de leur contraste. D'autre part, la phase est stable à un changement d'échelle jusqu'à 20%. Du fait de cette stabilité, l'utilisation de cette phase peut être mise en œuvre dans un contexte multi-échelle. Enfin, la phase est une variable continue qui permet d'obtenir des résultats sous-pixelliques, c'est-à-dire de précision supérieure à celle du signal.

Cependant la transformation de Fourier est globale : elle permet une localisation en fréquence et non pas en espace. C'est à dire elle ne permet pas de dire quelles fréquences appartiennent à quel point. Ce problème est connu sous le nom du principe d'incertitude et s'énonce de la manière suivante :

$$\Delta x \, \Delta\omega \geq constante$$

où x est la variable spatiale et ω la variable fréquentielle. Les dispersions Δx et $\Delta\omega$ représentent respectivement l'incertitude spatiale et fréquentielle. Ce principe montre donc que l'on ne peut pas être à la fois précis en espace et en fréquence. En fait, si un filtre est très précis en espace il l'est très peu en fréquence, et réciproquement. Pour remédier à ce problème et minimiser à la fois l'incertitude en espace et en fréquence, il est préférable

d'utiliser un fenêtrage. Gabor [Gab 46] a proposé d'utiliser un fenêtrage gaussien et a démontré qu'un tel fenêtrage est optimal pour obtenir une bonne précision à la fois en fréquence et en espace. La transformation de Gabor est donc la convolution du signal par un filtre dont l'expression est la suivante :

$$\mathcal{G}_{\sigma\omega_x\omega_y}(x,y) = e^{i(\omega_x x+\omega_y y)}\frac{1}{\sigma 2\pi}e^{-\frac{x^2+y^2}{2\sigma^2}}$$

La transformée de Gabor permet d'adapter l'analyse fréquentielle à la rapidité des changements de l'image, donc aux fréquences de l'image. Cependant, il faut pour ce faire déterminer la taille de la fenêtre à utiliser. Celle-ci correspond à la résolution que l'on veut obtenir. Ce choix fixe complètement la dispersion en fréquence de la description obtenue. Il est donc préférable d'effectuer une décomposition multi-échelle afin d'obtenir une description riche du signal.

Le filtre de Gabor est souvent utilisé pour les problèmes de stéréo-correspondance ou pour réaliser la vergence d'une tête stéréoscopique [Wes 92, Fle 91, San 88]. Ce filtre permet d'obtenir une information locale de la phase du signal et sert ainsi à estimer la disparité entre deux images. Dans le cadre du calcul d'appariements, une égalité de phase entre deux points de deux images différentes signifie une grande probabilité qu'il s'agisse de points à apparier, modulo le fait qu'une même valeur de phase peut apparaître plusieurs fois. Ainsi, pour qu'une égalité de phase corresponde à un appariement de façon certaine, une approche multi-résolution est nécessaire. Toutefois cette mesure est locale en fréquence et directionelle en espace. Aussi, une telle approche résiste uniquement à des rotations et des changements d'échelle faibles. Pour résoudre ce problème, Wu [Wu 95] a proposé une implémentation qui utilise des filtres de Gabor dans plusieurs directions et à plusieurs échelles.

Morlet , Grossmann et Meyer [Mor 83, Mey 91] ont construit une théorie reposant sur l'idée de caractériser un signal par différentes échelles et différentes résolutions : cette théorie est à l'origine des ondelettes. L'idée principale des ondelettes est que l'apparition de hautes fréquences est de faible durée en espace. Cette idée se justifie par l'hypothèse que les hautes fréquences correspondent à des discontinuités du signal et doivent donc être considérées uniquement de manière locale. Mallat [Mal 89] a étendu ces travaux dans le domaine de la vision par ordinateur au cas des signaux discrets. La transformée en ondelettes de la fonction f est de la forme suivante :

$$CWT_h(\tau,a) = \frac{1}{\sqrt{C_h}}\frac{1}{a}\int f(x)h^*(\frac{x-\tau}{a})dx \tag{3.1}$$

où

- a est un facteur d'échelle
- τ est une translation
- $\frac{1}{\sqrt{C_h}}$ est une constante utilisée pour la normalisation en énergie
- $h(x)$ est une fonction continue quelconque appelée fonction génératrice de l'ondelette

Les ondelettes nécessitent une répartition logarithmique en espace et en fréquence. La figure 3.3 illustre cet espacement. Cette répartition logarithmique est équivalente à avoir

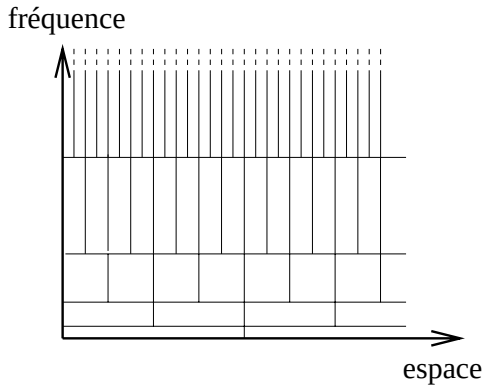

FIG. 3.3 – *L'espacement en espace et en fréquence est logarithmique pour une décomposition en ondelettes.*

des filtres dont la bande passante relative est constante :

$$\frac{\Delta\omega}{\omega} = constante$$

Les résultats obtenus par la transformation de Gabor ainsi que par les ondelettes dépendent de la taille du voisinage sur lequel sont effectués les calculs. Ceci n'est pas le cas pour la transformation de Wigner. Sa formulation est la suivante (cf. [Jac 91]) :

$$W_I(x, y, \omega_x, \omega_y) = \iint R_I(x, y, \alpha, \beta) e^{-i(\alpha\omega_x + \beta\omega_y)} d\alpha d\beta$$

avec

$$R_I(x, y, \alpha, \beta) = I(x + \frac{\alpha}{2}, y + \frac{\beta}{2}) I^*(x - \frac{\alpha}{2}, y - \frac{\beta}{2})$$

Cette transformation permet donc un calcul de la fréquence en tout point. Malheureusement, cette représentation est difficile et lourde à calculer.

3.1.3 Moments

Les moments permettent également de caractériser un signal. Théoriquement, étant donné un signal bidimensionnel $I(x, y)$, le moment d'ordre $(p + q)$ est défini dans le cas discret de la manière suivante :

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

Hu [Hu 62] a utilisé pour la première fois les moments pour des problèmes de reconnaissance en vision par ordinateur. Il a par ailleurs démontré l'existence de combinaisons de moments qui sont invariantes en translation, rotation et changement d'échelle.

Teague [Tea 80] a introduit la notion de moments orthogonaux pour caractériser une image. Pour ce faire, il utilise la théorie des polynômes orthogonaux et introduit la notion de moments de Zernike. Ces moments sont indépendants et peuvent être facilement construits pour un ordre quelconque. D'autres moments orthogonaux sont les moments de Legendre basés sur les polynômes du même nom. Teh [Teh 88] a montré que les moments de Zernike sont les moins sensibles au bruit et les moins redondants en information. Une

comparaison similaire a été menée par Kim [Kim 94] entre les moments de Zernike et ceux de Hu. Les résultats de cette comparaison confirment la supériorité des moments de Zernike pour différencier deux modèles.

Parmi les travaux plus récents, Van Gool [Goo 96] a présenté un ensemble de moments jusqu'au deuxième ordre qui sont à la fois invariants aux changement affines et aux changements d'intensité.

3.1.4 Autres caractérisations

Il existe d'autres méthodes de caractérisation d'un signal que celles présentées précédemment. Par exemple, Bigun [Big 95, Big 94] propose l'utilisation d'un système non linéaire de coordonnées. Son système de base doit satisfaire l'équation de Laplace et il doit être conjugué. Puis il cherche à tourner son système de coordonnées de telle manière que les isophotes dans ce nouveau repère soient les plus parallèles possible. Il utilise ces coordonnées pour caractériser des textures et pour reconnaître des objets.

Weiss [Wei 92] quant à lui propose une méthode pour calculer localement des invariants affines et projectifs. Pour ce faire, il utilise une représentation implicite de la courbe des contours. Pour un point donné il définit un voisinage puis il approxime localement la courbe passant par ce point. Toutefois il est difficile de calculer une représentation implicite à des ordres élevés. Il utilise un système de coordonnées canonique qui est localement défini par les propriétés de la forme. Un exemple simple est l'utilisation de la tangente et de la normale dans le cas des transformations rigides. Dans le cas général il utilise des courbes osculatrices pour obtenir l'invariance aux transformations.

Parmi les caractérisations possibles, il faut aussi citer une méthode simple, mais répandue : un point est décrit par les valeurs des pixels voisins. On stocke donc les niveaux de gris directement dans un vecteur. La comparaison entre des vecteurs de niveaux de gris définit une mesure de ressemblance entre des points. La manière dont est effectuée cette comparaison définit différentes variantes. La plus simple est la "SSD" (Sum of Squared Differences) qui prend la somme des carrés des différences entre les vecteurs. Une mesure plus élaborée est la "ZNCC" (Zero-Mean Cross-Correlation) qui normalise les vecteurs de niveaux de gris par rapport à la moyenne et à la variance avant d'effectuer une corrélation.

3.2 Introduction aux invariants

Afin de rendre la méthode de caractérisation locale invariante aux différentes transformations de l'image considérées, cette section présente une introduction à la théorie des invariants. Pour plus d'information le lecteur pourra se référer à [Mun 92b] et à [Gro 92].

3.2.1 Définition théorique d'un invariant

D'une manière générale un invariant est une propriété qui est constante pour un ensemble de fonctions. La définition théorique des invariants sous sa forme algébrique est due à Hilbert.

Définition 3.2 *Étant donnés deux ensembles E et F, un ensemble T de transformations de E dans F et I une fonction dont l'ensemble de départ est F, I est invariante par T si et seulement si:*

$$\forall e \in E \; \forall t, t' \in T \;\; I(t(e)) = I(t'(e))$$

Dans la cadre de la vision par ordinateur, trois types de transformations particulières nous intéressent : les transformations de la scène tridimensionnelle vers l'image, les transformations de l'image et les transformations qui opèrent sur le signal de l'image (les changements de luminosité). Le problème est de calculer des invariants pour ces différentes transformations.

3.2.2 Calcul des invariants

Il existe deux types de méthodes pour calculer les invariants d'un problème donné : les méthodes infinitésimales et les méthodes par généralisation et contrainte.

Méthodes infinitésimales

Le calcul d'invariants par les méthodes infinitésimales reposent sur les groupes de Lie.

Définition 3.3 *Un groupe de Lie est un ensemble qui est à la fois une sous-variété de $\mathbb{R}^n$ ou de $\mathbb{C}^n$ et un groupe tel que la multiplication et l'inversion sont continues.*

Les groupes de Lie sont des ensembles de fonctions paramétrées dont les paramètres définissent une structure de groupe. Dans le cadre de la définition 3.2 précédente, ces fonctions vont de E dans E et l'ensemble des transformations est un groupe noté G.

Pour un élément e de E donné, on définit son orbite $\mathcal{O}(e)$ comme l'ensemble des images de e par toutes les transformations du groupe G :

Définition 3.4

$$\mathcal{O}(e) = \{e' \in E \mid \exists g \in G \; e' = g(e)\}$$

est l'orbite de e selon G.

En fait, la relation $\mathcal{R}(e, e')$ *ssi* $\exists g \in G \mid e' = g(e)$ est une relation d'équivalence puisque G a une structure de groupe. L'orbite $\mathcal{O}(e)$ est donc la classe d'équivalence de l'élément e. L'ensemble des orbites des éléments de E forme une partition de E.

L'utilisation de groupes de Lie permet de calculer théoriquement des invariants par résolution d'équations différentielles. En effet, les invariants sont des fonctions analytiques f_i constantes sur les orbites, mais qui peuvent distinguer ces différentes orbites. Comme ces fonctions f_i sont constantes sur chaque orbite, leur gradient est donc orthogonal aux espaces tangents aux orbites. Soit f_j une fonction invariante dont le gradient est noté $\vec{\nabla} f_j$, et soit $V_i(e)$ une base de l'espace tangent à l'orbite de e. On a alors :

$$\forall j \; \forall i \; \forall e \quad \vec{\nabla} f_j . V_i(e) = 0$$

La résolution de cette équation différentielle permet d'obtenir des invariants.

Méthode par généralisation et contrainte

La résolution de l'équation différentielle présentée à la section précédente s'avère parfois délicate. Gros [Gro 92] a donc proposé d'utiliser une méthode classique pour calculer des invariants : pour chercher des invariants associés à un problème donné, on généralise le problème à résoudre. Ensuite on calcule les invariants pour le problème généralisé et on exprime le fait que le problème de départ en est un cas particulier. Le problème de cette méthode est de trouver une généralisation du problème de départ pour laquelle on sache calculer des invariants.

Mirbach [Sch 95] a proposé une méthode de généralisation qui exprime les solutions de l'équation différentielle sous la forme d'une intégrale calculée sur toute l'image. Pour une transformation donnée, il propose en fait de calculer la moyenne des valeurs sur une orbite associée à cette fonction.

3.2.3 Dénombrement des invariants

Dans le cadre des groupes de Lie, le théorème suivant permet de connaître le nombre d'invariants indépendants pour un problème donné. Avant de donner ce théorème, il est nécessaire de définir la notion d'invariants indépendants. Soit I un invariant pour une configuration x, alors pour toute fonction f, $f(I(x))$ est un invariant pour x. À partir d'un invariant, il est donc possible de générer une infinité d'invariants. Cependant, les dérivées partielles de tous ces invariants sont linéairement dépendantes. D'où la définition d'invariants indépendants :

Définition 3.5 *Des invariants sont dits indépendants si leurs dérivées partielles sont linéairement indépendantes.*

Soit E un espace vectoriel, et soit G un groupe de Lie opérant sur cet espace, alors le nombre n d'invariants indépendants est :

$$n = dimE - (dimG - \min_{e \in E}(dimG_e))$$

où G_e est le *groupe d'isotropie* ou *groupe stabilisateur* de e :

$$G_e = \{g \in G \mid g(e) = e\}$$

3.2.4 Théorème de Burns

Dans le cas de la vision par ordinateur, les ensembles E et F peuvent être différents. L'ensemble des transformations ne forme alors plus un groupe. Dans ce cas, les orbites ne forment plus une partition : elles peuvent se croiser et la relation $\mathcal{R}(e, e')$ n'est plus une relation d'équivalence. Dans ce cas, il n'existe pas d'invariant. Ceci a été énoncé par Burns [Bur 90], puis également par Moses [Mos 92] et Clemens [Cle 90] :

Théorème 3.1 *Dans le cas des configurations de n points et des projections perspectives, affines ou orthogonales, quel que soit l'entier n, les orbites se croisent de telle manière que toute fonction constante sur les orbites est constante sur tout l'ensemble d'arrivée. Il n'y a donc pas d'invariants pour ce problème exception faite des fonctions constantes.*

Dans le cas général, la recherche d'invariants pour les transformations de la scène tridimensionnelle vers l'image est donc vaine. Toutefois, il existe des invariants pour des classes géométriques d'objet 3D, comme l'ont montré Zisserman et al. [Zis 95]. Ces classes incluent les surfaces de révolution, les tubes, les objets symétriques ainsi que les polyèdres. Il est également possible d'approximer les transformations perspectives par des quasi-invariants comme nous le verrons à la section suivante.

3.2.5 Quasi-invariants

L'apparition des quasi-invariants dans la communauté en vision remonte à la fin des années 1960 et est due à Binford. Dans [Bin 93] il reprend et précise la définition des quasi-invariants. Il démontre également que les invariants au groupe des similitudes image sont des quasi-invariants pour une transformation perspective. En effet, la théorie des quasi-invariants permet de définir des invariants au premier ordre dans le cas des transformations perspectives.

Soit g une fonction de E dans F qui définit une relation d'équivalence sur E. Une fonction f de E dans F est un quasi-invariant de g en $e \in E$ si f est localement constante sur les classes d'équivalence de E et localement équivalente à g en e. En termes plus mathématiques, la définition d'un quasi-invariant est :

Définition 3.6 *Soit G un ensemble de transformation, la fonction f est quasi-invariante pour une transformation $g \in G$ si son développement de Taylor est constant au second ordre pour l'ensemble de transformations considérées et si son développement au premier ordre est égal au développement de g.*

Le dénombrement des quasi-invariants pour un problème donné est difficile. De même, vérifier qu'une fonction est un quasi-invariant est souvent complexe. Par contre, les quasi-invariants se révèlent souvent plus stables que les invariants et ils apportent une solution dans le cas où il n'existe pas d'invariants.

3.3 Invariance et transformations de l'image

Dans cette section les invariants obtenus pour une caractérisation basée sur les dérivées sont présentés. Les différentes transformations considérées sont une rotation image, un changement d'échelle, un changement de luminosité et un changement de point de vue. Obtenir des descriptions invariantes est également possible pour les autres caractérisations vues en section 3.1. Par exemple, dans le cas d'une description fréquentielle, la transformée de Fourier-Mellin [Gra 91, Rub 91] permet de rendre une transformation de Fourier invariante aux rotations image et aux changements d'échelle. De même, la méthode proposée par Hu [Hu 62] rend invariants les moments aux rotations et changements d'échelle.

3.3.1 Rotation image

Dans cette section, nous présentons deux approches différentes pour obtenir une invariance à la rotation. La première est d'utiliser des mesures différentielles invariantes au groupe des déplacements. Cette caractérisation repose sur des combinaisons de dérivées invariantes à une rotation image. D'une manière théorique, de tels invariants ont été proposé par Kœnderink [Koe 87] et Romeny et al. [Sal 92, Flo 93, Rom 94b, Rom 94a]. La

deuxième approche pour obtenir une caractérisation invariante au groupe des déplacements consiste à utiliser le principe des filtres ajustables. Dans ce deuxième cas, les dérivées sont ajustées dans la direction du gradient et ainsi invariantes à une rotation image.

Invariants différentiels

À partir du jet local défini à la section 3.1, Koenderink ainsi que Romeny proposent de calculer des invariants pour le groupe des déplacements $SO(2)$. Pour ce faire ils ont repris des résultats mathématiques, formulés entre autres par Hilbert [Hil 93]. Ils soulignent également la nécessité d'implémenter le calcul des dérivées de manière stable pour pouvoir calculer ces invariants à un ordre élevé.

Nous utilisons l'ensemble d'invariants différentiels jusqu'au troisième ordre. Ces invariants sont regroupés dans un vecteur noté $\vec{\mathcal{V}}$. La première partie de ce vecteur est constituée d'un ensemble complet et irréductible d'invariants différentiels jusqu'au deuxième ordre :

$$\vec{\mathcal{V}}[0..4] = \begin{bmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ji} \end{bmatrix} = \begin{bmatrix} L \\ L_x L_x + L_y L_y \\ L_{xx} L_x L_x + 2 L_{xy} L_x L_y + L_{yy} L_y L_y \\ L_{xx} + L_{yy} \\ L_{xx} L_{xx} + 2 L_{xy} L_{xy} + L_{yy} L_{yy} \end{bmatrix} \tag{3.2}$$

Les L_i sont les éléments du jet local défini par la définition 3.1. L représente par exemple la fonction de luminance convoluée avec une gaussienne. La formulation de cette première partie du vecteur est donnée en notation d'Einstein et en coordonnées cartésiennes. En notation cartésienne, les indices x et y représentent respectivement la dérivation par rapport aux variables x et y, par exemple $L_{xy} = \frac{\partial^2}{\partial_x \partial_y} L$. En notation d'Einstein, un indice i signifie la sommation des dérivations par rapport à l'ensemble des variables :

$$L_i = \sum_i L_i = L_x + L_y \text{ et } L_{ij} = \sum_i \sum_j L_{ij} = L_{xx} + L_{xy} + L_{yx} + L_{yy}.$$

On peut constater que la deuxième composante de ce vecteur est la magnitude du gradient et la quatrième le Laplacien. Il est possible de calculer les invariants pour différentes tailles σ de la gaussienne, car ils sont définis à partir des L_i.

La deuxième partie du vecteur est constitué d'un ensemble complet d'invariants du troisième ordre. Ces invariants en notation d'Einstein sont :

$$\vec{\mathcal{V}}[5..8] = \begin{bmatrix} \varepsilon_{ij}(L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{iij} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ -\varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix} \tag{3.3}$$

où ε_{ij} représente le tenseur canonique anti-symétrique : $\varepsilon_{12} = -\varepsilon_{21} = 1$ et $\varepsilon_{11} = \varepsilon_{22} = 0$.

En notation cartésienne nous obtenons donc :

$$\vec{\mathcal{V}}_i[5..8] =$$

$$\begin{bmatrix} L_{xxx} L_y L_y L_y + 3 L_{xyy} L_x L_x L_y - 3 L_{xxy} L_x L_y L_y - L_{yyy} L_x L_x L_x \\ L_{xxx} L_x L_y L_y + L_{xxy}(-2 L_x L_x L_y + L_y L_y L_y) + L_{xyy}(-2 L_x L_y L_y + L_x L_x L_x) + L_{yyy} L_x L_x L_y \\ L_{xxy}(-L_x L_x L_x + 2 L_x L_y L_y) + L_{xyy}(-2 L_x L_x L_y + L_y L_y L_y) - L_{yyy} L_x L_y L_y + L_{xxx} L_x L_x L_y \\ L_{xxx} L_x L_x L_x + 3 L_{xxy} L_x L_x L_y + 3 L_{xyy} L_x L_y L_y + L_{yyy} L_y L_y L_y \end{bmatrix} \tag{3.4}$$

Jet ajustable

À partir du jet local défini à la section 3.1, il est possible de calculer les dérivées dans une direction donnée. Pour être invariant à la rotation, cette direction est par exemple la direction du gradient. Nous donnons par la suite, les formules pour calculer jusqu'à l'ordre 3, les dérivées dans une direction θ donnée (cf. [Fre 91]). Les $L_x, L_y, \ldots$ sont les éléments du jet local.

$$\begin{aligned}
L'(\theta) &= L_x \cos(\theta) + L_y \sin(\theta) \\
L''(\theta) &= L_{xx} \cos^2(\theta) + 2L_{xy} \sin(\theta)\cos(\theta) + L_{yy} \sin^2(\theta) \\
L'''(\theta) &= L_{xxx} \cos^3(\theta) + 3L_{xxy} \cos^2(\theta)\sin(\theta) + 3L_{xyy} \sin^2(\theta)\cos(\theta) + L_{yyy} \sin^3(\theta)
\end{aligned}$$

Dans ces formules, la dérivée d'ordre n dans une direction donnée dépend des $n+1$ dérivées d'ordre n. Pour représenter de façon complète l'ensemble des dérivées à un ordre donné n, il faut utiliser $n+1$ dérivées directionelles correspondant à $n+1$ directions $\theta_{n,i}$ $i = 0 \ldots n$. Pour des raisons de stabilité, les $n+1$ directions $\theta_{n,i}$ utilisées doivent être espacées régulièrement. Afin d'obtenir des dérivées indépendantes de la rotation existante entre deux images, la direction $\theta_{n,0}$ doit de plus être rapportée à l'image. Si cette direction correspond à la direction du gradient, les orientations sont alors : $\theta_{n,i} = i\pi/(n+1) + \theta_g$ où $\theta_g = \arctan(L_y/L_x)$. Le calcul de cette direction θ_g est une source d'instabilité des méthodes utilisant les jets ajustables.

Normalisation en taille de l'image

Pour obtenir effectivement des dérivées invariantes en rotation, la forme des pixels doit être carrée. Sinon la rectangularité des pixels introduit une anisotropie qui fausse le calcul des dérivées. Il faut donc normaliser l'image. Ceci est fait par interpolation linéaire sur les colonnes de l'image en utilisant un facteur de réduction égal au facteur "α_v/α_u". Ce facteur de normalisation "α_v/α_u" représente le ratio entre la largeur et la longueur d'un pixel. Différentes expérimentations ont montré que ce facteur est stable et peu dépendant du calibrage.

3.3.2 Changement d'échelle

Un changement d'échelle peut être dû soit à un changement de la distance entre la caméra et l'objet soit à un changement de la longueur focale de l'objectif (dans le cas d'un zoom). Nous noterons dans la suite un changement d'échelle par α de manière à le distinguer de la taille σ de la gaussienne utilisée pour effectuer les calculs de dérivation.

Dans cette section des invariants théoriques à un changement d'échelle sont d'abord présentés. Il est ensuite montré que de tels invariants ne sont pas valable dans le contexte du jet local où les dérivées sont calculées sur un support. De ce fait, il est nécessaire d'utiliser une approche multi-échelle.

Invariants à un changement d'échelle

Étant donnée une fonction f, un changement d'échelle α peut être décrit par un changement de variable : $f(x) = g(u)$ où $g(u) = g(u(x)) = g(\alpha x)$. De cette relation découlent les relations suivantes entre f et g :

$$f^{(n)}(x) = \alpha^n g^{(n)}(u) \quad \text{où } f^{(n)}(x) \text{ représente la dérivée } n\text{-ième de } f. \tag{3.5}$$

L'équation 3.5 montre que les dérivées n-ièmes de f et de g sont égales à un facteur multiplicatif α^n près. À partir du quotient de deux dérivées il est donc possible d'éliminer ce facteur α^n. Des invariants théoriques à un changement d'échelle sont donnés par l'équation suivante :

$$\frac{\left[f^{(n)}(x)\right]^{\frac{k}{n}}}{f^{(k)}(x)}$$

Cependant, des résultats expérimentaux en prenant des images de tableaux de maître à différentes échelles ont montré que de tels invariants sont peu stables à un changement d'échelle supérieur à 20%. Les résultats présentés à la section 4.3.3 du chapitre suivant ont montré que les invariants à la rotation étaient eux aussi robustes à un changement d'échelle de 20%. Cette robustesse est cohérente avec les observations faites par Fleet dans [Fle 91] dans le contexte des filtres de Gabor.

En fait, les invariants à l'échelle n'apportent pas de stabilité supplémentaire. Ceci est dû au fait que le calcul numérique est effectué sur un support. En effet, dans le cas où les dérivées sont calculées par convolution avec les dérivées de la gaussienne, l'équation 3.5 précédente se réécrit de la manière suivante :

$$\int_{-\infty}^{+\infty} I_1(\vec{x}) G_{i_1 \ldots i_n}(\vec{x}, \sigma) d\vec{x} = \alpha^n \int_{-\infty}^{+\infty} I_2(\vec{u}) G_{i_1 \ldots i_2}(\vec{u}, \sigma\alpha) d\vec{u} \tag{3.6}$$

où les $G_{i_1 \ldots i_2}$ représentent les dérivées de la fonction gaussienne définie par l'équation 2.2.

Cette équation montre l'importance du support (σ à gauche et $\alpha\sigma$ à droite) sur lequel sont effectués les calculs. Ce support doit être adapté au changement d'échelle pour calculer effectivement un invariant. Ceci est à l'origine des méthodes multi-échelle telle que celle que nous allons présenter à la section suivante.

Approche multi-échelle

Dans la littérature il existe de nombreuses approches multi-échelle. Parmi les premières approches, il faut citer les pyramides qui ont été proposées par Burt [Bur 81] et Crowley [Cro 81, Cro 84]. Toutefois ces approches effectuent un sous-échantillonnage de l'image et ne sont donc pas adaptées à notre problème. En effet, notre approche est basée sur la notion d'espace d'échelle où un paramètre continu définit l'échelle. Cette notion a été introduite par Witkin [Wit 83] et Koenderink [Koe 84]. Plus tard, Lindeberg [Lin 94] a étendu et résumé leur approche.

L'espace d'échelle permet de calculer les invariants à une échelle donnée. Il est cependant impossible de calculer les invariants à toutes les échelles. La discrétisation de l'espace d'échelle est donc nécessaire. De nombreux auteurs ont proposé une discrétisation par octave ou par demi-octave. Avec un tel pas de discrétisation la caractérisation obtenue s'est révélée imprécise et instable. Puisque notre caractérisation est robuste à un changement d'échelle jusqu'à 20% (cf. section 4.3.3), nous avons choisi un pas de discrétisation qui garantit qu'entre deux échelles consécutives, le changement est inférieur à 20%. De manière à être résistant à un changement d'échelle jusqu'à un facteur 2, les différentes échelles

retenues ont pour valeur : 0.48, 0.58, 0.69, 0.83, 1, 1.2, 1.44, 1.73, 2.07. Nous effectuons donc les calculs pour des différentes échelles, c'est-à-dire pour différentes tailles σ de la gaussienne. Ceci nous permet de réaliser une approche multi-échelle.

L'intégration des invariants différentiels présentés à la section 3.3.1 dans un cadre multi-échelle permet d'obtenir une caractérisation robuste au groupe des similitudes.

3.3.3 Changement de luminosité

La caractérisation doit également être robuste à un changement de luminosité. Il existe plusieurs possibilités pour modéliser un changement de luminosité. Par la suite trois modèles de transformations de niveaux de gris sont présentés : une translation, une transformation affine et une transformation monotone. Pour chacun de ces modèles on définit les invariants correspondants.

Translation des niveaux de gris

Une translation des niveaux de gris se modélise par :

$$\tilde{I}(x,y) = I(x,y) + b$$

Il est facile de voir que par simple dérivation, le facteur b s'élimine et par conséquent les invariants différentiels, à part la moyenne des intensités lumineuses sont invariants à un tel changement. Le vecteur $\vec{\mathcal{V}}$ sans la composante $\vec{\mathcal{V}}[0]$ est un invariant. Il est dans la suite référencé par $\vec{\mathcal{V}}_T$.

Transformation affine des niveaux de gris

Une transformation affine des niveaux de gris se modélise par :

$$\tilde{I}(x,y) = aI(x,y) + b$$

Une telle transformation modifie les dérivées du signal de la manière suivante : $\tilde{I}^{(n)}(x,y) = aI^{(n)}(x,y)$. N'importe quel quotient de deux dérivées est donc invariant à une transformation affine de la luminance. Il y a différentes manières de rendre le vecteur $\vec{\mathcal{V}}$ invariant à une transformation affine. Nous avons choisi de diviser par la puissance adéquate de la magnitude du gradient :

$$\vec{\mathcal{V}_{\mathcal{A}}}[0..2] = \begin{bmatrix} \frac{L_i L_{ij} L_j}{(L_i L_i)^{3/2}} \\ \frac{L_{ii}}{(L_i L_i)^{1/2}} \\ \frac{L_{ij} L_{ji}}{L_i L_i} \end{bmatrix} \text{ et } \vec{\mathcal{V}_{\mathcal{A}}}[3..6] = \frac{1}{(L_i L_i)^2} \begin{bmatrix} \varepsilon_{ij}(L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{iij} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ -\varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix} \quad (3.7)$$

Transformation monotone de la luminosité

Un changement de luminosité peut également être modélisé par une fonction monotone et donc inversible. L'inversibilité de la fonction (ou sa stricte monotonie) est nécessaire

pour éviter une perte d'information par rapport à celle contenue dans l'image de niveaux de gris.

Florack [Flo 94] montre que le jet local permet également de calculer des invariants par rapport à n'importe quelle transformation inversible de la luminosité. Il fait remarquer que les isophotes ne sont pas modifiés sous l'action d'une transformation inversible de luminosité. Jusqu'au deuxième ordre il existe deux invariants indépendants, notamment la courbure des isophotes κ et la courbure des lignes de plus grande pente μ :

$$\kappa = \frac{\varepsilon_{ij}\varepsilon_{kl}L_iL_{jk}L_l}{(L_mL_m)^{3/2}} = \frac{2L_xL_yL_{xy} - L_x^2L_{yy} - L_y^2L_{yy} - L_y^2L_{xx}}{L_x^2 + L_y^2}$$

$$\mu = \frac{\varepsilon_{ij}L_jL_kL_{ik}}{(L_mL_m)^{3/2}} = \frac{L_{xy}(L_yL_y - L_xL_x) + L_xL_y(L_{xx} - L_{yy})}{L_x^2 + L_y^2}$$

Berthod et al. [Ber 94] présente une famille d'invariants à une transformation monotone de la luminosité. Ces invariants sont basés sur les orientations du gradient et ses dérivées partielles.

3.3.4 Autres transformations image

Il est possible de calculer des invariants différentiels pour d'autres types de transformation image, par exemple pour le groupe des transformations affines.

Dans ce cas, une possibilité est d'utiliser une transformation affine pour transformer la conique définie par une équation de la forme $\vec{x}^T L_{ij}\vec{x} = Cste$ en cercle et d'effectuer les calculs dans ce repère normalisé. Ceci est équivalent à calculer des invariants à partir de l'inverse de la matrice L_{ij}. On obtient alors l'ensemble suivant d'invariants jusqu'au troisième ordre :

$$\mathcal{V}_{\mathcal{AFFIN}\vec{\mathcal{E}}}[0..5] = \begin{bmatrix} L \\ L_iL_j(L^{-1})^{ij} \\ L_{ijl}L_lL_mL_n(L^{-1})^{il}(L^{-1})^{jm}(L^{(}-1))^{kn} \\ L_{ijk}L_l(L^{-1})^{ij}(L^{-1})^{kl} \\ L_{ijk}L_{lmn}(L^{-1})^{ij}(L^{-1})^{kl}(L^{-1})^{mn} \\ L_{ijk}L_{lmn}(L^{-1})^{il}(L^{-1})^{jm}(L^{-1})^{kn} \end{bmatrix}$$

Toutefois, comme dans le cas d'un changement d'échelle, pour calculer de tels invariants, il est nécessaire de tenir compte du support de calcul. Étendre directement l'approche multi-échelle est possible, mais extrêmement coûteux : il faut calculer les invariants pour différents supports, chacun correspondant à un jeu de paramètres donnés, sachant qu'il y a trois paramètres à prendre en compte. Après évaluation, nous n'avons pas retenu ces invariants dans notre approche à cause de la difficulté de leur mise en œuvre.

3.3.5 Changement de point de vue

Les invariants à une rotation image sont calculés dans un cadre multi-échelle. On obtient donc une caractérisation invariante au groupe des similitudes dans l'image. Binford [Bin 93] a montré que de tels invariants sont des quasi-invariants à une transformation perspective, c'est-à-dire qu'ils sont localement invariants à une telle transformation. Notre caractérisation est donc robuste à des transformations perspectives. Ceci est confirmé au chapitre suivant qui évaluent notre caractérisation (cf. section 4.3.5).

3.4 Évaluation de la caractérisation

La méthode d'évaluation de la caractérisation et les résultats obtenus sont présentés à la section 4.3 du chapitre suivant. Cette évaluation est effectuée dans un cadre d'appariement basé sur la caractérisation présentée. Le critère d'évaluation est le taux d'appariements corrects. En effet, un taux élevé montre le pouvoir discriminant de la caractérisation que nous avons choisie. Ce taux est calculé séparément pour les différentes transformations considérées, notamment une rotation image, un changement d'échelle, un changement de luminosité et un changement de point de vue. Un taux constant montre alors l'invariance à la transformation considérée. Par ailleurs, la caractérisation est représentée par un vecteur comportant 9 composantes. L'utilité de chacune de ces composantes est prouvée à la section 4.3.8.

3.5 Conclusion

Ce chapitre a présenté un état de l'art des méthodes de caractérisation d'une image. La théorie des invariants nous a permis d'introduire une caractérisation invariante aux différentes transformations. Celle-ci est basée sur les dérivées gaussiennes du signal. L'utilisation de combinaisons de ces dérivées permet d'être invariant à une rotation image. Puisqu'il n'existe pas d'invariant à l'échelle qui soit numériquement applicable, nous proposons d'utiliser un cadre multi-échelle pour obtenir une robustesse à un changement d'échelle. Une robustesse à un changement de luminosité peut être obtenue de deux manières différentes. La première manière consiste à ne pas utiliser la moyenne d'intensité des niveaux de gris comme composante du vecteur de caractéristiques. La deuxième repose sur des invariants spécifiques à un changement de luminosité. Le nombre de ces invariants est cependant réduit.

Nous avons choisi une caractérisation basée sur les dérivées. Toutefois, d'autres choix auraient été possibles et une étude comparative devra être menée. Ceci constitue une première extension de ce travail. Une autre extension possible de ce chapitre est d'utiliser la couleur. Nous avons pu voir dans le cas d'un changement de luminosité que nous ne pouvons pas utiliser la moyenne des intensités lumineuses autour d'un point. Toutefois, l'information de luminance est représentative d'un point. Les invariants colorimétriques permettront donc de garder cette information de luminance, sous une forme invariante à un changement de luminosité.

Chapitre 4

Appariement entre images

Ce chapitre utilise les points d'intérêt du chapitre 2 et la caractérisation du signal du chapitre 3 pour réaliser l'étape finale de notre algorithme : l'appariement. De très nombreuses méthodes ont été proposées pour trouver une solution au problème d'appariement entre images. La section 4.1 en présente un état de l'art. Les méthodes existantes sont soit basées sur des grandeurs géométriques soit sur des grandeurs photométriques. Les grandeurs géométriques s'avèrent instables et peu discriminantes pour des images réelles et les grandeurs photométriques sont peu invariantes aux transformations image couramment observées. L'utilisation faite à la section 4.2 de la caractérisation locale et invariante permet d'apporter une solution robuste qui remédie à ces défauts. Ceci est confirmé par les résultats présentés à la section 4.3.

4.1 État de l'art

4.1.1 Appariement basé sur des données photométriques

La méthode de mise en correspondance la plus ancienne est la corrélation du signal (voir par exemple [Fau 92a] pour une comparaison des méthodes, voir aussi [Zha 89] ou [Zab 94] pour une corrélation binaire rapide). La corrélation peut être interprétée comme une caractérisation élémentaire du signal autour d'un point : tout point est caractérisé par l'ensemble des valeurs de niveaux de gris sur un voisinage. Pour qu'une telle méthode réussisse, l'image doit avoir été prise dans des conditions voisines pour que les fenêtres de corrélation se superposent correctement par simple translation. Si une rotation sensible de l'image est intervenue, alors la méthode échoue. Il faut noter cependant l'existence de travaux qui permettent, lorsque la géométrie épipolaire est connue, de compenser les transformations géométriques et de plus d'obtenir des mises en correspondance au niveau sous-pixellique ([Ack 84], [Rem 94], [Lot 94]).

Il est également possible de devenir robuste en rotation en utilisant des mesures de corrélation dans plusieurs directions comme par exemple Hu [Hu 94]. Ceci nécessite la discrétisation de l'espace suivant plusieurs directions et reposent sur une hypothèse de linéarité entre deux directions adjacentes. Cette hypothèse est difficilement vérifiable. En

outre, le calcul dans plusieurs directions est coûteux.

La mesure de corrélation conduit à un certain nombre d'erreur de mise en correspondance. Pour améliorer les résultats obtenus par corrélation Deriche et al. [Der 94] proposent l'utilisation de caractéristiques du signal : la direction du gradient, la courbure et la disparité qui est supposée en dessous d'un certain seuil.

Pour diminuer les erreurs d'appariement et le coût de la corrélation, certaines méthodes utilisent des points particuliers où l'image présente un contenu informatif. Par exemple, Zhang et al. [Zha 95] ont appliqué la corrélation aux points d'intérêt. Afin d'améliorer les résultats obtenus, pour un couple de points appariés, les points dans des voisinages respectifs de même taille doivent également se correspondre. De plus ces points du voisinage doivent avoir la même position relative dans les deux images. Cette position relative repose sur une mesure de distance.

4.1.2 Appariement à partir de données géométriques

D'autres méthodes de mise en correspondance entre deux images ont cherché à effectuer la mise en correspondance à partir de données de nature géométrique. Ainsi bon nombre d'auteurs ont cherché à apparier deux images à partir de leurs lignes de contraste. On peut dans ce cas obtenir une structure de contours (essentiellement la structure du graphe [Hor 90] extrait de ces contours) qui capture la structure globale de la scène. Un critère de ressemblance globale à partir de cette structure est alors utilisé pour la mise en correspondance à partir de techniques d'optimisation comme la relaxation [Lon 86], ou alors de techniques combinatoires comme la recherche de cliques maximales [Hor 89]. Il faut noter cependant que dans toutes ces approches, la combinatoire reste si forte qu'il faut la contraindre par l'usage de la géométrie épipolaire afin de rester dans des temps d'exécution raisonnables.

Il est également possible d'utiliser les contours pour calculer des informations locales. En théorie, à un point d'un contour courbe peut être associé un invariant qui permet de discriminer ce point [Wei 91]. De fait ce genre de méthode est difficile à mettre en œuvre parce que trop locale sur le contour. En revanche, si une partie suffisante du contour est visible, des méthodes semi-locales [Die 94], [Rot 92] permettent de calculer des invariants qui caractérisent le morceau de contour observé. On peut ainsi faire une mise en correspondance grossière. Une telle application directe n'a cependant pas été menée à notre connaissance.

Toutes les méthodes de mise en correspondance utilisant les contours partagent un défaut commun : l'utilisation des contours. En effet, ce type de méthode n'est applicable que si la segmentation des images en contours est correctement faite. Sur des scènes aux structures simples, ces méthodes ont prouvé leur applicabilité, mais elles restent néanmoins difficiles à maîtriser.

4.2 Algorithme d'appariement

4.2.1 Principe de l'appariement

En conclusion de l'état de l'art précédent, les grandeurs géométriques s'avèrent instables et peu discriminantes pour des images réelles. Ces méthodes repose sur des caractéristiques préalablement extraites de l'image, c'est-à-dire sur une description symbolique.

Leur instabilité provient donc des erreurs lors de l'extraction de la description symbolique et leur manque de discriminance du fait qu'une partie de l'information, l'information de niveaux de gris, n'est pas utilisée. Les méthodes basées sur des grandeurs photométriques sont par nature plus discriminantes car moins symboliques. Elles sont en effet plus proches du signal. Par contre leur défaut principale est qu'elles ne sont pas invariantes à des types de transformations qui peuvent être pris en compte par les méthodes géométriques.

Dans la suite nous présentons une solution robuste qui remédie à ces défauts. L'algorithme d'appariement est basé sur la caractérisation locale du signal présentée au chapitre précédent. Cette caractérisation est invariante aux transformations image. Ceci permet d'être à la fois discriminant et invariant aux transformations images. Pour pouvoir apparier les vecteurs d'invariants composant cette caractérisation, il faut d'abord introduire une distance qui permet de les comparer. Il est ensuite possible de choisir dans deux images les vecteurs correspondant au même point physique en utilisant le principe de l'appariement croisé. Pour augmenter la robustesse des contraintes semi-locales sont ajoutées.

4.2.2 Distance entre images

Pour apparier deux points, il faut mettre en correspondance leur vecteur d'invariants $\vec{\mathcal{V}}$. Le problème essentiel est de décider si deux vecteurs sont similaires.

L'utilisation de la distance euclidienne pour comparer deux vecteurs d'invariants est peu judicieuse puisque les différentes composantes d'un vecteur peuvent être corrélées et qu'elles n'ont pas le même ordre de grandeur. Par exemple, la première composante de $\vec{\mathcal{V}}$ représente la moyenne de luminance autour d'un point. Elle est donc contenue dans l'intervale $[0, 255]$. À titre de comparaison, la troisième composante représente le Laplacien de la fonction de luminance. L'ordre de grandeur de cette composante est d'environ 0.1. Nous modélisons donc les composantes par des variables aléatoires gaussiennes et nous utilisons la distance de Mahalanobis pour comparer deux vecteurs d'invariants. Cette distance statistique utilise la matrice de covariance Λ des composantes et prend ainsi en compte la différence de magnitude des composantes ainsi que leur corrélation éventuelle :

$$d_M(\vec{b}, \vec{a}) = \sqrt{(\vec{b} - \vec{a})^T \Lambda^{-1} (\vec{b} - \vec{a})}$$

En seuillant cette distance, il est possible de décider si deux vecteurs sont similaires. En outre, le carré de la distance de Mahalanobis est une variable aléatoire qui suit une distribution du χ^2. La fonction racine carré étant une bijection de $\mathbb{R}^+$ dans $\mathbb{R}^+$, il est possible d'utiliser une table de cette distribution pour seuiller la distance et rejeter les appariements qui ont la plus grande probabilité d'être faux. Ces appariements correspondent aux plus grandes valeurs de la distance.

La qualité des résultats obtenus avec cette distance dépend de la représentativité de la matrice de covariance. Cette matrice doit tenir compte du bruit des images, des variations d'éclairage et de l'imprécision en position des points d'intérêt. Un changement de seulement un pixel de cette position perturbe de façon importante la valeur des invariants. Un calcul théorique de cette matrice est difficile puisque la forme du signal autour d'un point d'intérêt est quelconque. Un tel calcul est uniquement possible si l'on restreint les points utilisés à des coins. Cette matrice a donc été estimée de façon empirique. Étant donné un point d'intérêt, il a été suivi sur une séquence d'images. La matrice de covariance a alors été calculée pour ce point en utilisant l'ensemble des vecteurs d'invariants calculés pour

chacune des images. Afin d'obtenir une matrice représentative de la variété des points possibles, ceci a été réalisé pour plusieurs points sur plusieurs séquences avec des scènes différentes. La moyenne de toutes les matrices obtenues a ensuite été utilisée comme la matrice de covariance globale. Afin d'obtenir une matrice représentative, les séquences utilisées sont relativement variées et tous les points de ces séquences ont été utilisées. Cette procédure s'est avérée suffisante. Toutefois, une étude plus approfondie reste à mener pour savoir le nombre de points à considérer et la variété des séquences à utiliser. De même, il peut être intéressant d'étudier les changements de la matrice en fonction des points utilisés.

4.2.3 Procédure d'appariement

Le processus de mise en correspondance est basé sur le principe d' " appariement-croisé ". C'est un algorithme d'appariement très simple. La section suivante complète cet algorithme par l'ajout de contraintes et l'utilisation d'un seuillage statistique.

Ayant calculé des vecteurs de caractéristiques $\vec{\mathcal{V}}$ pour chaque point d'intérêt, l'algorithme recherche d'abord pour chaque point d'intérêt de la première image le point le plus semblable dans la deuxième image. Pour décider si deux points sont semblables, on compare leur vecteur de caractéristiques en utilisant la distance de Mahalanobis (cf. 4.2.2). Le processus est ensuite interverti. On obtient donc deux listes de paires de points appariés. Les appariements retenus sont les paires de points qui se sont choisies mutuellement. La figure 4.1 illustre le principe de cette mise en correspondance.

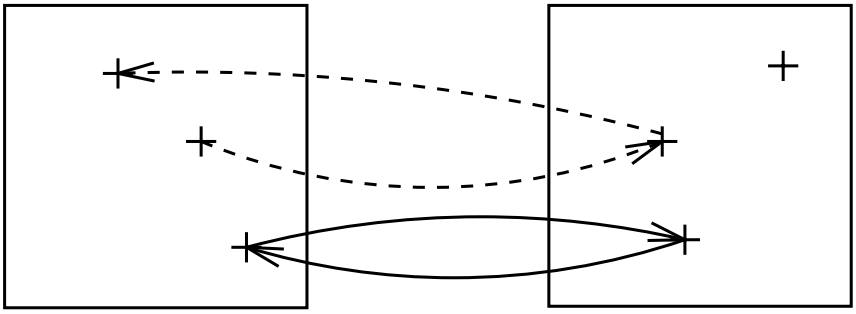

FIG. 4.1 – *Principe de la mise en correspondance par " appariement-croisé "*

Pour rendre robuste la mise en correspondance à un changement d'échelle, le vecteur d'un point d'intérêt de la première image est comparé avec les vecteurs d'un point de la deuxième image calculés à plusieurs tailles σ de la gaussienne. Les tailles σ utilisées ont été précisées à la section 3.3.2. La taille σ pour laquelle les vecteurs se ressemblent le plus est retenue comme estimation du facteur d'échelle entre les deux images. La même procédure est effectuée symétriquement pour chaque point de la deuxième image, on obtient ainsi une deuxième estimation du changement d'échelle. Une paire de points n'est alors retenue comme appariement que si elle a été retrouvée deux fois et si les tailles σ se correspondent.

4.2.4 Contraintes semi-locales

En présence de bruit, un vecteur de caractéristiques donné peut être impliqué dans plusieurs appariements. Un nombre important d'appariements possibles ainsi que la simi-

larité de certains vecteurs de caractéristiques accroît encore la probabilité qu'un vecteur vote pour plusieurs appariements.

Califano [Cal 94] ainsi que Rao [Rao 95] ont suggéré que l'utilisation de vecteurs de caractéristiques de grande dimensionnalité diminue cette probabilité d'erreur. Dans notre cas, ceci signifie augmenter l'ordre de dérivation des caractéristiques ce qui est de fait peu praticable. D'autre part, l'ajout de vecteurs calculés à différentes échelles, comme l'a proposé Rao dans [Rao 95] rend l'approche multi-échelle difficile à mettre en œuvre.

Un autre moyen de diminuer la probabilité de fausses correspondances est de filtrer les mises en correspondance par l'addition d'une contrainte de cohérence basée sur le voisinage. Nous avons opté pour cette dernière solution : l'utilisation de contraintes semi-locales. Les contraintes détaillées dans la suite sont des contraintes de voisinage et de nature géométrique. Ces contraintes s'avèrent particulièrement utiles dans le cadre d'appariement entre une image et une base d'images. Leur influence est montrée à la section 5.2.3 du chapitre suivant.

Contrainte de voisinage

La contrainte de voisinage prend en compte la configuration relative des points d'intérêt. Ainsi, à chaque point P_j dans une image sont associés les p points d'intérêt les plus proches. Un point d'une image est alors représenté par son vecteur d'invariants et ces p points voisins :

$$(\vec{\mathcal{V}}_j, P_{j,1}, \ldots P_{j,p})$$

Si le vecteur de caractéristiques $\vec{\mathcal{V}}_j$ issu de l'image est apparié avec le vecteur $\vec{\mathcal{V}}_k$ d'une seconde image, tout ou partie des p points d'intérêt voisins de P_j doivent correspondre à des voisins de P_k. Ceci est illustré par la figure 4.2 et s'écrit formellement :

$$\vec{\mathcal{V}}_{j,m} \text{ et } \vec{\mathcal{V}}_{k,n} \text{ se correspondent pour quelques } (m,n)$$

$$\text{où } \vec{\mathcal{V}}_{j,m} \text{ vecteur de } P_{j,m} \text{ et } P_{j,m} \text{ un des } p\text{-voisins de } P_j$$

$$\text{et } \vec{\mathcal{V}}_{k,n} \text{ vecteur de } P_{k,n} \text{ et } P_{k,n} \text{ un des } q\text{-voisins de } P_k$$

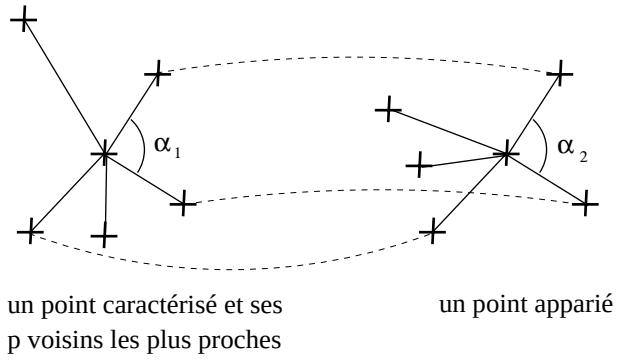

FIG. 4.2 – *Une contrainte de voisinage est ajoutée au processus de mise en correspondance en stockant les p voisins les plus proches d'un point. Durant la phase d'appariement, il est imposé qu'au moins 50% de ces voisins se correspondent pour que la correspondance soit validée.*

Imposer que tous les p voisins les plus proches de P_j correspondent aux voisins de P_k revient à imposer qu'il n'y a aucun bruit de détection[1]. Afin d'être robuste à ces inévitables bruits, il a donc été imposé qu'au moins 50% des voisins de P_j correspondent à des voisins de P_k. Ce seuil de 50% est arbitraire et serait certainement à évaluer, mais les résultats obtenus n'ont pas nécessité d'approfondissement. L'utilisation de cette contrainte augmente le taux d'appariements corrects en réduisant significativement la probabilité des fausses mises en correspondance (cf. figures 5.4 du chapitre suivant).

Contrainte géométrique

L'utilisation de contraintes géométriques permet une vérification supplémentaire des appariements trouvés. Dans le cadre de ce travail, une contrainte basée sur la conservation des angles a été utilisée. Une telle conservation est vérifiée dans le cas des similitudes entre images. La contrainte de voisinage définie dans la section précédente contraint l'appariement entre deux points (voir la figure 4.2). Dans le cas des similitudes image, l'angle défini par deux voisins d'un point doit être constant pour toutes les vues de ce point. Dans le contexte de la figure 4.2, ceci s'écrit $\alpha_1 = \alpha_2$. Il faut se rappeler que chaque point P_j d'une image est caractérisé par son vecteur de caractéristiques $\vec{\mathcal{V}}_j$ et par ses voisins $\{P_{j,n}\}_{n=1}^{p}$.

Ce point P_j est mis en correspondance avec le point P_k d'une seconde image si et seulement si :

$$\vec{\mathcal{V}}_{j,m}, \vec{\mathcal{V}}_{k,n} \text{ se correspondent pour quelques } (m,n)$$

et

$$\|\alpha_{k,n} - \alpha_{j,m}\| < \alpha_t \text{ pour les } (m,n) \text{ considérés}$$

où $\alpha_{j,m}$ est l'angle défini entre deux voisins mis en correspondance et α_t est un seuil angulaire donné. Nous avons utilisé 20 degrés pour ce seuil. L'ajout de cette contrainte augmente encore le taux d'appariements corrects (cf. figures 5.5 du chapitre suivant).

4.3 Évaluation de l'appariement

Le but de cette section est d'évaluer le processus d'appariement. Cette évaluation permet de montrer le pouvoir discriminant de notre caractérisation et son invariance aux différentes transformations. L'évaluation est effectuée séparément pour les différentes transformations considérées, notamment une rotation image, un changement d'échelle, un changement de luminosité et un changement de point de vue. La stabilité au bruit de la caméra a également été testée. Ensuite, quelques résultats pour des transformations complexes sont présentés. La section 4.3.8 quant à elle montre l'importance relative des différentes composantes du vecteur dans un processus de mise en correspondance et prouve l'apport des invariants du troisième ordre.

4.3.1 Cadre d'évaluation

Cette section présente le cadre utilisé pour évaluer l'appariement. Elle présente le critère d'évaluation retenu, la méthode d'évaluation automatique appliquée, les séquences d'images utilisées ainsi que quelques détails d'implémentation.

1. Par bruit de détection, on entend les erreurs de détections dues aux imperfections du détecteur utilisé ainsi que celles dues au bruit de l'image.

Critère d'évaluation

Étant données deux images, on souhaite que tous les points détectés soient correctement appariés. Le critère d'évaluation choisi est donc le nombre d'appariements corrects par rapport au nombre d'appariements trouvés :

$$\frac{nombre\ d'appariements\ corrects}{nombre\ d'appariements\ trouvés}$$

Pour calculer le nombre d'appariements corrects, il est indispensable de connaître la relation entre les deux images à apparier. Le paragraphe suivant explique comment déterminer de manière automatique si un appariement donné est correct.

Évaluation automatique

L'évaluation automatique d'un appariement est différente suivant que la scène observée est plane ou non. Dans le cas d'une scène plane, la méthode d'évaluation utilisée est similaire à celle développée à la section 2.4 du chapitre 2. Deux images d'une scène plane sont reliées par une homographie. Étant donné un appariement de points entre deux images (un couple de points), l'homographie existant entre ces images permet directement de connaître la validité de cet appariement. Dans ce qui suit, les homographies sont calculées d'une manière robuste à partir des appariements calculés. Ce calcul repose sur une méthode de moindres carrés médians et est donc robuste jusqu'à une proportion de 50% de faux appariements.

Dans le cas d'une scène non-plane, il n'existe pas d'homographie entre deux images. Cependant, il existe une contrainte entre deux vues quelconques d'une scène, à savoir la relation épipolaire représentée par la matrice fondamentale F : à un point a d'une image correspond une droite épipolaire F_a dans la seconde image. Un appariement (a, b) peut alors être évalué comme correct si le point b appartient à la droite épipolaire correspondant au point a : $F_a * b = 0$, soit ${}^t bFa = 0$. Une telle évaluation n'est pas exempte d'erreur puisque deux points d'un appariement faux peuvent vérifier la contrainte épipolaire. Cependant, la probabilité d'un tel événement est suffisamment faible pour que cet estimateur fournisse une bonne évaluation du nombre d'appariements corrects. Dans ce qui suit, la contrainte épipolaire entre deux images est calculée à partir des appariements trouvés. Ce calcul repose sur une méthode de moindres carrés médians (cf. [Zha 95, Bou 95]).

Séquences utilisées

L'évaluation de la mise en correspondance a été effectuée sur deux scènes planes et sur quelques scènes tridimensionnelles. Les figures 4.3 et 4.4 montrent les deux scènes planes. Elles sont référencées dans la suite par "Sanja" et "Van Gogh". On peut constater la nature différente de ces deux images : l'image "Sanja" contient surtout des contours nets et l'image "Van Gogh" contient beaucoup de texture. Les figures 4.17, 4.18, 4.20 présentent les scènes tridimensionnelles utilisées.

Détails d'implémentation

Les invariants utilisés pour les expériences qui suivent sont donnés par les équations 3.2 et 3.3. La taille de la gaussienne utilisée pour l'ensemble des expériences est de 3 sauf si

spécifié autrement. Dans le cas d'un changement de luminosité, nous avons comparé ces invariants aux invariants à un changement de luminosité (cf. section 3.3.3). D'autre part, les expériences ont été effectuées sans utiliser de contraintes semi-locales.

4.3.2 Rotation image

Cette section montre les résultats obtenus pour une rotation image. Pour cette expérimentation, nous avons suivi le même procédé expérimental que celui utilisé pour l'évaluation des détecteurs de points d'intérêt : pour chaque scène, une séquence d'images a été prise en faisant tourner la caméra approximativement autour de l'axe optique de son objectif. La figure 4.3 montre les résultats obtenus pour une paire d'images de la scène "Sanja". L'angle de rotation entre les deux images est de 133 degrés. Les croix blanches indiquent les appariements corrects et les noires les faux appariements. Le pourcentage d'appariements corrects est de 88.52% La figure 4.4 montre les résultats pour la scène "Van Gogh". Le pourcentage d'appariements corrects est alors de 74.2%. Il faut noter que parmi les faux appariements certains sont dus à la nature répétitive des motifs.

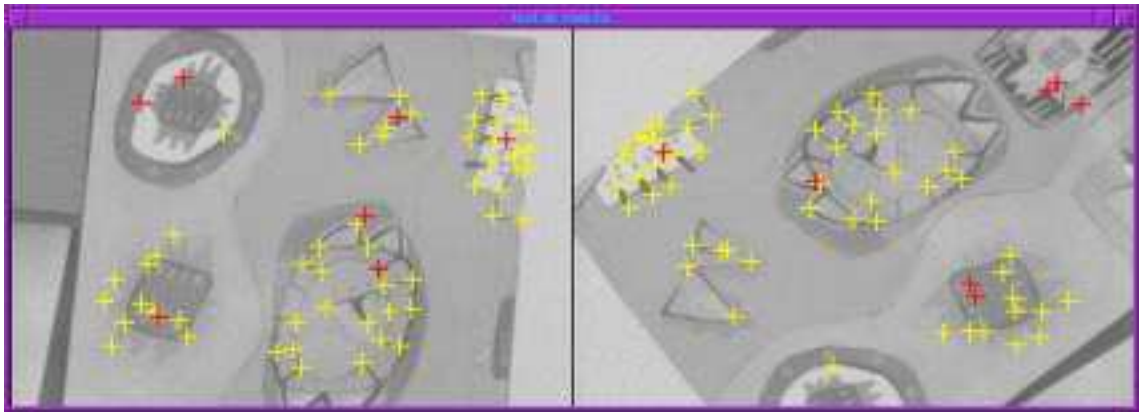

FIG. 4.3 – 88.52% *d'appariements corrects pour une rotation de* 133.0 *degrés.*

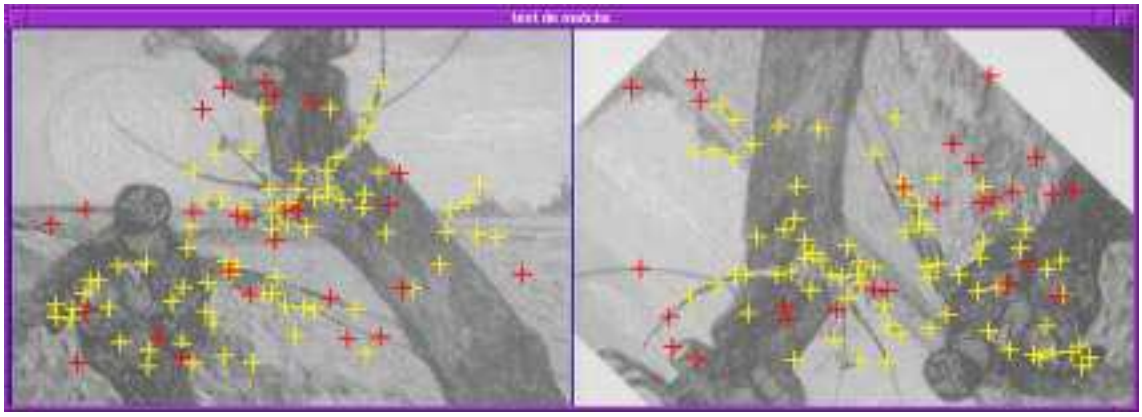

FIG. 4.4 – 74.26% *d'appariements corrects pour une rotation de* 134.8 *degrés.*

La figure 4.5 trace le pourcentage d'appariements corrects en fonction de l'angle de rotation pour la scène "Van Gogh". Ce pourcentage a été calculé pour deux détecteurs

différents de points d'intérêt : Heitger et HarrisPrécis (se reporter à la section 2.2). Ce pourcentage a également été évalué à partir des points précis projetés par l'homographie. Les courbes diffèrent en fonction du détecteur utilisé. Pour les points précis, le taux de mise en correspondance est approximativement de 100%. Les résultats obtenus pour le détecteur de HarrisPrécis sont presque aussi bons. Le détecteur de Heitger donne des résultats nettement moins bons.

Cette expérience montre deux choses. Premièrement, elle montre que la caractérisation proposée permet de distinguer les différents points et qu'elle est invariante à la rotation. Deuxièmement, on peut observer que les résultats obtenus dépendent fortement de la répétabilité du détecteur utilisé. Ceci est dû à la localité de la caractérisation : l'instabilité du détecteur conduit à caractériser deux pixels voisins, mais non identiques. Les caractérisations obtenues sont alors différentes. En effet, on peut remarquer la très forte similarité des résultats obtenus avec le détecteur de Heitger avec la courbe 2.7 de répétabilité de ce détecteur. Des résultats équivalents ont été obtenus sur la scène "Sanja" et sont présentés à l'annexe B sur la figure B.1.

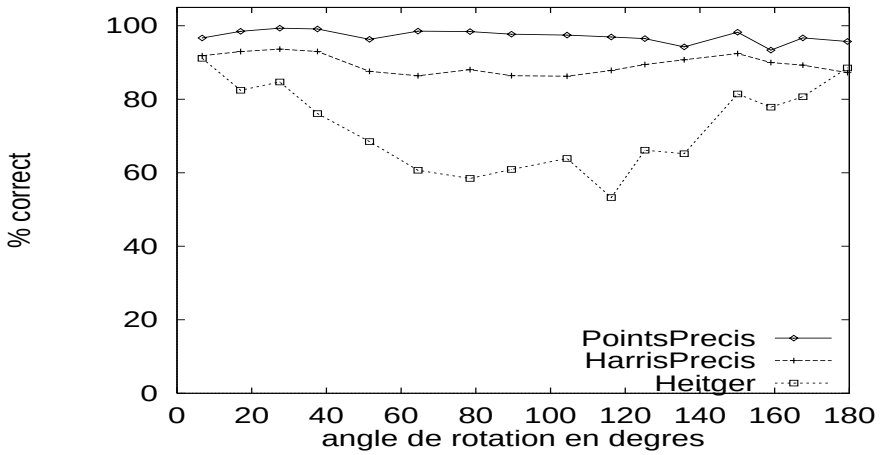

FIG. 4.5 – *Pourcentage d'appariements corrects pour la séquence rotation image et la scène "Van Gogh". Les trois courbes correspondent aux différents détecteurs de points d'intérêt utilisés : Heitger, HarrisPrécis et les points précis.*

Les résultats obtenus avec le détecteur de Heitger peuvent être améliorés en utilisant des tailles σ de gaussienne plus importantes. La figure 4.6 montre les résultats obtenus avec ce détecteur pour différentes tailles σ. Plus cette taille est importante et meilleurs sont les résultats. Ceci est dû au fait que plus cette taille augmente et moins la caractérisation est locale, donc moins sensible à une erreur de précision. L'incertitude en position du détecteur utilisé peut donc être compensée en utilisant une taille σ plus importante. Des résultats similaires ont été obtenus pour la scène "Sanja". Ils sont présentés à l'annexe B sur la figure B.2.

4.3.3 Changement d'échelle

Cette section montre la robustesse de notre algorithme d'appariement par rapport à un changement d'échelle. Pour ce faire, une séquence d'images a été prise pour chaque scène en faisant varier la longueur focale de notre objectif (un zoom). Les paragraphes suivants

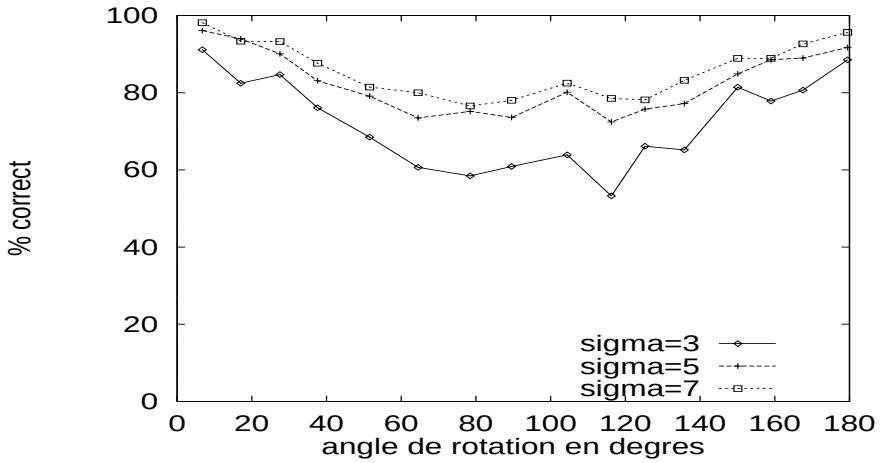

FIG. 4.6 – *Pourcentage d'appariements corrects pour la séquence rotation image et la scène "Van Gogh". Le détecteur utilisé est celui de Heitger. Les différentes courbes correspondent à différentes tailles σ de gaussienne.*

montrent la robustesse de la caractérisation lorsque l'approche multi-échelle est utilisée, puis la stabilité des invariants différentiels lors d'un changement d'échelle. Ces expériences montrent la validité de l'approche multi-échelle et permettent de déterminer l'espacement nécessaire entre deux échelles.

Approche multi-échelle

Dans ce paragraphe les invariants différentiels sont utilisés dans un cadre multi-échelle. La figure 4.7 montre les résultats obtenus pour une paire d'images de la scène "Sanja". Le changement d'échelle entre ces deux images est d'un facteur 1.5. Les croix blanches indiquent les appariements corrects et les noires les faux appariements. Le pourcentage d'appariements corrects est de 90.91%. La figure 4.8 montre les résultats pour la scène "Van Gogh" pour un facteur d'échelle de 1.5. Le pourcentage d'appariements corrects est alors de 70.31%.

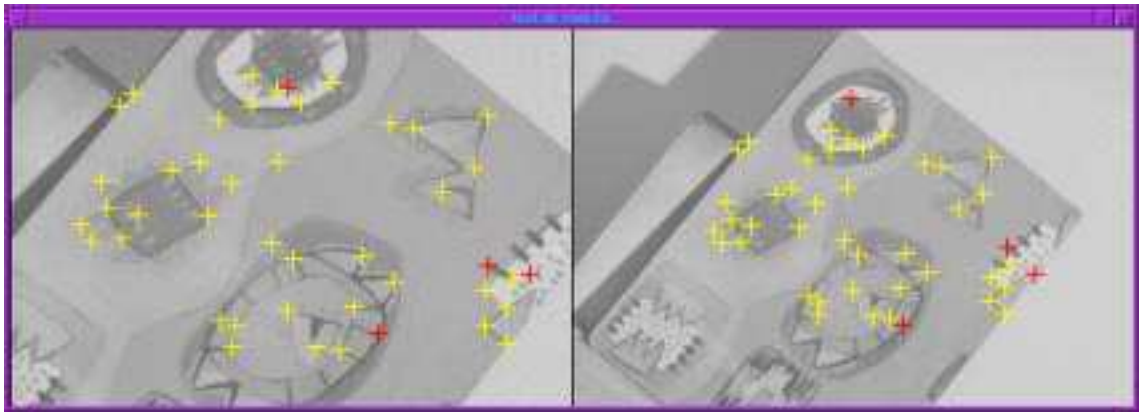

FIG. 4.7 – *Le taux d'appariements corrects est de* 90.91% *pour un facteur d'échelle de* 1.5.

La figure 4.9 trace le pourcentage d'appariements corrects en fonction du facteur

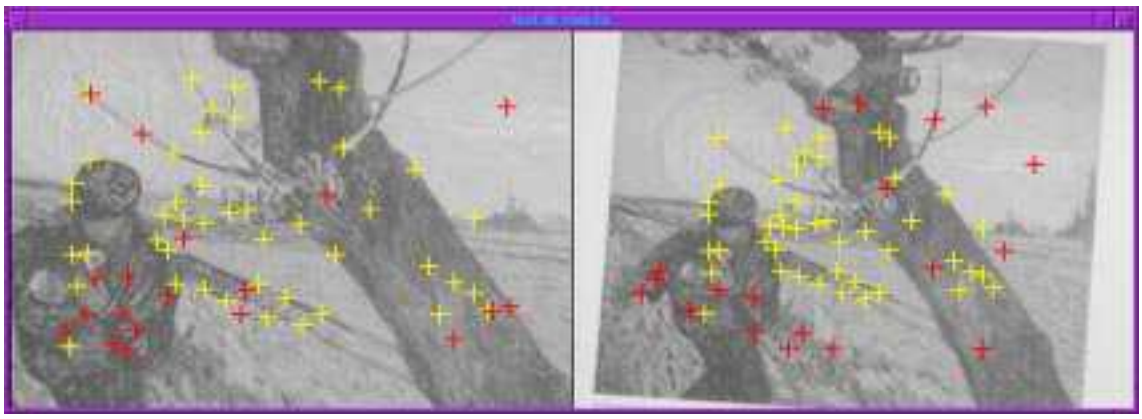

FIG. 4.8 – *Le taux d'appariements corrects est de* 70.31% *pour un facteur d'échelle de* 1.5.

d'échelle pour la scène "Van Gogh". Ce pourcentage a été calculé pour deux détecteurs différents de points d'intérêt : Heitger et HarrisPrécis (se reporter à la section 2.2). Ce pourcentage a également été évalué à partir des points précis projetés par l'homographie. La taille moyenne σ de la gaussienne utilisée est de 5 pour cette expérimentation, ce qui correspond à des tailles utilisées commençant à 2.5.

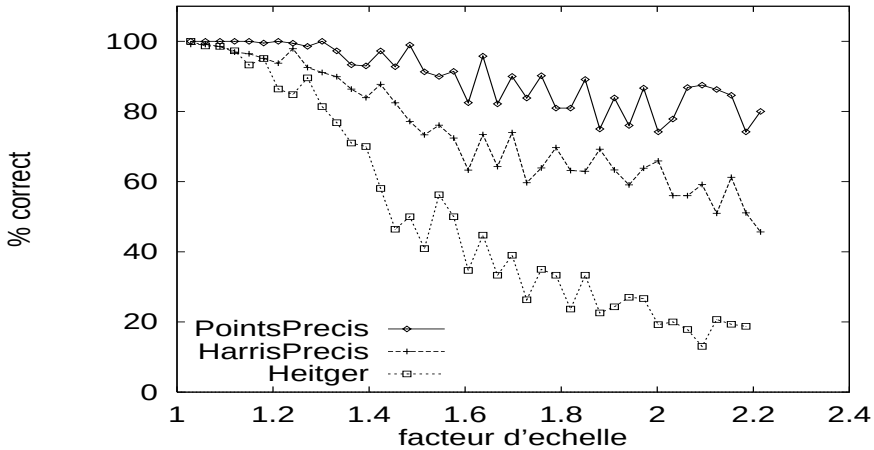

FIG. 4.9 – *Pourcentage d'appariements corrects pour la séquence changement d'échelle et la scène "Van Gogh" en utilisant une approche multi-échelle. Les trois courbes correspondent aux différents détecteurs de points d'intérêt utilisés : Heitger, HarrisPrécis et les points précis. La taille moyenne* σ *de la gaussienne utilisée est de* 5.

Cette expérience montre qu'on obtient des résultats robustes à un changement d'échelle. En se reportant à la figure 2.10, on peut également observer que les résultats obtenus dépendent plus de la répétabilité du détecteur utilisé que de la caractérisation. Des résultats similaires ont été obtenus pour la scène "Sanja" et sont présentés dans l'annexe B à la figure B.3.

Robustesse à un changement d'échelle

Dans ce paragraphe les invariants différentiels sont utilisés sans approche multi-échelle. La figure 4.10 montre que les résultats se dégradent rapidement au-dessus d'un facteur d'échelle de 1.2. Ceci confirme l'espacement choisi pour notre approche multi-échelle. Des résultats similaires ont été obtenus pour la scène "Sanja". Ils sont présentés à l'annexe B sur la figure B.2.

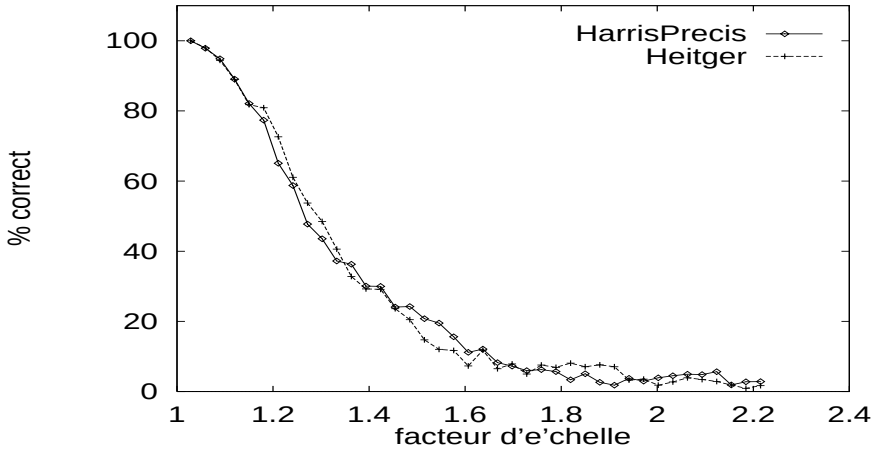

FIG. 4.10 – *Pourcentage d'appariements corrects pour la séquence changement d'échelle et la scène "Van Gogh" sans utiliser une approche multi-échelle. Les deux courbes correspondent aux différents détecteurs de point d'intérêt utilisés : Heitger et HarrisPrécis. La taille σ de la gaussienne utilisée est de 5.*

4.3.4 Changement de luminosité

Cette section présente la robustesse de notre algorithme d'appariement à un changement de luminosité. Pour ce faire nous avons pris des séquences de test pour deux types de changements de luminosité, notamment un changement uniforme et un changement complexe. Dans le cas d'un changement uniforme, uniquement l'intensité de la luminosité varie et dans le cas d'un changement complexe la source lumineuse est déplacée.

Changement uniforme de luminosité

Pour obtenir un tel changement, des séquences d'images ont été prises en changeant l'ouverture de l'objectif. Pour mesurer ce changement nous avons introduit à la section 2.5.4 la notion de "niveau de gris relatif". Cette mesure est le rapport des moyennes des niveaux de gris d'une image de la séquence par rapport à une image de référence. L'image de référence est une image " au milieu " de la séquence, c'est-à-dire une image qui est ni très sombre ni très claire. La figure 2.9 montre deux images de la séquence, une très sombre avec un niveau de gris relatif de 0.6 et une claire avec un niveau de gris relatif de 1.7.

La figure 4.11 montre les résultats obtenus en utilisant différents vecteurs d'invariants, notamment $\vec{\mathcal{V}}$, $\vec{\mathcal{V}}_T$ et $\vec{\mathcal{V}}_A$. Le détecteur utilisé est celui de HarrisPrécis. On peut observer que le vecteur $\vec{\mathcal{V}}$ est peu robuste à des changements de luminosité. Le vecteur $\vec{\mathcal{V}}_T$ est légèrement plus robuste à de tels changements. Il repose sur une modélisation par translation

des niveaux de gris. On peut observer qu'une telle modélisation n'est pas suffisante. Par contre si l'on modélise le changement de luminosité par une transformation affine, on obtient de très bons résultats, voir la courbe pour le vecteur $\vec{\mathcal{V}}_A$. En effet, le pourcentage d'appariements corrects est supérieur à 80% pour un niveau de gris relatif allant de 0.5 à 1.7. Des résultats similaires ont été obtenus pour la scène "Sanja" et sont présentés sur la figure B.5 de l'annexe B.

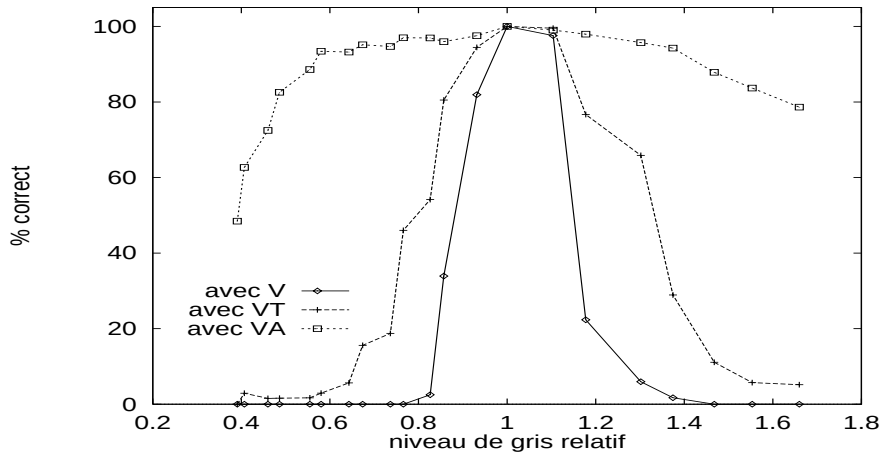

FIG. 4.11 – *Pourcentage d'appariements corrects pour la séquence changement uniforme de luminosité et la scène "Van Gogh". Les trois courbes correspondent aux différents vecteurs d'invariants utilisés :* $\vec{\mathcal{V}}$, $\vec{\mathcal{V}}_T$ *et* $\vec{\mathcal{V}}_A$.

Changement complexe de luminosité

Pour obtenir un tel changement, une séquence d'images a été prise en déplaçant la source lumineuse le long d'un cercle entre approximativement -45 degrés et 45 degrés. La figure 2.14 montre trois images de la séquence. L'image 0 est prise pour la position de la source lumineuse la plus à droite sert d'image de référence pour nos tests.

La figure 4.12 montre les résultats obtenus en utilisant différents vecteurs d'invariants, notamment $\vec{\mathcal{V}}$, $\vec{\mathcal{V}}_T$ et $\vec{\mathcal{V}}_A$. Le détecteur utilisé est celui de HarrisPrécis. On peut observer que si l'on modélise le changement de luminosité par une transformation affine, on obtient de très bons résultats, même dans le cas d'une image saturée. En effet, le pourcentage d'appariements corrects est toujours supérieur à 80%. Par contre, le pourcentage d'appariement corrects sans utiliser d'invariant à la luminosité est très mauvais dès que l'image est un peu saturée (cf. courbe avec $\vec{\mathcal{V}}$). En outre, quelque soit le vecteur utilisé plus l'image est saturée plus les résultats se dégradent.

4.3.5 Changement de point de vue

Pour obtenir un tel changement, une séquences d'images a été prise en déplaçant la caméra le long d'un cercle approximativement entre -50 degrés et 50 degrés. La figure 2.16 montre trois images de la séquence. L'image 7 pour laquelle la caméra est positionnée en face du tableau sert d'image de référence pour nos tests.

La figure 4.13 montre les résultats obtenus pour un changement de point de vue pour la scène "Van Gogh". Le détecteur utilisé est celui de HarrisPrécis. Les bons résultats

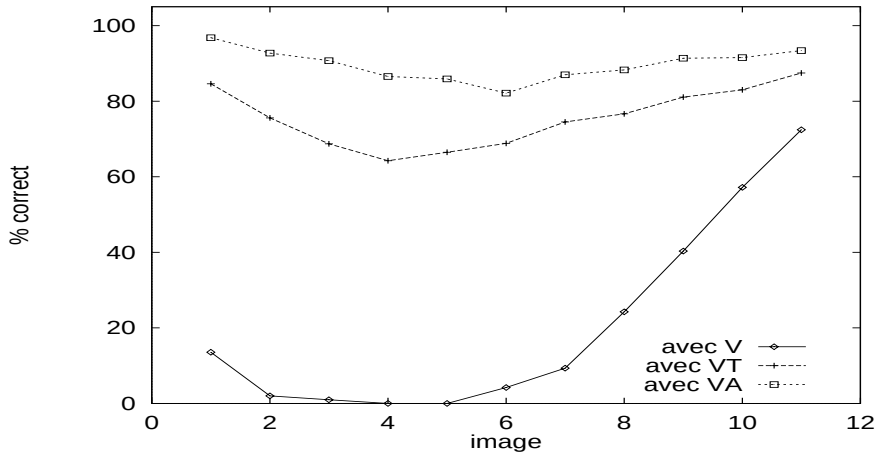

FIG. 4.12 – *Pourcentage d'appariements corrects pour la séquence changement complexe de luminosité et la scène "Van Gogh". Les trois courbes correspondent aux différents vecteurs d'invariants utilisés :* $\vec{\mathcal{V}}$, $\vec{\mathcal{V}}_{\mathcal{T}}$ *et* $\vec{\mathcal{V}}_{\mathcal{A}}$.

obtenus sont dus au fait que nos invariants sont robustes à un changement de point de vue. En effet, notre caractérisation est invariante aux similitudes image et est donc de fait quasi-invariante aux transformations perspectives, c'est-à-dire robuste à de telles transformations. Ceci explique les bons résultats obtenus même pour des changements importants de point de vue.

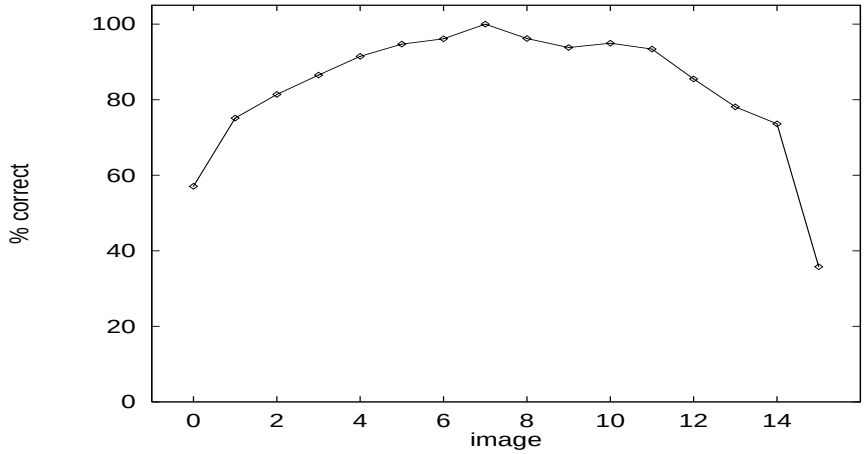

FIG. 4.13 – *Pourcentage d'appariements corrects pour la séquence changement de point de vue et la scène "Van Gogh".*

4.3.6 Bruit de la caméra

La figure 4.14 montre la robustesse de notre caractérisation au bruit de la caméra. Pour cette expérimentation, une séquence statique d'images a été prise. Cette figure montre les résultats obtenus pour la scène "Van Gogh". Le taux d'appariements corrects avoisine 100%. La figure B.6 de l'annexe B montre les résultats obtenus pour la scène "Sanja".

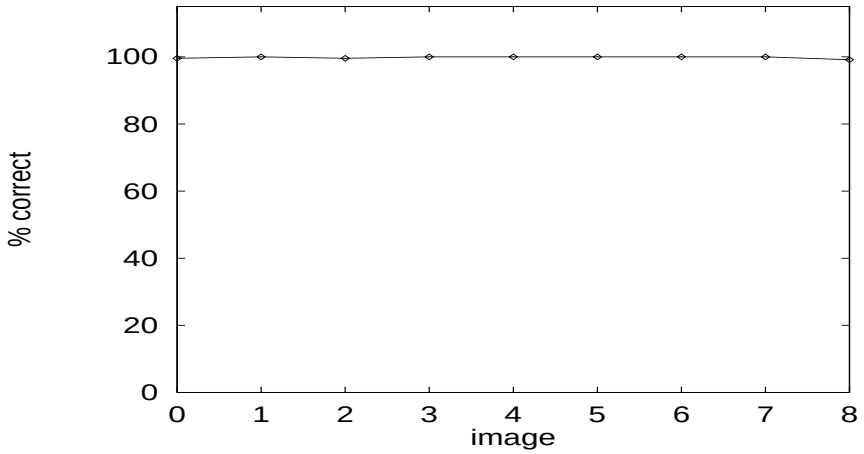

FIG. 4.14 – *Pourcentage d'appariements corrects pour la séquence bruit de la caméra et la scène "Van Gogh".*

4.3.7 Transformations complexes

Les sections précédentes ont examiné la robustesse de la caractérisation proposée séparément pour chaque transformation image et uniquement pour des scènes planes. Cette section présente les résultats obtenus pour la combinaison d'une rotation et d'un changement d'échelle ainsi que les résultats obtenus pour des scènes tridimensionnelles.

Combinaison d'une rotation et d'un changement d'échelle

Ce paragraphe présente la robustesse de notre caractérisation à la combinaison d'une rotation et d'un changement d'échelle La figure 4.15 montre les résultats obtenus pour la scène "Sanja" dans le cas d'une rotation de 108.8 degrés et d'un changement d'échelle d'un facteur de 1.5. Comme précédemment, les croix blanches indiquent les appariements corrects et les noires les appariements faux. Le taux d'appariements corrects est de 82.35%. La figure 4.16 montre les résultats obtenus pour la scène "Van Gogh" dans le cas d'une rotation de 99.8 degrés et d'un changement d'échelle d'un facteur 1.5. Le pourcentage d'appariements corrects est de 68.42%. Ces résultats prouvent que la méthode proposée permet de gérer les similitudes image.

Scènes tridimensionnelles

Ce paragraphe présente les résultats obtenus pour des scènes tridimensionnelles. L'évaluation des résultats a été réalisée en utilisant la méthode décrite à la section 4.3.1. La figure 4.17 montre les résultats obtenus sur une scène d'extérieur complexe. La transformation entre les deux images est relativement faible, cependant la présence de motifs répétitifs rend la mise en correspondance particulièrement difficile. Le pourcentage d'appariements corrects est de 84.34%. Il s'agit d'un bon résultat vue la difficulté de la scène. On peut d'ailleurs remarquer que certains des faux appariements sont dus aux motifs répétitifs.

La figure 4.18 montre les résultats obtenus sur une autre scène. La transformation entre les deux images est constituée d'une rotation scène, d'une rotation image et d'un changement d'échelle. Le pourcentage d'appariements corrects est de 80%. Là encore, les

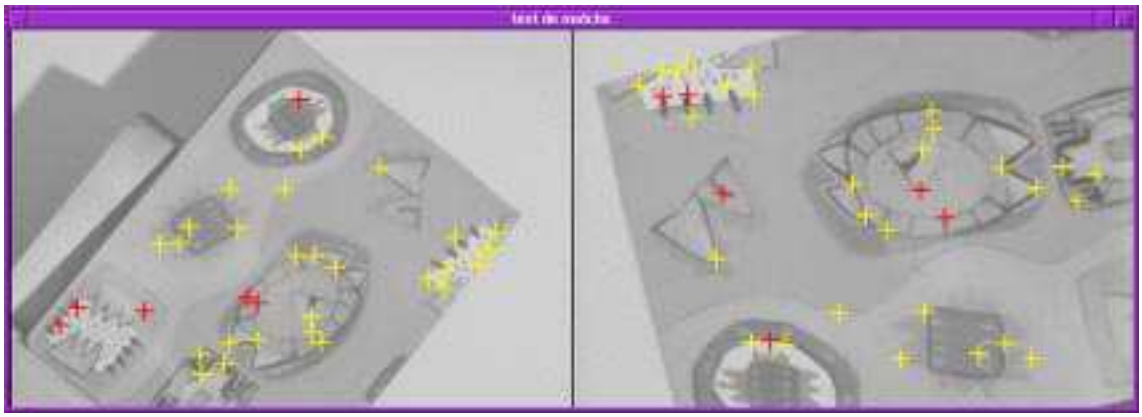

FIG. 4.15 – *Le pourcentage d'appariements corrects pour une rotation de* 108.8 *degrés et un changement d'échelle d'un facteur de* 1.5 *est de* 82.35%.

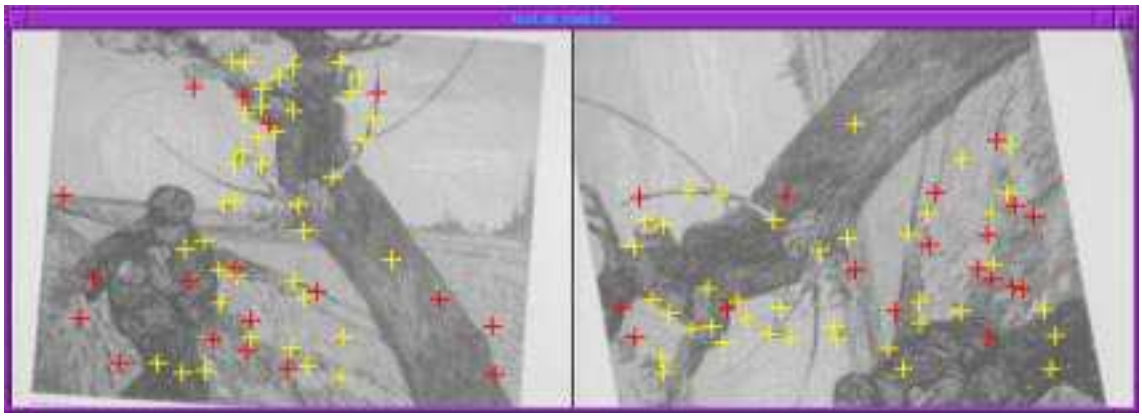

FIG. 4.16 – *Le pourcentage d'appariements corrects pour une rotation de* 98.8 *degrés et un changement d'échelle d'un facteur de* 1.5 *est de* 68.42%.

FIG. 4.17 – *Le pourcentage d'appariements corrects est de* 84.34% *sur cette scène.*

faux appariements sont essentiellement dus aux motifs répétitifs contenus dans la scène. La figure 4.19 montre le champ de déplacement calculé à partir des appariements corrects de la figure 4.18. Une méthode classique basée sur la corrélation échoue complètement sur une telle scène.

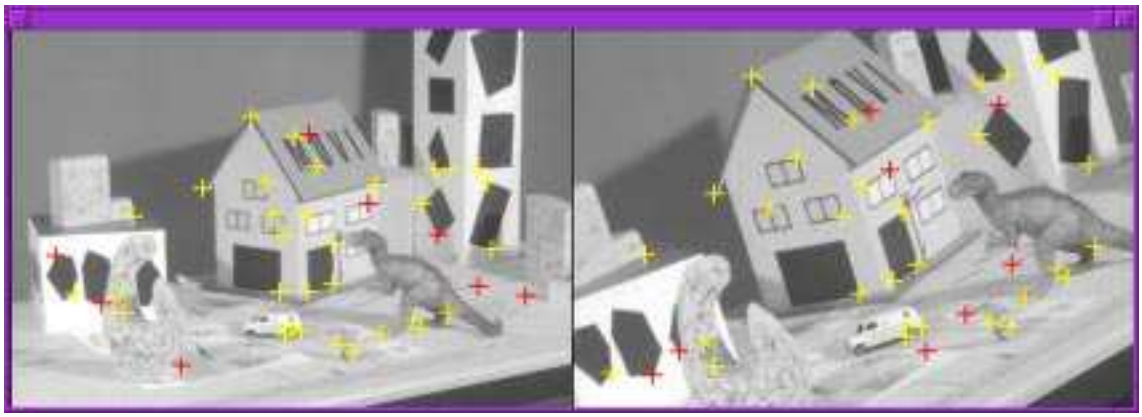

FIG. 4.18 – *Le pourcentage d'appariements corrects est de* 80.0% *sur cette scène.*

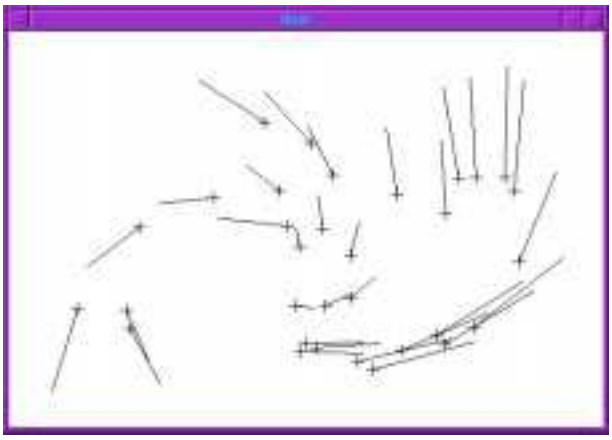

FIG. 4.19 – *Champ de déplacement pour les appariements corrects de la figure 17.*

La figure 4.20 montre les résultats obtenus pour une scène d'intérieur. La transformation entre les deux images consiste en une rotation scène, une rotation image et un changement d'échelle. Le taux d'appariements corrects est de 84.9%. La figure 4.21 montre le champ de déplacement calculé à partir des appariements évalués comme corrects. On peut voir qu'un appariement a été faussement évalué. Ceci est dû à la méthode d'évaluation qui repose sur la contrainte épipolaire (cf. section 4.3.1).

Les résultats des sections précédentes ont déjà montré l'invariance aux rotations image, la robustesse aux changements d'échelle et la robustesse à un changement important de point de vue. Les résultats obtenus sur des scènes tridimensionnelles ne font donc que confirmer ces résultats. La qualité des résultats obtenus en présence d'un changement de point de vue pour ces scènes provient du fait que les invariants aux similitudes sont des quasi-invariants aux transformations perspectives.

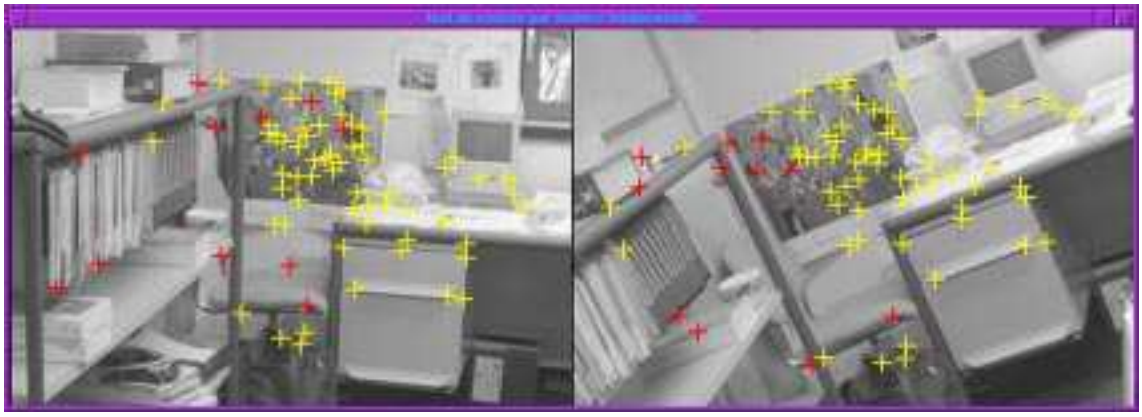

FIG. 4.20 – *Le pourcentage d'appariements corrects est de* 84.9% *sur cette scène.*

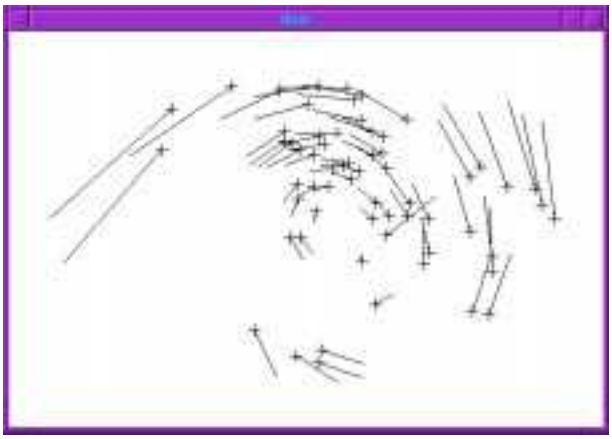

FIG. 4.21 – *Champ de déplacement pour les appariements corrects de la figure 19. Un appariement a été faussement évalué.*

4.3.8 Influence des différentes composantes du vecteur

Cette section montre l'influence des différentes composantes du vecteur de caractéristiques. Pour ce faire, deux expériences ont été réalisées. La première expérience évalue le taux d'appariement lorsque l'on utilise une seule composante du vecteur de caractéristiques. La deuxième montre le taux d'appariement obtenu en utilisant un sous-ensemble de composantes. Ces expériences permettent de montrer la discriminance de chacune des composantes, notamment des invariants de troisième ordre.

La figure 4.22 montre le taux d'appariement en fonction de la composante utilisée pour les scènes "Astérix" et "Van Gogh". Les valeurs de cette figure correspondent aux moyennes obtenues sur l'ensemble des images des séquences de bruit (cf. section 4.3.6). Les taux d'appariement sont supérieurs pour la scène "Van Gogh" ce qui s'explique par la texture de cette scène qui facilite le calcul d'appariement. En outre, pour la scène "Van Gogh" les composantes correspondant aux premières et deuxièmes dérivées permettent un meilleur taux d'appariement. Ceci est dû à la présence de texture. Dans ce cas les dérivées de troisième ordre sont moins stables que les dérivées d'ordre inférieur. En ce qui concerne la scène "Astérix", toutes les composantes permettent d'obtenir approximativement le même taux d'appariement.

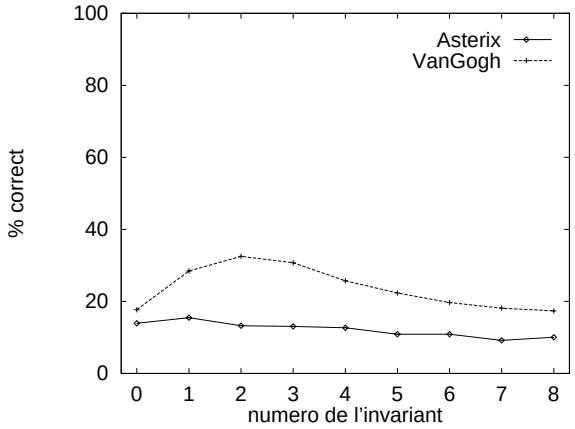

FIG. 4.22 – *Taux d'appariement en utilisant une composante du vecteur de caractéristiques.*

Le tableau 4.1 montre le taux d'appariement pour un sous-ensemble de composantes. La première colonne donne le sous-ensemble utilisé et les deux autres colonnes donnent les résultats pour les scènes "Van Gogh" et "Astérix". On peut observer que certains sous-ensembles permettent d'obtenir de très bons résultats. D'autre part, les invariants de troisième ordre apportent le même taux d'appariement que l'ensemble des invariants de premier et de deuxième ordre.

Ces deux expériences justifient l'utilisation des invariants de troisième ordre. Ces invariants s'avèrent suffisamment stables et apportent une information non négligeable.

4.4 Conclusion

Ce chapitre a montré l'avantage de notre méthode d'appariement par rapport aux méthodes classiques. L'algorithme d'appariement développé repose sur les points d'intérêt du chapitre 2 et sur la caractérisation proposée au chapitre 3. Avec un algorithme de

invariants	Astérix	VanGogh
0 ... 8	99.78	99.81
1 ... 8	99.55	99.67
0 ... 4	99.69	99.42
1 ... 4	97.78	98.91
2 ... 4	92.07	98.36
5 ... 8	97.30	99.21
0 , 1	82.38	92.99

TAB. 4.1 – *Taux d'appariement en fonction du sous-ensemble de composantes utilisées.*

mise en correspondance simple, nous obtenons un taux d'appariements corrects nettement supérieur à 50% ce qui suffit pour une estimation robuste de la transformation entre deux images. Ce taux d'appariement est dû à la discriminance de la caractérisation utilisée. De plus, cette caractérisation est invariante aux rotations images et robuste à un changement d'échelle. Ce chapitre a également prouvé que la qualité de l'appariement ne dépend pas uniquement de la caractérisation, mais aussi des points pour lesquels la caractérisation est calculée. Il est important que ces points soient répétables, car une telle répétabilité influence la stabilité de la caractérisation et donc la qualité des appariements obtenus.

Les expériences présentées ont montré de très bons résultats pour une rotation image et un changement d'échelle mais aussi pour un changement perspectif important. Ceci est dû au fait que notre caractérisation est robuste aux similitudes image et de fait robustes aux transformations perspectives. En effet, les invariants aux similitudes image sont des quasi-invariants aux transformations perspectives. Ceci a été confirmé par les résultats obtenus sur des scènes tridimensionnelles. L'utilisation d'invariants à la luminosité permet en outre d'obtenir de bons résultats pour des changements uniformes et des changements complexes de luminosité.

Chapitre 5

Recherche d'image

L'idée de ce chapitre est de considérer la recherche d'une image comme un problème d'appariement d'une image avec une base d'images. Ainsi, la recherche d'une image est l'extension de l'appariement entre deux images présenté au chapitre précédent. Le problème est donc d'apparier l'image recherchée avec chacune des images-modèles, c'est-à-dire avec l'ensemble des images stockées dans la base. Ceci nécessite l'introduction d'un mécanisme de ressemblance permettant de déterminer quelle image de la base est la plus ressemblante à l'image recherchée. Ceci est réalisé à la section 5.2 par un algorithme de vote. La robustesse de cet algorithme est augmentée par l'utilisation de contraintes semi-locales qui permettent d'accroître la discriminance de la caractérisation utilisée. Ceci s'avère nécessaire du fait du grand nombre de points contenus dans la base d'images. D'autre part, une comparaison image par image fait accroître la complexité rapidement. Afin d'éviter des temps de recherche trop importants, un mécanisme d'accès rapide est introduit à la section 5.3. Les résultats présentés à la section 5.4 montrent les performances atteintes pour l'appariement d'une image à une base d'images et le gain obtenu par rapport aux méthodes existantes.

5.1 État de l'art

Cette section présente un état de l'art des méthodes de recherche d'images, c'est-à-dire d'appariement d'une image avec un ensemble d'images modèles. Il n'est pas toujours facile de séparer les méthodes existantes dans la littérature en méthodes de recherche d'images et en méthodes de reconnaissance d'objets 3D qui seront présentées dans le chapitre suivant. Le critère que nous avons retenu pour effectuer une telle séparation est de vérifier quel type d'information a été reconnu : s'agit-il d'une image 2D ou d'un objet 3D? Les méthodes utilisant uniquement des images 2D sont référencées dans cet état de l'art. Il faut toutefois constater que les méthodes de recherche d'images peuvent parfois être étendues à la reconnaissance d'objet 3D (cf. section 6.1).

5.1.1 Recherche basée sur les données photométriques

Un premier groupe de méthodes d'appariement entre une image et une base d'images utilise l'information de luminance d'un objet, c'est à dire son aspect comme signature de l'objet. Dans ce contexte, la première idée a été d'utiliser la couleur. Swain [Swa 91] a démontré que les histogrammes de couleur peuvent être utilisés pour indexer et mettre en correspondance des objets. Le plus gros défaut de cette approche est son manque de robustesse vis-à-vis de changements de luminosité. Plusieurs auteurs ont accru les performances de la technique initiale de mise en correspondance par histogramme de couleur en introduisant des mesures plus ou moins sensibles à des changements de luminosité. Ainsi, Funt [Fun 95] a proposé d'utiliser la distribution de rapports de couleur et a démontré que ces rapports fournissent une constante de couleur pour un objet. Slater [Sla 96] a démontré que les moments de la distribution des couleurs sont invariants à un changement de luminosité dans l'hypothèse d'un modèle linéaire de réflexion lumineuse. Nayar [Nay 93] et Nagao [Nag 95], quant à eux ont utilisé des invariants photométriques basés sur des rapports de réflection. Enfin, il est également possible d'utiliser des histogrammes de filtres locaux comme l'a fait Schiele [Sch 96].

Une autre approche de la reconnaissance d'image à partir d'informations photométriques est celle de Turk [Tur 91]. Sa méthode utilise une grande collection d'images qui est décomposée en composantes principales. Les composantes correspondant aux plus grandes valeurs propres représentent des formes génériques. Turk a utilisé cette méthode pour reconnaître des visages. Cette approche a été appliquée par Murase [Mur 95] pour reconnaître des images d'objets quelconques. Les avantages de cette méthode sont sa rapidité, sa généralité et sa robustesse à de petites occultations. Par contre, elle nécessite de centrer les images et elle n'est pas robuste aux rotations images ni aux occultations importantes.

Des approches assez similaires à celle suivie dans notre travail sont celles de Rao [Rao 95], Wu [Wu 95] et Lades [Lad 93]. Ces auteurs ont utilisé des mesures locales basées sur l'image de niveaux de gris. Le signal est caractérisé localement par des filtres ajustables dans le cas de Rao et par des transformées de Gabor dans le cas de Wu et de Lades. Ces filtres sont calculés sur une grille qui est centrée sur l'objet par un calcul simple du centre de l'objet. Rao utilise une grille circulaire tandis que Wu et Lades utilisent une grille rectangulaire. Le positionnement de la grille est difficile à réaliser dès que l'objet est présenté devant un arrière-plan complexe. En outre, ces méthodes ne permettent pas de reconnaître un objet à partir d'images d'une portion de cet objet. Ceci provient du fait que la grille ne peut pas être positionnée si uniquement une partie de l'objet est donnée. D'autre part, l'utilisation d'une grille implique que certains points de celle-ci sont peu voire aucunement représentatifs de l'objet. Rao [Rao 95] propose donc d'utiliser des vecteurs de caractéristiques comportant jusqu'à 45 composantes. Pour ce faire, cette caractérisation est calculée dans un contexte multi-échelle. Pour reconnaître un objet en présence d'un changement d'échelle, il faut translater les vecteurs de caractéristiques afin de trouver les sous-ensembles de composantes qui se correspondent. La mise en correspondance est alors faite sur une partie de la caractérisation et les résultats se dégradent de façon significative.

5.1.2 Recherche basée sur des données géométriques

Un deuxième groupe de méthodes d'appariement entre une image et une base d'images utilise des données géométriques telles que segments, jonctions et ellipses. De telles don-

nées sont extraites préalablement des images et l'appariement est effectué en utilisant uniquement ces données. Ces méthodes d'appariement reposent donc sur des donnés symboliques même si elles utilisent ces données pour calculer des grandeurs numériques. Un certain nombre d'approches sont basées sur le paradigme suivant : calcul d'hypothèse et vérification. Durant la première phase, des caractéristiques sont extraites à partir de l'image à reconnaître, puis elles sont associées aux caractéristiques des modèles 2D contenus dans une base. La recherche exhaustive de tous les modèles existant dans la base engendre un coût de calcul polynômial. La contribution majeure de différents systèmes de reconnaissance a été de contrôler et de diminuer la complexité de la phase d'appariement. Par exemple, Ayache et Faugeras [Aya 86] utilisent une évaluation récursive d'hypothèses. Lamdan [Lam 88] a proposé d'utiliser des méthodes d'indexation et de hachage pour obtenir une accélération significative. Dans le cas de l'indexation, la mise en correspondance des caractéristiques et la recherche d'un modèle de la base sont remplacés par un mécanisme de " look-up table ". Dans un contexte similaire, le groupe d'Oxford a utilisé des invariants projectifs comme élément d'indexation (voir par exemple [Rot 93]). Dans le cas d'objet 2D de tels invariants peuvent être calculés pour n'importe quel objet.

D'autres méthodes de recherche d'image sont basées sur la transformée de Hough. Elles choisissent le modèle en recherchant un point d'accumulation dans l'espace des transformations (cf. par exemple Ballard [Bal 81]). Grimson [Gri 90] a toutefois démontré qu'une telle approche est peu robuste au bruit de l'image. En effet, en présence d'un tel bruit, il n'est plus possible de distinguer entre deux modèles différents. Pour remédier à ce problème, Gros [Gro 95] utilise des invariants aux similitudes et vote uniquement dans l'espace de Hough si ces invariants se correspondent. Ceci réduit le nombre de votes dans l'espace de Hough et rend ainsi la distinction des différents modèles possible.

5.2 Algorithme de recherche

5.2.1 Principe de la recherche

En conclusion de l'état de l'art précédent, les méthodes basées sur des données géométriques permettent de traiter uniquement des objets relativement simples. Dans le cas d'objets plus complexes, le calcul de grandeurs géométriques devient instable. De plus, ces méthodes sont peu discriminantes, car elles sont basées sur des données symboliques. Toutefois, ces méthodes sont locales et donc robustes aux occultations. De plus, elles sont invariantes aux différentes transformations.

Les méthodes basées sur les données photométriques sont par nature plus discriminantes car moins symboliques. Elles sont en effet plus proches et donc plus représentatives de l'objet considéré. Ces méthodes permettent de reconnaître des objets qui ne peuvent pas être traités par les méthodes géométriques. Le grand inconvénient des méthodes proposées jusqu'à présent dans la littérature est qu'elles sont toutes globales. Cette globalité signifie que l'information est calculée sur toute l'image contrairement aux méthodes locales qui reposent sur un ensemble d'informations calculé à plusieurs endroits de l'image. Le désavantage des méthodes globales par rapport aux méthodes locales est leur manque de robustesse par rapport à des perturbations locales de l'image. Les méthodes globales sont donc par exemple peu robuste à des occultations. De plus, les méthodes basées sur les données photométriques sont peu invariantes aux transformations images couramment observées.

Dans la suite nous présentons une solution robuste qui remédie à ces deux défauts des méthodes basées sur les données photométriques. Cette solution est une extension directe de la méthode d'appariement présentée au chapitre précédent. Cette méthode est locale et invariante aux transformations image usuelles. L'idée est de considérer la recherche d'une image comme un problème d'appariement d'une image à une base d'images. Cependant, étant donnés une image et un ensemble d'images, le problème n'est plus de savoir quels points se correspondent mais de retrouver l'image la plus semblable. Les sections suivantes vont donc montrer comment utiliser une procédure d'appariement pour déterminer la ressemblance de deux images. Ceci est illustré sur la figure 5.1 suivante. Étant capable de dire si deux images sont très semblables, un peu semblables ou différentes, il est alors possible d'identifier les deux images les plus ressemblantes.

5.2.2 Structure de la base d'images

Une base d'images contient un ensemble $\{M_k\}$ de modèles. Chaque modèle M_k est défini par un ensemble de vecteur d'invariants $\{\vec{\mathcal{V}}_{k,j}\}$ calculés aux endroits où des points d'intérêt ont été détectés pour les images du modèle. Durant la phase d'enregistrement dans la base, chaque vecteur $\vec{\mathcal{V}}_{k,j}$ est ajouté dans la base avec une référence explicite au numéro k du modèle pour lequel il a été calculé. Formellement, la base la plus simple est une table de n-uplets $(\vec{\mathcal{V}}_{k,j}, k)$ (cf. figure 5.2).

En outre, dans la base que nous venons de présenter, les éléments peuvent être insérés de façon incrémentale. Aucun ré-arrangement de la base n'est nécessaire après une insertion. La mise à jour de la base, ou simplement sa construction sont donc très rapides.

5.2.3 Mesure de ressemblance

Mettre en correspondance une image à une base d'images consiste à retrouver le modèle qui correspond à une image I donnée. Pour cette image, un ensemble de vecteurs d'invariants $\{\vec{\mathcal{V}}_j\}$ est calculé aux endroits où ont été extraits des points d'intérêt. Ces vecteurs sont alors comparés aux vecteurs $\vec{\mathcal{V}}_{k,l}$ de la base. Cette comparaison est faite en calculant la distance d_M de Mahalanobis entre $\vec{\mathcal{V}}_j$ et $\vec{\mathcal{V}}_{k,l}$: $d_M(\vec{\mathcal{V}}_j, \vec{\mathcal{V}}_{k,l}) = d_{j,k,l}\ \forall(j,k,l)$. Pour chaque vecteur $\vec{\mathcal{V}}_j$, les modèles $M_{k\star}$ pour lesquels la distance $d_{j,k\star,l}$ est inférieure à un seuil donné t sont retenus. Ce seuil est défini à partir de la distribution du χ^2 observée en calculant la matrice de covariance des vecteurs d'invariants (cf. section 4.2.2). Lorsque la distance $d_{j,k\star,l}$ est inférieure à ce seuil, $M_{k\star}$ est un modèle probable pour le vecteur de caractéristiques $\vec{\mathcal{V}}_j$. On dit aussi que le modèle $M_{k\star}$ est " sélectionné ". Étant donné un vecteur de caractéristiques, plusieurs modèles peuvent lui correspondre du fait de bruit ou d'ambiguïtés. Afin d'autoriser une telle incertitude quant à l'origine d'un vecteur de caractéristiques, l'ensemble $\{k_j^\star\}$ des modèles probables n'est pas réduit à un seul élément. Ainsi, un vecteur de caractéristiques peut être mis en correspondance avec plusieurs modèles. Le modèle correspondant à une image émerge du fait qu'un modèle correspond plus souvent que les autres aux vecteurs de caractéristiques de cette image.

De façon similaire à la transformée de Hough [Sha 78], l'idée d'un algorithme de vote est de sommer le nombre de fois qu'un modèle correspond à un vecteur de caractéristiques. Aussi, chaque fois qu'un modèle M_k est sélectionné, une table de votes T est mise à jour de manière à ce que la valeur $T(k)$ soit incrémentée. Notons qu'un point peut sélectionner n'importe quel modèle mais qu'il ne peut sélectionner qu'une seule fois un modèle donné.

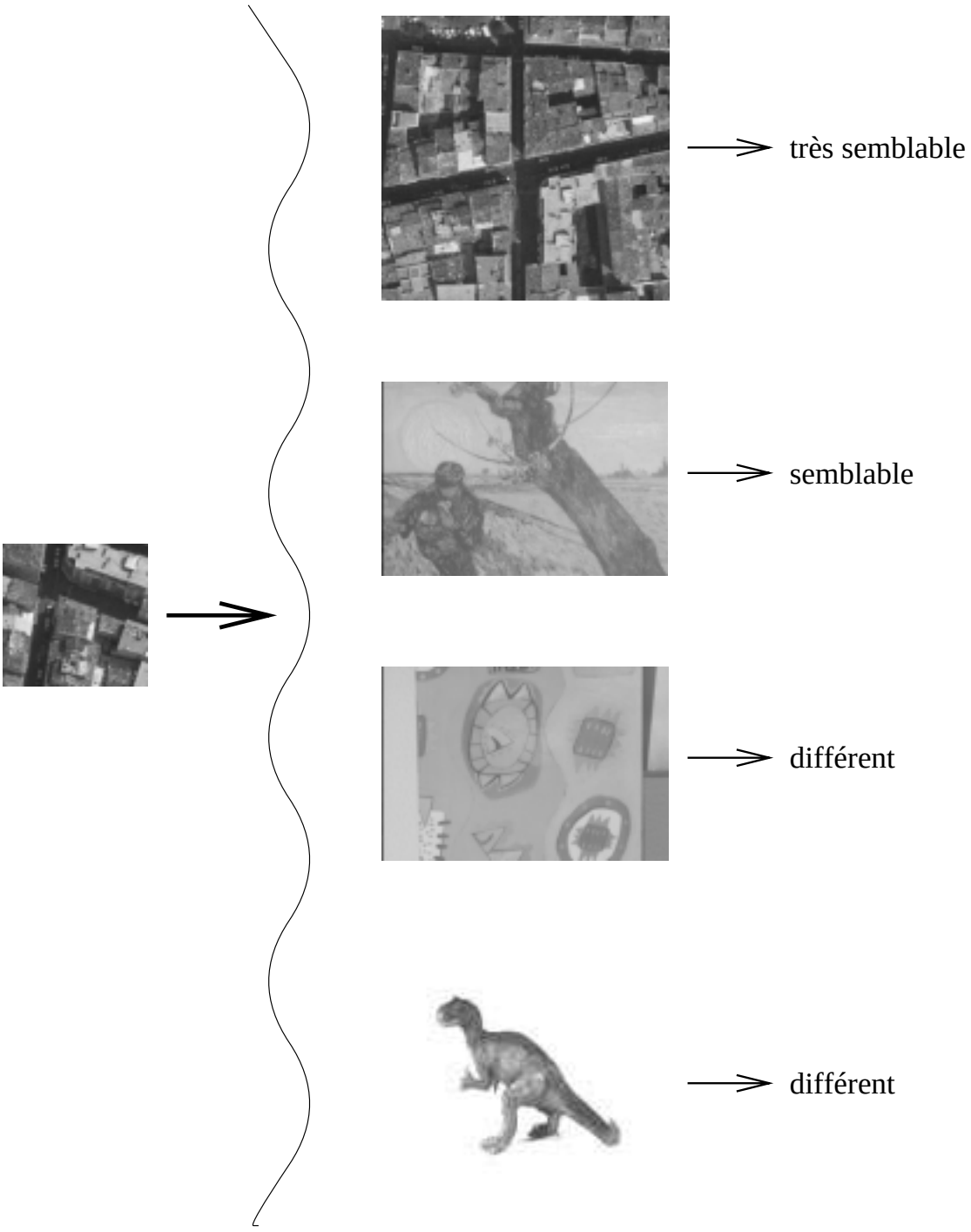

FIG. 5.1 – *Principe de la recherche d'image.*

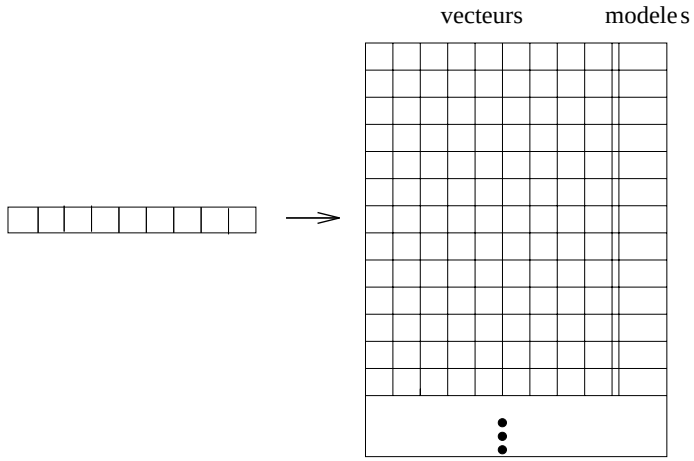

FIG. 5.2 – *Recherche d'un vecteur dans la base d'images.*

Le modèle le plus souvent sélectionné est considéré comme le meilleur représentant de l'image : l'image représente le modèle $M_{\hat{k}}$ pour lequel

$$\hat{k} = \operatorname*{argmax}_k T(k)$$

La figure 5.3 montre le contenu de la table T de votes sous la forme d'un histogramme. Pour cet exemple, la base contenait 100 images. L'image de numéro 0 a été correctement reconnue. Cependant, d'autres images ont obtenu un score de sélection du même ordre de grandeur.

Pour rendre le modèle à reconnaître plus distinct, on utilise des contraintes semi-locales. La figure 5.4 montre le contenu de la table T de votes sous la forme d'un histogramme lorsque la contrainte de voisinage est utilisée. Le modèle reconnu apparaît alors distinctement. Si l'on ajoute en plus la contrainte géométrique, le modèle reconnu apparaît encore plus distinctement (cf. figure 5.5).

Chacune des contraintes présentées diminue le nombre d'ambiguïtés lors de l'appariement. Ceci prouve le gain en robustesse obtenu grâce à l'utilisation de ces contraintes. En conséquence, le seuil t initial utilisé pour déterminer les appariements initiaux a un rôle beaucoup moins important.

5.2.4 Adaptation de l'approche multi-échelle

Il est aisé d'étendre l'algorithme présenté à la section précédente dans un contexte multi-échelle. Les résultats théoriques de la section 3.3.2 ont montré comment adapter le calcul des invariants à un changement d'échelle. Il existe alors trois possibilités pour mettre en œuvre une approche multi-échelle :

1. Étant donné une image à reconnaître, les invariants de cette image sont calculés à différentes échelles.

2. Les invariants sont calculés à différentes échelles pour tous les modèles de la base.

3. Les invariants sont calculés à différentes échelles à la fois pour l'image et pour l'ensemble des modèles de la base.

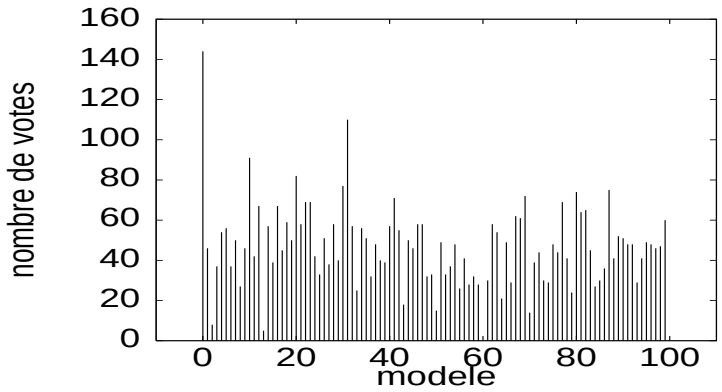

FIG. 5.3 – *Résultats de l'algorithme de vote.*

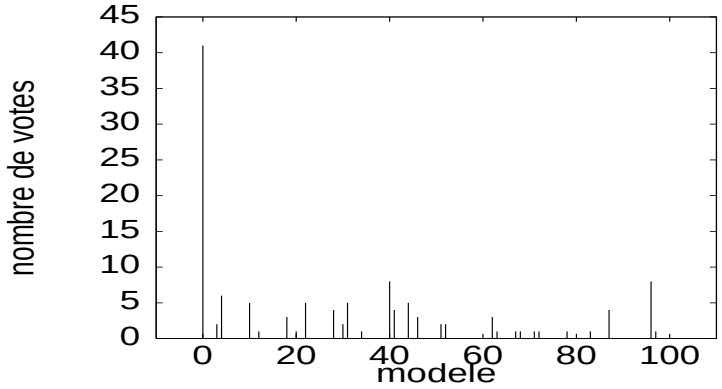

FIG. 5.4 – *Résultats de l'algorithme de vote lorsque la contrainte de voisinage est utilisée. Cette figure est directement comparable à la figure 5.3. L'objet reconnu apparaît distinctement.*

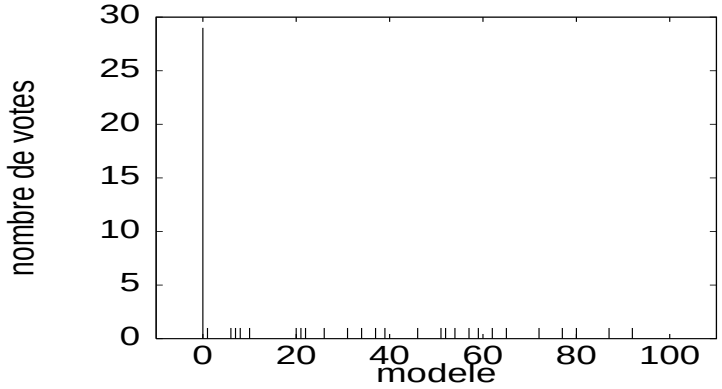

FIG. 5.5 – *Résultats de l'algorithme de vote lorsque les contraintes de voisinage et géométrique sont utilisées. Cette figure est directement comparable aux figures 5.3 et 5.4*

La dernière possibilité est la plus robuste puisqu'elle permet la réalisation d'une phase de vérification au travers des échelles considérées. Cependant, elle est très lourde à mettre en œuvre autant par le temps d'exécution qu'elle nécessite que par la mémoire qu'elle requiert. La première possibilité est la moins coûteuse en ressource. C'est celle qui a été retenue dans le cadre de notre approche.

Il faut cependant noter qu'elle peut introduire des erreurs par ambiguïtés. Cela survient par exemple lorsque les invariants calculés à une échelle inadéquate votent pour un faux modèle. Ce problème est toutefois résolu par l'utilisation des contraintes semi-locales présentées à la section 4.2.4 précédente. En effet, ces contraintes semi-locales incluent implicitement une contrainte d'échelle puisque les invariants calculés en un point et les invariants de son voisinage sont calculés à la même échelle. Ainsi, si deux points ainsi que leur voisinage respectif se correspondent, alors la contrainte d'échelle est vérifiée. Les résultats expérimentaux présentés en section 5.4 montrent qu'une approche multi-échelle est effective dès que les contraintes semi-locales sont utilisées.

5.3 Indexation

Le temps pour apparier une image à une base d'images dépend de façon linéaire du nombre d'images de la base. Dans le cas où l'on veut pouvoir traiter des base avec plus de mille images, la recherche est donc très lente. Il est donc nécessaire de développer un mécanisme d'indexation rapide.

5.3.1 Changement de base

La distance de Mahalanobis est fort peu pratique pour mettre en œuvre une technique de mise en correspondance rapide par indexation en utilisant par exemple une table multi-dimensionnelle de hachage. Ceci est dû au fait que les composantes d'un vecteur de caractéristiques peuvent être corrélées. En fait, l'ensemble des vecteurs qui peuvent correspondre à un vecteur donné se situe dans un ellipsoïde à neuf dimensions, centré autour de ce vecteur. En outre, les axes principaux de cet ellipsoïde sont en général non parallèles aux axes canoniques des vecteurs. Ceci est illustré dans la cas de deux dimensions à gauche sur la figure 5.6.

FIG. 5.6 – *Lorsque l'on utilise la distance de Mahalanobis, l'ensemble des vecteurs qui peuvent correspondre à un vecteur donné se situe dans un ellipsoïde (figure de gauche). Après changement de base, il est possible d'utiliser la distance euclidienne. L'ensemble des vecteurs qui peuvent correspondre à un vecteur donné se trouve dans une sphère facilement englobée dans un cube (figure de droite).*

Il existe toutefois un changement de base qui rend possible l'utilisation de la distance euclidienne habituelle d_E pour comparer deux vecteurs d'invariants. Puisque la matrice de

covariance est réelle symétrique et semi-définie positive, il est possible de la décomposer de la manière suivante :

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

où P est orthogonale et D positive diagonale. De là, il se déduit que :

$$d_M(\vec{a}, \vec{b})^2 =$$

$$(\vec{b} - \vec{a})^T P^T \sqrt{D}\sqrt{D} P(\vec{b} - \vec{a}) =$$

$$\left[\sqrt{D}P(\vec{b} - \vec{a})\right]^T \left[\sqrt{D}P(\vec{b} - \vec{a})\right] =$$

$$\left[\sqrt{D}P\vec{b} - \sqrt{D}P\vec{a}\right]^T \left[\sqrt{D}P\vec{b} - \sqrt{D}P\vec{a}\right] =$$

$$d_E(\sqrt{D}P\vec{a}, \sqrt{D}P\vec{b})^2$$

Calculer la distance de Mahalanobis entre deux vecteurs d'invariants est donc équivalent à transformer ces deux vecteurs en les multipliant par la matrice $\sqrt{D}P$ puis à calculer la distance Euclidienne entre les deux vecteurs transformés. L'intérêt de ce changement de base est illustré sur la figure 5.6 : après changement de base, l'ensemble des vecteurs qui peuvent correspondre à un vecteur donné se situe dans une "sphère" à neuf dimensions centrée autour de ce vecteur et facilement englobée par un "cube".

5.3.2 Table de hachage multi-dimensionnelle

La complexité de l'algorithme de vote est de l'ordre $J \times K \times L$ pour calculer toutes les distances $d_{j,k,l}$ plus K pour trouver le modèle le plus vraisemblable, où J est le nombre de points de l'image recherchée, K le nombre d'images de la base et L le nombre moyen de points détectés par image de la base. Ce coût correspond au calcul de la similarité entre la caractérisation d'une image et l'ensemble des modèles de la base. Cette complexité peut être réduite en organisant la base sous la forme d'une table de hachage multi-dimensionnelle (cf. [Wol 90]). Pour ce faire, une fonction de $\mathbb{R}^9$ dans $\mathbb{R}^9$ est utilisée pour regrouper en catégories les invariants. Ceci revient à ne calculer la similarité de l'image qu'avec des modèles plausibles. Ceci est rendu possible par l'utilisation d'une distance. De plus, la section précédente a montré que deux vecteurs d'invariants peuvent être comparés en utilisant la distance Euclidienne. Ceci simplifie grandement la mise en œuvre d'une technique d'indexation.

Étant donné un vecteur d'invariants $\vec{\mathcal{V}}$, il est possible de définir le voisinage dans lequel se situent tous les modèles plausibles de ce vecteur. La figure 5.6 précédente illustre cette idée. Une technique d'indexation en permet une mise en œuvre en ordonnant les vecteurs d'invariants dans une table multi-dimensionnelle de hachage (cf. figure 5.7). Chaque dimension de cette table indexe une composante du vecteur de caractéristiques. La table réalise ainsi une partition de l'espace euclidien de dimension neuf. La mise en œuvre de cette table soulève deux problèmes majeurs qui sont la granularité du partitionnement et la dimension de la table. La granularité permet de répartir les valeurs d'une composante en plusieurs sous-ensembles. Elle réalise une partition (au sens mathématique) d'un axe défini par une composante. Elle correspond à l'espacement entre deux cases dans chaque tableau de la figure 5.7. La dimension de la table de hachage contrôle le nombre de composantes utilisées pour le partitionnement. C'est le nombre de tableaux de la figure 5.7.

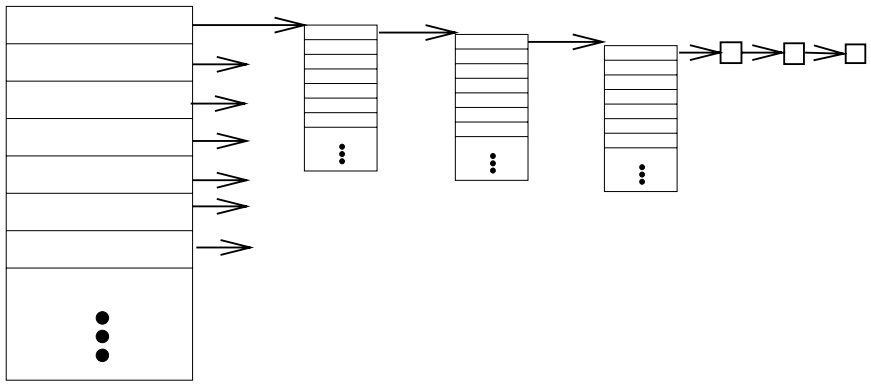

FIG. 5.7 – *Table de hachage multi-dimensionnelle. Pour cette illustration le nombre de dimensions est limité à 4.*

Nous avons observé qu'une dimensionnalité élevée avec une granularité grossière à chaque niveau accélère plus la recherche qu'une faible dimensionnalité avec une granularité fine à chaque niveau. Ceci peut facilement être expliqué par le fait qu'une dimensionnalité élevée permet une meilleure différenciation spatiale des vecteurs de caractéristiques. Dans notre mise en œuvre, nous avons donc arrêté le partitionnement dès que le nombre de points caractérisés pour une partition donnée est en dessous d'un seuil critique. Nous avons observé un optimum en temps de reconnaissance pour une seuil de trois points. En résumé, cette table de hachage réalise un " K-tree " sur l'espace Euclidien de dimension k. Il faut noter que la granularité n'est pas constante pour tous les axes mais définies pour chaque composante en fonction de la distribution de celle-ci.

5.4 Expérimentation

Cette section présente les résultats expérimentaux obtenus lors du calcul de mise en correspondance entre une image et une base d'images. Les expériences ont été menées sur une base d'images contenant plus de 1000 images Ces images sont de types différents, tels que des images de tableaux de maître, de scènes aériennes et d'objets tridimensionnels. La section 5.4.1 présente plus en détail cette base et les images tests utilisées. Le taux de reconnaissance dépasse les 99% pour des images tests prises sous différentes conditions. Il faut préciser qu'aucune des images tests n'est contenue dans la base. Pour les images contenues dans la base la reconnaissance est exempte de toute erreur. La section 5.4.2 illustre la recherche d'images et les expérimentations présentées en section 5.4.3 montrent la robustesse de la méthode vis-à-vis d'une rotation image, d'un changement d'échelle, de luminosité, de point de vue ainsi que dans le cas d'une visibilité partielle et de fouillis.

5.4.1 Cadre d'évaluation

Base utilisée

La base utilisée pour les expérimentations présentées dans cette section contient 1020 images. La figure 5.8 montre quelques images de cette base. Ces images se répartissent en 200 tableaux de maître, 100 scènes aériennes et 720 images d'objets tridimensionnels

dont 360 images de la base "Columbia". Certains des tableaux de maître sont présentés sur la première ligne de la figure 5.8, d'autres à la figure C.1 de l'annexe C. On peut voir quelques scènes aériennes sur la deuxième ligne de la figure 5.8 et aussi à la figure C.2 de l'annexe C. Ces images aériennes ont été fournies par la société Istar et en reste leur propriété. Les images d'objets tridimensionnels comprennent les images de la base "Columbia". Ces images sont présentées sur la troisième ligne. Des objets tridimensionnels propres à notre laboratoire sont présentés sur la quatrième ligne. Ces images présentent une grande variété. Cependant, si l'on considère les tableaux de maître (cf. figure C.1 de l'annexe C) ou encore les images aériennes (cf. figure C.2 de l'annexe C), on peut observer la grande similarité de certaines images. Cette similarité induit des ambiguïtés qui sont de très bons tests pour la robustesse et la discriminance de la méthode proposée. De plus, dans le cas des images aériennes, il faut noter la finesse des détails sur lesquels il faut s'accrocher pour différencier deux images.

Dans le cas d'objets planaires, un objet est représenté dans la base par une seule image. Ceci est également valable pour les objets quasiment plats, comme c'est le cas des images aériennes qui correspondent à des projections para-perspectives. C'est-à-dire que la profondeur relative de l'objet est (très) faible par rapport à la distance d'observation. En revanche, dans le cas des objets tridimensionnels, il est nécessaire de représenter ces objets par plusieurs images correspondant à différents points de vue. La question se pose alors de savoir combien d'images sont nécessaires pour représenter de façon complète un objet donné. Ce point sera abordé à la section 6.2 du chapitre suivant.

Images tests

Pour l'évaluation présentée aux sections suivantes, les images test sont des images réelles non stockées dans la base d'images. Par rapport aux images de la base, il s'agit d'images acquises après avoir fait subi à la caméra une rotation, un changement de longueur focale, d'ouverture ou de point de vue. Pour certaines images tests, la source lumineuse a également été déplacée. Ainsi les images tests présentent par rapport aux images de la base des rotations image, des changements d'échelle, de luminosité et de point de vue. La combinaison de ces transformations conduit en outre à des déformations de l'image plus complexes. D'autres images tests correspondent à l'observation d'une partie de la scène pour laquelle une image est stockée dans la base. Toutes les images test utilisées dans ce chapitre correspondent à des scènes planes ou à des transformations para-perspectives. C'est-à-dire que la distance d'observation est grande par rapport à la profondeur relative de la scène observée. En ce qui concerne les résultats pour des objets 3D le lecteur doit se rapporter au prochain chapitre.

5.4.2 Illustration de la recherche d'images

Quelques exemples de recherche

Les exemples suivants illustrent les conditions sous lesquelles on peut reconnaître correctement une image de la base. Pour les figures 5.9 et 5.10, l'image de droite a été correctement reconnue quelle que soit l'image de gauche considérée pour la recherche.

La figure 5.9 présente la reconnaissance d'un tableau de maître en présence d'une rotation image et/ou d'un changement d'échelle. Cette figure montre qu'il est également possible de reconnaître le modèle à partir de l'image d'un fragment du tableau.

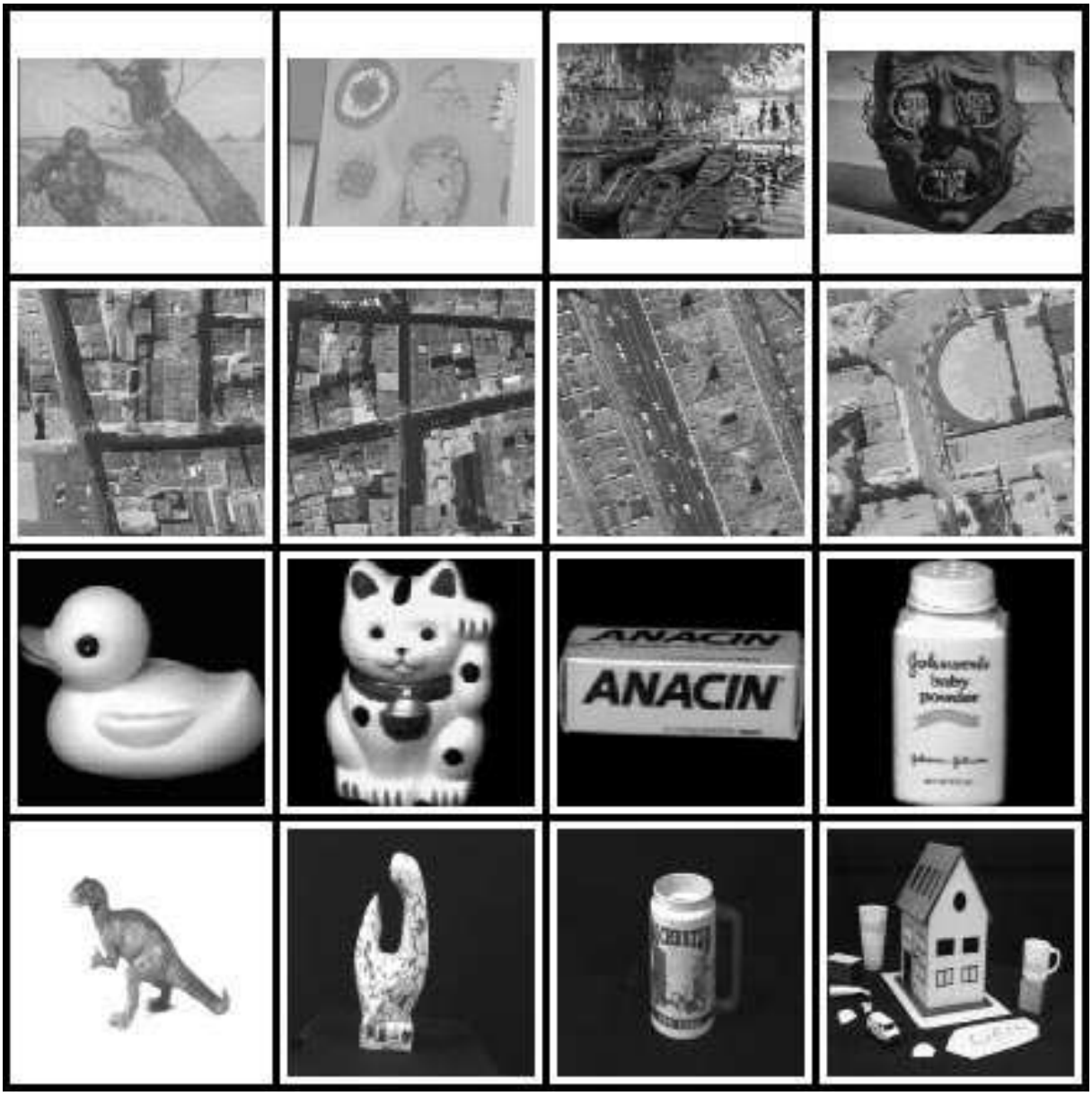

FIG. 5.8 – *Quelques images de notre base d'images. Cette base contient* 1020 *images.*

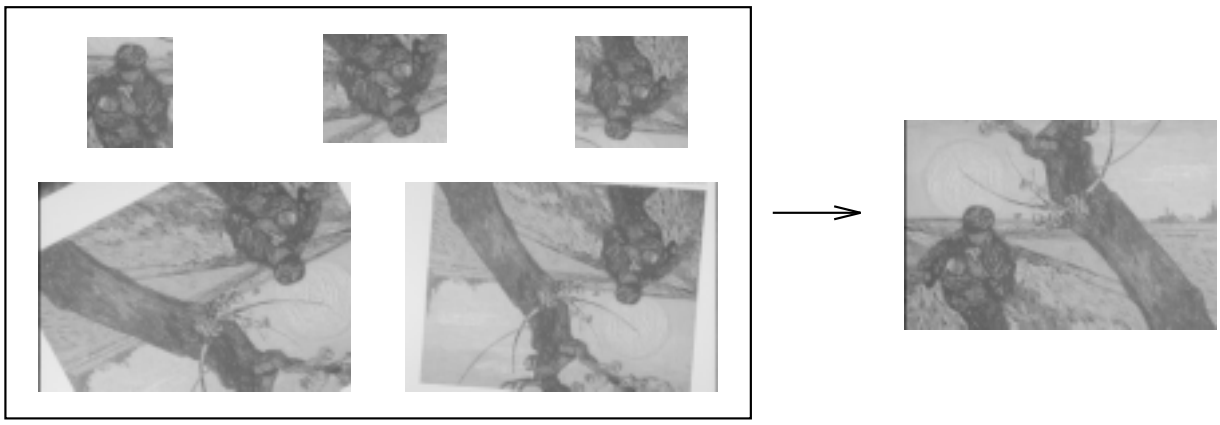

FIG. 5.9 – *Exemple de reconnaissance. L'image de droite a été correctement reconnue quelle que soit l'image de gauche utilisée pour la recherche.*

La figure 5.10 montre la reconnaissance d'une image aérienne en présence d'une rotation image et/ou d'un changement d'échelle. La reconnaissance est également possible en utilisant une partie d'image. En outre, un changement de point de vue engendre des déformations perspectives perceptibles d'autant plus fortement que la scène n'est qu'approximativement plane. On peut voir que les immeubles apparaissent de manières différentes. De même, l'écart temporel entre les deux prises d'images fait que des voitures se sont déplacé. La robustesse de la méthode a permis de résister à ces perturbations. Pour rappel, une seule image a été stockée dans la base.

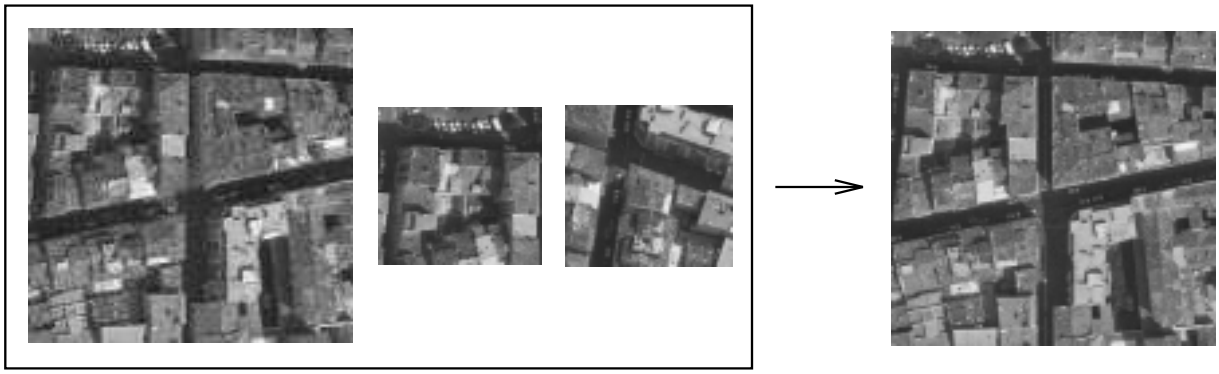

FIG. 5.10 – *Exemple de reconnaissance dans le cas d'images aériennes. L'image de droite a été correctement reconnue quelle que soit l'image de gauche considérée (images fournies par Istar).*

Points sélectionnés

Les contraintes utilisées permettent de sélectionner uniquement les points discriminants d'une image et permettent d'éliminer les points dus au bruit. Elles diminuent le nombre des fausses correspondances et réduisent le nombre total de mises en correspondance.

La figure 5.11 montre les points utilisés durant la phase d'appariement pour une rotation image. La rotation entre les deux images est de 152 degrés. L'image de gauche montre l'image recherchée et les points détectés qui ont été mis en jeu dans le processus

d'appariement. L'image de droite de la figure 5.11 présente le modèle reconnu ainsi que les points reconnus. L'image de la base est représentée par 377 points et 363 points d'intérêt ont été détectés sur l'image à reconnaître. Dans le processus d'appariement 116 points ont été utilisés.

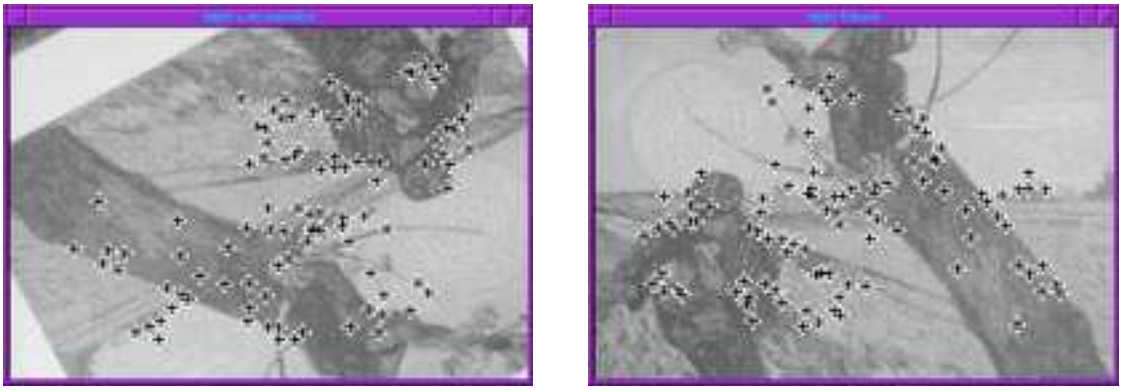

FIG. 5.11 – *Exemple d'une image recherchée et du modèle reconnu. Les croix symbolisent les points utilisés pendant la phase d'appariement. La transformation entre les deux images est constituée d'une rotation de* 152 *degrés.*

La figure 5.12 montre les points utilisés durant la phase de reconnaissance pour deux images entre lesquelles il y a une rotation de 185 degrés et un changement d'échelle de 1.3. Pour cet exemple, l'approche multi-échelle n'était pas utilisée. Dans le processus d'appariement 48 points ont été mis en correspondance. Une raison pour laquelle aussi peu de points ont été mis en correspondance est que l'approche multi-échelle n'est pas utilisée et donc le changement d'échelle n'est pas pris en compte. D'autre part un changement d'échelle rend la détection de points moins stable et donc plus de points doivent être éliminés.

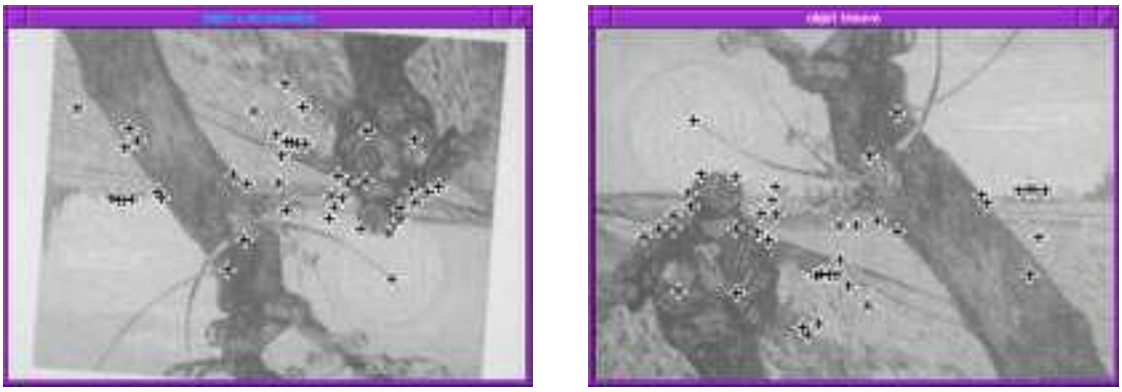

FIG. 5.12 – *Exemple d'une image recherchée et du modèle reconnu. Les croix symbolisent les points utilisés pendant la phase d'appariement. La transformation entre les deux images est constituée d'une rotation de* 185 *degrés et d'un facteur d'échelle de* 1.3.

De façon générale, il y a entre 20 et 150 points qui sont mis en correspondance. Ceci dépend de l'image considérée. Le faible nombre de points mis en correspondance comparé au nombre de points détectés illustre bien le rejet des points non discriminants et explique pourquoi une image caractérisée dans la base avec seulement 20 points peut être correc-

tement appariée alors que d'autres images de la base sont caractérisées par plus de 600 points.

5.4.3 Évaluation systématique de la recherche

Rotation image

FIG. 5.13 – *Quelques images de la séquence de rotation image du tableau "Sanja". L'image de droite est l'image stockée dans la base. Elle a été correctement reconnue quelle que soit l'image de gauche considérée.*

Pour tester l'invariance à une rotation image, plusieurs images ont été prises d'un même tableau de maître en faisant tourner la caméra approximativement autour de l'axe optique de son objectif. Ces images sont approximativement prises à intervalle régulier. La figure 5.13 présente quelques images d'une des séquences de rotation pour la scène "Sanja". L'image la plus à droite est celle contenue dans la base. Le taux de reconnaissance est de 100% pour les 40 images que contient cette séquence de rotation. Cette expérience montre l'invariance de la caractérisation vis-à-vis d'une rotation image. Cela démontre qu'il est suffisant de ne stocker qu'une seule image dans la base pour différentes rotations.

Changement d'échelle

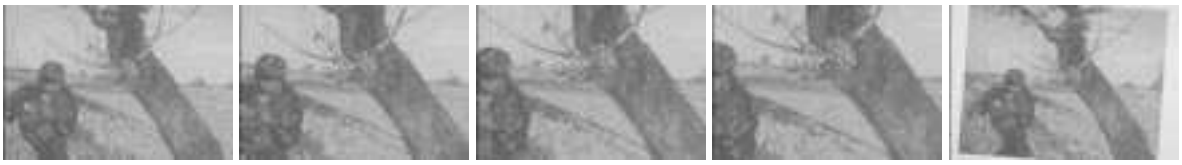

FIG. 5.14 – *Quelques images de la séquence changement d'échelle du "semeur" de Van Gogh. L'image de droite est l'image stockée dans la base. Elle a été correctement reconnue quelle que soit l'image de gauche considérée en utilisant l'approche multi-échelle.*

Pour tester la robustesse à un changement d'échelle, plusieurs images d'une même scène ont été prises en faisant varier le facteur de grossissement de l'objectif utilisé. Il s'agit d'un objectif à focale variable. La figure 5.9 présente quelques images de la séquence de changement d'échelle pour le tableau "le semeur" de Van Gogh. En utilisant l'approche multi-échelle, le taux de reconnaissance est de 100% jusqu'à un changement d'échelle de 2.2. Ceci montre la robustesse de notre caractérisation à un changement d'échelle. Toutefois, le facteur de changement a été limité à 2.2 par le détecteur de points d'intérêt utilisé.

Changement de luminosité

Dans la suite la robustesse à deux types de changement de luminosité sont examinés : un changement uniforme et un changement complexe. Dans le cas d'un changement uniforme,

FIG. 5.15 – *Quelques images des séquences changement uniforme et changement complexe de luminosité du "semeur" de Van Gogh. L'image de droite est l'image stockée dans la base. Elle a été correctement reconnue quelle que soit l'image de gauche considérée.*

uniquement l'intensité de la luminosité varie et dans le cas d'un changement complexe la source lumineuse est déplacée. La figure 5.15 montre des images du "semeur" de Van Gogh pour ces changements. Les images des deux séquences correspondant à ces types de changements ont toutes été reconnues correctement. Ceci confirme les résultats obtenus à la section 4.3.4 dans le cadre de l'appariement.

Changement de points de vue

Nous allons maintenant étudier la robustesse de notre méthode dans le cas d'un changement de point de vue. Cette étude est effectuée pour des images de tableau et pour des images aériennes.

Images de tableau

FIG. 5.16 – *Quelques images de la séquence changement de point de vue du "semeur" de Van Gogh. L'image de droite est l'image stockée dans la base. Elle a été correctement reconnue quelle que soit l'image de gauche considérée.*

La figure 5.16 montre quelques images de la séquence changement de point de vue du "semeur" de Van Gogh. L'image la plus à droite est l'image stockée dans la base. Les quatre autres images sont les images les plus extrême de la séquence. Toutes les images de la séquence ont été reconnues correctement exceptée l'image la plus à gauche. Si l'on compare ces résultats avec les résultats d'appariement obtenus sur cette séquence à la section 4.3.5, on peut noté que cette image est la seule pour laquelle le taux d'appariement est inférieur à 50%.

Images aériennes

La figure 5.17 montre à gauche l'image recherchée et à droite l'image correspondante contenue dans la base. On peut voir que la caméra s'est déplacée entre les deux vues. En plus, la scène est uniquement approximativement plane. Il y a donc des parties de maisons qui apparaissent et disparaissent. En plus des voitures ont bougé.

Pour ces images aériennes nous avons des images prises de 4 points de vue différents. Les 100 images prises du premier point de vue, noté "vue 1" sont stockées dans la base.

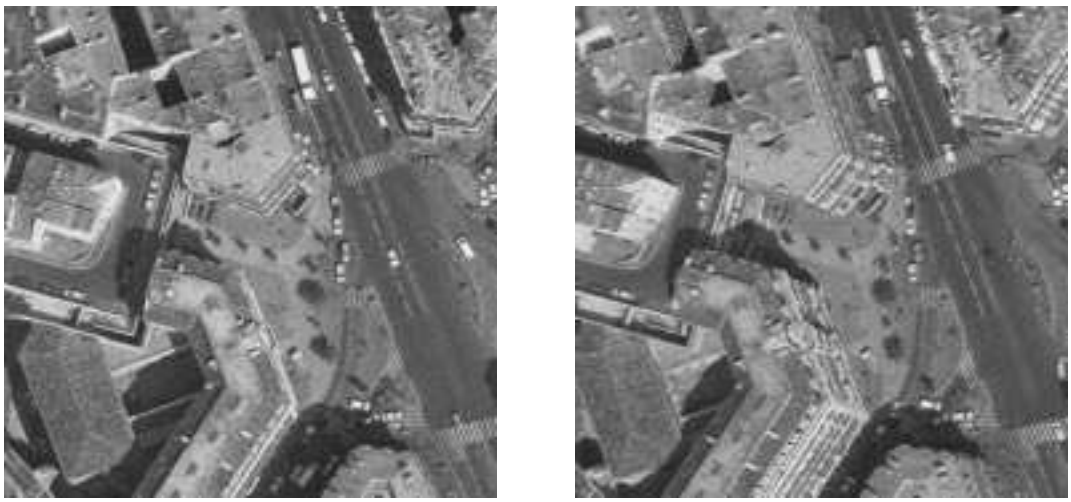

FIG. 5.17 – *À droite, une image de la vue 1 stockée dans la base; à gauche une image à reconnaître provenant de la vue 4. On peut remarquer les différences d'aspect des immeubles dues au changement de point de vue ainsi que le déplacement des véhicules (images fournies par Istar).*

Pour chacune de ces 100 images, nous avons utilisé 3 images tests prises à des points de vues différents, notés "vue 2", "vue 3", et "vue 4". Toutes ces images tests sont reconnues correctement exception faite de l'image du port qui ne contient que de l'eau.

Visibilité partielle

Les expérimentations présentées dans cette section ont pour but de montrer qu'il est possible de reconnaître une partie d'une image comme provenant de l'image entière. Des tests systématiques ont été effectués pour les images de tableaux de maîtres et pour les images aériennes. Pour ces tests, nous avons choisi aléatoirement des parties de taille relative entre 100% et 10% de l'image entière. Les parties contenant moins de 10 points d'intérêt ont été éliminées. Pour chaque taille relative 100 parties ont été tirées aléatoirement.

Images de tableau

FIG. 5.18 – *Exemples de parties d'images correctement retrouvées.*

La figure 5.18 montre quelques exemples de parties d'image de tableau pour lesquelles l'image entière est correctement retrouvée. L'ensemble des images tests ont été prises dans des conditions différentes des images de la base, notamment une rotation image et une translation. Jusqu'à 30% de taille relative, le taux de reconnaissance est de 100%. Pour une taille de 20%, nous obtenons 95% et pour une taille de 10% nous obtenons 90%. Les

images sur lesquelles la reconnaissance échouent sont les parties qui contiennent uniquement une texture répétitive, comme par exemple une pelouse ou un ciel peint par Monnet. Sur de telles parties le nombre de points d'intérêt est suffisant (supérieur à 10), mais les vecteurs de caractéristiques sont peu significatifs.

Images aériennes

En ce qui concerne les images aériennes nous possédons des images prises de 4 points de vue différents (cf. figure 5.17). Les images prises pour le premier point de vue sont stockées dans la base. Pour les images des trois autres points de vue, le taux de reconnaissance a été évalué en fonction de la taille relative. La figure 5.19 montre ces résultats pour les points de vue 2, 3 et 4 où le point de vue 4 est le plus éloigné du point de vue initial. On peut constater que plus les points de vue sont éloignés du point de vue stocké dans la base et moins bons sont les résultats. Pour le point de vue 2 nous obtenons presque toujours 100 % de taux de reconnaissance. En ce qui concerne le point de vue 3, le taux de reconnaissance est supérieur à 90% jusqu'à une taille relative de 20%. Pour une taille relative de 10%, le taux de reconnaissance se dégrade. Pour le point de vue 4, les résultats se dégradent à partir d'une taille relative de 40%. Ceci provient de l'importante déformation perspective entre les vues 1 et 4 et du déplacement des voitures.

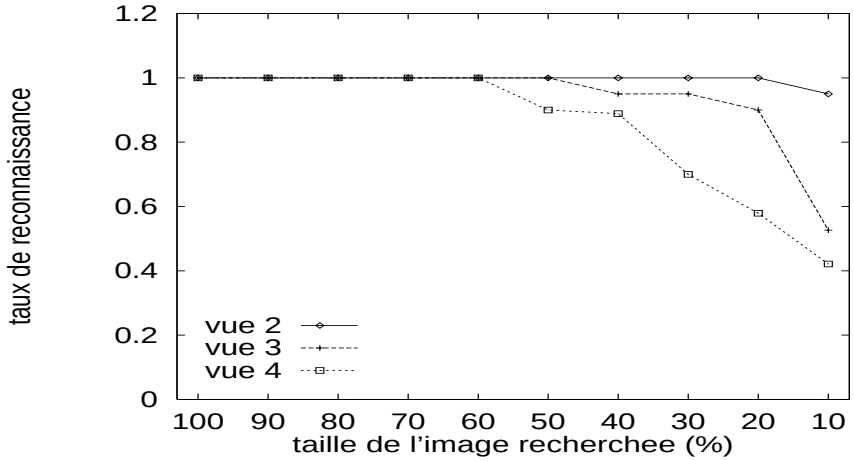

FIG. 5.19 – *Taux de reconnaissance en fonction de la taille relative de l'image recherchée pour une séquence d'images aériennes.*

La figure 5.20 montre les résultats pour la vue 4 si nous ne comptons pas uniquement le meilleur choix, mais également le deuxième puis le troisième choix. On peut observer que ceci améliore les résultats obtenus.

En conclusion de ces tests, les images de la base ont été reconnues à partir d'images n'en représentant qu'une partie. Vue la taille de la base d'images, ce résultat ne peut s'expliquer que par la discriminance des caractérisations et des contraintes semi-locales utilisées. Les applications possibles de la reconnaissance d'une partie d'un modèle sont nombreuses. Par exemple, trouver la position d'un hôtel dans l'image d'une ville est une application potentielle.

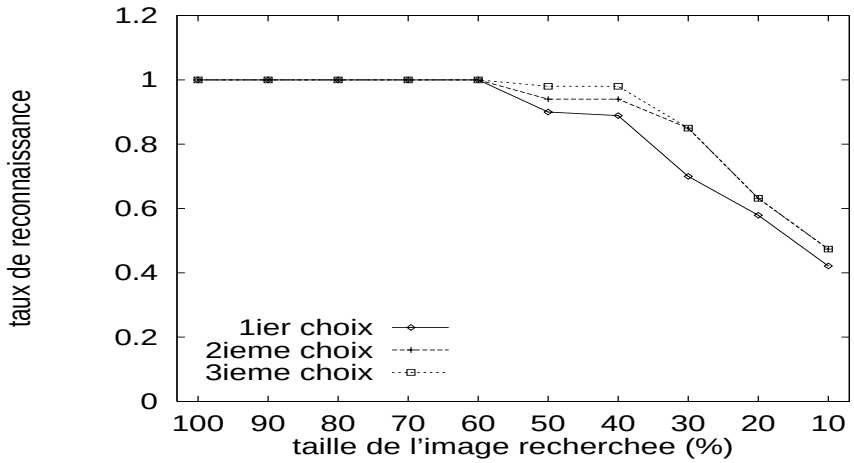

FIG. 5.20 – *Taux de reconnaissance en fonction de la taille relative de l'image recherchée pour une séquence d'images aériennes. Les différentes courbes correspondent au nombre de choix considérés.*

5.4.4 Temps de recherche

Dans la suite, le temps de recherche est détaillé selon les différentes étapes de notre algorithme. Les chiffres donnés correspondent au temps de calcul nécessaire sur une station de travail Sparc 10 sans configuration particulière.

Le temps de détection des points d'intérêt dépend uniquement de la taille de l'image. Pour une image 512 x 512 ce temps est de 7 secondes.

En ce qui concerne la caractérisation et la structuration des points d'intérêt, le temps dépend du nombre de points extraits. Pour une image sur laquelle 100 points ont été trouvés, la caractérisation prend 8 secondes et la structuration $1/20^{ème}$ de secondes.

Enfin, la recherche d'images basée sur la technique d'indexation présentée à la section 5.3 dépend du nombre de points extraits, mais aussi du nombre de points stockés dans la base. Pour évaluer de façon théorique la dépendance de la taille de base, il faudra étudier la distribution statistique des invariants. Ici l'évaluation a été effectuée en utilisation notre base de 1020 images contenant 154030 points. Pour rechercher 100 points dans cette base, il faut en moyenne cinq secondes.

En conclusion, le temps de recherche est au total de 20 secondes pour une image 512x512 et 100 points détectés. La rapidité de cette recherche pourrait être encore facilement accrue par la parallélisation de l'algorithme puisque la caractérisation et l'indexation sont faits de façon indépendante pour chaque point d'intérêt.

5.5 Conclusion

Dans ce chapitre, le problème de la recherche d'une image a été considéré comme un problème d'appariement d'une image à une base d'images. L'approche proposée est basée sur des données photométriques et permet de s'affranchir des deux défauts principaux de ces méthodes : elles sont en général globales et non invariantes aux transformations image. Pour ce faire, des invariants locaux de niveaux de gris sont calculés aux points d'intérêt qui représentent des points caractéristiques du signal. En fait, l'utilisation des points d'intérêt permet de rendre locale la méthode de recherche d'image, car ces points sont détectés

automatiquement et dépendent uniquement de l'image considérée.

Pour retrouver une image dans une base d'images, il est nécessaire de retrouver parmi plus d'un millier d'images celle qui ressemble le plus à une image donnée. Il ne s'agit plus alors de mettre en correspondance un point d'une image avec un autre parmi quelques centaines de points issus d'une seconde image mais d'apparier un point d'une image avec quelques centaines de millier de points issus de plusieurs images. Dans ce contexte les points d'une image donnée sont appariés avec plusieurs points de la base. Un algorithme de vote permet de faire émerger l'image de la base la plus ressemblante. Cet algorithme fait ressortir la cohérence globale des différents appariements.

Cet algorithme permet de reconnaître des images en présence de rotation, de changement d'échelle, de changement de luminosité, de changement de point de vue (limité), de fouillis et de visibilité partielle. L'utilisation de contraintes semi-locales permet en outre d'accroître de façon importante le taux de reconnaissance. Ces contraintes reposent sur le voisinage des points et sur une information géométrique d'angle entre points voisins.

Un autre problème soulevé par la recherche d'une image dans une base d'images est le coût de recherche. La transformation des vecteurs d'invariant dans une base tenant compte de la corrélation des différentes composantes et de leur ordre de grandeur permet de développer une technique d'indexation. La méthode ainsi proposée permet de rechercher une image parmi plus de mille en moins de cinq secondes sans matériel particulier. En outre le taux de reconnaissance est supérieur à 99%.

Chapitre 6

Modélisation 2D d'objet 3D

Ce chapitre étend la méthode d'appariement entre une image et une base d'images présentée au chapitre précédent à la reconnaissance d'objet. Pour ce faire, un objet 3D est modélisé par un ensemble d'images. Un état de l'art des différentes méthodes de modélisation d'objet est présenté à la section 6.1. La section 6.2 expose le principe de notre méthode de modélisation et montre quelles images utiliser pour modéliser un objet 3D. Les résultats de reconnaissance obtenus en utilisant une telle modélisation sont présentés en section 6.3 : un objet peut être reconnu même s'il est présenté au milieu d'une scène complexe ou s'il n'est que partiellement visible (cas des occultations jusqu'à 50% de l'objet).

Toutefois une telle modélisation ne contient aucune information 3D. Pour obtenir ce type d'information, il est nécessaire d'attacher des données symboliques 3D aux images de la base. L'ajout de ces données ainsi que leur calcul pour une image recherchée sont exposés à la section 6.4. Les résultats présentés à la section 6.5 montrent les données symboliques retrouvées et la précision atteinte.

6.1 État de l'art

Cette section présente différentes méthodes de modélisation d'objets. Un objet 3D peut être modélisé par un modèle géométrique. Il est également possible de représenter un objet 3D par des graphes d'aspect. Enfin, un objet 3D peut être modélisé par un ensemble d'images 2D. Nous présentons maintenant chacune de ces méthodes et montrons comment la modélisation influence l'étape de reconnaissance.

6.1.1 Modèle géométrique 3D

Les modèles géométriques d'un objet 3D sont basés sur des caractéristiques géométriques telles qu'arêtes, jonctions, ellipses, surfaces et volumes. Un modèle 3D de l'objet à reconnaître est établi à partir de ces caractéristiques. Ces modèles sont souvent basés sur

des modèles CAO[1]. Parmi les modèles CAO existants, la modélisation par un fil de fer consiste en une liste de jonctions et de connexions entre ces jonctions. La géométrie par construction de solides (Constructive Solid Geometry), quant à elle, modélise un objet par des opérations ensemblistes à partir de primitives volumiques. La représentation par occupation spatiale décrit le volume occupé par l'objet 3D et la représentation par enveloppe surfacique (B-Rep.) modélise un objet par des morceaux de surface. Besl et al. [Bes 85] et Chin et al. [Chi 86] présentent un état de l'art des modèles géométriques et des systèmes de reconnaissance d'objet basés sur de tels modèles. Ces systèmes mettent en correspondance une image 2D et un modèle géométrique 3D. Il existe de nombreux travaux pour effectuer une telle mise en correspondance. Ces travaux peuvent être partitionnés en méthodes basées sur un mécanisme de prédiction/vérification, sur la transformée de Hough et sur l'utilisation d'un arbre d'interprétation.

Les systèmes basés sur un mécanisme de prédiction/vérification mettent en correspondance quelques caractéristiques du modèle avec quelques caractéristiques de l'image. Ceci permet un calcul initial de la transformation modèle - image. Cette transformation est utilisée pour projeter les autres caractéristiques du modèle sur l'image et ensuite vérifier la correspondance avec les caractéristiques de l'image. Dans le cas d'objets polyédriques, Huttenlocher et al. [Hut 90] et Lowe et al. [Low 86] ont développé une telle approche. Bolles et al. [Bol 86] et Faugeras et al. [Fau 86] ont développé des approches similaires dans le cas des images de profondeur. Toutefois, la recherche exhaustive de tous les modèles existant dans la base engendre un coût de calcul exponentiel. La contribution majeure de différents systèmes de reconnaissance a été de contrôler et de diminuer la complexité de la phase d'appariement. Par exemple, Bolles et al. [Bol 86] utilisent un arbre de recherche.

Kriegman et al. [Kri 90] ont étendu une telle approche pour des modèles courbes. La représentation implicite de ces courbes dans l'image est paramétrisée par la position et l'orientation de l'objet. Le calcul de ces paramètres se réduit au problème d'ajustement entre le contour théorique et les points contour dans l'image. La vérification est effectuée en comparant les erreurs d'ajustement pour les différents modèles.

D'autres travaux, comme par exemple Mundy et al. [Mun 90], calculent les transformations entre les primitives image détectées et les primitives des données CAO. Ils utilisent ensuite la transformée de Hough dans l'espace des paramètres de ces transformations pour trouver le point d'accumulation. Ce point d'accumulation donne à la fois le modèle correspondant et la transformation entre l'image et le modèle.

Les arbres d'interprétation contiennent toutes les combinaisons possibles entre les primitives détectées dans l'image et les primitives du modèle. Ces combinaisons sont organisées dans un arbre, par exemple le premier niveau de l'arbre contient les combinaisons entre une primitive extraite et les primitives du modèle. Cet arbre donne lieu à un énorme espace de recherche. Il est donc indispensable d'introduire des contraintes supplémentaires qui évitent le parcours exhaustif de l'arbre. Brooks [Bro 83] a par exemple développé une telle approche. En plus son approche permet d'utiliser des contraintes avec un intervalle de confiance dans le cas d'objets génériques. Grimson et al. [Gri 87, Gri 89] ont utilisé des arbres d'interprétation dans le cas d'images de profondeur. Dans le cas des images de niveau de gris leur approche est limitée aux objets 2D.

Ces représentations symboliques d'un objet sont séduisantes à l'esprit mais cepen-

1. CAO : Conception Assistée par Ordinateur. Cette conception permet l'automatisation de processus de conception et de manufacture.

dant éloignées de la réalité des images manipulées en vision par ordinateur. En effet, ces représentations sont à la fois trop simples pour modéliser des objets non manufacturés et difficiles à apparier avec les résultats des algorithmes d'extraction de primitives existants. Chen et Mulgaonkar [Che 91] ont montré l'inadéquation des modèles CAO dans un contexte de vision par ordinateur : les contours extraits à partir d'images de synthèse issues d'un modèle CAO ne coïncident pas avec les contours extraits de l'image réelle. En fait, après de nombreuses expérimentations avec des objets relativement simples, Chen et al. ont conclu ne pas être capables de prédire à partir des données CAO ce qui devrait être perçu dans les images réelles. Ceci a été observé malgré des algorithmes de génération d'image très sophistiqués. Les segments attendus n'étaient que très rarement les segments détectés. Malgré les succès limités obtenus, il est apparu au sein de la communauté que la vision reposant sur des données CAO n'est pas suffisamment robuste. Certains auteurs ont même affirmé que le problème d'appariement entre une image et un modèle CAO n'a pas de solution générale puisque les niveaux de représentations entre le modèle CAO et ce qui peut être détecté dans les images sont trop différents. On peut se reporter à la discussion ayant eu lieu au workshop sur la vision utilisant des modèles CAO [Sha 91]. L'une des principales raisons de cet échec réside dans le fait que les images ne reflètent pas directement les informations abstraites CAO des objets : l'image 2D d'intensité est trop différente de la structure 3D abstraite. En outre de telles représentations ne sont pas disponibles pour de nombreux objets naturels tels que les arbres, et elles ne le sont pas non plus pour des objets manufacturés tels que les tableaux de maîtres (si l'on excepte certaines œuvres cubistes). Pour cette raison, il apparaît nécessaire de modéliser les objets perçus à partir de descriptions 2D.

6.1.2 Graphe d'aspect

Les graphes d'aspect ont été introduits par Kœnderink [Koe 79]. Ils permettent de modéliser un objet par un ensemble de descriptions 2D. Pour ce faire, les aspects topologiques 2D des droites perçues dans les images sont utilisés. Ces graphes peuvent être calculés de façon exacte à partir d'un modèle théorique de l'objet. La figure 6.1 montre par exemple un graphe d'aspect pour une sphère facettisée. Il existe de nombreux travaux pour calculer des graphes exacts. Ces travaux diffèrent par la catégorie de l'objet considéré et le type de projection considérée. Dans le cas d'objets polyédriques, on peut par exemple citer Stewman [Ste 88] et Gigus [Gig 91]. Eggert [Egg 89] et Kriegman [Kri 89] ont calculé des graphes d'aspect pour des solides de révolution. En ce qui concerne des objets plus complexes, Rieger [Rie 87] et Petitjean [Pet 92] ont proposé des solutions.

Il est également possible de construire des graphes d'aspect à partir d'un nombre fini de vues synthétisées d'un objet. En effet, Herbert [Heb 85] et Ikeuchi [Ike 88] n'utilisent pas de modèle exact de l'objet mais une centaine de vues synthétiques pour calculer leur graphe d'aspect. Cette approche simplifie les graphes obtenus mais ne tient pas compte du fait que certaines vues ne diffèrent que par des détails mineurs qui sont en fait indétectables. De bons résultats ont été obtenus dans le cas d'objets simples.

En conclusion, les graphes d'aspect utilisent un modèle théorique de l'objet et présentent de ce fait un nombre de désavantages (cf. par exemple l'article de Bowyer [Bow 91]). Comme dans le cas des modèles géométriques les graphes d'aspect requièrent une très bonne segmentation des images. Un modèle est défini par des caractéristiques théoriques qui doivent être détectées à partir des images. Rien ne garantit que ces caractéristiques

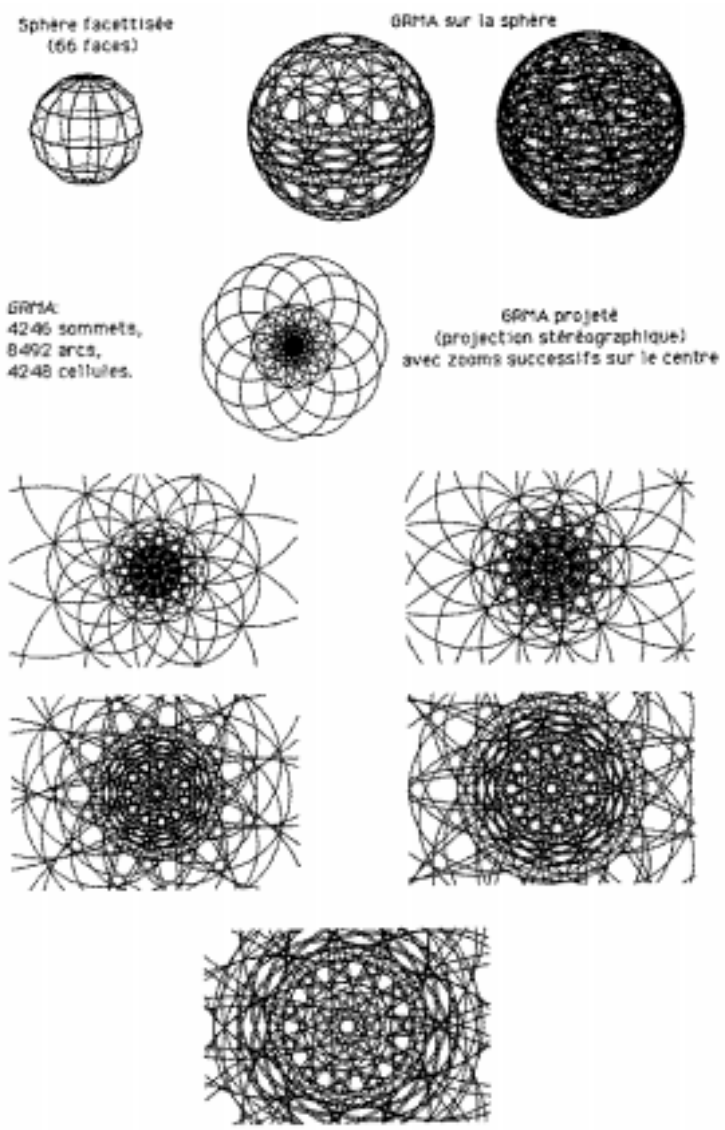

FIG. 6.1 – *Construction du graphe d'aspects d'une sphère facettisée (66 faces). Cette figure est extraite de la thèse de Degott [Deg91].*

soient effectivement détectées. En outre, les graphes d'aspect sont volumineux et complexes puisqu'ils prennent en compte les détails des objets modélisés. À chaque catastrophe (apparition ou disparition d'une arête), un nouvel aspect est créé. Pour un objet réel même simple, le graphe d'aspect obtenu peut capturer nombre d'aspects sans importance et être ainsi extrêmement complexe (cf. figure 6.1). Cependant, cette représentation reste pauvre puisque uniquement la nature topologique d'un objet est capturée.

6.1.3 Ensemble d'images

L'approche consistant à utiliser des images pour modéliser un objet 3D est relativement nouvelle. L'idée principale d'une telle approche est de ne plus utiliser de modèles théoriques éloignés des images, mais d'utiliser des images caractéristiques pour représenter un objet. Ces images sont dans la suite référencées par images-modèles.

Nayar et al. [Nay 93] proposent de modéliser un objet 3D à partir d'images 2D de niveaux de gris. Comme images-modèles ils utilisent un ensemble d'images régulièrement espacées. À partir de ces images, un espace de vecteurs propres est construit. Pour une nouvelle image, une décomposition dans cet espace est effectuée ce qui permet de reconnaître l'objet.

Gros [Gro 95] utilise des primitives extraites des images-modèles, notamment des segments et des jonctions de segments. Pour déterminer les images nécessaires pour représenter un objet 3D, il prend un grand nombre d'images de cet objet. Il utilise ensuite un algorithme de clustering pour décider quelles images il faut garder comme images représentatives de l'objet. Le désavantage de son approche est que les segments ne peuvent être extraits que pour des objets relativement simples comme des polyèdres. De même, Gdalyahu [Gda 96] utilise les contours des images-modèles pour représenter un objet. Cette approche souligne la nécessité de choisir de "bonnes" vues pour représenter l'objet. De telles vues sont intrinsèquement plus stables et plus représentatives de l'objet. En outre, avec de telles vues, la métrique utilisée pour comparer deux images a peu d'influence sur les résultats obtenus.

Enfin, pour des classes spécifiques d'objets 3D il est possible de caractériser un objet à partir d'une seule vue. Comme Zisserman [Zis 95] l'a montré, il existe des invariants projectifs pour un certain nombre de classe d'objet 3D. Ces invariants peuvent donc être extraits à partir d'une vue d'un tel objet. En utilisant ces invariants, il est ensuite possible de reconnaître l'objet à partir de n'importe quel point de vue.

6.2 Modélisation à partir d'images 2D

6.2.1 Principe

Cette section présente la méthode retenue pour modéliser un objet 3D à partir d'images 2D. Chaque image 2D (image-modèle) représente en fait un "aspect" de l'objet. Deux motivations importantes sont à l'origine d'une telle modélisation. Premièrement, les images permettent de représenter l'information réellement utilisable lors de la phase de reconnaissance et non pas une information abstraite difficilement détectable comme dans le cas des modèles géométriques ou des graphes d'aspect. Deuxièmement, il n'y a pas d'invariant générique 3D comme nous avons pu le voir à la section 3.2. Il est donc impossible de modéliser un objet 3D quelconque par une seule vue. Par contre, une modélisation à partir

de plusieurs vues est possible.

Pour mettre en œuvre une telle modélisation, se pose le problème de déterminer quelles images sont nécessaires pour représenter un objet 3D. Pour cela il faut déterminer l'ensemble minimal d'images qui représente l'objet. Il est donc nécessaire de savoir si deux images sont proches. Pour ce faire, le critère utilisé est l'appariement. Si le taux d'appariements correct entre deux images est élevé, ceci signifie qu'elles peuvent être représentées par la même image-modèle.

La modélisation présentée par la suite utilise des images-modèles qui sont espacées régulièrement. Il s'agit d'une approximation grossière de la réalité, car les images se ressemblent plus ou moins selon le point de vue. Il est en dehors des objectifs de ce travail d'étudier la possibilité d'utiliser une distribution non uniforme dépendant de la forme de l'objet 3D à modéliser comme cela a été fait dans [Gda 96] et dans [Gro 95].

Dans la suite nous allons présenter des exemples de modélisation d'objet sur un cercle. Cette modélisation est ensuite étendue de façon théorique à la modélisation sur une sphère.

6.2.2 Exemple d'une modélisation sur un cercle

Pour modéliser un objet 3D différentes images d'un objet sont prises en déplaçant la caméra sur un cercle centré sur l'objet à modéliser (ou en faisant tourner l'objet sur lui-même, ce qui est équivalent). Dans la suite nous calculons les images modèles pour l'objet "Dinosaure" (cf. figure 6.2) et l'objet "Main Abstraite" (cf. figure 6.4).

FIG. 6.2 – *Quelques images de l'objet "Dinosaure" contenues dans la base. L'espacement entre deux images consécutives est de 20 degrés.*

Pour l'objet "Dinosaure", 18 images espacées de 20 degrés en position sont suffisantes pour obtenir un modèle complet. Nous avons obtenu cette valeur de façon expérimentale en utilisant 36 vues de l'objet "Dinosaure" espacées de 10 degrés. Pour ce faire nous calculons le nombre de votes obtenus pour chacune de ces vues pour une image test se trouvant géométriquement entre les vues 1 et 2. La figure 6.3 montre l'histogramme de votes obtenus. On peut voir que le nombre de votes est important pour les vues 1 et 2 et très nettement inférieur pour les autres vues. Les vues 1 et 2 sont donc suffisamment ressemblantes pour qu'il ne soit pas nécessaire de les stocker toutes les deux dans la base. Ceci montre que le nombre de vues peut être réduit par un facteur d'au moins 2 sans diminuer la qualité du processus de mise en correspondance. 18 vues espacées de 20 degrés sont donc bien suffisantes pour représenter l'objet "Dinosaure".

Toutefois, le nombre de vue nécessaire dépend de la complexité de l'objet. Pour des objets moins complexes, telle la "Main Abstraite", uniquement 9 vues se sont révélées suffisantes pour définir le modèle de l'objet (cf. figure 6.4).

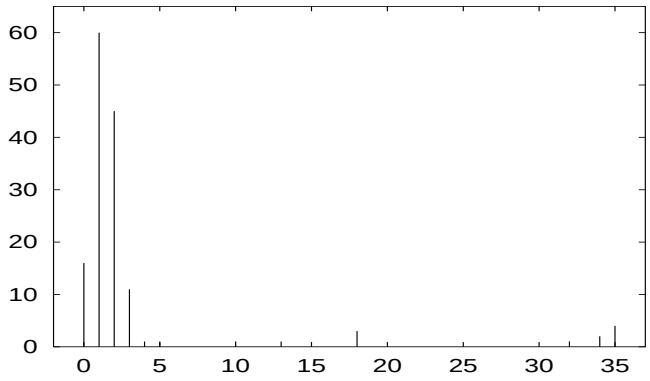

FIG. 6.3 – *Résultats du nombre de votes pour 36 vues de l'objet "Dinosaure". L'image à reconnaître est proche des images 1 et 2.*

FIG. 6.4 – *Quelques images de l'objet "Main Abstraite" contenues dans la base. L'espacement entre deux images consécutives est de 40 degrés.*

6.2.3 Extension à la modélisation sur une sphère

La modélisation d'un objet sur une sphère de vue est une extension directe de la modélisation sur un cercle. Nous n'avons pas pu tester une telle modélisation à cause de sa difficulté de mise en œuvre sans outil robotique adapté. Toutefois, il est possible d'estimer de façon théorique combien de vues sont nécessaires pour une telle modélisation. Si l'on considère un espacement régulier de 20 degrés ce qui s'est avèré suffisant dans la section précédente, un centaine de vues sont suffisante pour modéliser un objet. Ceci ne présentera pas de difficulté particulière. Toutefois, une telle modélisation accroît de façon importante le nombre d'images contenues dans la base d'images. Nous verrons dans les perspectives à la section 7.3 les solutions envisagées pour surmonter ce problème.

6.3 Résultats de reconnaissance

Ayant modélisé un objet 3D par des images 2D la reconnaissance d'objet se traduit par une recherche d'images. La procédure de recherche d'image développée au chapitre précédent est donc utilisée. À la section 6.3.1 nous allons d'abord montrer quelques exemples qui illustrent que la reconnaissance est possible en présence d'arrière-plan complexe et d'occultation. Ensuite, la complexité du problème est illustrée à la section 6.3.2. Pour ce faire les points détectés et les points sélectionnés pour la reconnaissance sont visualisés. Enfin, une évaluation systématique est présentée à la section 6.3.3.

6.3.1 Quelques exemples de reconnaissance

L'objet tridimensionnel "Dinosaure" est reconnu correctement à partir des images de gauche de la figure 6.5. On peut voir que l'objet est correctement reconnu en présence de rotation, de changement d'échelle, de changement de arrière-plan et d'occultation. En outre, le changement de point de vue entre l'image recherchée et l'image reconnue correspond à un angle de 10 degrés. Ceci correspond au plus grand angle possible entre une image et une image-modèle puisque l'espacement entre deux images-modèles est de 20 degrés. La figure 6.5 montre de plus que l'on a retrouvé l'image la plus proche de la base ce qui donne une estimation de l'attitude.

FIG. 6.5 – *Exemple de reconnaissance d'un objet tridimensionnel. L'image de droite a été correctement reconnue quelle que soit l'image de gauche considérée.*

La figure 6.6 montre le résultats de reconnaissance pour un deuxième objet "Main Abstraite". L'objet est correctement reconnu en présence d'un arrière-plan complexe.

6.3.2 Points sélectionnés

Cette section illustre la difficulté de la reconnaissance en présence d'un arrière-plan complexe. Pour ce faire, les points d'intérêt sont visualisés pour l'exemple de la figure 6.6. La figure 6.7 montre les points d'intérêt détectés sur l'image recherchée. Elle illustre la complexité du problème de la reconnaissance : il y a nettement plus de points d'intérêt trouvés sur l'arrière-plan que sur l'objet en lui-même. Toutefois, la discriminance de la caractérisation et l'utilisation des contraintes semi-locales permettent d'éliminer les points détectés sur l'arrière-plan. Ceci est confirmé par la figure 6.8 qui montre les points d'intérêt

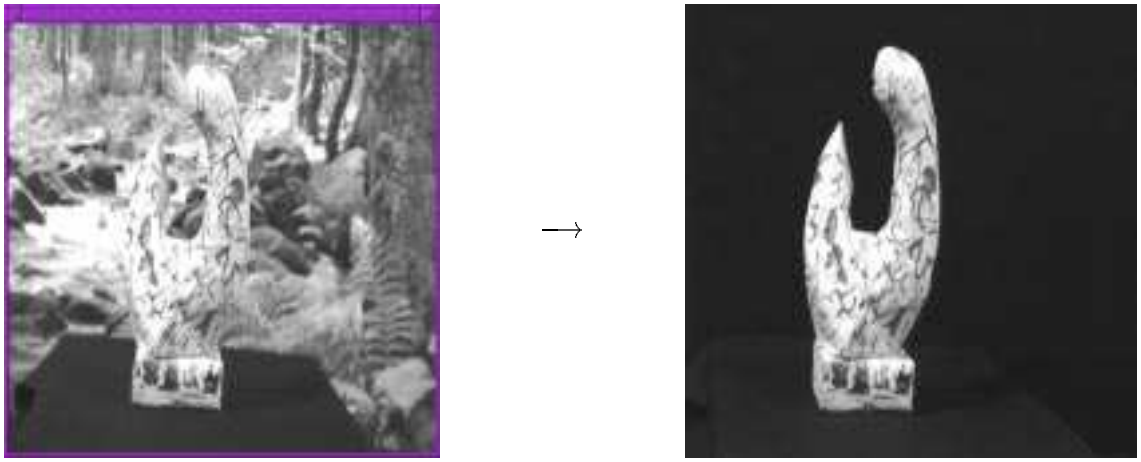

FIG. 6.6 – *Exemple de reconnaissance d'un objet tridimensionnel. L'image de droite a été correctement reconnue à partir de l'image de gauche.*

appariés lors du processus de reconnaissance. De plus, les appariements trouvés permettent de localiser l'objet dans une scène complexe.

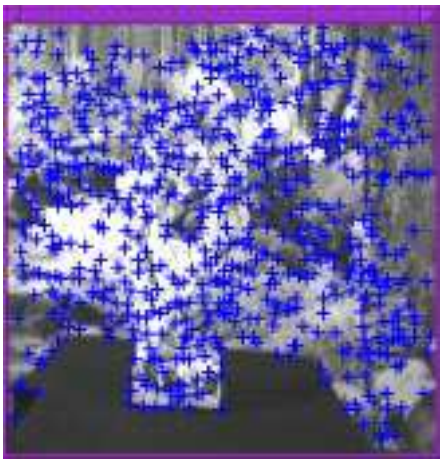

FIG. 6.7 – *Points d'intérêt détectés sur l'image recherchée de la figure 6.6*

6.3.3 Évaluation systématique

Cette section présente les résultats de reconnaissance d'objets 3D à partir de n'importe quel point de vue. Par simplicité chaque objet est représenté dans la base par 18 vues espacées de 20 degrés. Pour l'évaluation systématique les images tests ont été prises en faisant tourner les objet par pas de 10 degrés. De plus, des images tests donné ont été prises sous des angles de vue différents de ceux utilisés pour les images stockées dans la base. Le taux de reconnaissance pour un ensemble de 720 images tests est de 99.86%.

Notre base d'images contient également les images de la base de Columbia. Cette base, malgré ses défauts, sert de base de tests à plusieurs systèmes de reconnaissance. Il

FIG. 6.8 – *Illustration des points appariés lors du processus de reconnaissance. L'image de droite a été correctement reconnu à partir de l'image de gauche.*

était donc intéressant de tester notre méthode uniquement sur ces images. En utilisant les mêmes conditions expérimentales que les autres chercheurs, le taux de reconnaissance a été de 100% ce qui est le même taux que celui obtenu par d'autres chercheurs. Par rapport à Rao et al. [Rao 95] qui utilisent également une caractérisation locale, nous utilisons moins de points qu'eux et la dimension de nos caractéristiques est moindre. Pour mémoire, Rao utilise des vecteurs qui possèdent 45 composantes, alors que notre caractérisation ne contient que 9 composantes. De plus, Rao fixe la position des points caractérisés sur une grille. La sélection automatique des points à caractériser et l'utilisation d'invariants permet donc d'obtenir les mêmes résultats que Rao avec des coûts nettement inférieurs. Ceci montre la représentativité de la caractérisation retenue.

6.4 Localisation de données symboliques 3D

La modélisation à partir d'images 2D ne contient aucune information symbolique 3D. Notre seule information sont les points appariés. Toutefois, ces points sont la plupart du temps non significatifs dans un contexte d'interprétation. De plus, ces points varient du fait de l'instabilité inhérente au processus de vision. Nous proposons donc d'ajouter les données symboliques dont on a besoin aux images-modèles contenues dans la base. Cette section présente comment ajouter ces données puis comment les retrouver pour une image recherchée.

6.4.1 Ajout de données symboliques

Les données symboliques peuvent être ajoutées n'importe où dans l'image. Nous proposons de stocker les coordonnées 2D correspondant à ces données dans un fichier attaché à l'image modèle. L'ajout peut se faire à la main ou de façon semi-automatique par projection des données CAO : ayant repéré quelques caractéristiques particulières, il est possible de calculer la matrice de projection perspective 3D-2D et ensuite l'intégralité des données CAO peut être projetée sur l'image modèle. Ces données symboliques peuvent être des points, des lignes, des ellipses, des axes de symétrie ou autres. Bien évidemment il faut

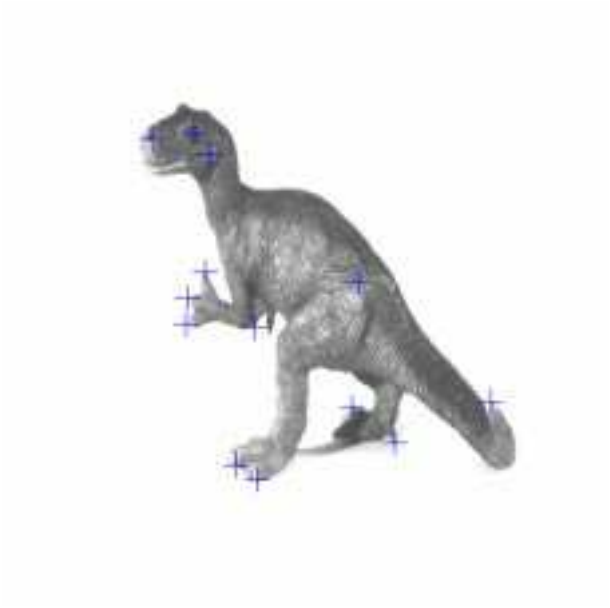

FIG. 6.9 – *Données symboliques pour l'objet "Dinosaure".*

avoir une cohérence entre les données symboliques des différentes images-modèles d'un même objet. Dans le cadre de ce travail cette cohérence a été assurée manuellement.

La figure 6.9 montre les données symboliques définies pour l'objet "Dinosaure". On a marqué des points intéressants : l'œil, les doigts, le bout de la queue, etc. Dans le cas d'une tasse, on a ajouté les axes de symétrie et les ellipses caractéristiques (cf. figure 6.10).

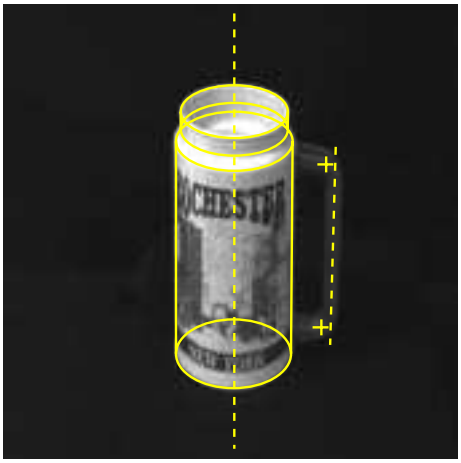

FIG. 6.10 – *Données symboliques pour une tasse.*

Les données symboliques étant localisées dans les images, il est facile de relier ces positions à tout type d'information symbolique ou numérique pertinente. Ce peut être le nom d'un point ou ses coordonnées dans un repère particulier, ou des informations sur les matériaux, etc. En particulier, ce type d'information peut permettre de connaître la position relative des points ce qui est important pour beaucoup d'applications (téléguidage assisté par ordinateur ou asservissement visuel).

6.4.2 Identification des informations symboliques

Ayant ajouté des données symboliques, le problème est maintenant de les retrouver dans une image inconnue. De fait, les appariements utilisés lors de la recherche d'image sont connus. La connaissance de relations point à point entre deux images ne permet malheureusement pas d'établir une information projective suffisamment riche pour retrouver de façon univoque la position des données symboliques. En revanche, il existe une relation univoque point à point entre tout triplet de points se correspondant. Nous utilisons donc trois images : en plus de l'image recherchée et de l'image reconnue, nous utilisons une autre image-modèle de la base. Ceci impose de calculer des correspondances entre les images-modèles d'un même objet. Ceci est naturellement effectué hors ligne. On connaît dans ces conditions les appariements entre trois images ce qui permet de calculer la relation trilinéaire existant entre ces images. Cette relation permet de déterminer la position d'un point dans une image si l'on connaît les positions de ce point dans deux autres images. Dans notre cas, les données symboliques ont été définies pour les deux images-modèles de la base. En utilisant la relation trilinéaire on peut donc retrouver ces mêmes données dans l'image recherchée. Ceci est exposée schématiquement sur la figure 6.11. Nous allons maintenant présenter les équations importantes de la relation trilinéaire et montrer comment la mettre en œuvre dans notre cas.

Relation trilinéaire

La géométrie projective a montré qu'il existe une relation point à point entre trois images. On peut se reporter à [Sha 94, Fau 95, Mun 92a, Fau 92b]. Dans la suite, cette relation sera notée $\mathcal{T}$. L'équation 6.1 présente une forme de cette relation (il en existe en fait quatre). Cette équation exprime la contrainte entre les coordonnées (x, y), (x', y') et (x'', y'') des projections p, p' et p'' d'un point P dans trois images. Étant donné un ensemble de correspondances entre ces trois images, il est alors possible de calculer les paramètres $\alpha_{[1..18]}$ de cette équation.

$$\left\{\begin{array}{lcl} \alpha_1 + \alpha_2 x + \alpha_3 x'' + \alpha_4 y + \alpha_5 y' + & & \\ \alpha_6 x x'' + \alpha_7 y y' + \alpha_8 x y' + \alpha_9 x'' y + & & \\ \alpha_{10} x'' y' + \alpha_{11} x'' y y' + \alpha_{12} x x'' y' & = & 0 \\ \alpha_{13} + \alpha_{14} x + \alpha_{15} y + \alpha_{16} y' + \alpha_3 y'' + & & \\ \alpha_{17} y y' + \alpha_9 y y'' + \alpha_{10} y' y'' + \alpha_{18} x y' + & & \\ \alpha_6 x y'' + \alpha_{12} x y' y'' + \alpha_{11} y y' y'' & = & 0 \end{array}\right. \tag{6.1}$$

Il existe plusieurs méthodes pour calculer la relation trilinéaire. La méthode utilisée ici est une variante de la méthode proposée par [Bob 96] qui est basée sur une formulation géométrique présentée dans [Bea 94] et [Har 92]. La méthode proposée repose sur l'expression d'une reconstruction implicite du modèle observé.

Le calcul de la relation trilinéaire est très sensible aux fausses correspondances. Il est donc nécessaire de rejeter celles-ci. Notre méthode de calcul utilise une méthode de moindres carrés médians. Ceci autorise jusqu'à 50% de fausses correspondances. D'autre part, le rejet effectué par la méthode de calcul introduit une contrainte de cohérence globale sur les données mises en correspondance.

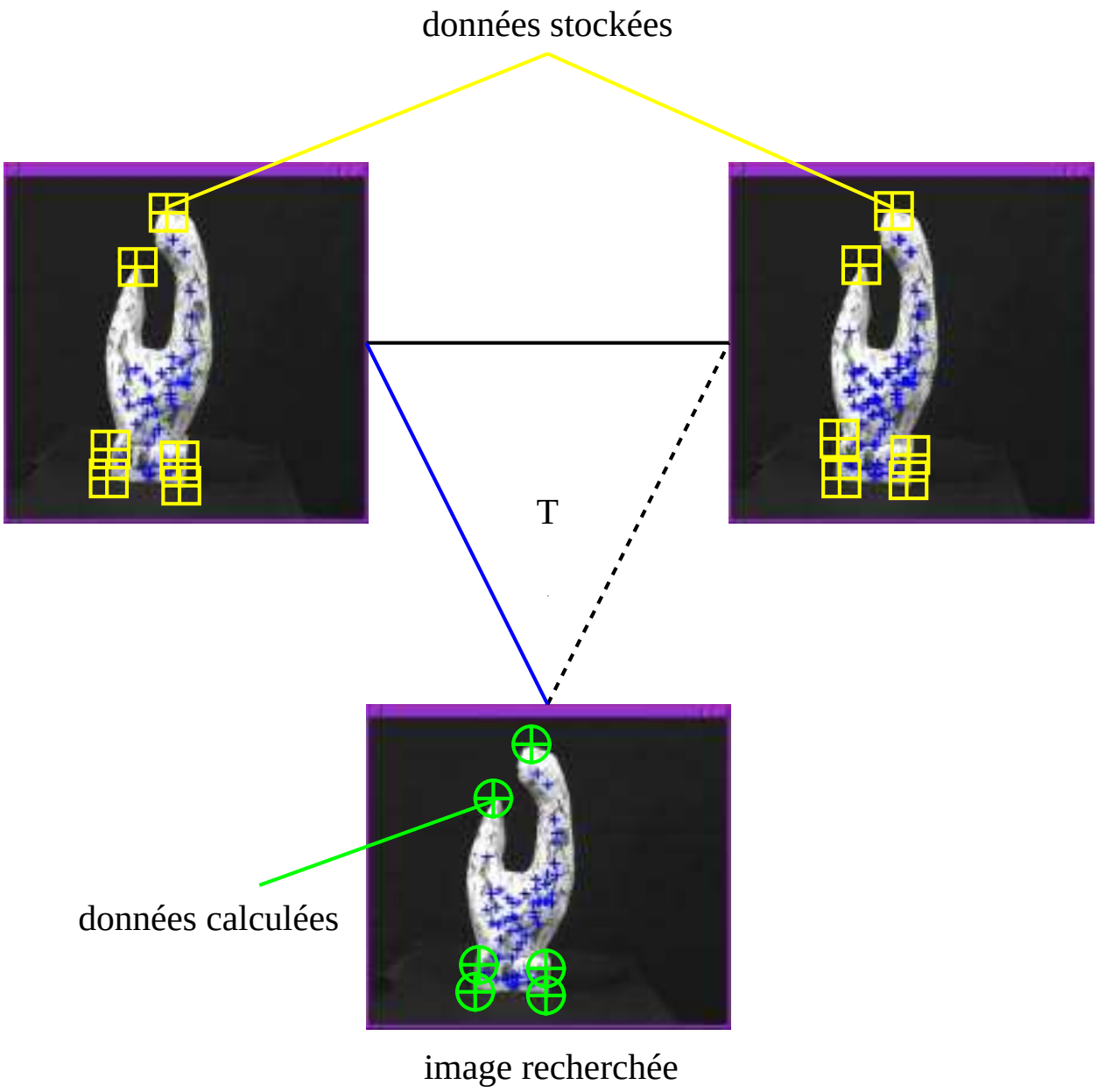

FIG. 6.11 – *Étant données deux images-modèles de la base pour lesquelles les données symboliques sont connues, la relation trilinéaire $\mathcal{T}$ permet de retrouver ces données sur l'image recherchée.*

Détermination de la position des données symboliques

Étant donnée une relation trilinéaire entre trois images et un ensemble de points correspondants entre deux de ces images, les positions de ces points peuvent être directement calculées dans la troisième image. Il est donc possible de calculer les positions des données symboliques dans une image si on connaît les positions de ces données dans deux autres images. La figure 6.11 illustre le principe de ce calcul. Ce calcul est détaillé par la suite.

Soient $\mathcal{I}'$ et $\mathcal{I}''$ deux images-modèles pour lesquelles on connaît la position des données symboliques et soit $\mathcal{I}$ une image pour laquelle on recherche ces données. En utilisant la relation trilinéaire $\mathcal{T}$, c'est-à-dire les coefficients $\alpha_{[1..18]}$, les positions correspondantes peuvent être calculées dans l'image $\mathcal{I}$ à partir de l'équation 6.2. Cette équation montre que le calcul d'une position (x, y) est obtenue par la résolution d'un système linéaire.

$$\left\{\begin{array}{lcl} (\alpha_1 + \alpha_3 x'' + \alpha_5 y' + \alpha_{10} x'' y') + & & \\ (\alpha_2 + \alpha_6 x'' + \alpha_8 y' + \alpha_{12} x'' y') x + (\alpha_4 + \alpha_9 x'' + \alpha_7 y' + \alpha_{11} x'' y') y & = & 0 \\ (\alpha_{13} + \alpha_{16} y' + \alpha_3 y'' + \alpha_{10} y' y'') + & & \\ (\alpha_{14} + \alpha_{18} y' + \alpha_6 y'' + \alpha_{12} y' y'') x + (\alpha_{15} + \alpha_{17} y' + \alpha_9 y'' + \alpha_{11} y' y'') y & = & 0 \end{array}\right. \tag{6.2}$$

Il faut noter que si les coordonnées tridimensionnelles de ces données symboliques sont connues, il est alors facile de déterminer l'attitude de l'objet dans le repère de la caméra.

Étant donnés les appariements entre une image recherchée et l'image modèle lui correspondant, il faut retrouver les données symboliques sur l'image recherchée. La figure 6.9 à la section précédente présente les données symboliques attachées à l'image-modèle "Dinosaure".

6.5 Résultats de localisation

Cette section présente des résultats de localisation de données symboliques. La figure 6.12 montre les données symboliques retrouvées pour une image recherchée. Si l'on compare ces données aux données ajoutées à l'image-modèle (cf. figure 6.9), on peut voir que ces données ont été retrouvées correctement. En outre, la précision des données symboliques est bonne : la distance moyenne entre les points retrouvés et les points réels est de 0.23 pixel. Cette précision a été évaluée en mesurant la distance entre les données retrouvées et les données détectées manuellement. La figure 6.13 atteste de la précision de la position retrouvée par un agrandissement de l'œil du "Dinosaure" sur l'image recherchée. La méthode robuste de calcul de la relation trilinéaire a en outre permis de rejeter deux faux appariements.

Il est également possible de retrouver les données symboliques en présence d'occultations pourvu que suffisamment d'appariements restent pour qu'on puisse déterminer la relation trilinéaire. Pour montrer ce fait, nous avons caché la tête du "Dinosaure". La figure 6.14 montre les données symboliques retrouvées à partir de cette image. L'image de gauche est l'image utilisée pour la recherche sur laquelle sont positionnées les données retrouvées. L'image de droite dévoile l'occultation et montre la précision des résultats obtenus.

La figure 6.15 montre la récupération de données symboliques non ponctuelles pour une tasse. L'image recherchée contient la tasse devant un arrière-plan complexe et pour une position différente de celle de l'image modèle. Pour définir les ellipses dans les images-modèles

FIG. 6.12 – *Données symboliques retrouvées sur l'objet "Dinosaure".*

FIG. 6.13 – *Agrandissement de l'œil du "Dinosaure". Cette figure atteste de la précision avec laquelle les données sont positionnées sur l'image recherchée.*

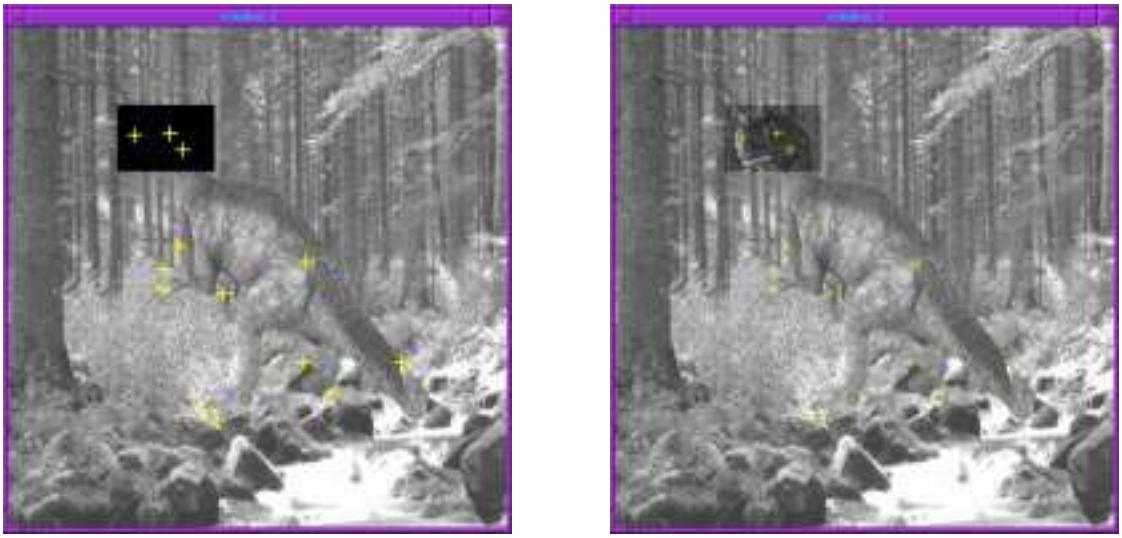

FIG. 6.14 – *Exemple de récupération de données symboliques en présence d'occultations. L'image de gauche est l'image utilisée pour la recherche sur laquelle sont positionnées les données retrouvées. L'image de droite dévoile l'occultation et montre la précision des résultats obtenus.*

8 points ont été utilisés par ellipse. Pour chaque ellipse, ces 8 points ont été projetés sur l'image recherchée afin de calculer les équations de l'ellipse sur l'image recherchée.

FIG. 6.15 – *Données symboliques retrouvées sur l'objet Tasse.*

6.6 Conclusion

La modélisation 3D à partir d'images 2D permet de reconnaître un objet 3D en présence d'occultations, de scènes comportant un arrière-plan complexe et dans le cas de positions différentes. Comme la reconnaissance se traduit par une recherche d'image, on peut appliquer l'algorithme développé au chapitre précédent. La reconnaissance est donc

robuste dans les mêmes conditions et applicable pour n'importe quel type d'images. En outre, la modélisation proposée repose sur des images, c'est-à-dire sur des données réelles. En effet, nous utilisons ce qui *est perçu* (détecté) réellement dans les images et pas ce qui *doit être vu* dans les images (et qui correspond à ce qu'un humain voudrait voir détecté). Nous nous affranchissons ainsi des problèmes classiques de la modélisation, notamment de celui de détecter dans les images des données abstraites, comme les données CAO ou les données utilisées pour les graphes d'aspect.

Toutefois, la modélisation à partir d'images 2D présente le désavantage de ne pas contenir d'information symbolique 3D. On connaît les appariements entre les deux images, mais ces appariements ne correspondent pas à des données symboliques de l'objet. Il est donc impossible de retrouver par exemple l'anse d'une tasse. De plus, pour reconnaître correctement un objet il n'est pas nécessaire de détecter tous les points contenus dans la base, mais uniquement un sous-ensemble. On n'est donc pas sûr de retrouver un point d'intérêt particulier. En effet, l'information symbolique doit être indépendante du processus d'appariement. Nous proposons donc d'ajouter des données symboliques aux images-modèles. Ensuite ces données symboliques sont projetées sur l'image recherchée en utilisant la relation trilinéaire calculée à partir des appariements.

Une amélioration possible de la méthode proposée serait d'ajouter un mécanisme d'apprentissage ou "clustering" pour le choix des images-modèles. Ceci permettrait un espacement non régulier de ces images pour la représentation d'un objet. Ceci tiendrait compte du fait que les images-modèles se ressemblent plus ou moins suivant le point de vue utilisé pour observer l'objet, et qu'elles sont plus ou moins stables. Il s'agit là d'une extension détaillée dans les perspectives de ce travail (voir le chapitre 7 suivant).

Ayant retrouvé l'image-modèle correspondant à l'image r

Chapitre 7

Conclusion et perspectives

Dans cette thèse a été développée une méthode d'appariement entre images qui est robuste et capable de s'affranchir des limites des méthodes existantes. En outre, cette méthode a permis de poser de façon originale le problème de la perception et de la modélisation d'objets tridimensionnels. Ces deux points constituent les contributions majeures de notre travail. Les sections suivantes présentent tout d'abord un résumé de notre travail et exposent ensuite les perspectives ouvertes.

7.1 Une méthode d'appariement robuste

Notre méthode d'appariement est robuste : elle permet de filtrer les différents types d'erreurs qui apparaissent pendant l'étape de prétraitement d'une image. Les origines de cette robustesse sont multiples. Premièrement, la distance de Mahalanobis permet de tenir compte de l'incertitude des vecteurs de caractéristiques qui peut apparaître à cause du bruit dans les images ou de l'instabilité de la position des points d'intérêt. Deuxièmement, des contraintes de cohérence semi-locale permettent un filtrage des appariements obtenus, qui est résistant aux erreurs de détection. Troisièmement, un algorithme de vote fait émerger une cohérence globale des détections et des appariements calculés. Ceci permet de s'affranchir du bruit de fond des images et augmente encore la robustesse de la détection.

La méthode d'appariement développée est basée sur une caractérisation locale du signal de niveaux de gris. Cette caractérisation est calculée aux points d'intérêt. De ce fait, elle représente une information très riche. Nous avons vu qu'à partir de seulement quelques vecteurs de caractéristiques d'une image, il est possible d'identifier l'image correspondante dans une base d'images. La caractérisation utilisée est également invariante pour le groupe des similitudes d'images et permet d'apparier des images ayant subi de telles transformations. Le groupe des similitudes absorbe au premier ordre les variations dues à des changements de points de vue lors d'une projection perspective (cf. le papier sur les quasi-invariants de Binford [Bin 93]). Notre approche est donc robuste à une telle transformation.

L'utilisation de points d'intérêt permet de manipuler des familles très générales d'images et d'objets, parce qu'ils peuvent être extraits à partir de n'importe quel type de scène. En

outre, les résultats obtenus ne sont pas conditionnés à la détection de segments ni même de contours. Toutefois, dans le cas d'images ne comportant que des objets sans nuance de texture et avec des contours francs, l'apport de notre méthode est nettement moins significative.

7.2 Une modélisation 3D pour la reconnaissance

La méthode d'appariement développée dans ce travail permet une modélisation tridimensionnelle à partir d'images bidimensionnelles. L'idée principale est de rester le plus "proche" possible des données. On "apprend" donc des modèles à partir de données de même type que celles qui seront à reconnaître, c'est-à-dire à partir d'images et non de données artificielles, de type CAO par exemple. Dans ce but, il est nécessaire de déterminer les images-modèles qui représentent un objet donné. Uniquement un sous-ensemble de toutes les vues possibles est nécessaire pour représenter un objet, car les caractéristiques utilisées sont des quasi-invariants des transformations perspectives.

Toutefois, une telle modélisation n'inclut pas d'information de nature tridimensionnelle. Nous proposons donc d'ajouter aux images modèles des données symboliques 3D, comme par exemple des axes de symétrie, des points particuliers, des contours, etc. Pour toute nouvelle image d'un objet représenté dans la base il est ensuite possible de "retrouver", c'est-à-dire de localiser cette information sans mettre en jeu d'algorithme de détection. Cette localisation est réalisée en couplant les appariements obtenus au tenseur trilinéaire liant trois images. Un calcul robuste de ce tenseur permet un positionnement précis de l'information symbolique 3D dans la nouvelle image. Cette approche permet de traiter des objets pour lesquels il n'était pas possible jusqu'à présent d'obtenir une modélisation en utilisant les algorithmes actuels de vision par ordinateur. Ces résultats autorisent la manipulation automatisée d'objets observés par exemple en commande référencée vision (cf. [Esp 92]).

7.3 Perspectives

L'approche développée doit être considérée comme une première réalisation. De nombreuses extensions sont possibles. Il faudrait tout d'abord devenir plus robuste à des changements complexes de luminosité. Ensuite on souhaite pouvoir traiter des bases de taille plus grande. Gérer des objets génériques, c'est-à-dire des classes d'objets, est également un but à atteindre. Résoudre ces problèmes au moins partiellement permet de mettre en place un grand nombre d'applications.

7.3.1 Changement complexe de luminosité

Concernant la caractérisation des points d'intérêt, il faut rendre cette caractérisation plus indépendante des conditions d'éclairage (ou la plus invariante possible). Nous avons pu voir dans le cas d'un changement de luminosité que nous ne pouvons pas utiliser la moyenne des intensités lumineuses autour d'un point. Toutefois, l'information de luminance est représentative d'un point. Dans ce contexte, l'étude de la couleur est une voie intéressante. Pour ce faire, il reste à étudier comment les indices colorimétriques peuvent s'affranchir de l'incidence de la source lumineuse. La réponse n'est pas simple : ainsi en

couleur peu saturée, seule la luminance apporte une information significative ; à l'inverse en couleur saturée, la simple information colorimétrique contient une grande partie de l'information. On peut espérer qu'en ajoutant des invariants colorimétriques un gain significatif sera obtenu. Les invariants colorimétriques permettront de garder cette information de luminance, sous une forme invariante à un changement de luminosité.

7.3.2 Large base d'images

Obtenir une caractérisation plus discriminante est certainement un premier pas pour pouvoir traiter de larges bases d'images. Pour ce faire, il est envisageable d'utiliser la couleur, d'ajouter d'autres descripteurs et d'intégrer l'utilisation d'autres primitives. Toutefois, il est également nécessaire d'organiser la base pour pouvoir traiter de larges quantités de données. On peut prévoir une modélisation plus compacte d'un objet, la pondération par des probabilités et une vérification supplémentaire par une contrainte de cohérence globale. Enfin la généralisation, c'est-à-dire la catégorisation des images, permettra de partitionner la base d'images.

Utilisation de la couleur

L'utilisation de la couleur apporte de l'information supplémentaire. Dans le cas d'images en noir et blanc nous avons une information par point de l'image par rapport à trois informations dans le cas d'images couleur. Ceci permet donc d'obtenir des vecteurs de caractéristiques plus longs et donc plus discriminants. Il est ainsi possible de différencier plus d'objets.

Ajout d'autres descripteurs

Une autre piste pour devenir plus discriminant est l'utilisation d'autres descripteurs. Ceci permettra d'enrichir la caractérisation des points d'intérêt. Notre caractérisation repose sur la décomposition du signal dans une base de fonctions : les invariants sont des compositions des projections du signal sur une base de dérivées de gaussiennes. Il semble intéressant de comparer la caractérisation choisie avec d'autres caractérisations, c'est-à-dire avec d'autres bases de fonctions. Des exemples d'autres caractérisations ont été présentés à la section 3.1.

Intégration avec d'autres primitives de l'image

On peut également utiliser d'autres primitives, comme par exemple les segments. Ceci apporte une information supplémentaire et indépendante qui enrichit la caractérisation. Une réalisation possible est de combiner les vecteurs d'invariants avec des invariants géométriques entre point et segment.

Utiliser d'autres primitives présente comme deuxième avantage de pouvoir traiter tout type de scène. Pour beaucoup d'images, notamment des images qui contiennent de la texture, le choix des points d'intérêt est valable. Toutefois, pour des scènes simples, comportant par exemple les objets polyédriques, le nombre de points d'intérêt est limité. Dans ce cas, l'identification devient plus difficile. L'utilisation d'autres caractéristiques, comme les segments, s'avère donc nécessaire.

Modélisation plus compacte d'un objet 3D

Dans ce travail un objet 3D est modélisé par des images. Réduire le nombre d'images par objet augmente le nombre d'objets qui peuvent être modélisés. On aimerait donc modéliser un objet de la manière la plus compacte possible.

Pour l'instant les vues sont équi-réparties. Un tel espacement n'est pas optimal, car il y a des zones d'un objet qui sont plus stables que d'autres. En utilisant un espacement non équi-réparti, on obtiendra une représentation plus compacte. Pour ce faire, il faut être capable de réunir dans un groupe (clustering) des images voisines, et de là en tirer les caractéristiques stables. Ceci suit l'approche de modélisation commencée par Patrick Gros [Gro 95]. Dans la mesure où notre algorithme de mise en correspondance permet d'induire une distance, cette extension ne devrait pas poser de difficulté majeure.

Pondération par des probabilités

Parmi les vecteurs de caractéristiques stockés dans la base, un certain nombre sont représentatifs de plusieurs objets. D'autres décrivent uniquement un objet. Une pondération par des probabilités conditionnelles permet de donner moins d'importance à des vecteurs peu discriminants. Pour développer une telle méthode, nous proposons d'utiliser la distribution des invariants de la base d'image pour connaître la discriminance d'un invariant donné. Si à un invariant la fonction de densité des invariants est faible, cela signifie que l'invariant apparaît peu souvent dans la base et qu'il est donc représentatif d'un objet. Au contraire, un invariant correspondant à une valeur importante de la fonction de densité est partagé par plusieurs objets. L'inverse de la fonction de densité peut donc être utilisée comme facteur de pondération dans l'algorithme de recherche.

Vérification par cohérence globale

La méthode développée dans ce travail utilise des invariants locaux et des contraintes semi-locales. Nous avons vu, lors de la reconnaissance d'objet 3D et du calcul de la relation trilinéaire entre 3 images, qu'il est possible d'éliminer des erreurs d'appariement par l'utilisation d'une contrainte globale. Dans le cas de la reconnaissance d'objets 3D, cette contrainte est implicitement contenue dans la relation trilinéaire et dans l'utilisation d'une méthode statistique robuste qui rejette les outliers. Nous proposons comme perspective d'utiliser une contrainte globale entre deux images pour pouvoir éliminer des faux appariements. Ceci permettra d'augmenter le taux de reconnaissance et de distinguer deux objets très similaires par leur structure locale.

7.3.3 Généralisation

La généralisation permet de décrire des concepts, comme par exemple un visage, un chien ou une fleur. Pour ce faire, on peut évidemment stocker toutes les images possibles qui représentent un concept, mais ceci est coûteux et rarement exhaustif. Il s'agit donc de trouver des descripteurs qui représentent un concept.

La figure 7.1 illustre ceci pour l'exemple des visages. On peut voir que les nez de différentes personnes se ressemblent plus que l'œil et le nez. Ce fait permet d'apprendre des descripteurs locaux. Ceci peut se faire à partir de nos invariants en utilisant une distance adaptée qui est basée sur des matrices de covariance spécifiques à la variabilité

FIG. 7.1 – *Illustration du principe de la généralisation*

des invariants pour un type de point. On obtient ainsi une distance spécifique par type permettant de savoir de quel type est un point donné. Ceci permet de regrouper tous les points d' un type donné. Apprendre des descripteurs locaux est également possible à partir de critères locaux de texture ou de couleur. Il faudra cependant veiller à rester robuste comme cela a été notre ligne directrice dans ce travail ; cela signifie en particulier ne pas se reposer sur la segmentation en région ; ce type de segmentation est par expérience la plus fragile de toutes.

Enfin il faut relier toutes ces informations et modéliser leurs dépendances géométriques. Pour résoudre ce problème on peut s'imaginer d'ajouter des relations globales ou probabilistes entre ces descripteurs, comme par exemple les réseaux bayésiens.

En conclusion, la généralisation est un problème de recherche à long terme qui a de nombreuses applications. Elle permet par exemple de partitionner une base d'images ce qui est important de le contexte de large base d'images. Ainsi pour rechercher une nouvelle image, on détermine que c'est un visage et ensuite on recherche uniquement dans la sous-base des visages. La généralisation rend également possible l'interrogation de base d'images.

7.3.4 Applications

Nous présentons dans cette section deux applications qui nous semblent particulièrement intéressantes : l'interrogation de larges bases d'images et la modélisation d'une scène 3D par des images. Pour pouvoir interroger des base d'images, il est indispensable d'avoir résolu le problème de la généralisation, c'est-à-dire d'être capable de définir des mesures de similarité. Et pour pouvoir modéliser une scène 3D, il faut être capable de traiter de grandes bases d'images.

Interrogation de larges bases d'images

Le but est de pouvoir répondre à des requêtes par analogie comme par exemple : “je souhaite voir les images qui ont tel aspect”. On veut par exemple trouver dans une large base d'images des images qui contiennent des visages qui ont des yeux noirs.

Il existe aujourd'hui des méthodes apportant des solutions partielles à ce problème. Un premier type de méthodes se base sur les histogrammes de couleur. Cependant de telles méthodes s'avèrent insuffisantes, car il s'agit de méthodes globales dont le pouvoir

discriminant est limité. Il est par exemple impossible avec ces méthodes de distinguer entre un champ de fleurs rouges et un camion de pompier. En outre, l'aspect d'un objet ne se résume pas à sa couleur qui peut varier. D'autres méthodes utilisent la texture des objets. Ces méthodes reposent sur la distribution statistique de textures particulières dans l'image. Toutefois, l'inconvénient majeur de ces méthodes est qu'elles procèdent par une mesure globale ; ceci limite énormément le domaine d'application. Un autre type de méthodes d'interrogation de bases d'image utilise des systèmes basés sur l'information textuelle. De tels systèmes permettent de retrouver facilement les images associées à une information particulière. Toutefois, le texte est ajouté à priori et souvent peu représentatif. En outre, l'ajout doit se faire de façon manuelle et est donc coûteux.

Nous proposons d'indexer par le contenu des images. Cette recherche doit être basée sur une mesure de ressemblance intégrant la notion de généralisation. En outre, une interrogation doit se faire en interaction avec l'utilisateur qui définit dynamiquement les critères de sa recherche.

Modélisation de scène 3D

La modélisation d'une scène 3D à partir d'images bidimensionnelles est une extension de la modélisation compacte d'un objet 3D. On peut imaginer de modéliser un espace 3D par une collection d'images et ensuite d'utiliser cette collection d'images comme base de représentation pour se déplacer dans l'espace. Avec des outils capables de retrouver une posture à partir de milliers d'images, on peut espérer se positionner par rapport à des points d'observation utilisés lors de l'apprentissage. Reste alors à voir comment on peut déterminer les positions spatiales en fonction de l'image qui lui était présentée, et comment la combiner avec les images voisines. Ceci peut être appliqué à des tâches de positionnement relatif.

Bibliographie

[Ack 84] F. Ackermann. Digital image correlation : performance and potential application in photogrammetry. *Photogrammetric Record*, 64(11): 429–439, 1984.

[Asa 86] H. Asada et M. Brady. The curvature primal sketch. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(1): 2–14, 1986.

[Aya 86] N. Ayache et O. Faugeras. HYPER : a new approach for the recognition and positioning of 2D objects. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(1): 44–54, 1986.

[Bal 81] D. H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2): 111–122, 1981.

[Bau 96] C. Bauckhage et C. Schmid. Evaluation of keypoint detectors. Rapport technique, INRIA, 1996.

[Bea 78] P.R. Beaudet. Rotationally invariant image operators. Dans *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pages 579–583, 1978.

[Bea 94] P. Beardsley, A. Zisserman et D. Murray. Sequential update of projective and affine structure from motion. Rapport technique 2012/94, University of Oxford, 1994.

[Ber 94] M. Berthod et G. Giraudon. About grey-level invariants. Vérsion préliminaire non publiée, 1994.

[Bes 85] P.J. Besl et R.C. Jain. Three-dimensional object recognition. ACM *Computing Surveys*, 17(1): 75–145, 1985.

[Big 94] J. Bigün et M.H. du Buf. N-folded symmetries by complex moments in gabor space and their application to unsupervised texture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1): 80–87, 1994.

[Big 95] J. Bigün. Pattern recognition in images by symmetries and coordinate transformations. *Computer Vision and Image Understanding*, 1995. Soumis.

[Bin 93] T.O. Binford et T.S. Levitt. Quasi-invariants : theory and exploitation. Dans *Proceedings of* DARPA *Image Understanding Workshop*, pages 819–829, 1993.

[Bob 96] P. Bobet, J. Blanc et R. Mohr. Aspects cachés de la trilinéarité. Dans *Actes du 10ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle*, pages 137–146, 1996.

[Bol 86] R. C. Bolles et R. Horaud. 3DPO : a three-dimensional part orientation system. *International Journal of Robotics Research*, 5(3): 3–26, 1986.

[Bou 95] B. Boufama et R. Mohr. Epipole and fundamental matrix estimation using the virtual parallax property. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 1030–1036, 1995.

[Bow 91] K. Bowyer. Why aspect graphs are not (yet) practical for computer vision. Dans *Proceedings of the* IEEE *Workshop on Direction on automated CAD-based Vision*, pages 97–104, 1991.

[Bra 87] M. Brady. Seeds of perception. Dans *Proceedings of the 3rd Alvey Vision Conference*, pages 259–265, 1987.

[Bra 94] P. Brand et R. Mohr. Accuracy in image measure. Dans *Proceedings of the* SPIE *Conference on Videometrics III*, volume 2350, pages 218–228, 1994.

[Bra 95] P. Brand. *Reconstruction tridimensionnelle d'une scène à partir d'une caméra en mouvement: de l'influence de la précision.* Thèse de doctorat, Université Claude Bernard, Lyon I, 1995.

[Bro 83] R. A. Brooks. Model-based three-dimensional interpretations of two-dimensional images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 5(2): 140–150, 1983.

[Bur 81] P. J. Burt. Fast filter transforms for image processing. *Computer Graphics and Image Processing*, 16: 20–51, 1981.

[Bur 90] J. B. Burns, R. Weiss et E. M. Riseman. View variation of point set and line segment features. Dans *Proceedings of* DARPA *Image Understanding Workshop*, pages 650–659, 1990.

[Cal 94] A. Califano et R. Mohan. Multidimensional indexing for recognizing visual shapes. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 16(4): 373–392, 1994.

[Can 86] J. Canny. A computational approach to edge detection. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(6): 679–698, 1986.

[Che 91] C.H. Chen et P.G. Mulgaonkar. CAD-based feature-utility measures for automatic vision programming. Dans *Proceedings of the* IEEE *Workshop on Direction in Automated CAD-Based Vision*, pages 106–114, 1991.

[Chi 86] R. T. Chin, H. Smith et S. C. Fralick. Model-based recognition in robot vision. *ACM Computing Surveys*, 18(1): 67–108, 1986.

[Cle 90] D. J. Clemens et D. W. Jacobs. Model-group indexing for recognition. Dans *Proceedings of* DARPA *Image Understanding Workshop*, pages 604–613, 1990.

[Cot 94] J. C. Cottier. *Extraction et appariements robustes des points d'intérêt de deux images non étalonnées.* Stage de maîtrise, 1994.

[Cro 81] J. L. Crowley. *A representation for visual information.* Thèse de doctorat, Carnegie-Mellon University, 1981.

[Cro 84] J. L. Crowley et A. C. Parker. A representation for shape based on peaks and ridges in the difference of low pass transform. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 6(2): 156–170, 1984.

[Der 90] R. Deriche et G. Giraudon. Accurate corner detection : an analytical study. Dans *Proceedings of the 3rd International Conference on Computer Vision*, 1990.

[Der 93a] R. Deriche. Recursively implementing the gaussian and its derivatives. Rapport technique, INRIA, 1993.

[Der 93b] R. Deriche et T. Blaszka. Recovering and characterizing image features using an efficient model based approach. Dans *Proccedings of the Conference on Computer Vision and Pattern Recognition*, pages 530–535, 1993.

[Der 93c] R. Deriche et G. Giraudon. A computational approach for corner and vertex detection. *International Journal of Computer Vision*, 10(2): 101–124, 1993.

[Der 94] R. Deriche, Z. Zhang, Q.-T. Luong et O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. Dans *Proceedings of the 3rd European Conference on Computer Vision*, pages 567–576, 1994.

[Die 94] M. Van Diest, L. Van Gool, T. Moons et E. Pauwels. Projective invariants for planar contour recognition. Dans *Proceedings of the 3rd European Conference on Computer Vision*, pages 527–534, 1994.

[Dre 82] L. Dreschler et H.-H. Nagel. Volumetric model and 3D trajectory of a moving car derived from monocular tv frame sequences of a street scene. *Computer Graphics and Image Processing*, 20: 199–228, 1982.

[Egg 89] D. Eggert et K. Bowyer. Computing the orthographic projection aspect graph of solid of revolution. Dans *Proceedings of the* IEEE *Workshop on Interpretation of 3D Scenes*, pages 102–108, 1989.

[Esp 92] B. Espiau, F.Chaumette et P. Rives. A new approach to visual servoing in robotics. IEEE *Transactions on Robotics and Automation*, 8(3): 313–326, 1992.

[Fau 86] O. Faugeras et M. Hebert. The representation, recognition, and locating of 3-D objects. *International Journal of Robotics Research*, 5: 27–52, 1986.

[Fau 92a] O. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat et Z. Zhang. Quantitative and qualitative comparisons of some area and feature-based stereo algorithms. Dans *Robust Computer Vision*, pages 1–26, 1992.

[Fau 92b] O. Faugeras, Q.-T. Luong et S. J. Maybank. Camera self-calibration: theory and experiments. Dans *Proceedings of the 2nd European Conference on Computer Vision*, pages 321–334, 1992.

[Fau 95] O. Faugeras et B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 951–956, 1995.

[Fle 91] D. J. Fleet, A. D. Jepson et M. R. M. Jenkin. Phase-base disparity measurement. *Computer Vision Graphics and Image Processing*, 53(2): 198–210, 1991.

[Flo 93] L. M. J. Florack. *The syntactical structure of scalare images.* Thèse de doctorat, Universiteit Utrecht, 1993.

[Flo 94] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink et M. A. Viergever. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4: 171–187, 1994.

[För 87] W. Förstner et Gülch. A fast operator for detection and precise location of distinct points, corners and circular features. Dans *Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, 1987.

[För 94] W. Förstner. A framework for low level feature extraction. Dans *Proceedings of the 3rd European Conference on Computer Vision*, 1994.

[Fre 91] W. T. Freeman et E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern and Machine Intelligence*, 13(9): 891–906, 1991.

[Fun 95] B. Funt et G. Finlayson. Color constant color indexing. *IEEE Transactions on Pattern and Machine Intelligence*, 17(5):522–529, 1995.

[Gab 46] D. Gabor. Theory of communication. *Proceedings of Inst. Elec. Eng.*, 93(26): 429–441, 1946.

[Gda 96] Y. Gdalyahu et D. Weinshall. Measures for silhouettes resemblance and representative silhouettes of curved objects. Dans *Proceedings of the 4th European Conference on Computer Vision*, pages 363–375, 1996.

[Gig 91] Z. Gigus, J. Canny et R. Seidel. Efficiently computing and representing aspect graphs of polyhedral objects. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13(6): 542–551, 1991.

[Gir 91] G. Giraudon et R. Deriche. On corner and vertex detection. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 650–655, 1991.

[Goo 96] L. Van Gool, T. Moons et D. Ungureanu. Affine / photometric invariants for planar intensity patterns. Dans *Proceedings of the 4th European Conference on Computer Vision*, pages 642–651, 1996.

[Gra 91] A. E. Grace et M. Spann. A comparison between Fourier-Mellin descriptors and moment based features for invariant object recognition using neural networks. *Pattern Recognition Letters*, 12: 635–643, 1991.

[Gri 87] W. E. .L. Grimson et T. Lozano-Perez. Localizing overlapping parts by searching the interpretation tree. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 9:(4) 469–482, 1987.

[Gri 89] W. E. L. Grimson. On the recognition of curved objects. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(3): 632–643, 1989.

[Gri 90] W. E. L. Grimson et D. P. Huttenlocher. On the sensitivity of the Hough transform for object recognition. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 12(3): 225–274, 1990.

[Gro 92] P. Gros et L. Quan. Projective invariants for vision. Rapport technique RT 90 IMAG - 15 LIFIA, Grenoble, 1992.

[Gro 95] P. Gros. Matching and clustering: Two steps towards object modelling in computer vision. *International Journal of Robotics Research*, 14(6): 633–642, 1995.

[Mor 83] A. Grossmann et J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math.*, 15: 723–736, 1984.

[Har 88] C. Harris et M. Stephens. A combined corner and edge detector. Dans *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.

[Har 92] R. Hartley. Invariants of points seen in multiple images. Rapport technique, G.E. CRD, Schenectady, 1992.

[Heb 85] M. Hebert et T. Kanade. The 3D profile method for object recognition. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1985.

[Hei 92] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans et O. Kuebler. Simulation of neural contour mechanism: from simple to end-stopped cells. *Vision Research*, 32(5): 963–981, 1992.

[Hil 93] D. Hilbert. Ueber die vollen Invariantensystemen. *Math. Annalen*, 42: 313–373, 1893.

[Hor 89] R. Horaud et T. Skordas. Stereo correspondence through feature grouping and maximal cliques. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 11(11): 1168–1180, 1989.

[Hor 90] R. Horaud, T. Skordas et F. Veillon. Finding geometric and relational structures in an image. Dans *Proceedings of the 1st European Conference on Computer Vision*, pages 374–384, 1990.

[Hu 62] M. K. Hu. Visual pattern recognition by moment invariants. IEEE *Transactions on Information Theory*, 8: 179–187, 1962.

[Hu 94] X. Hu et N. Ahuja. Feature extraction and matching as signal detection. *International Journal of Pattern Recognition and Artificial Intelligence*, 8(6): 1343–1379, 1994.

[Hut 90] D. P. Huttenlocher et S. Ullman. Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, 5(2): 195–212, 1990.

[Ike 88] K. Ikeuchi et T. Kanade. Applying sensor models to automatic generation of object recognition programs. Dans *Proceedings of the 2nd International Conference on Computer Vision*, pages 228–237, 1988.

[Jac 91] L. Jacobson et H. Wechsler. Derivation of optical flow using a spatiotemporal-frequency approach. *Computer Vision, Graphics and Image Processing*, 38: 29–65, 1991.

[Kim 94] W.-Y. Kim et P. Yuan. A practical pattern recognition system for translation, scale and rotation invariance. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 391–396, 1994.

[Kit 82] L. Kitchen et A. Rosenfeld. Gray-level corner detection. *Pattern Recognition Letters*, 1: 95–102, 1982.

[Koe 79] J. J. Koenderink et A. V. van Doorn. The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32: 211–216, 1979.

[Koe 84] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50: 363–396, 1984.

[Koe 87] J. J. Koenderink et A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55: 367–375, 1987.

[Kri 89] D. J. Kriegman et J. Ponce. Computing exact aspect graphs of curved objects: Solids of revolution. Dans *Proceedings of the* IEEE *Workshop on 3D Scene*, pages 116–122, 1989.

[Kri 90] D. J. Kriegman et J. Ponce. On recognizing and positioning curved 3-D objets from image contours. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 12(12): 1127–1137, 1990.

[Lad 93] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P . Würtz et W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transanctions on Computers*, 42(3): 300–311, 1993.

[Lam 88] Y. Lamdan, J. T. Schwartz et H. J. Wolfson. Object recognition by affine invariant matching. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 335–344, 1988.

[Lin 94] T. Lindeberg. *Scale-Space Theory in Computer Vision.* Kluwer Academic Publishers, 1994.

[Lon 86] P. Long et G. Giraudon. Stereo matching based on contextual line-region primitives. Dans *Proceedings of the 8th International Conference on Pattern Recognition*, 1986.

[Lot 94] J. L. Lotti et G. Giraudon. Adaptive window algorithm for aerial image stereo. Dans *Proceedings of the 12th International Conference on Pattern Recognition*, pages 701–703, 1994.

[Low 86] D. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, pages 355–395, 1986.

[Mal 89] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7): 674–693, 1989.

[Med 87] G. Medioni et Y. Yasumoto. Corner detection and curve representation using cubic B-splines. *Computer Vision, Graphics and Image Processing*, 39(1): 267–278, 1987.

[Mey 91] Y. Meyer. Ondelettes et fonctions splines. Dans *Sem. Equations aux Dérivées Partielles*, École Polytechnique, Paris, 1986.

[Mok 86] F. Mokhtarian et A. Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 8(1): 34–43, 1986.

[Mor 79] H. Moravec. Visual mapping by a robot rover. Dans *Proceedings of the 6th International Joint Conference on Artifical Intelligence*, pages 598–600, 1979.

[Mor 81] H. Moravec. Rover visual obstacle avoidance. Dans *Proceedings of the 7th International Joint Conference on Artifical Intelligence*, pages 785–790, 1981.

[Mos 92] Y. Moses et S. Ullman. Limitations of non model–based recognition. Dans *Proceedings of the 2nd European Conference on Computer Vision*, pages 820–828, 1992.

[Mun 90] J. L. Mundy et A. J. Heller. The evolution and testing of a model-based object recognition system. Dans *Proceedings of the 3rd International Conference on Computer Vision*, pages 268–282, 1990.

[Mun 92a] J. L. Mundy et A. Zisserman. Projective geometry for machine vision. Dans *Geometric Invariance in Computer Vision*, chapitre 23, pages 463–519. MIT Press, 1992.

[Mun 92b] J. L. Mundy et A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[Mur 95] H. Murase et S. K. Nayar. Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14: 5–24, 1995.

[Nag 83] H.-H. Nagel. Displacement vectors derived from second order intensity variations in image sequences. *Computer Vision, Graphics and Image Processing*, 21: 85–117, 1983.

[Nag 95] K. Nagao. Recognizing 3D objects using photometric invariant. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 480–487, 1995.

[Nay 93] S. K. Nayar et R. M. Bolle. Computing reflectance ratios from an image. *Pattern Recognition*, 26(10): 1529–1542, 1993.

[Nob 88] J. A. Noble. Finding corners. *Image and Vision Computing*, 6(2): 121–128, 1988.

[Per 95] P. Perona. Deformable kernels for early vision. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(5): 488–499, 1995.

[Pet 92] S. Petitjean, J. Ponce et D. J. Kriegman. Computing exact aspect graphs of curved objects : algebraic surfaces. *International Journal of Computer Vision*, 9(3): 231–255, 1992.

[Rao 95] R. P. N. Rao et D. H. Ballard. Object indexing using an iconic sparse distributed memory. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 24–31, 1995.

[Rei 95] D. Reisfeld, H. J. Wolfson et Y. Yeshurun. Context-free attentional operators : the generalized symmetry transform. *International Journal of Computer Vision*, 14: 119–130, 1995.

[Rem 94] P. Remagnino, P. Brand et R. Mohr. Correlation techniques in adaptative template matching with uncalibrated cameras. Dans *Vision Geometry III, SPIE's international symposium on photonic sensors & control for commercial applications*, volume 2356, pages 252–253, 1994.

[Rie 87] J. H. Rieger. On the classification of views of piecewise smooth objects. *Image and Vision Computing*, 5(2): 91–97, 1987.

[Roh 90] K. Rohr. Über die Modellierung und Identifikation charakteristischer Grauwertverläufe in Realwertbildern. Dans *12. DAGM-Symposium Mustererkennung*, 1990.

[Roh 92] K. Rohr. Recognizing corners by fitting parametric. *International Journal of Computer Vision*, 9(3): 213–230, 1992.

[Rom 94a] B. M. ter Haar Romeny. *Geometry-Driven Diffusion in Computer Vision.* Kluwer Academic Publishers, 1994.

[Rom 94b] B. M. ter Haar Romeny, L. M. J. Florack, A. H. Salden et M. A. Viergever. Higher order differential structure of images. *Image and Vision Computing*, 12(6): 317–325, 1994.

[Ros 92] L. Rosenthaler, F. Heitger, O. Kuebler et R. von der Heydt. Detection of general edges and keypoints. Dans *Proceedings of the 2nd European Conference on Computer Vision*, pages 78–86, 1992.

[Rot 92] C.A. Rothwell, A. Zisserman, D. A. Forsyth et J. L. Mundy. Canonical frames for planar object recognition. Dans *Proceedings of the 2nd European Conference on Computer Vision*, pages 757–772, 1992.

[Rot 93] C.A. Rothwell. Hierarchical object descriptions using invariants. Dans *Proceeding of the* DARPA–ESPRIT *Workshop on Applications of Invariants in Computer Vision*, pages 287–303, 1993.

[Rub 91] J. Rubinstein, J. Segman et Y. Zeevi. Recognition of distorted patterns by invariance kernels. *Pattern Recognition*, 24(10): 959–967, 1991.

[Sal 92] A. H. Salden, B. M. ter Haar Romeny, L. M. J. Florack, M. A. Viergever et J. J. Koenderink. A complete and irreducible set of local orthogonally invariant features of 2-dimensional images. Dans *Proceedings of the 11th International Conference on Pattern Recognition*, pages 180–184, 1992.

[San 88] T. Sanger. Stereo disparity computation using gabor filters. *Biological Cybernetics*, 2(59): 405–418, 1988.

[Sch 95] H. Schulz-Mirbach. Anwendung von Invarianzprinzipien zur Merkmalgewinnung in der Mustererkennung. Thèse de doctorat, Technische Universität Hamburg, 1995.

[Sch 96] B. Schiele et J. L. Crowley. Object recognition using multidimensional receptive field histograms. Dans *Proceedings of the 4th European Conference on Computer Vision*, pages 610–619, 1996.

[Sha 78] S. D. Shapiro. Feature space transforms for curve detection. *Pattern Recognition*, 10(3): 129–143, 1978.

[Sha 84] M.A. Shah et R. Jain. Detecting time-varying corners. *Computer Vision, Graphics and Image Processing*, 28: 345–355, 1984.

[Sha 91] L. Shapiro et K. Bowyer. *Proceedings of the* IEEE *Workshop on Directions in Automated CAD-Based Vision*, 1991.

[Sha 94] A. Shashua. Trilinearity in visual recognition by alignment. Dans *Proceedings of the 3rd European Conference on Computer Vision*, pages 479–484, 1994.

[Sla 96] D. Slater et G. Healey. The illumination-invariant recognition of 3D objects using color invariants. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 18(2): 206–210, 1996.

[Ste 88] J. Stewman et K. Bowyer. Creating the perspective projection aspect graph of polyhedral objects. Dans *Proceedings of the 2nd International Conference on Computer Vision*, pages 494–500, 1988.

[Swa 91] M. Swain et D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 32(11):11–32, 1991.

[Tea 80] M. R. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70: 920–930, 1980.

[Teh 88] C.-H. Teh et R. T. Chin. On image analysis by the methods of moments. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 10(4): 496–513, 1988.

[Tor 86] V. Torre et T. A. Poggio. On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2): 147–163, 1986.

[Tur 91] M. A. Turk et A. P. Pentland. Face recognition using eigenfaces. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[Wan 92] H. Wang et J. M. Brady. Corner detection with subpixel accuracy. Rapport technique OUEL 1925/92, Dept. Engineering Science, Oxford University, 1992.

[Wei 91] I. Weiss. Noise-resistant invariant of curves. Dans *Proceeding of the* DARPA–ESPRIT *Workshop on Applications of Invariants in Computer Vision*, pages 319–344, 1991.

[Wei 92] I. Weiss. Noise resistant projective and affine invariants. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 115–121, 1992.

[Wes 92] C. J. Westelius, H. Knutsson et J. Wiklund. Robust vergence control using scale-space phase information. Rapport technique LiTH-ISY-I-1363, Linköpings tekniska högskola, Department of Electrical Engineering, 1992.

[Wit 83] A. P. Witkin. Scale-space filtering. Dans *Proceedings of the 8th International Joint Conference on Artifical Intelligence*, pages 1019–1023, 1983.

[Wol 90] H. J. Wolfson. Model-based object recognition by geometric hashing. Dans *Proceedings of the 1st European Conference on Computer Vision*, pages 526–536, 1990.

[Wu 95] X. Wu et B. Bhanu. Gabor wavelets for 3D object recognition. Dans *Proceedings of the 5th International Conference on Computer Vision*, pages 537–542, 1995.

[Zab 94] R. Zabih et J. Woodfill. Non-parametric local transforms for computing visual correspondance. Dans *Proceedings of the 3rd European Conference on Computer Vision*, pages 151–158, 1994.

[Zha 89] J. Zhao. *Extraction d'information tridimensionnelle par stéréovision.* Thèse de doctorat, Université Paul Sabatier, Toulouse, 1989.

[Zha 95] Z. Zhang, R. Deriche, O. Faugeras et Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78: 87–119, 1995.

[Zis 95] A. Zisserman, D. A. Forsyth, J. L. Mundy, C. Rothwell, J. Liu et N. Pillow. 3D object recogntion using invariance. *Artificial Intelligence*, 78(1–2): 239–288, 1995.

[Zun 83] O.A. Zuniga et R.M. Haralick. Corner detection using the facet model. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 30–37, 1983.

Annexe A

Répétabilité des points d'intérêt sur la scène "Astérix"

Dans cette annexe, les détecteurs de points d'intérêt sont évalués pour la scène "Astérix". Cette annexe donne des résultats supplémentaires au chapitre 2. Dans la suite le taux de répétabilité est donné pour les différentes transformations considérées.

Rotation image

FIG. A.1 – *À gauche l'image de référence pour la séquence rotation image et à droite l'image avec une angle de rotation de 154 degrés.*

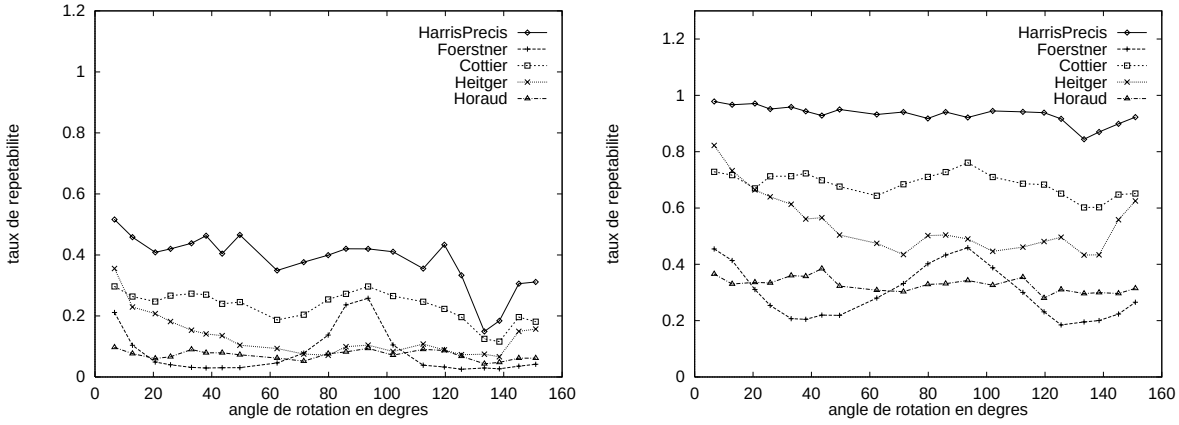

FIG. A.2 – *Taux de répétabilité pour la séquence rotation image et la scène "Astérix". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

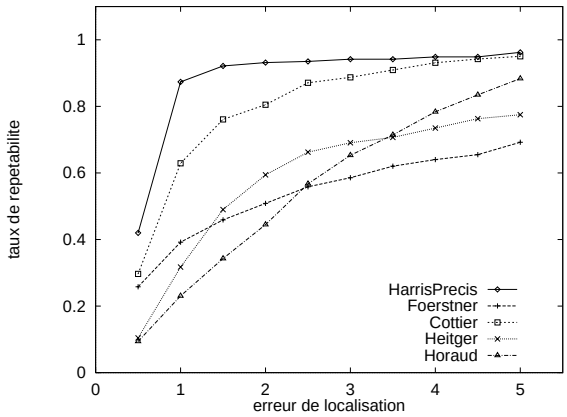

FIG. A.3 – *Taux de répétabilité pour un angle de rotation de 93 degrés et la scène "Astérix".*

Changement d'échelle

FIG. A.4 – *À gauche l'image de référence pour la séquence changement d'échelle et à droite la dernière image de cette séquence. Le changement d'échelle entre les deux est de 4.1.*

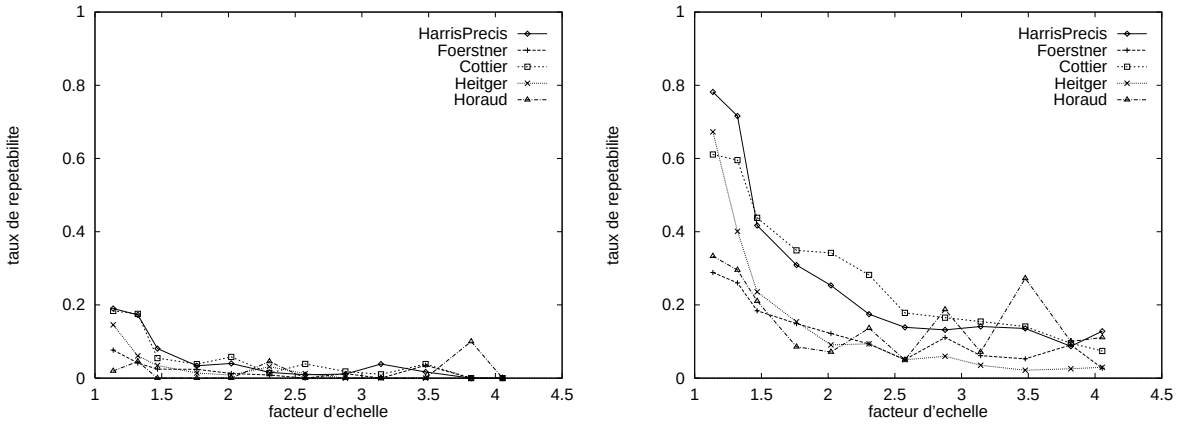

FIG. A.5 – *Taux de répétabilité pour la séquence changement d'échelle et la scène "Astérix". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

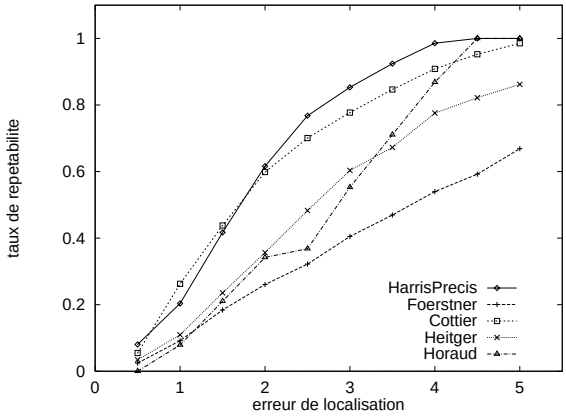

FIG. A.6 – *Taux de répétabilité pour un changement d'échelle de 1.5 et la scène "Astérix".*

Changement de la luminosité

FIG. A.7 – *À gauche l'image de la séquence changement uniforme de luminosité avec un niveau de gris relatif de 0.6 et à droite l'image avec un niveau de gris relatif de 1.5.*

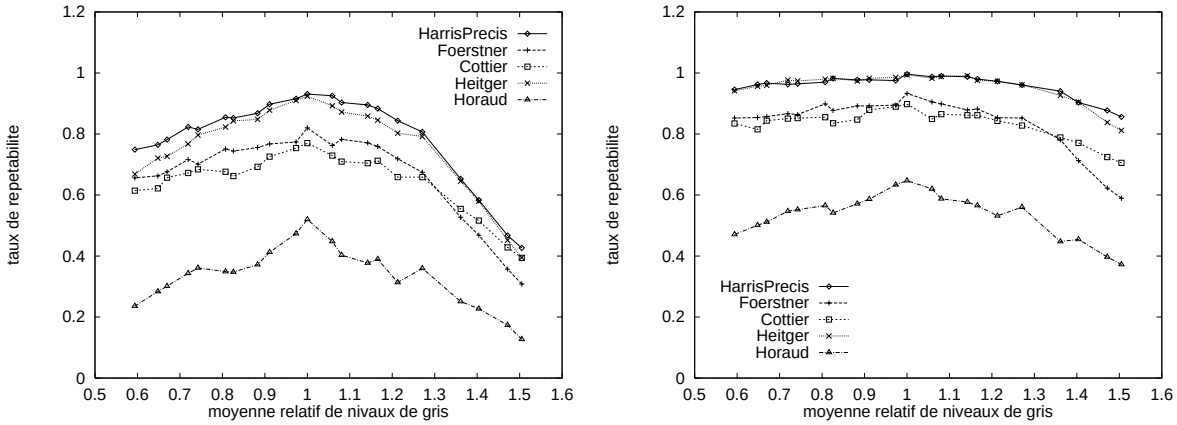

FIG. A.8 – *Taux de répétabilité pour la séquence changement uniforme de luminosité et la scène "Astérix". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

Bruit

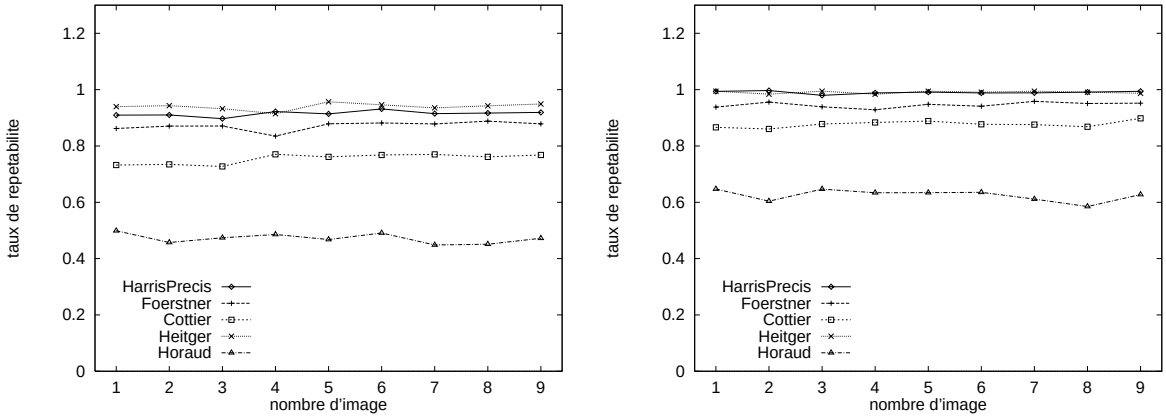

FIG. A.9 – *Taux de répétabilité pour la séquence bruit de la caméra et la scène "Astérix". Pour le graphe de gauche $\varepsilon = 0.5$ et pour le graphe de droite $\varepsilon = 1.5$.*

Comparaison Harris et HarrisPrécis

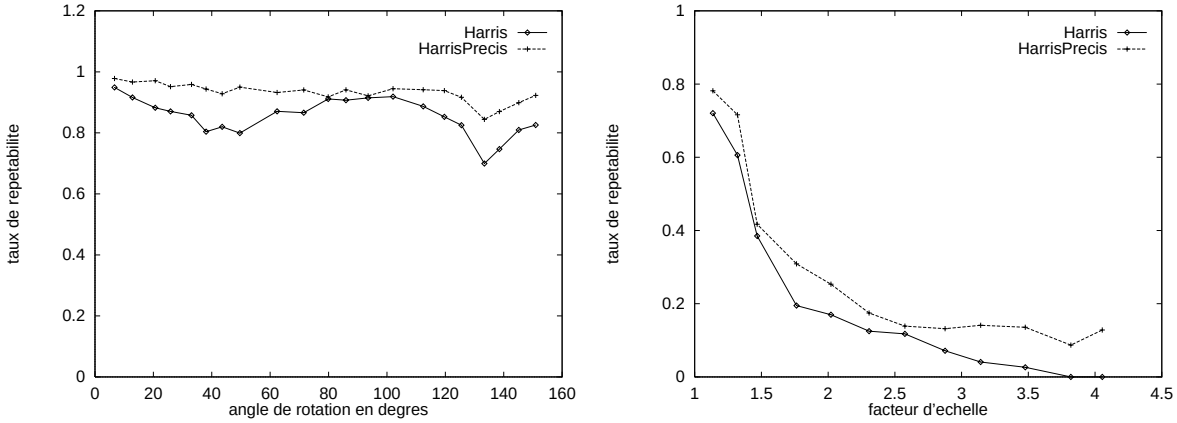

FIG. A.10 – *Comparaison de Harris et HarrisPrécis. À gauche pour la séquence rotation image et à droite pour la séquence changement d'échelle. La scène utilisée est "Astérix" et $\varepsilon = 1.5$.*

Adaptation à l'échelle de HarrisPrécis

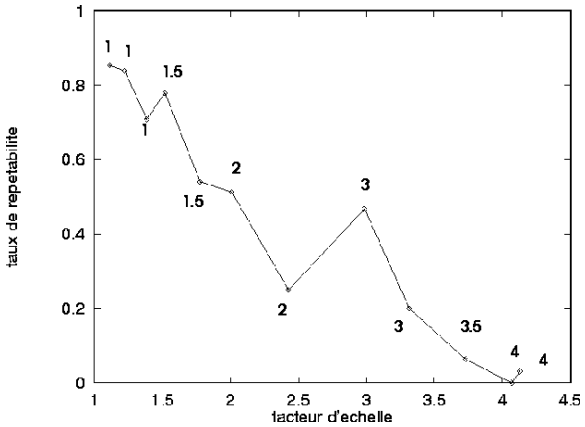

FIG. A.11 – *Adaptation de HarrisPrécis à un changement d'échelle. La scène utilisée est "Van Gogh" et $\varepsilon = 1.5$.*

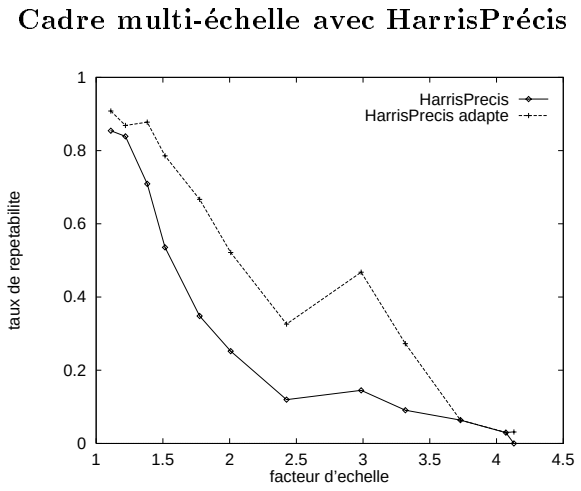

FIG. A.12 – *Utilisation d'un cadre multi-échelle pour HarrisPrécis. La scène utilisée est "VanGogh" et $\varepsilon = 1.5$.*

Annexe B

Évaluation de l'appariement pour la scène "Sanja"

Dans cette annexe l'appariement est évalué pour la scène "Sanja". Cette annexe donne des résultats supplémentaire au chapitre 4. Dans la suite le taux d'appariements corrects est donné pour les différentes transformations considérées.

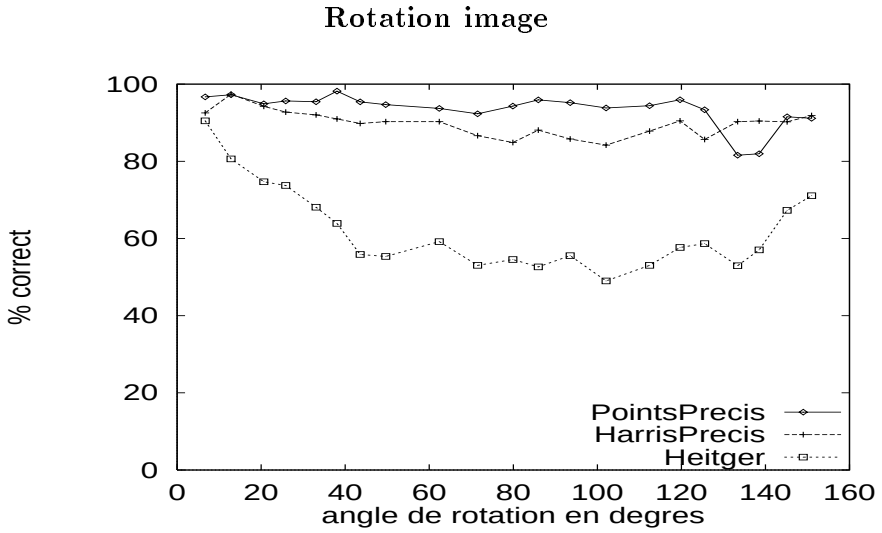

FIG. B.1 – *Pourcentage d'appariements corrects pour la séquence rotation image et la scène "Sanja". Les trois courbes correspondent aux différents détecteurs de points d'intérêt utilisés : Heitger, HarrisPrécis et les points précis.*

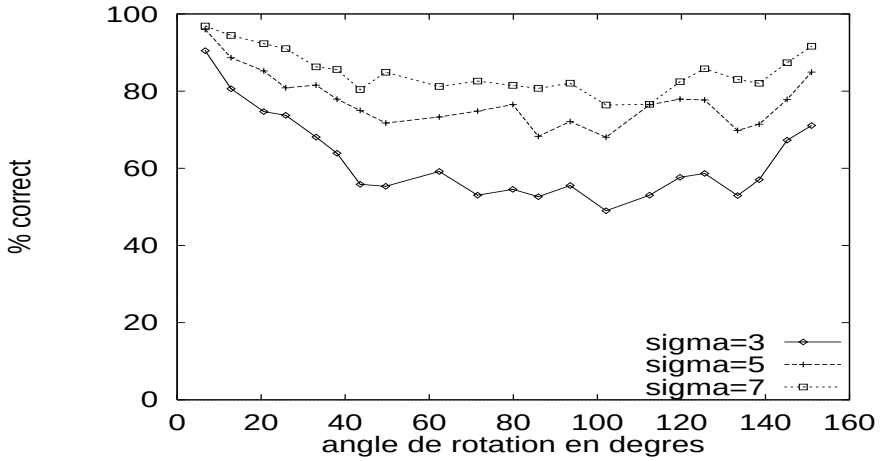

FIG. B.2 – *Pourcentage d'appariements corrects pour la séquence rotation image et la scène "Sanja". Le détecteur utilisé est Heitger. Les différentes courbes correspondent à différentes tailles σ de gaussienne.*

Changement d'échelle

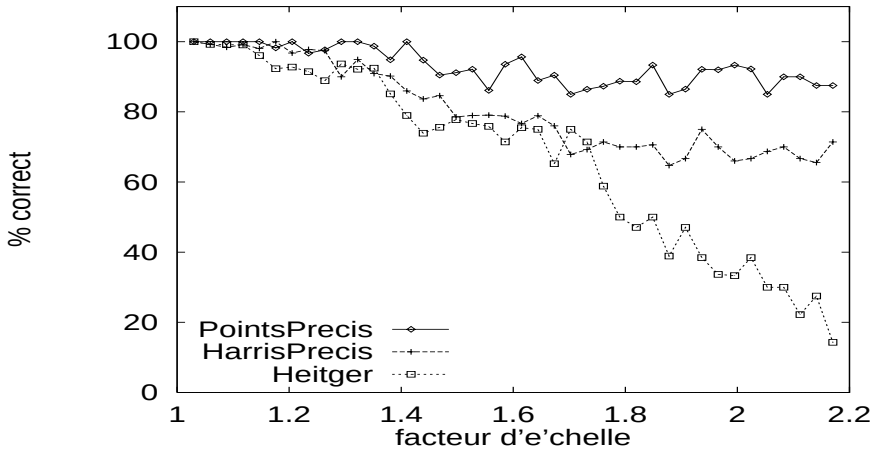

FIG. B.3 – *Pourcentage d'appariements corrects pour la séquence changement d'échelle et la scène "Sanja" en utilisant une approche multi-échelle. Les trois courbes correspondent aux différents détecteurs de points d'intérêt utilisés : Heitger, HarrisPrécis et les points précis. La taille moyenne de la gaussienne utilisée est de 5.*

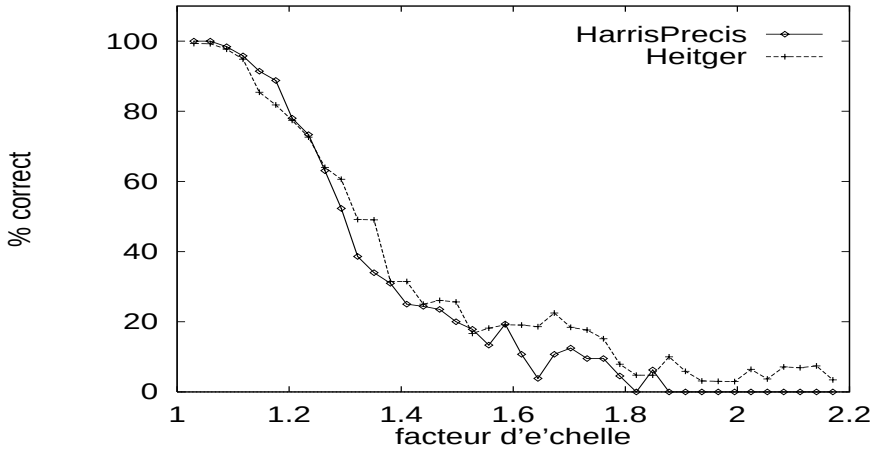

FIG. B.4 – *Pourcentage d'appariements corrects pour la séquence changement d'échelle et la scène "Sanja" sans utiliser une approche multi-échelle. Les deux courbes correspondent aux différents détecteurs de points d'intérêt utilisés : Heitger et HarrisPrécis. La taille de la gaussienne utilisée est de* 5.

Changement de la luminosité

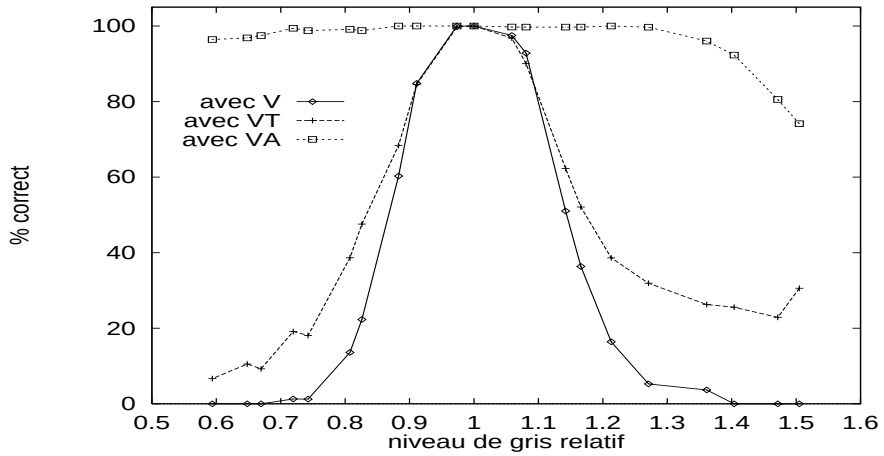

FIG. B.5 – *Pourcentage d'appariements corrects pour la séquence changement uniforme de luminosité et la scène "Sanja". Les trois courbes correspondent aux différents vecteurs d'invariants utilisés :* $\vec{\mathcal{V}}$, $\vec{\mathcal{V}}_{\mathcal{T}}$ *et* $\vec{\mathcal{V}}_{\mathcal{A}}$.

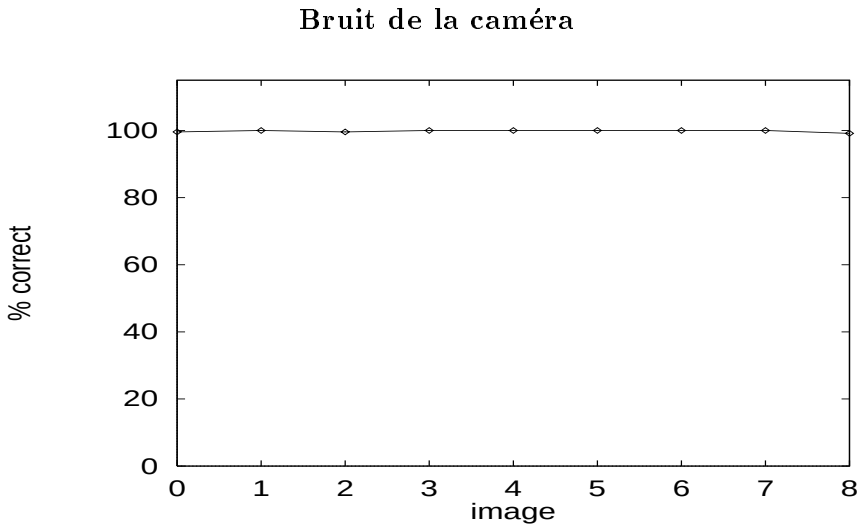

FIG. B.6 – *Pourcentage d'appariements corrects pour la séquence bruit de la caméra et la scène "Sanja".*

Annexe C

Quelques images de la base

Cette annexe présente quelques images de la base utilisée pour nos expériences. La figure C.1 montre quelques images de tableaux et la figure C.2 quelques images aériennes. Plus de détails sur la base sont donné à la section 5.4.1.

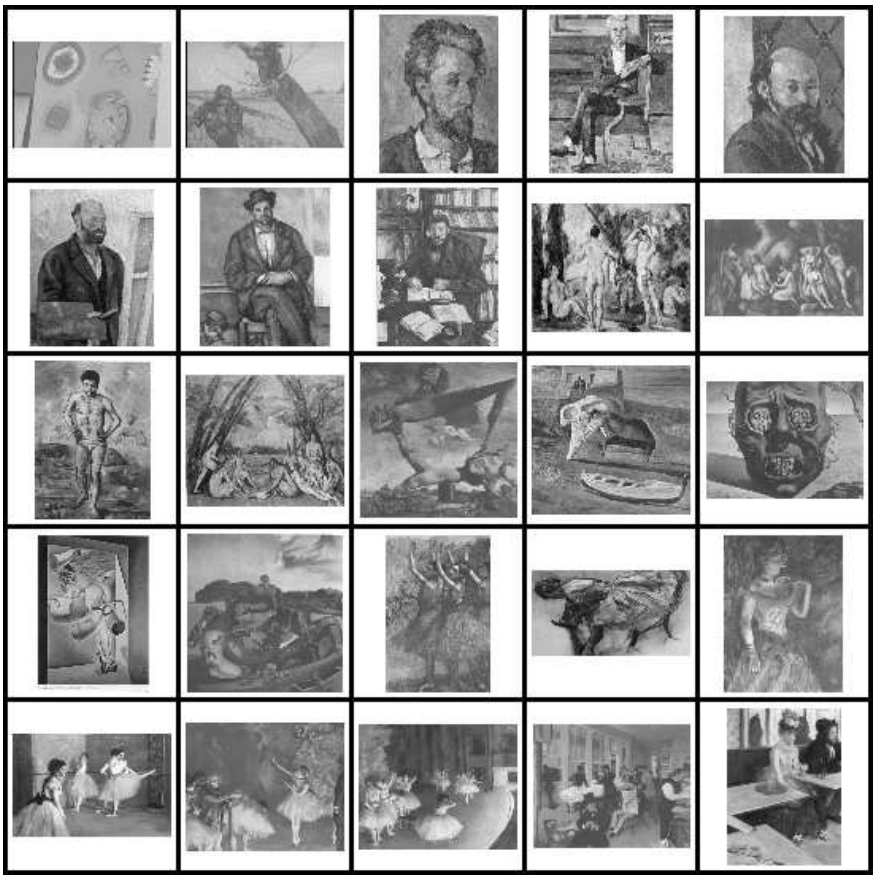

FIG. C.1 – Quelques images de tableaux de notre base d'images.

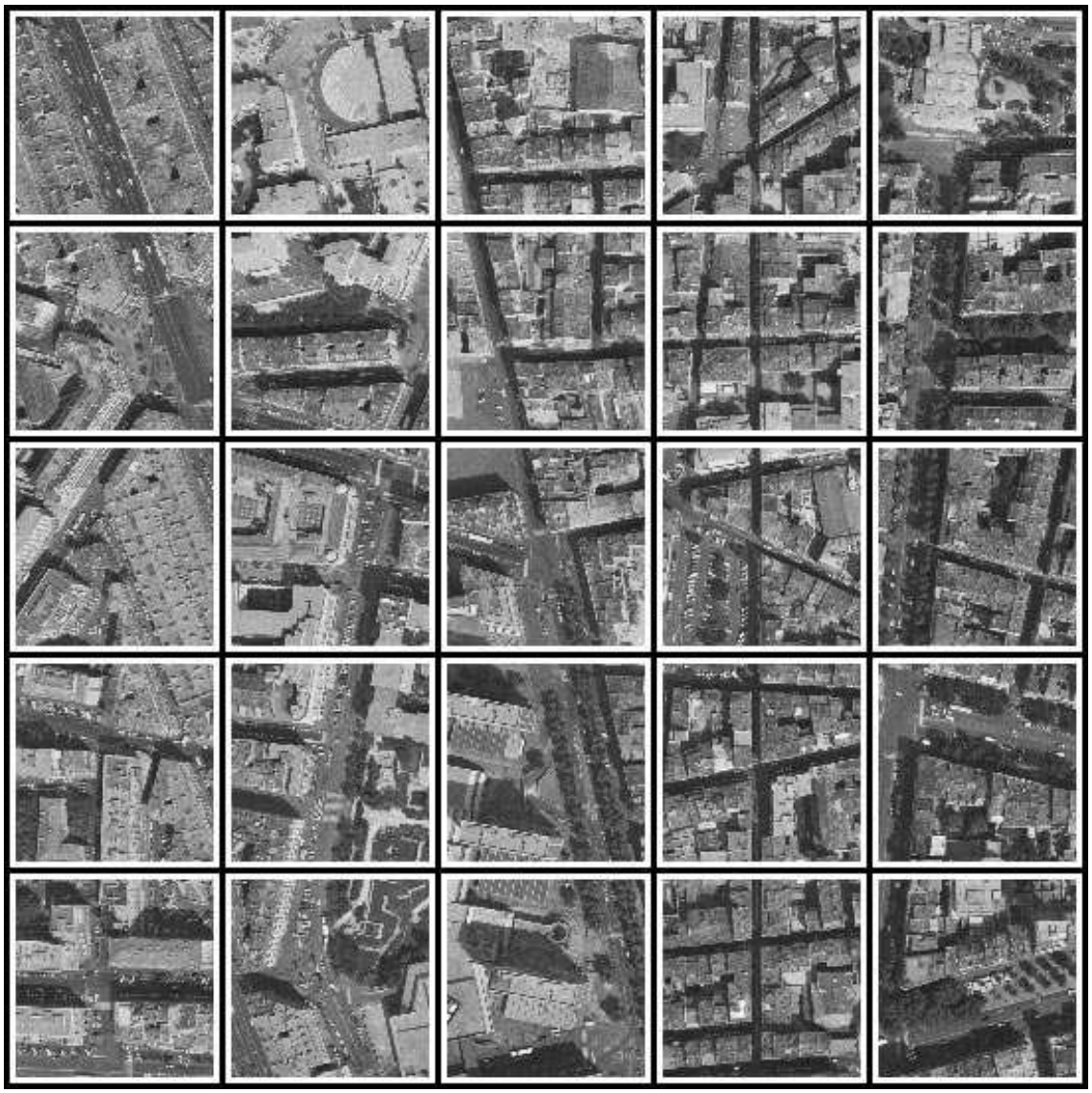

FIG. C.2 – Quelques images aériennes de notre base d'images (propriété d'Istar).

Appariement d'images par invariants locaux de niveaux de gris
Application à l'indexation d'une base d'objets

Cette thèse s'inscrit dans le domaine de l'appariement, un sujet fondamental en vision par ordinateur. Ce domaine recouvre des problèmes variés allant de celui de l'appariement entre deux images à celui de l'appariement d'une image et un modèle CAO. Notre approche permet d'apparier des objets s'ils sont observés dans des scènes complexes, s'ils sont partiellement visibles et s'ils sont aperçus de points de vue différents. Cette méthode est étendue à l'interrogation de bases d'images et à la reconnaissance d'objets.

Notre approche est basée sur une caractérisation locale des niveaux de gris d'une image. Cette caractérisation est calculée en des points particuliers des images : les points d'intérêt. Ces points sont détectés automatiquement et sont représentatifs de l'objet observé. De ce fait, la caractérisation obtenue représente une information très riche. De plus, elle est invariante pour le groupe des similitudes image et permet d'apparier des images ayant subi de telles transformations. Comme le groupe des similitudes absorbe au premier ordre les variations dues à un changement de point de vue lors d'une projection perspective, notre représentation est quasi-invariante et donc robuste à une telle transformation.

La solution présentée a été appliquée à la recherche d'une image dans une volumineuse base d'images. Comme la multiplicité des correspondances ne permet plus d'avoir directement de réponse satisfaisante, une méthode statistiquement robuste fait émerger la solution. D'autre part, pour effectuer une recherche rapide dans une large base un mécanisme d'indexation a été développé.

La recherche d'image a été étendue à la reconnaissance d'objet à partir d'une seule image. Pour ce faire, un objet 3D est modélisé par une collection d'images représentatives de l'objet. Pour obtenir une information 3D, des données symboliques sont ajoutées aux différents aspects de l'objet stockés dans la base. La relation trilinéaire permet alors de retrouver ces données sur une image recherchée.

Mots clés : vision par ordiateur, appariements d'images, recherche d'images, reconnaissance d'objets, points d'intérêt, invariants locaux de niveaux de gris.

Image matching by local greyvalue invariants
Applied to indexing an object database

This thesis concerns matching, a fundamental subject in computer vision. Matching covers a variety of problems such as matching two images or matching an image with a CAD model. Our approach allows objects to be matched if they are observed in complex scenes, partially occluded or seen from different viewpoints. The method is extended to image database consultation and object recognition.

Our approach is based on a local characterization of the greyvalue signal. This characterization is calculated at particular "points of interest". These are detected automatically and are representative of the observed object. Therefore, the characterization obtained has a high information content. In addition, it is invariant to the similarity group of transformations in the image and allows images that have undergone such transformations to be matched. To first order, the similarity group absorbs variations of perspective viewpoint changes, so our representation is quasi-invariant and therefore robust to such transformations.

The method has been applied to the retrieval of images from a large database. When there are many images there are typically many possible matches for any given point, so a robust statistical technique has been developed to find the corresponding image. To reduce the amount of computation required for a large database and make rapid retrieval possible, an indexing mechanism has been developed.

Our image retrieval scheme has been applied to 3D object recognition from a single image. Each object is modeled by a set of images taken from different viewpoints chosen to be representative of the object. To obtain 3D information, the different aspects of the objects stored in the database are annotated with symbolic data. The trilinearity constraint allows this data to be localized in the image.

Keywords : computer vision, image matching, image retrieval, object recognition, interest points, local greyvalue invariants.