Reconnaissance d'objets utilisant des histogrammes multidimensionnels de champs réceptifs

Bernt Schiele

▶ **To cite this version:**

Bernt Schiele. Reconnaissance d'objets utilisant des histogrammes multidimensionnels de champs réceptifs. Interface homme-machine [cs.HC]. Institut National Polytechnique de Grenoble - INPG, 1997. Français. NNT : . tel-00004962

HAL Id: tel-00004962

https://theses.hal.science/tel-00004962

Submitted on 20 Feb 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée par

Bernt SCHIELE

pour obtenir le grade de DOCTEUR

de l'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

(Arrêté ministériel du 30 mars 1992)

spécialité INFORMATIQUE

Reconnaissance d'Objets utilisant des Histogrammes Multidimensionnels de Champs Réceptifs

---

Thèse soutenue le 16 juillet 1997

Composition du jury :

Président :	Roger Mohr
Rapporteurs :	Shimon Edelman
	Henri Maître
Examinateurs :	Hans Burkhardt
	James L. Crowley (Directeur de thèse)

Thèse préparée au sein du laboratoire GRAVIR - IMAG
I.N.P. de Grenoble, 46, av. Félix Viallet, 38031 Grenoble Cedex, FRANCE

Acknowledgments

I wish to express my sincere gratitude to all those people who have contributed in different ways to the work that is presented in this thesis.

First of all I would like to thank my advisor and co–worker, Prof. James L. Crowley, who provided me motivation, a stimulating research environment and who introduced me to the research community. I am also grateful to Roger Mohr, Shimon Edelman, Henri Maître and Hans Burkhardt for the interest in my work and for being members of the thesis jury.

For friendship and research collaboration, thanks to Stephen Jones, Guido Appenzeller and Frank Wallner. I am especially grateful to Steve who has patiently corrected the English version of the thesis.

Many thanks to many people for smoothing the rough edges of my French and in particular to Bénédicte, Beatrice, Serge, Jérôme, Christophe, Vincent, Augustin, Alexandre, Delphine, Claude et Vincent le deuxième.

Thanks to all the current and past members of the research group PRIMA, Alvaro, Bénédicte, Bruno, Claude, Claus, Christophe, Cordelia, Frank, Jérôme, Jim, Olivier, Philippe, Patrice, Patrick, Steve, Vincent and Vincent le deuxième, for the friendly, fun and international atmosphere. It has been a pleasure working in Grenoble. Special thanks to Philippe for his help and discussions in the first months.

No words are adequate to express the author's debt to his parents, his sister, his two brothers and their families for their love and support. The work is dedicated to the author's brother Ralf who has been always a source of encouragement.

Finally I would like to thank you, Silke, for your patience, love and support and for just being.

Bernt Schiele,
Grenoble, 21 July 1997

Abstract

During the last few years, there has been a growing interest in object recognition techniques directly based on images, each corresponding to a particular appearance of the object. Representations of objects, which use only information of images are called *appearance based* models. The interest in such representation schemes is due to their robustness, speed and success in recognizing objects.

The thesis proposes a framework for the statistical representation of the appearances of 3D objects. The representation consists of a probability density function over a set of robust local shape descriptors which can be extracted reliably from images. The object representation is therefore learned automatically from sample images. Multidimensional receptive field histograms are introduced for the approximation of the probability density function. A main result of the thesis is that such a representation scheme based on local object descriptors provides a reliable means for object representation and recognition.

Different recognition algorithms are proposed and experimentally evaluated. The first recognition algorithm by histogram matching can be seen as a generalization of the color indexing scheme of Swain and Ballard. The second recognition algorithm calculates probabilities for the presence of objects only based on multidimensional receptive field histograms. The most remarkable property of the algorithm is that it relies on neither correspondence nor figure ground segmentation. Experiments show that this algorithm is capable of recognizing 100 objects in cluttered scenes. The third recognition algorithm incorporates several viewpoints in an active recognition framework in order to solve ambiguities inherent in single view recognition schemes.

The thesis also proposes visual classes as a general framework for appearance based object classification. Classification has been proven difficult for arbitrary objects due to instabilities of invariant representations. The proposed concepts for extraction, representation and recognition of visual classes provide a general framework for object classification.

From an abstract point of view, the thesis aims to push the limits of the appearance based paradigm without using neither figure ground segmentation nor correspondence. The active object recognition method allows the consistent recognition of objects in 3D and therefore overcomes the limits of single view recognition. The appearance based classification framework based on the concept of visual classes will serve for future research.

Table des matières

Preface

In the classical approach to image analysis approach [Mar 78, Kan 78, Ros 84, Nag 92] *image features* such as edges or *image regions* such as texture regions are extracted from the image. High level *feature groups* may be obtained by *grouping* these basic image features. This approach hypothesizes the identity and the pose of the object in the scene by calculating *feature correspondence* between the feature groups and the features of the object model.

The principal difficulty with this classical approach is that the process of determining feature correspondence has a complexity which is exponential with the number of extracted image features. Furthermore the extraction and grouping processes which produce image features are unstable, producing broken and spurious features which compound the complexity of correspondence.

In order to make the problem tractable, the number of extracted features must be reduced. This implies the use of salient – meaning discriminant – features. Because of the exponential complexity, only a relatively small number of image features can be used so that each image feature must be highly discriminant. Due to the tradeoff between robustness of the feature extraction and the discriminant power of features, the process of feature extraction tends to be unstable. Furthermore, the saliency of image features depends on the object classes employed making the techniques suitable only for particular object classes such as geometric objects.

The above limitations of the classical approach to image analysis require a paradigm shift in computer vision: the object's identity and the object's pose are estimated directly from measurements which can be calculated reliably from the image. The process of estimating the object's identity and the object's pose has a complexity which can be linear with the number of image measurements. This implies that a large number of image measurements may be used and therefore that robust image measurements can be chosen. In this context, the model of an object is given by a representation of image measurements which can be learned automatically from sample images. These techniques are called appearance based methods since each of the represented images corresponds to a particular appearance of the object.

The advantage of appearance based methods is that they can use robust image measurements

and that they can avoid feature correspondence. From an abstract point of view, these techniques calculate *object correspondences* between the image and the object models. This calculation of object hypotheses might be used as a pre–step of the classical image analysis approach: the hypothesized object can serve as a priori knowledge in order to reduce the complexity of the processes of correspondence and grouping.

Different appearance based object recognition techniques have been proposed: examples include the alignment scheme of Huttenlocher and Ullman [Hut 87], which relies on point correspondence of a small number of salient features, the eigenpicture approach [Sir 87, Tur 91a, Mur 95], which assumes the detection or the segmentation of the object, and the aspect graph [Koe 79, Fau 92, Egg 93], which is so far only applied to geometric objects.

The color indexing approach of Swain and Ballard [Swa 91] uses directly the color distribution of objects for recognition. Their approach has been shown to be remarkably robust to changes in the object's orientation, changes of the scale of the object, partial occlusion or changes of the viewing position. This approach is an attractive method for object recognition, because of its simplicity, speed and robustness. However, its reliance on object color and, to a lesser degree, light source intensity make it inappropriate for many recognition problems.

The focus of our work has been to develop a technique similar to color indexing using local descriptions of an object's shape provided by a vector of linear neighborhood operators. The first part of the thesis is therefore concerned with the definition of a statistical object representation framework based on local neighborhood operators. The principal aim is to develop fast and robust recognition techniques using the defined statistical object representation. The applicability of the techniques is shown experimentally on different databases each containing up to 100 objects. In order to overcome the limitations of the classical image analysis approach the thesis examines recognition without reliance on pre–segmentation and feature correspondence.

The speed and the robustness of appearance based object recognition approaches comes with a price: appearance based approaches use directly image measurements for recognition. Images and therefore appearances are recognized rather than objects. Due to this fact, any appearance based approach has to be evaluated with respect to the principal challenges of appearance based models. The main challenges are the recognition of objects in the presence of partial occlusion, the recognition of 3D objects and the classification of objects. The second part of the thesis extends the application of the defined statistical object representation framework to manage these three challenges.

Chapitre 1

Introduction et motivations

Au cours des dernières années, l'intérêt pour des algorithmes de reconnaissance fondés sur l'apparence a considérablement augmenté. Ces algorithmes utilisent directement des informations bidimensionnelles des images. A partir des images d'objets ces approches construisent des *modèles fondés sur l'apparence*, car chaque image représentée correspond à une apparence particulière d'un objet. La fiabilité, la vitesse et le taux de reconnaissance élevé de ces techniques en constituent les intérêts majeurs. Le succès de ces méthodes est considérable pour la reconnaissance de visages, dans le contexte de l'interface homme–machine et pour l'accès à des bases d'images par leurs contenus.

Cette thèse propose une technique où les objets sont représentés par des statistiques sur des opérateurs locaux et robustes. On veut montrer qu'une telle représentation fondée sur l'apparence est fiable et extrêmement discriminante pour la reconnaissance d'objets.

La motivation initiale de cette étude était la reconnaissance rapide d'objets par la méthode des histogrammes de couleurs. Cette méthode utilise les statistiques de couleurs comme modèle d'objets. Notre but était de généraliser cette approche en modélisant des objets par les statistiques de leurs caractéristiques locales. La technique généralisée – que l'on appelle *histogrammes multidimensionnels de champs réceptifs* – permet de discriminer un grand nombre d'objets. Néanmoins, les faiblesses de cette approche sont liées aux *défis des modèles fondés sur l'apparence*. Ces défis sont cités plus bas et examinés dans cette thèse.

L'intérêt principal de cette thèse est le développement d'un modèle de représentation d'objets qui utilise les statistiques de vecteurs de champs réceptifs. Plusieurs algorithmes de reconnaissance d'objets sont proposés. En particulier, un algorithme probabiliste est défini: il ne

dépend pas de la correspondance entre les images de test et les objets de la base de données. Des expériences obtiennent des taux de reconnaissance élevés en utilisant le modèle de représentation proposé. De plus, les défis généraux des modèles fondés sur l'apparence sont pris en considération par des extensions de notre technique.

La section 1.1 justifie l'utilisation d'une formulation probabiliste pour la reconnaissance d'objets. Puis la reconnaissance d'objets est définie comme cas d'étude de la vision par ordinateur (section 1.2). L'approche des histogrammes de couleurs est brièvement discutée comme inspiration initiale de notre travail (section 1.3). Les défis principaux des modèles fondés sur l'apparence (section 1.4) vont être examinés pendant cette étude. La dernière section 1.5 résume chaque chapitre de la thèse.

1.1 Motivation pour une formulation probabiliste

Cette thèse propose l'utilisation d'une formulation probabiliste pour la reconnaissance d'objets. Cette formulation permet d'incorporer dans le processus de reconnaissance des informations issues des connaissances a priori, du contexte ou d'autres capteurs. L'intégration de ces informations supplémentaires peut être effectuée sans modification de l'algorithme de reconnaissance. En général, l'utilisation d'une formulation probabiliste offre les avantages suivants :

- intégration des incertitudes
- souplesse de décision
- incorporation d'informations qui peuvent être indépendantes du contenu de l'image

Par définition, les statistiques permettent l'intégration des incertitudes. Cette intégration peut se faire à différents niveaux comme, par exemple, la modélisation de capteurs, la modélisation de données incomplètes et la décision. Les incertitudes peuvent et doivent être intégrées dans un cadre statistique.

La souplesse de décision est donnée par le caractère probabiliste des résultats dans le contexte statistique. Les résultats de la reconnaissance peuvent être formulés comme des probabilités pour chaque objet. Des décisions "dures" sont caractérisées par un choix binaire entre présence et absence d'un objet. Si une décision dure est désirée, elle peut être obtenue en appliquant un seuil approprié aux probabilités d'objets.

Nous pensons que de nombreuses décisions ne dépendent pas seulement du contenu de l'image, du signal lui–même, mais, aussi du contexte ou d'autres connaissances. Une formulation probabiliste est particulièrement adaptée pour l'intégration de ces connaissances. En outre, toute source d'information peut être intégrée, comme, par exemple, des informations provenant d'autres capteurs.

En conclusion, la modélisation et la reconnaissance statistique peuvent constituer un cadre intéressant et puissant. Néanmoins le problème fondamental de quantité suffisante de données pour l'estimation des fonctions de densités probabilistes limite souvent le succès des algorithmes statistiques. Ici, la représentation statistique résout ce problème en négligeant les informations topologiques des caractéristiques locales. En particulier, le chapitre 4 développe la représentation statistique d'objets en utilisant les histogrammes multidimensionnels de champs réceptifs.

1.2 Reconnaissance d'objets comme cas d'étude

Pour notre étude nous avons choisi le problème de la reconnaissance d'objets, car celui-ci peut être vu comme cas d'étude général de la vision par ordinateur. On peut identifier plusieurs degrés de liberté du problème de la reconnaissance d'objets :

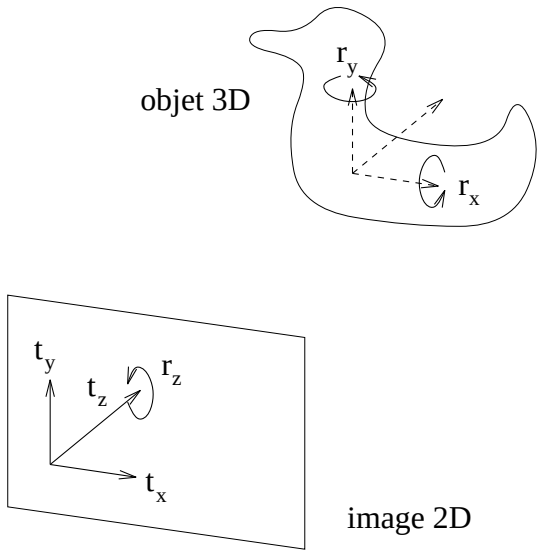


FIG. 1.1 – *Différentes composantes de la rotation et de la translation d'un objet 3D*

Transformation similaire dans le plan de l'image : on peut identifier trois degrés de liberté en translation (t_x, t_y et t_z) et un degré de liberté en rotation (r_z) (voir figure 1.1).

Transformation 3D d'un objet : il existe deux degrés de liberté supplémentaires en rotation (r_x et r_y) par rapport à la transformation similaire (voir figure 1.1).

Changements de la scène : ces changements incluent des occultations partielles et des changements du fond de la scène.

Conditions d'enregistrement : elles changent avec les variations de l'éclairage et avec les différentes perturbations comme le bruit de signal, les erreurs de discrétisation et le flou. Généralement, ces changements ne peuvent pas être contrôlés ou prédis.

Ces degrés de liberté sont discutés en détail dans le contexte de la représentation statistique d'objets (section 4.1). L'importance pour l'évaluation d'un algorithme de reconnaissance de ces degrés de liberté justifie leur utilisation comme cadre de référence de l'étude. La structure du chapitre 6 se base sur ces degrés de liberté.

Nous différencions l'*identification d'objets* de la *classification d'objets*, deux parties du problème général de la reconnaissance d'objets. L'identification d'objets consiste à reconnaître des

objets préalablement vus par le système. Par contre, la classification d'objets consiste à généraliser en dehors de la base des objets représentés et préalablement vus. Le deuxième problème de la classification d'objets est plus intéressant même si l'identification d'objets est souvent suffisante. Dans la suite, on utilise les termes *reconnaissance d'objets* et *identification d'objets* comme synonymes. Le terme *classification d'objets* est utilisé pour la généralisation en dehors de la base de données.

1.3 Motivation pour les histogrammes multidimensionnels de champs réceptifs

Swain et Ballard [Swa 91] ont développé une technique d'identification d'objets fondée sur des histogrammes de couleurs. Le principe de base est de comparer un histogramme de couleurs d'une région d'image avec un histogramme d'une apparence d'un objet. Leur technique est remarquablement fiable par rapport aux changements d'orientation d'objet, aux variations d'échelle d'objet, à l'occultation partielle et aux variations de point de vue. Même les changements de la forme d'objet ne dégrade pas toujours la performance de leur méthode. Néanmoins, le inconvénient principal de leur méthode est sa sensibilité aux variations d'éclairage. Plusieurs auteurs ont amélioré la performance de la technique en utilisant des mesures moins sensibles aux variations de la luminosité (voir section 2.1).

La simplicité, la vitesse et la fiabilité de la comparaison des histogrammes de couleurs constituent les intérêts majeurs de cet algorithme de reconnaissance d'objets. De plus, la méthode ne dépend pas de la mise en correspondance entre l'image de test et les modèles d'objets. Néanmoins, ses dépendances sur les couleurs d'objets et les variations de l'éclairage rendent la technique inadaptée pour de nombreux problèmes de reconnaissance. La motivation initiale de notre travail était alors de développer une technique similaire, qui remplace la couleur par des descripteurs locaux des formes d'objets. De tels descripteurs sont donnés, par exemple, par des vecteurs de champs réceptifs locaux. La stabilité par rapport aux changements d'échelle et de rotation de l'algorithme original de Swain et Ballard est due à l'utilisation de la couleur. Par contre, la robustesse par rapport aux changements de point de vue et à l'occultation partielle est due à l'utilisation de la *comparaison des histogrammes*. Il est alors logique d'exploiter la puissance de la comparaison des histogrammes pour la reconnaissance en utilisant des histogrammes de descripteurs locaux de forme. Le descripteur le plus général est donné par un vecteur multidimensionnel d'opérateurs locaux, ou champs réceptifs.

La première partie de la thèse (chapitres 3 à 6) décrit la généralisation de la comparaison des histogrammes de couleurs à la comparaison des *histogrammes multidimensionnels de champs réceptifs*.

1.4 Défis principaux des modèles fondés sur l'apparence

Cette thèse emploie le terme *modèle fondé sur l'apparence* pour des techniques qui utilisent seulement des informations 2D pour la représentation et la reconnaissance d'objets. Le modèle le plus connu s'applique à l'analyse des composantes principales des images. Des approches utilisant des descripteurs 2D et notre méthode utilisant les statistiques des opérateurs locaux emploient aussi des modèles fondés sur l'apparence. L'avantage principal de ces approches est la

fiabilité et la vitesse de l'extraction des informations 2D. Ces techniques peuvent être appelées *centrées sur l'observateur* ou *fondées sur l'image*. Dans ce contexte, un objet 3D est représenté par une collection de modèles 2D de différentes apparences de l'objet.

Le contraire des techniques de modèles fondés sur l'apparence sont les techniques utilisant des modèles 3D. Un objet 3D est souvent représenté par un seul modèle 3D centré sur l'objet. Typiquement, ces modèles sont plus simples et moins coûteux au niveau de la mémoire que des modèles fondés sur l'apparence. Néanmoins, le inconvénient principal de ces méthodes est l'instabilité de l'extraction des informations 3D à partir des images 2D.

Les avantages et inconvénients de ces deux approches sont complémentaires. C'est à dire, qu'il est impossible de juger, en général, l'un supérieur à l'autre. Cette thèse emploie un modèle fondé sur l'apparence. Ce choix est justifié surtout par la robustesse et la vitesse qui peuvent être obtenues par une telle méthode. Par conséquent, il faut considérer et examiner les problèmes principaux de l'application d'une technique qui utilise des modèles fondés sur l'apparence. On peut énumérer les *défis principaux de l'application d'un modèle fondé sur l'apparence* :

- reconnaissance en présence de changement de point de vue
- occultation partielle des objets
- reconnaissance d'objets 3D à partir des images 2D
- classification d'objet comme généralisation en dehors de la base d'objets
- consommation de mémoire pour la représentation d'objets

Chaque point est traité séparément. En particulier, les chapitres et sections suivants peuvent être cités :

- la section 6.5 montre la stabilité de la représentation par des histogrammes multidimensionnels de champs réceptifs par rapport aux changements de point de vue.
- le chapitre 7 propose un algorithme probabiliste de reconnaissance considérant l'occultation partielle. Des résultats expérimentaux montrent qu'une petite portion d'un objet visible est suffisante pour la reconnaissance de 103 objets.
- le chapitre 8 propose un algorithme de la reconnaissance active utilisant seulement des apparences d'objets. L'idée de base est l'utilisation de plusieurs points de vue qui permet une reconnaissance d'objets en 3D.
- le chapitre 9 introduit le concept des classes visuelles comme cadre général pour la classification d'objets. Un algorithme de reconnaissance des classes visuelles selon le maximum de vraisemblance est proposé. Des expériences appliquent cet algorithme dans le contexte de l'accès à une base d'images.
- la consommation de mémoire est analysée dans la section 6.7. Dans le futur, la consommation de mémoire peut être réduite en appliquant les concepts développés pour la classification d'objets et en utilisant des techniques de réduction de dimensionalité classiques.

Ces chapitres et sections indiquent que les défis principaux des modèles fondés sur l'apparence structure la deuxième partie de la thèse, c'est à dire les chapitres 7 à 9.

1.5 Sommaire des chapitres de la thèse

La suite résume chaque chapitre de cette thèse.

Le chapitre 2 décrit brièvement quelques techniques qui ont été source d'inspiration pour différents aspects de la thèse. En particulier, on discute les approches des histogrammes de couleurs. Cela inclue la technique originale proposée par Swain et Ballard et d'autres méthodes proposées pour l'augmentation de la stabilité par rapport aux variations de l'éclairage. On doit noter la popularité de l'approche pour l'accès aux bases d'images par leurs contenus. Le chapitre décrit également plusieurs techniques de reconnaissance fondées sur des caractéristiques locales car cette thèse généralise la technique des histogrammes de couleurs à des histogrammes de vecteurs d'opérateurs locaux. Comme nous nous intéressons à une formulation probabiliste, le chapitre introduit plusieurs méthodes statistiques pour la reconnaissance d'objets.

Le chapitre 3 est consacré à la discussion des caractéristiques locales. Les dérivées Gaussiennes sont très populaires, car elles sont connues, robustes, et ont même une justification physiologique. Les filtres de Gabor sont généralement plus coûteux en calcul mais utilisés souvent dans le contexte d'analyse de textures. On décrit également des descripteurs de couleur invariants par rapport aux changements de luminosité et couleur de l'éclairage. Le chapitre discute aussi des techniques de normalisation qui peuvent rendre résistant les descripteurs locaux aux bruits et aux variations de luminosité. La stabilité des techniques de normalisation est examinée par rapport au bruit Gaussien additif (section 3.2.4). La section 5.3 examine la robustesse des techniques de normalisation par rapport aux variations de luminosité.

Le chapitre 4 développe une représentation statistique générale d'objets. Chaque degré de liberté de la reconnaissance d'objets, comme introduit en section 1.2, est discuté et considéré d'une façon appropriée. Un ensemble d'histogrammes multidimensionnels de champs réceptifs pour chaque objet est proposé comme approximation de la représentation statistique d'objets. Ce cadre statistique est interprété comme modèle général d'objets fondé sur des descripteurs locaux. Cette interprétation permet le développement d'une analogie entre la reconnaissance d'objets et la théorie d'information. Cette analogie peut être appliquée, par exemple, pour évaluer un ensemble de descripteurs par la "transinformation".

Le chapitre 5 introduit différentes fonctions de comparaison d'histogrammes. La fonction de comparaison de l'approche originale, l'intersection $\cap$, possède des limitations dans le contexte plus général des histogrammes multidimensionnels de champs réceptifs. Le chapitre définit et analyse différentes fonctions de comparaison comme les statistiques χ^2, les distances quadratiques et les intersections modifiées. La complexité et les caractéristiques de chaque fonction sont discutées. De plus, la deuxième partie du chapitre analyse la stabilité des fonctions par rapport au bruit Gaussien, au flou, à la rotation de l'image et aux variations de luminosité.

Le chapitre 6 applique les fonctions de comparaison à la reconnaissance d'objets. On décrit d'abord un exemple de reconnaissance de 261 objets. Ensuite, les différents degrés de liberté de la reconnaissance d'objets sont considérés. En particulier, le chapitre décrit la prise en compte de la rotation de l'image et du changement d'échelle. La section 6.5 est consacrée à l'aspect

important de la reconnaissance en présence de changement de point de vue. La fin du chapitre discute la consommation de mémoire par des histogrammes multidimensionnels.

Le chapitre 7 étend l'application des histogrammes multidimensionnels à la reconnaissance probabiliste d'objets. L'algorithme probabiliste est capable de reconnaître des objets à partir d'une petite partie visible d'objet augmentant la fiabilité par rapport à l'occultation partielle. Des résultats de reconnaissance sont décrits pour une base de 103 objets en présence de rotation d'image, de changement d'échelle et de changement de point de vue. A partir de ces résultats on propose une approche d'une "table de hash dynamique" utilisant des régions d'image comme indexes de la table de hash. Ce dernier algorithme reconnaît les objets dans des scènes compliquées.

Les deux chapitres suivants proposent deux extensions (ou applications) des histogrammes multidimensionnels de champs réceptifs pour la reconnaissance active d'objet et pour la classification d'objets. Comme les expériences de ces deux derniers chapitres ne sont pas exhaustives, on doit regarder ces deux chapitres comme des perspectives de la thèse.

Le chapitre 8 adopte des stratégies hypothèse–test pour la reconnaissance active d'objets dans une seule image 2D et pour plusieurs images. Pour le cas d'une seule image 2D un détecteur des points d'intérêt général est développé. Ce détecteur peut être utilisé aussi par d'autres algorithmes de reconnaissance. Le deuxième algorithme de reconnaissance active utilise le concept de la transinformation pour évaluer les points de vue les plus discriminants d'un objet. En déplaçant la caméra vers ces points de vue discriminants, il est possible de vérifier l'hypothèse d'objet calculée à un autre point de vue. Comme l'algorithme utilise à chaque point de vue seulement des informations 2D, cet algorithme permet la reconnaissance d'objets 3D à partir de plusieurs images 2D. Des résultats expérimentaux soulignent la propriété de l'algorithme à reconnaître des objets qui sont similaires en 3D.

Le chapitre 9 propose le concept des classes visuelles comme cadre général pour la classification d'objets. Les classes visuelles sont définies à partir des similarités d'objets en 2D et/ou en 3D. Ces similarités peuvent être dérivées de la représentation statistique d'objets. Le chapitre propose une technique selon le maximum de vraisemblance pour la reconnaissance des classes visuelles. La technique est appliquée pour obtenir des images visuellement similaires d'une base d'images.

Le chapitre 10 conclut les résultats principaux et donne quelques perspectives de la thèse.

Chapitre 2

État de l'art

Ce chapitre décrit brièvement quelques références qui ont inspiré cette étude. Malgré son titre, il ne fournit pas un résumé exhaustif des algorithmes de reconnaissance d'objets (voir par exemple [Obj 96, Pop 95, Gri 92, Gri 91]). Cette étude s'intéresse particulièrement aux modèles d'objets estimés automatiquement à partir d'exemples d'images d'objets.

La motivation initiale de cette thèse était la reconnaissance rapide et fiable d'objets par les méthodes des histogrammes de couleurs. La première section décrit ces techniques. Afin de généraliser le concept des histogrammes de couleurs, le vecteur de couleurs est remplacé par un vecteur multidimensionnel de champs réceptifs. Ces vecteurs sont donnés par les descripteurs de voisinages locaux. Ainsi la section 2.2 introduit quatre algorithmes de reconnaissance fondés sur les vecteurs de descripteurs locaux. Étant donné l'importance du choix de ces caractéristiques locales, le chapitre 3 entier est dédié à ce sujet. D'un point de vue plus abstrait, les histogrammes de vecteurs de descripteurs locaux peuvent servir comme approximation d'une densité probabiliste de vecteurs de descripteurs locaux. La dernière section 2.3 résume donc trois algorithmes statistiques de reconnaissance d'objets.

2.1 Techniques fondées sur des histogrammes de couleurs

La section suivante 2.1.1 introduit la technique des histogrammes de couleurs [Swa 91] et quelques extensions visant l'indépendance de l'éclairage. La section 2.1.2 décrit une technique de reconnaissance qui combine les informations géométriques avec celles des couleurs. La dernière section 2.1.3 montre l'importance des histogrammes de couleurs dans le contexte de l'accès aux

bases d'images par leur contenu.

2.1.1 Indexation par la couleur

Swain et Ballard [Swa 90, Swa 91, Swa 93a] ont proposé une technique selon laquelle chaque objet est représenté par un histogramme de couleur (c'est à dire par une approximation de la distribution de ses couleurs). Les objets sont identifiés par la comparaison de l'histogramme de couleurs d'une région de l'image à un histogramme de couleurs d'un exemple de l'objet (voir chapitre 5 pour les fonctions de comparaison d'histogrammes). Leur technique est remarquablement fiable par rapport aux changements d'orientation d'objet, aux variations d'échelle, aux occultations partielles et aux changements de point de vue. Même des changements de la forme d'un objet ne dégradent pas forcément la performance de leur méthode. L'inconvénient principal de leur méthode est sa sensibilité à la couleur et à l'intensité de l'éclairage. En outre, il est impossible de représenter toutes les classes d'objets uniquement par la distribution de leurs couleurs. Plusieurs auteurs ont amélioré la technique des histogrammes de couleurs en utilisant des mesures de couleurs plus indépendantes aux changements d'éclairage (voir plus bas et [Fun 95, Hea 94, Enn 95]).

La technique des histogrammes de couleurs est une méthode bien adaptée au problème de la reconnaissance d'objets grâce à sa simplicité, à sa rapidité et à sa fiabilité. En outre, l'utilisation des histogrammes de couleurs ne nécessite ni une segmentation d'objet ni un modèle géométrique explicite. Un objet est décrit simplement par son histogramme de couleurs. En opposition à ces avantages, la dépendance à la couleur de l'objet, à l'intensité et à la couleur de l'éclairage, rend cette technique inappropriée à de nombreux problèmes de reconnaissance. La technique de Swain et Ballard est fiable aux changements d'échelle et de rotation grâce à l'utilisation de la couleur. La robustesse par rapport aux changements de point de vue et aux occultations partielles est due à l'utilisation de la *comparaison d'histogrammes*. Ainsi il est évident d'exploiter les possibilités de la comparaison d'histogrammes afin d'effectuer une reconnaissance fondée sur des histogrammes de propriétés de formes locales. La méthode la plus générale pour mesurer ces propriétés consiste en un vecteur d'opérateurs linéaires de voisinage local, c'est à dire en un vecteur multidimensionnel de champs réceptifs.

La sensibilité des couleurs à l'intensité et à la couleur de l'éclairage rend le choix de la représentation de la couleur critique. Swain et Ballard proposent l'utilisation d'une variante de la représentation de couleurs "opposées" décrite par l'équation suivante [Bal 82, Swa 90]:

$$\begin{pmatrix} rg \\ by \\ wb \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & -1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} R \\ V \\ B \end{pmatrix} \tag{2.1}$$

Cette transformation linéaire de l'espace RVB est souvent considérée comme la perception humaine des couleurs. Dans le contexte des histogrammes de couleurs cette transformation linéaire permet une discrétisation moins fine de l'"intensité" wb par rapport aux autres axes. Les résolutions proposées des axes consistent en 8 cellules pour l'axe wb et en 16 cellules pour les deux autres axes. C'est à dire que chaque histogramme contient 2048 cellules. Pour autant, la transformation linéaire ne réduit pas la sensibilité de manière significative par rapport à l'intensité de couleurs. Swain et Ballard proposent donc l'application d'un algorithme de la constance

de couleurs [1] avant le calcul des histogrammes afin de réduire la sensibilité aux changements d'éclairage.

Comme il est mentionné plus haut, plusieurs auteurs ont amélioré la technique originale en envisageant une réduction de la sensibilité aux changements d'éclairage. Deux techniques performantes ont été proposées par Healey/Slater [Hea 94] et Funt/Finlayson [Fun 95]. Healey et Slater calculent des invariants de moment de l'histogramme entier de couleurs. Ces invariants sont fondés sur le modèle linéaire de dimension finie de couleurs permettant la modélisation de changements de l'intensité d'éclairage par une transformation linéaire d'histogrammes. Les résultats expérimentaux sont persuasifs [Hea 94]. Par contre la méthode est globale car elle suppose un changement global et constant de l'intensité pour toute l'image. Cette supposition rend la technique inadaptée pour de nombreuses situations [Col 96].

Funt et Finlayson [Fun 95] utilisent les dérivées de logarithmes de canaux de couleurs afin de fournir des caractéristiques invariantes aux changements d'éclairage. La supposition fondamentale est un éclairage localement constant. Les invariants de couleurs sont fondés sur un modèle approximatif de couleurs (modèle de coefficients) permettant de calculer des invariants locaux de couleurs. Malgré la simplification extrême du modèle, les invariants obtenus sont plus appropriés pour la reconnaissance que d'autres invariants de couleurs [Col 96] comme les invariants proposés par Nagao [Nag 95]. Le plus grand avantage des invariants proposés est leur calcul local. Ce calcul ne suppose pas un changement uniforme de l'éclairage de l'image entière. Notons que les histogrammes des invariants proposés ne représentent pas directement l'information de couleur, mais plutôt les relations entre des régions de couleurs voisines.

Une extension de la méthode originale d'indexation par la couleur est proposée par Ennesser et Medioni [Enn 93, Enn 95] : en utilisant des histogrammes locaux de couleurs d'une image test, leur algorithme est capable d'extraire les régions de l'image qui ont une forte probabilité de contenir un objet spécifique. Des expérimentations montrent la capacité de l'algorithme à détecter des objets spécifiques dans des scènes complexes.

Hunke, Schiele et Waibel [Hun 94, Sch 95c, Sch 95b] utilisent la normalisation du vecteur de couleur par la luminosité fournissant un moyen fiable de détecter des couleurs de peau. Ce détecteur peut servir pour trouver des visages et des mains. La simplicité de la technique permet la construction d'un système temps réel de suivi de visages sous conditions variées.

2.1.2 Indexation par la couleur et information géométrique

Plusieurs auteurs ont combiné l'approche des histogrammes de couleurs avec différents types d'informations géométriques pour la reconnaissance. Les exemples incluent [Sla 95, Mat 95]. Cette section décrit un système inspiré par des réseaux de neurones et appelé SEEMORE [Mel 96].

SEEMORE utilise 102 différents canaux de caractéristiques. Les réponses de chaque canal sont calculées par l'échantillonage et la sommation de sorties d'un opérateur local du champ visuel (plus précisément de la région d'image pré–segmentée). Les 102 sommes de canaux de caractéristiques sont les entrées principales d'un classificateur par plus proches voisins [2]. Les canaux de caractéristiques se classent en cinq catégories : 23 canaux circulaires de couleurs (représentées par la tonalité chromatique [3] et la saturation), 11 canaux d'intensités aux grosses

1. color constancy algorithm

2. nearest–neighbour classifier

échelles, 12 canaux circulaires et orientés de régions d'intensités, 40 canaux de contours généralisés (comme les jonctions, les courbes et les paires de contours parallèles et obliques) et 16 canaux de texture fondés sur des filtres de Gabor (utilisant l'énergie et l'orientation relative aux différentes échelles et orientations).

Des résultats convainquants sont décrits à partir d'une base de 100 objets pré–segmentés de types variés. Le résultat le plus intéressant est une certaine capacité de généralisation en dehors de la base de données. La représentation des couleurs par la tonalité chromatique et la saturation requiert des conditions d'éclairage constantes. En conséquence, les résultats de l'algorithme se dégradent en présence de changements d'éclairage. De plus, les 102 canaux de caractéristiques ont été choisis manuellement.

2.1.3 Accès aux bases d'image

Récemment, un domaine de recherche très actif s'intéresse à l'accès aux bases d'images et/ou de vidéos par leur contenu. Les groupes de recherche à MIT [Pen 96, Pic 96], Stanford [Tom 94, Gui 96], Université de Californie Berkeley [Bel 97] et IBM [Fli 95] sont parmi les groupes les plus actifs. Souvent, la couleur est un indice employé pour l'accès aux grandes bases de données. Le système de Berkeley et celui de IBM emploient directement les histogrammes de couleurs pour l'indexation. Ricard, au MIT, propose une technique d'histogrammes pour le triage d'images à "coup d'oeil" [Gor 94, Pic 96].

L'intérêt de la couleur provient du fait que beaucoup d'images contiennent des couleurs caractéristiques. L'intérêt des techniques d'histogrammes de couleurs est justifié par le faible coût de calcul et de mémoire. Il est alors intéressant d'analyser les histogrammes de champs réceptifs comme une généralisation des histogrammes de couleurs dans le contexte de l'accès aux bases d'images.

2.2 Reconnaissance d'objets fondée sur des descripteurs locaux

Cette section décrit brièvement quatre approches de reconnaissance utilisant des descripteurs locaux. La section 2.2.1 décrit une technique de hachage géométrique fondée sur des caractéristiques de points. Cette technique est extensible à des caractéristiques plus générales. La section 2.2.2 introduit un système de l'université de Rochester pour la reconnaissance temps–réel par une mémoire iconique distribuée grossièrement [4]. La section 2.2.3 décrit un algorithme profitant des avantages des deux précédents : certainement un système parmi les plus performants d'aujourd'hui. La dernière section 2.2.4 introduit brièvement un système de reconnaissance fondée sur une mémoire à deux stades qui peut être interprétée comme une implémentation efficace de la transformation de Hough.

2.2.1 Hachage géométrique

Lamdan et al. [Lam 88c, Lam 88b] décrivent un algorithme général de reconnaissance utilisable pour des scènes ayant des recouvrements et des objets partiellement occultés. Les objets

3. hue

4. iconic sparse distributed memory

sont modélisés par une représentation des ensembles de points d'intérêt, invariante à une transformation affine. Pour réduire le temps et la complexité de reconnaissance, toutes les combinaisons possibles de points d'intérêt sont mémorisées dans une table de hachage. La reconnaissance consiste à extraire des ensembles de points d'intérêt d'une scène, à indexer la table de hachage et à voter pour les objets. La reconnaissance est alors une mise en correspondance de points (ou plus généralement de caractéristiques).

La première étape de l'algorithme est donnée par un opérateur de points d'intérêt. Le choix de cet opérateur n'est pas important car la suite de l'algorithme utilise seulement les coordonnées de points d'intérêt. En supposant l'extraction de m points d'intérêt, un triplet non collinéaire de points d'intérêt est choisi comme base affine. Les coordonnées des autres $m-3$ points d'intérêt sont calculées et mémorisées comme entrées de la table de hachage (avec une référence d'objet correspondant). Pour réduire la complexité de la reconnaissance, toutes les combinaisons de triplets non collinéaires sont utilisées comme bases affines. C'est à dire la complexité totale de cette première étape d'apprentissage est de l'ordre de $O(m^4)$.

Pour la reconnaissance d'objets le même opérateur de points d'intérêt est appliqué pour l'extraction de n points d'intérêt d'une image test. Un triplet arbitraire, non collinéaire, est employé comme base affine et les coordonnées des $n-3$ autres points sont utilisées pour voter pour les objets (ou plus précisément pour un triplet particulier et son modèle d'objet). La complexité d'une étape de votes est $O(n)$. Si aucun objet n'a obtenu un nombre suffisant de votes après une étape, il est possible de choisir un autre triplet non collinéaire pour une autre étape de votes. La complexité, dans le pire cas (rejet d'une image test), est de l'ordre de $O(n^4)$ en essayant tous les triplets possibles. Si un modèle d'objet a accumulé un nombre suffisant de votes, il peut être vérifié par une mise en correspondance de l'image test et du modèle d'objet [Lam 88c]. (L'approche est généralisée par [Lam 88b] en utilisant comme transformation une similitude et une transformation projective : pour le premier cas deux points seulement définissent une base. Dans le cas projectif, quatre points définissent une base projective. La complexité est alors $O(m^{k+1})$, avec $k=2$ et $k=4$.)

La propriété la plus intéressante de l'algorithme est que le modèle de représentation est strictement *local* : un ensemble de points d'intérêt est, par sa définition, local (seulement une petite région d'image est employée). Comme le modèle d'un objet consiste en une collection de points d'intérêt le modèle lui–même est local. Un ensemble de points d'intérêt peut être interprété comme une caractéristique de haut niveau, utilisée pour l'indexation. L'algorithme peut être donc interprété comme une mise en correspondance de caractéristiques.

Un autre avantage est que l'algorithme ne reconnaît pas seulement l'identité d'un objet mais également sa position dans l'image (exprimée en coordonnées du triplet).

Plusieurs auteurs ont tenté de réduire la complexité (ou au moins le temps moyen de reconnaissance) de l'algorithme. D'un côté, plus il y a de points utilisés pour la construction d'un ensemble de points d'intérêt, plus cet ensemble devient discriminant. D'un autre côté, la complexité augmente avec le nombre k car elle est donnée par $O(m^{k+1})$ pour la phase d'apprentissage et $O(n^{k+1})$ dans le pire des cas de reconnaissance. Ce dilemme provient des coordonnées (affines) de points utilisées dans la structure locale autour d'un point d'intérêt (ou la classification d'un point).

[Lam 88a, Wol 90] remplace les points d'intérêt par les caractéristiques d'intérêt, codant ainsi plus d'informations. Néanmoins, ces caractéristiques doivent être invariantes par rapport à la transformation considérée (par exemple une transformation affine), limitant considérablement

le choix des caractéristiques.

[Gri 90] analyse théoriquement la sensibilité des techniques de hachage géométrique. Le résultat principal est que la probabilité d'un vote faux positif augmente considérablement, même en présence de bruit modéré perturbant les points de données. Un schéma amélioré de votes est proposé par [Rig 93]. Les travaux plus récents incluent [Beb 95] et [Lam 96].

2.2.2 Mémoire iconique distribuée grossièrement

Une technique proposée par Wixson, Rao et Ballard [Bal 93, Rao 95b, Rao 95a] représente les objets (ou les régions d'objets) par un vecteur "iconique" de caractéristiques à hautes dimensions. Rao et Ballard [Rao 95b] soulignent les propriétés favorables d'un espace à hautes dimensions pour la mise en correspondance. Ces espaces donnent dans la plupart des cas une réponse unique pour un objet particulier. Ils suivent Karneva [Rao 95a] en argumentant qu'un vecteur iconique à hautes dimensions d'un objet peut être perturbé par du bruit important sans qu'il soit confondu avec un vecteur d'un autre objet.

Les vecteurs de caractéristiques à hautes dimensions $r(x,y)$ (centrés sur une position d'image (x,y)) consistent en 45 réponses de neuf filtres Gaussiens à cinq échelles différentes ($9 \times 5 = 45$) :

$$r(x,y) = \begin{pmatrix} r_{\sigma_1}(x,y) \\ r_{\sigma_2}(x,y) \\ r_{\sigma_3}(x,y) \\ r_{\sigma_4}(x,y) \\ r_{\sigma_5}(x,y) \end{pmatrix} \text{ avec } \quad r_{\sigma_i}(x,y) = \begin{pmatrix} G^{\sigma_i}_{1,0} \\ G^{\sigma_i}_{1,\frac{\pi}{2}} \\ G^{\sigma_i}_{2,0} \\ G^{\sigma_i}_{2,\frac{\pi}{3}} \\ G^{\sigma_i}_{2,\frac{2\pi}{3}} \\ G^{\sigma_i}_{3,0} \\ G^{\sigma_i}_{3,\frac{\pi}{4}} \\ G^{\sigma_i}_{3,\frac{\pi}{2}} \\ G^{\sigma_i}_{3,\frac{3\pi}{4}} \end{pmatrix} \star I(x,y) \quad i = 1,2,\ldots,5 \tag{2.2}$$

Pour rendre le vecteur $r(x,y)$ invariant aux rotations dans le plan image, les réponses de filtres sont normalisées : en utilisant l'orientation correspondant à la réponse maximale de la première dérivée Gaussienne comme direction de référence, toutes les composantes du vecteur sont orientées dans cette direction (en appliquant l'orientabilité des dérivées Gaussiennes, voir section 3.1.1 pour les détails).

Pour l'indexation, les vecteurs de caractéristiques $r(x,y)$ de différents objets sont stockés dans une version généralisée de la mémoire distribuée grossièrement de Karneva. Cette mémoire est appropriée à l'énorme espace d'adresses couvert par les vecteurs de caractéristiques à 45 dimensions. Pendant l'apprentissage et la reconnaissance, un objet est segmenté du fond (en utilisant la disparité zéro d'une paire d'images stéréo). Les réponses de vecteurs sont calculées par le système de traitement d'image MV200 en pipeline qui permet le calcul de convolution en temps–réel. L'accès à la mémoire est aussi réalisé par le système MV200 accélérant la reconnaissance.

Un inconvénient du vecteur de caractéristiques $r(x,y)$ proposé est son support relativement large : les réponses du vecteur $r(x,y)$ sont calculées à partir d'un noyau de 8×8 pixels à cinq différents niveaux d'une pyramide d'images. Comme chaque niveau de la pyramide réduit la

taille de l'image par deux, le support total d'un seul vecteur $r(x, y)$ est de l'ordre de 128×128 pixels. Le vecteur $r(x, y)$ n'est pas local car il couvre $\frac{1}{16}$ d'une image 512×512 pixels. Pour un calcul plus local de $r(x, y)$ le support de vecteurs doit être réduit. Ceci compromet le caractère unique de réponses de vecteurs.

Le caractère global du vecteur de caractéristiques rend l'approche sensible aux occultations partielles. [Bal 94] introduit un algorithme à part pour le traitement des occultations partielles. Cet algorithme suppose une segmentation de l'objet. L'idée principale est la reconstruction approximative d'une région de l'image par une transformation inverse d'un seul vecteur de caractéristiques $r(x, y)$. En appliquant le masque des parties occultées à la région d'image reconstruite, une estimation du vecteur $\hat{r}(x, y)$ est obtenue, correspondant à l'estimation d'occultations. Le vecteur $\hat{r}(x, y)$ peut enfin être comparé à l'observation dans l'image.

2.2.3 Reconnaissance comme problème de correspondance

Un système de reconnaissance d'objets parmi les meilleurs a été proposé par Schmid et Mohr [Sch 96h, Sch 96i]. Leur algorithme consiste en trois étapes : détection de points d'intérêt, caractérisation de points d'intérêt par les vecteurs de descripteurs locaux et stockage de chaque vecteur dans une table de hachage. Dans un sens, cette technique est une synthèse des deux précédentes : représentation locale par une table de hachage et description puissante de structures locales par des vecteurs de caractéristiques locales.

Le détecteur de points d'intérêt est une version modifiée du "détecteur de Harris". Il réduit les données d'image entière à un nombre de points d'intérêt. Chaque point d'intérêt d'une image est représenté par un vecteur à neuf dimensions de caractéristiques invariantes aux rotations d'image. Ces caractéristiques sont fondées sur des dérivées Gaussiennes (jusqu'à un ordre de trois) et ont été proposées par Koenderink [Koe 87]. Les réponses de vecteurs de tous les points d'intérêt sont stockées dans une table de hachage indexée par le vecteur à neuf dimensions.

L'application majeure de la technique est la mise en correspondance d'une image test et des images mémorisées dans la table de hachage. De plus, l'approche est utilisable pour la reconnaissance d'objets (ou d'images) interprétée comme problème de correspondance. En appliquant le détecteur de points d'intérêt à une image test et en calculant les réponses de vecteurs de ces points, l'algorithme vote pour les différentes images (ou objets). En ajoutant aux réponses de vecteurs des invariants géométriques entre les différents points, le schéma de votes devient plus sélectif. Une autre possibilité d'amélioration est l'utilisation d'un schéma de votes probabilistes récemment proposé par Mohr et al. [Moh 97].

Des résultats expérimentaux convainquants sont obtenus pour une base de plusieurs centaines d'objets. Néanmoins, le point le moins fort de la méthode est l'application d'un détecteur de point d'intérêt. Le succès de la méthode dépend de la répétabilité du détecteur pour différentes images et différentes conditions d'enregistrement (par exemple changements d'éclairage et changements d'échelles). Cette répétabilité est difficile à obtenir. De plus, la méthode telle qu'elle est présentée dans [Sch 96h], est optimisée pour l'accès rapide à un seul vecteur de caractéristiques (table de hachage). C'est à dire que la représentation entière d'une image (d'un objet) n'est pas directement accessible. Il est alors difficile de généraliser à partir des images stockées et de trouver des similarités globales entre différents objets et/ou différentes images.

2.2.4 Reconnaissance fondée sur une mémoire à deux stades

Nelson [Nel 95, Nel 96] propose une méthode qui combine une mémoire associative avec une technique de combinaison d'évidence ressemblant à une transformation de Hough. La méthode est fondée sur des clés robustes (appelés semi–invariantes par Nelson) associées à une hypothèse d'objet, à une évidence et à une configuration d'objet. La justification principale est la simplicité et la généralité d'une représentation fondée sur une mémoire.

La méthode utilise une mémoire associative. L'accès à la mémoire par une clé évoque des hypothèses associées pour les identités et les configurations d'objets qui pourraient correspondre à cette clé. Il est alors nécessaire que les clés puissent être extraites de façon robuste et qu'elles contiennent des informations suffisantes pour la spécification d'une configuration d'un objet. La deuxième étape de la méthode utilise une deuxième mémoire associative indexée par la configuration (et l'identité) d'objets. Cette deuxième mémoire maintient une estimation probabiliste de chaque hypothèse et les statistiques de l'occurrence des clés de la première mémoire (les statistiques sont utilisées pour améliorer l'algorithme de votes). L'idée est similaire à une transformation multidimensionnelle de Hough mais réduit les degrés de liberté considérablement de l'espace de transformations.

Le choix des clés est particulièrement important pour cette approche : les clés doivent être locales, donnant ainsi une robustesse par rapport aux occultations et aux recouvrements. En outre, la robustesse de clés permet la réduction de la mémoire. Les clés doivent être alors robustes (pour l'extraction) mais aussi discriminantes. Ces deux dernières contraintes (robuste à extraire et discriminante) compliquent le choix des clés. Nelson propose la construction de caractéristiques de plus haut niveau (discriminantes) à partir de caractéristiques plus simples en les groupant par des heuristiques (équivalent à un groupement perceptif). Dans [Nel 95], des chaînes de trois segments connectés sont utilisées pour la représentation d'objets polyédriques (en utilisant les relations de longueurs et d'angles de segments). [Nel 96] utilise aussi les plaques de courbes[5] pour l'indexation qui contiennent les chaînes de segments comme cas spécial.

Les expérimentations de [Nel 96] obtiennent une reconnaissance entre 70% et 95% pour une base de 7 objets en fonction du degré d'occultation et de recouvrement. La raison principale d'échec est la mauvaise performance de l'extraction de caractéristiques de bas niveau. De plus, le nombre de points de vue par objet était de l'ordre de 100 (correspondant à la sphère entière de vue en utilisant une distance de 20 degrés).

2.3 Reconnaissance statistique d'objets

Des méthodes statistiques sont fréquemment employées en vision par ordinateur. Les applications incluent la segmentation d'image [Can 86, Bel 89] et la modélisation probabiliste du problème de la mise en correspondance [Bre 93, Hut 95]. Les modèles statistiques sont aussi appliqués pour la localisation et la reconnaissance d'objets. Néanmoins, les objets eux–mêmes sont rarement représentés par une densité probabiliste. Cette thèse s'intéresse à l'acquisition automatique d'un modèle statistique d'objet.

Trois approches sont décrites brièvement. Toutes les trois apprennent un modèle statistique d'objet à partir d'images. L'approche d'images propres (section 2.3.1) emploie directement la distribution de valeurs de pixels d'images. La section 2.3.2 décrit une application de l'algorithme

5. curve patches

de maximisation de l'espérance[6] pour l'estimation de la densité probabiliste de caractéristiques de points. La section 2.3.3 introduit un algorithme d'apprentissage non supervisé d'un graphe probabiliste d'objets.

La difficulté majeure de l'apprentissage d'une densité probabiliste d'objet est le manque de données d'images. Comme les réseaux artificiels de neurones sont capables de généraliser à partir d'un ensemble de données, quelques systèmes de reconnaissance parmi les meilleurs sont fondés sur ces techniques [Pog 90]. [Shv 90, Ede 93] analysent théoriquement la capacité d'apprentissage de modèles d'objets à partir d'images. Les réseaux artificiels de neurones ne sont pas considérés dans cette thèse.

2.3.1 Approches d'images propres

Récemment de nombreux chercheurs [Sir 87, Kir 90, Tur 91a, Tur 91b, Mur 93, Mur 95, Mog 95, Ohb 96] utilisent la transformation de Karhunen–Loeve [Fuk 90] pour le calcul d'*images propres* dans le contexte de la reconnaissance d'objets. L'idée générale est l'interprétation d'une image avec $n = n_1 \times n_2$ pixels comme un vecteur ϕ à n dimensions. Ayant L images interprétées comme des vecteurs $\phi_i (i = 1, 2, \ldots L)$ à n dimensions, la matrice de covariance C est donnée par (avec $\overline{\phi} = \frac{1}{L} \sum_{i=1}^{L} \phi_i$) :

$$C = \sum_{i=1}^{L} (\phi_i - \overline{\phi})(\phi_i - \overline{\phi})^T \tag{2.3}$$

Supposons les vecteurs propres u_i et les valeurs propres λ_i correspondantes (voir [Fuk 90, Pre 92] et [Tur 91a] pour une re–formulation algébrique réduisant la complexité de calcul de vecteurs propres). De plus, supposons que $i = 1, 2, \ldots L - 1 : \lambda_i \geq \lambda_i + 1$. Une base optimale (dans le sens d'erreur de moindres carrés) à K dimensions est donnée par les premiers K vecteurs propres $u_1, u_2, \ldots, u_K$. La transformation de Karhunen–Loeve (ou analyse en composantes principales) est une technique classique de reconnaissance des formes. Néanmoins, elle est une des meilleures approches récentes de reconnaissance d'objets.

Dans le contexte de la reconnaissance des visages, Turk et Pentland [Tur 91a] et plus récemment Pentland et al. [Pen 94, Mog 95] ont réussi à appliquer la technique d'images propres aux grandes bases de visages. Un système temps–réel de reconnaissance d'objets 3D est implémenté par Murase et Nayar [Mur 95] fondé sur "l'espace propre modulaire" [7].

L'avantage majeur de la technique et de la représentation d'une image par un petit nombre de coefficients qui peuvent être stockés et recherchés de façon efficace. Malgré son succès, la méthode possède deux inconvénients principaux : premièrement, le modèle d'objets est global rendant l'approche sensible aux occultations partielles. Deuxièmement, tous les changements de valeurs de pixels, causés par une translation, par un changement d'échelle, par une rotation d'objet ou par une variation de l'éclairage, modifient les coefficients de la représentation de l'image. Deux méthodes sont employées pour traiter ces changements : la première normalise chaque image avant de la projeter sur l'espace propre et la seconde calcule l'espace propre en considérant tous les changements possibles. Il existe des fonctions puissantes de normalisation

6. Expectation–Maximization algorithm

7. "modular eigenspace"

pour des cas spécifiques comme la reconnaissance de visages. Néanmoins, il est difficile en général de supposer une telle fonction pour des objets 3D arbitraires. Étant donné ces difficultés, Murase et Nayar proposent par exemple une segmentation de l'objet du fond avant la projection sur l'espace propre.

Le calcul d'images de Fisher (remplaçant des images propres) était proposé par Belhumeur et al. [Bel 96] pour la reconnaissance de visages. Leur approche est plus appropriée aux changements de luminosité de l'éclairage que la technique des images propres. [Ohb 96] calcule des images propres à partir de petites régions d'images rendant la méthode plus robuste aux recouvrements et aux occultations partielles. L'approche d'images propres a été aussi appliquée au suivi d'objets 3D [Bla 96], à la reconnaissance de gestes [Mar 97] et à la transmission d'images pour une vidéo conférence [Cro 96]. En conclusion, l'approche d'images propres a réussi dans plusieurs domaines de la vision par ordinateur.

2.3.2 Apprentissage, localisation et identification statistique d'objets

Hornegger et Niemann [Hor 95, Hor 96] proposent une représentation d'objets par une densité probabiliste de caractéristiques d'objets. L'apprentissage et la localisation d'objets sont formulés comme des problèmes d'estimation de paramètres. Dans leur formulation, les caractéristiques sont interprétées comme des variables aléatoires. Un objet est alors représenté par une densité probabiliste d'un ensemble de caractéristiques. [Hor 95] utilise les coordonnés de points comme caractéristiques. Chaque transformation d'un objet (par exemple rotation, translation, changement d'échelle, changement de point de vue) est représentée par un paramètre de la densité probabiliste. Les auteurs s'intéressent, en particulier, au problème de la reconnaissance d'objets 3D à partir d'images 2D, et au problème de l'estimation d'un modèle 3D à partir des images 2D. A cela s'ajoute le problème de la projection de l'espace modèle sur le plan image.

Pendant l'apprentissage non supervisé (sans supposition de correspondance), la densité probabiliste d'un objet 3D est estimée à partir des images 2D en utilisant l'algorithme de maximisation de l'espérance. Cet algorithme est adéquat pour ce type de problème d'estimation incomplète. La densité probabiliste est représentée par une mélange paramétrique de distributions Gaussiennes multi–variables. Ce modèle est justifié dans le contexte de coordonnées de points utilisés comme caractéristiques. Malheureusement, la généralisation de la méthode à des ensembles de caractéristiques plus généraux n'est pas donnée.

Malgré la formulation élégante de l'approche, peu de résultats expérimentaux sont décrits. Dans [Hor 95], seulement quatre objets 2D et deux objets 3D sont employés. De plus, le temps de reconnaissance est important car la procédure d'identification est formulée comme un problème d'estimation de paramètres (problème d'optimisation),

2.3.3 Acquisition automatique de modèles d'objet

Plusieurs auteurs [Pop 93, Bei 94, Beb 95, Pop 95, Pop 96] s'intéressent à l'acquisition automatique de modèles d'objets. Les deux approches, citées dans cette section, suivent la même philosophie: en utilisant des caractéristiques simples (par exemple des lignes, des coins) des caractéristiques de plus haut niveau sont construites par groupement perceptif. Fondé sur ces caractéristiques et sur des ensembles de caractéristiques, [Bei 94] propose l'apprentissage d'une table de hachage probabiliste pour l'indexation. [Pop 93] propose l'apprentissage de relations entre différentes caractéristiques. Un modèle d'objet (et une image test) est représenté par

un graphe où les noeuds correspondent aux caractéristiques et les arcs représentent les relations entre les caractéristiques. En particulier, la probabilité d'observation d'une caractéristique apprise est utilisée pendant la reconnaissance pour un schéma d'alignement probabiliste. Notons que la densité probabiliste est obtenue par une estimation non paramétrique (estimation par fenêtre de Parzen [Fuk 90]). Pour chaque aspect 2D d'un objet, un graphe probabiliste est automatiquement appris. Un objet 3D est alors représenté par une collection de graphes probabilistes.

Comme dans la section précédente 2.3.2, la technique n'est appliquée qu'à un petit nombre d'objets.

2.4 Conclusion

Le chapitre a résumé quelques algorithmes de reconnaissance d'objets intéressants dans le cadre de cette étude. Trois types d'approches ont été décrites : les techniques des histogrammes de couleurs ont été introduites car leur rapidité et leur fiabilité sont la source initiale de motivation pour la réalisation de cette thèse. Comme la représentation d'objets dans cette thèse se base sur les statistiques de descripteurs locaux, d'autres techniques fondées sur des descripteurs locaux ont été décrites. Les méthodes statistiques de reconnaissance sont citées afin de mettre en évidence les avantages d'une formulation statistique de la représentation et de la reconnaissance d'objets.

Le chapitre suivant est dédié aux caractéristiques locales comme les dérivées Gaussiennes, les filtres de Gabor et les informations de couleurs. Cette thèse représente les objets par les statistiques de vecteurs de tels descripteurs locaux.

Chapitre 3

Descripteurs locaux

Tous les systèmes de reconnaissance s'appuient, de façon explicite ou implicite, sur certaines caractéristiques. Le choix des caractéristiques employées est délicat et dépend de facteurs comme les classes d'objets considérées, les caractéristiques des capteurs, le contexte et la tâche à accomplir. Ce choix se base souvent sur un compromis entre la précision et la généralité des caractéristiques. Malgré la connaissance de ces difficultés, le choix est souvent arbitraire et manuel. Une des contributions de cette thèse est une mesure quantitative permettant la comparaison de différents ensembles de caractéristiques pour un ensemble donné d'objets (voir l'application de la théorie de l'information à la reconnaissance d'objets dans la section 4.2.3).

Danse cette thèse nous proposons une formulation générale pour la représentation d'objets par une densité probabiliste d'un ensemble de descripteurs locaux d'apparences d'objets (voir section 4.1). Visant la généralité ces descripteurs locaux ne sont pas restreints à un type singulier d'objets ou à un ensemble spécial de descripteurs. Il est néanmoins nécessaire de formuler des contraintes minimales pour les descripteurs locaux. Une contrainte fondamentale est la *localité* des caractéristiques. En opposition aux caractéristiques locales, les caractéristiques globales sont sensibles aux occultations partielles et à toute perturbation locale de l'image (comme par exemple les réflections spéculaires). Une autre contrainte fondamentale est la *robustesse* des caractéristiques. La robustesse est une contrainte moins forte que l'invariance qui est généralement difficile à obtenir. Trois catégories de caractéristiques peuvent ainsi être distinguées :

caractéristiques invariantes : elles sont considérées constantes en présence de certaines transformations (par exemple une transformation affine ou orthographique),

caractéristiques équivariantes : leurs valeurs sont données en fonction d'une certaine transformation,

caractéristiques robustes : leurs valeurs varient peu en présence de certaines transformations.

Le contrainte d'invariance des caractéristiques est la plus efficace car elle réduit l'ensemble des valeurs possibles de caractéristiques. Si le calcul de caractéristiques invariantes est raisonnable (c'est à dire si elles sont stables et discriminantes) elles doivent être employées. Classiquement, les caractéristiques invariantes restreignent les classes d'objets. Elles s'appuient souvent sur le calcul de dérivées d'ordre supérieur (problème d'instabilité) et/ou elles ont un caractère global (problème d'occultation partielle). Chacune de ces contraintes limite la généralité de notre technique. En conséquence, le besoin d'invariance doit être affaibli.

Les caractéristiques équivariantes sont données par les dérivées Gaussiennes qui sont équivariantes par rapport à la rotation dans le plan image et au changement d'échelle. Néanmoins, l'équivariance de caractéristiques n'est par toujours accessible. La notion de robustesse de caractéristiques est plus générale. Notre argument est que de nombreuses caractéristiques, appropriées pour la reconnaissance d'objets, peuvent être calculées de façon robuste et peuvent coder l'information discriminante. Notre intérêt se dirige principalement vers des caractéristiques robustes (et si possible des caractéristiques équivariantes) qui peuvent être calculées localement et qui sont robustes par rapport au bruit dans l'image, au flou, à la rotation de l'image et au changement d'échelle.

La section 3.1.1 introduit les dérivées Gaussiennes, leur orientabilité[1] par rapport aux rotations d'image et leur équivariance par rapport aux changements d'échelles. Les dérivées Gaussiennes sont souvent employées en vision par ordinateur. Leur popularité provient de leur généralité (les images propres d'un grand nombre de régions d'images ressemblent aux Gaussiens [Rao 95b]), de leur capacité de modéliser les réponses de neurones [You 86] et de l'existence d'une implémentation recursive [Der 93]. De plus les dérivées Gaussiennes (comme les filtres de Gabor) sont robustes par rapport aux changements d'échelle d'approximativement $\pm 20\%$ [Sch 96h]. Les expérimentations menées dans cette thèse emploient souvent les dérivées Gaussiennes étant donné leur robustesse, leur orientabilité et leur équivariance par rapport à l'échelle.

Les filtres de Gabor (section 3.1.2) possèdent les mêmes propriétés (de robustesse, d'orientabilité et l'équivariance par rapport à l'échelle) que les dérivées Gaussiennes. Ils ont été employés lors de nos premières expérimentations. Comme les résultats de reconnaissance ont été pratiquement identiques à ceux obtenus avec les dérivées Gaussiennes, ces expérimentations ne sont pas décrites.

La couleur a été examinée pendant une étude liée à cette thèse [Col 96]. Cette étude a démontré que l'emploi des histogrammes de dérivées de logarithmes de couleurs constitue un moyen fiable pour la reconnaissance en présence de changements de l'intensité d'éclairage et même de la couleur de l'éclairage. La section 3.1.3 résume brièvement ces invariants.

La section 3.2 décrit des procédures de normalisation utilisables pour stabiliser les réponses de filtres par rapport au bruit et aux changements de l'éclairage. La robustesse de ces techniques de normalisation par rapport au bruit Gaussien additif est examinée dans la section 3.2.4. Des expérimentations sur leur robustesse par rapport aux changements de l'intensité de l'éclairage sont données dans la section 5.3.

1. steerability

3.1 Caractéristiques locales

La section 3.1.1 décrit les dérivées Gaussiennes, leur orientabilité par rapport aux rotations d'image et leur équivariance par rapport aux changements d'échelle. Les filtres de Gabor et le choix de paramètres de filtres sont introduits dans la section 3.1.2. Dans la Section 3.1.3 nous discutons sur quelques invariants de l'information de couleur.

3.1.1 Dérivées Gaussiennes

Cette section introduit les caractéristiques locales fondées sur les dérivées Gaussiennes. Les dérivées Gaussiennes sont souvent utilisées et bien maîtrisées [Fre 91, Rao 95b]. En utilisant les dérivées Gaussiennes l'échelle peut être choisie explicitement. De plus les dérivées peuvent être "orientées" selon une rotation arbitraire: il est possible de calculer les dérivées Gaussiennes de l'ordre n et de l'orientation ϕ à partir d'une combinaison linéaire d'un nombre fini de dérivées Gaussiennes de l'ordre n. Cette section décrit les dérivées Gaussiennes, développe leur équivariance par rapport à l'échelle et résume leur orientabilité

Soit donnée une distribution Gaussienne $G^\sigma(x,y)$:

$$G^\sigma(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.4}$$

La première dérivée dans la direction x est donnée par:

$$G_x^\sigma(x,y) \quad = \quad -\frac{x}{\sigma^2} G^\sigma(x,y) \tag{3.5}$$

La première dérivée dans la direction $\vec{v} = (\cos\phi \ \sin\phi)^T$ est donnée par:

$$G_{1,\phi}^\sigma(x,y) = \frac{\partial}{\partial \vec{v}} G^\sigma(x,y) \tag{3.6}$$

De la même façon la dérivée Gaussienne de l'ordre n dans la direction $\vec{v} = (\cos\phi \ \sin\phi)^T$ est définie par:

$$G_{n,\phi}^\sigma(x,y) = \frac{\partial^n}{\partial \vec{v}^n} G^\sigma(x,y) \tag{3.7}$$

Dans cette thèse nous utilisons seulement les dérivées Gaussiennes de l'ordre 1 et 2. Des notations spéciales sont alors introduites pour les dérivées employées. L'axe de x est défini comme étant parallèle au vecteur $\vec{v} = (1\ 0)^T$ c'est à dire à $\phi = 0°$. L'axe de y est défini par $\phi = 90°$ et est alors parallèle au vecteur $\vec{v} = (0\ 1)^T$. Les premières dérivées dans les directions x et y sont données par:

$$G_x^\sigma(x,y) \quad = \quad G_{1,0°}^\sigma(x,y) = -\frac{x}{\sigma^2} G^\sigma(x,y) \tag{3.8}$$

$$G_y^\sigma(x,y) \quad = \quad G_{1,90°}^\sigma(x,y) = -\frac{y}{\sigma^2} G^\sigma(x,y) \tag{3.9}$$

Fondées sur ces premières dérivées, la norme $Mag(x,y)$ et la direction $Dir(x,y)$ de la première dérivée peuvent être définies :

$$Mag(x,y) = \sqrt{(G_x^\sigma(x,y))^2 + (G_y^\sigma(x,y))^2} \tag{3.10}$$

$$Dir(x,y) = \arctan \frac{G_y^\sigma(x,y)}{G_x^\sigma(x,y)} \tag{3.11}$$

Trois dérivées de deuxième ordre sont données par :

$$G_{xx}^\sigma(x,y) = (\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2})G^\sigma(x,y) \tag{3.12}$$

$$G_{xy}^\sigma(x,y) = (\frac{xy}{\sigma^4})G^\sigma(x,y) \tag{3.13}$$

$$G_{yy}^\sigma(x,y) = (\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2})G^\sigma(x,y) \tag{3.14}$$

Fondées sur les premières et deuxièmes dérivées, des caractéristiques locales invariantes à la rotation d'image peuvent être définies :

$$Lap(x,y) = G_{xx}^\sigma(x,y) + G_{yy}^\sigma(x,y) \tag{3.15}$$

$$G12(x,y) = G_{xx}^\sigma(G_x^\sigma)^2 + 2G_{xy}^\sigma G_x^\sigma G_y^\sigma + G_{yy}^\sigma(G_y^\sigma)^2 \tag{3.16}$$

$Lap(x,y)$ est l'opérateur Laplacien. La deuxième caractéristique locale va être appelée $G12(x,y)$ dans cette thèse car elle se base sur les premières et les deuxièmes dérivées. $G12(x,y)$ a été introduit par Koenderink [Koe 84] comme filtre invariant à la rotation (d'image) et utilisé par Schmid et Mohr [Sch 96i, Sch 96h] pour la reconnaissance d'objets. Trois caractéristiques locales invariantes à la rotation d'image sont utilisées : $Mag(x,y)$, $G12(x,y)$ and $Lap(x,y)$ (équations 3.10, 3.16 et 3.15).

Les paragraphes suivants introduisent l'équivariance des dérivées Gaussiennes par rapport à l'échelle et l'orientabilité.

Équivariance de dérivées Gaussiennes par rapport à l'échelle

Comme mentionné plus haut, les caractéristiques locales doivent être calculables pour des échelles arbitraires. Ce calcul n'est pas seulement possible pour les dérivées Gaussiennes mais aussi pour d'autres types de filtres comme les filtres de Gabor (voir section 3.1.2). Cette section introduit l'équivariance de dérivées Gaussiennes par rapport à l'échelle.

Étant donné une fonction bidimensionnelle $p(x,y)$ et une version de la même fonction à une échelle différente : $f(x,y) = p(sx,sy)$. La mathématique analytique donne :

$$f(x,y) = p(sx,sy) \tag{3.17}$$

$$\frac{\partial}{\partial x} f(x,y) = s\frac{\partial}{\partial x} p(sx,sy) \tag{3.18}$$

$$\vdots$$

$$\frac{\partial^n}{\partial x^n} f(x,y) = s^n \frac{\partial^n}{\partial x^n} p(sx,sy) \tag{3.19}$$

Ces équations permettent le calcul de la dérivée à l'ordre n de la fonction f à partir de la dérivée à l'ordre n de $p(sx, sy)$. Ce calcul suppose la connaissance exacte de la fonction p. Dans la vision par ordinateur, cette supposition n'est pas toujours possible. En utilisant les dérivées Gaussiennes, la dérivée à l'ordre n de $f(x, y) = p(sx, sy)$ peut être calculée à partir de $p(x, y)$. La propriété de l'équivariance par rapport à l'échelle est discutée pour la première dérivée. La première dérivée Gaussienne de f est définie par :

$$\frac{\partial}{\partial x} f(x, y) = G_x^\sigma(x, y) \star f(x, y) \tag{3.20}$$

où $G_x^\sigma(x, y)$ est la première dérivée Gaussienne (voir équation (3.8)) et l'opérateur $\star$ signifie la convolution. C'est à dire (en utilisant aussi l'équation (3.18)):

$$\begin{aligned} \frac{\partial}{\partial x} f(x, y) &= s \frac{\partial}{\partial x} p(sx, sy) && (3.21) \\ &= s G_x^\sigma(x, y) \star p(sx, sy) && (3.22) \\ &= s G_x^{\sigma s}(x, y) \star p(x, y) && (3.23) \end{aligned}$$

La dernière équation montre le calcul de la première dérivée de f à partir de la première dérivée de $p(x, y)$ appelée *adaptation de dérivées Gaussiennes à l'échelle*. De manière analogue l'équation de l'adaptation de la dérivée de l'ordre n est obtenue par :

$$\frac{\partial^n}{\partial x^n} f(x, y) = s^n G_{x^n}^{\sigma s}(x, y) \star p(x, y) \tag{3.24}$$

C'est à dire que la dérivée de l'ordre n de la fonction $f(x, y)$ peut être calculée directement à partir de la fonction $p(x, y)$ (si f est donné par : $f(x, y) = p(sx, sy)$). Pour l'emploi de cette propriété le facteur d'échelle s doit être connu, ce qui ne peut pas être supposé en général. Classiquement la dérivée est calculée pour différents facteurs s. De plus la région de support du calcul de la dérivée de p doit être adaptée. Cette adaptation est exprimée par l'adaptation de la déviation standard σs du filtre Gaussien.

Cette adaptation de dérivées Gaussiennes aux changements d'échelle par le facteur s est appelée *équivariance* de dérivées Gaussiennes par rapport à l'échelle. Plus globalement toutes les caractéristiques sont équivariantes à un changement singulier s'il existe un certain paramètre connecté directement au changement.

L'*équivariance par rapport à l'échelle* est valable pour d'autres caractéristiques que les dérivées Gaussiennes. La même propriété, par exemple, est vraie pour les filtres de Gabor étant donné leur enveloppe Gaussienne.

Orientabilité de dérivées Gaussiennes à la rotation d'image

Pour calculer la réponse d'un filtre (par exemple d'un filtre Gaussien) à une rotation arbitraire ϕ la version correspondante du filtre peut être calculée. Si la rotation n'est pas connue a priori ou si de nombreuses réponses du filtre pour différentes rotations doivent être calculées, le calcul des différentes versions du filtre requiert un temps de calcul important. Il est alors désirable de définir un ensemble fini de filtres de base et d'interpoler selon une règle permettant le

calcul des réponses du filtre à partir de l'ensemble de base. Pour la première dérivée Gaussienne une telle règle d'interpolation est connue [Fre 91] :

$$G^{\sigma}_{1,\phi} = \cos\phi G^{\sigma}_{x} + \sin\phi G^{\sigma}_{y} \tag{3.25}$$

Pour formaliser cette propriété Freeman et Adelson [Fre 91] utilisent une fonction bidimensionnelle $F(x, y)$ et définissent $F^{\theta}(x, y)$ comme une version de $F(x, y)$ tournée de l'angle θ. Une fonction est *orientable*[2] si elle peut être décrite comme une version tournée d'elle–même :

$$F^{\phi}(x, y) = \sum_{j=1}^{J} k_j(\phi) F^{\theta_j}(x, y) \tag{3.26}$$

Cette équation est appelée la contrainte d'orientabilité. J correspond au nombre de fonctions d'interpolation $k_j(\theta)$ et les F^{θ_j} forment un ensemble fini de fonctions tournées de base. Il existe deux questions importantes : combien de fonctions d'interpolation sont nécessaires et comment les obtenir. Pour formuler deux théorèmes la fonction $F(x, y)$ est réécrite en coordonnées polaires $r = \sqrt{(x^2 + y^2)}$ et $\rho = \arg(x, y)$:

$$F(r, \rho) = \sum_{n=-N}^{N} a_n(r) e^{in\rho} \tag{3.27}$$

Le premier théorème [Fre 91] déclare que la contrainte d'orientabilité (équation 3.26) est satisfaite pour les fonctions extensibles à la forme de l'équation 3.27 si et seulement si les fonctions d'interpolation $k_j(\phi)$ sont les solutions de :

$$\begin{pmatrix} 1 \\ e^{i\phi} \\ \vdots \\ e^{iN\phi} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{i\theta_1} & e^{i\theta_2} & \cdots & e^{i\theta_J} \\ \vdots & \vdots & \ddots & \vdots \\ e^{iN\theta_1} & e^{iN\theta_2} & \cdots & e^{iN\theta_J} \end{pmatrix} \begin{pmatrix} k_1(\phi) \\ k_2(\phi) \\ \vdots \\ k_J(\phi) \end{pmatrix} \tag{3.28}$$

Par le choix des angles θ_j les fonctions d'interpolation sont données comme solutions de l'équation 3.28. Un deuxième théorème [Fre 91] déclare que le nombre minimal de fonctions d'interpolation est T, avec T le nombre de coefficients $a_n(r)$ non–zéro de l'équation 3.27. Les réponses théoriques aux deux questions (combien de fonctions d'interpolation sont nécessaires et comment obtenir ces fonctions) sont alors connues.

Ces deux théorèmes sont applicables aux dérivées Gaussiennes. En utilisant le deuxième théorème, le nombre minimal de fonctions d'interpolation de dérivées Gaussiennes de l'ordre n est $n + 1$. Les fonctions d'interpolation dépendent des angles θ_j choisis. Un choix approprié consiste en une distribution uniforme des angles entre $0°$ et $180°$ (Un autre choix des θ_j peut être justifié par la séparabilité de filtres dans les directions x et y. Voir pour les détails [Fre 91]).

2. steerable

A côté des fonctions d'interpolation de premières dérivées Gaussiennes (voir équation 3.25 : $\theta_1 = 0°, \theta_2 = 90°, k_1(\phi) = \cos(\phi), k_2(\phi) = \sin(\phi)$) les fonctions suivantes d'interpolation sont obtenues pour les dérivées Gaussiennes de l'ordre deux (en utilisant $\theta_1 = 0°$, $\theta_2 = 60°$ et $\theta_3 = 120°$) :

$$k_j(\phi) = \frac{1 + 2\cos(2(\phi - \theta_j))}{3} \text{ pour } j = 1, 2, 3 \tag{3.29}$$

$$G^{\sigma}_{2,\phi} = k_1(\phi)G^{\sigma}_{2,0°} + k_2(\phi)G^{\sigma}_{2,60°} + k_3(\phi)G^{\sigma}_{2,120°} \tag{3.30}$$

3.1.2 Filtres de Gabor

Les filtres de Gabor sont des filtres compacts réglés par une bande de fréquences. Les filtres de Gabor sont définis par une enveloppe Gaussienne modulée par un cosinus et par un sinus imaginaire donnant une paire de filtres constituée en un filtre pair et en un filtre impair [Gab 46]. La réponse d'un filtre de Gabor peut être représentée par une partie réelle et une partie imaginaire. Ce nombre complexe peut être exprimé en coordonnées polaires comme la norme et la direction.

Une paire de filtres de Gabor est compacte dans les domaines de l'espace et de la fréquence. Pour nos expérimentations la formulation de fonctions de Gabor de Daugman est employée [Dau 93] :

$$g(x, y) = e^{-\pi\left(\frac{(x-x_0)^2}{\alpha^2} + \frac{(y-y_0)^2}{\beta^2}\right)} e^{-2\pi i(u_0(x-x_0)+v_0(y-y_0))} \tag{3.31}$$

où (x_0, y_0) sont les coordonnés centrales du filtre, (α, β) sont les déviations standards pour déterminer la largeur et la longueur, et (u_0, v_0) spécifient les modulations dans les directions x et y avec la fréquence spatiale $\omega_0 = \sqrt{u_0^2 + v_0^2}$ et la direction $\theta_0 = \arctan(v_0/u_0)$. La fonction $G(u, v)$ de transfert du domaine de Fourier est donnée par :

$$G(u, v) = e^{-\pi\left((u-u_0)^2\alpha^2 + (v-v_0)^2\beta^2\right)} e^{-2\pi i(x_0(u-u_0)+y_0(v-v_0))} \tag{3.32}$$

L'avantage majeur des filtres de Gabor est le choix libre de la fréquence (c'est à dire de l'échelle) et de la largeur de la bande du filtre.

Pour le choix de paramètres d'un filtre 1D de Gabor, Westelius [Wes 92] propose le réglage de la déviation standard α et de la fréquence spatiale u_0. Ces deux paramètres déterminent la taille spatiale et la largeur de bande du filtre. Étant donné une certaine fréquence spatiale u_0, Westelius choisit le rayon du support de fréquence telle que la fonction est suffisamment petite pour $u = 0$ (la composante DC). Il définit la relation entre la composante DC et le maximum du filtre qui doit être plus petite qu'un seuil choisi P_{DC}. Pour la généralisation de cette relation à 2D nous définissons $\alpha = \beta$. La relation est alors donnée par ($(x_0, y_0) = (0, 0)$) :

$$\frac{G(0,0)}{G(u_0, v_0)} \leq P_{DC} \Rightarrow \alpha \geq \frac{\sqrt{-\ln P_{DC}}}{\sqrt{\pi}\omega_0} \tag{3.33}$$

Le support spatial et le support fréquentiel d'une fonction de Gabor sont tous les deux infinis. Pour définir un filtre numérique il faut échantillonner et limiter le support spatial de la fonction de Gabor. La limite de la taille spatiale est choisie tel que l'amplitude du filtre est plus petite qu'un seuil P_{cut} pour toutes les positions en dehors de la limite. Le rayon $R = \sqrt{r_x^2 + r_y^2}$ du support spatial est donc :

$$\frac{\|g(r_x, r_y)\|}{\|g(0,0)\|} \leq P_{cut} \Rightarrow R \geq \alpha \frac{\sqrt{-\ln P_{cut}}}{\sqrt{\pi}} \tag{3.34}$$

Notons que le rayon R affecte la composante DC du signal. C'est à dire qu'il faut vérifier la composante DC après la troncature du support du filtre.

En utilisant l'équation d'Euler ($e^{i\omega x} = \cos \omega x + i \sin \omega x$) l'équation 3.31 peut être réécrite sous forme polaire :

$$g(x,y) = Re(g(x,y)) + Im(g(x,y)) \tag{3.35}$$

$$Re(g(x,y)) = \cos\left(-2\pi\left(u_0(x-x_0) + v_0(y-y_0)\right)\right) e^{-\pi\left(\frac{(x-x_0)^2}{\alpha^2} + \frac{(y-y_0)^2}{\beta^2}\right)} \tag{3.36}$$

$$Im(g(x,y)) = i \sin\left(-2\pi\left(u_0(x-x_0) + v_0(y-y_0)\right)\right) e^{-\pi\left(\frac{(x-x_0)^2}{\alpha^2} + \frac{(y-y_0)^2}{\beta^2}\right)} \tag{3.37}$$

La partie réelle $Re(g(x,y))$ peut être interprétée comme une deuxième dérivée et la partie imaginaire $Im(g(x,y))$ comme une première dérivée (dans la direction de θ_0).

3.1.3 Couleur

La couleur est un des descripteurs locaux le plus intéressant et le moins employé pour la reconnaissance d'objets. La couleur est intéressante car pour de nombreuses classes d'objets (par exemple des fleurs, des arbres, des fruits, des rues, des visages) elle peut servir comme indice principal pour l'identification et la classification. Malheureusement, elle est relativement instable en présence de changements d'éclairage (comme de changements de la luminosité et de la couleur d'éclairage). C'est probablement pour cette raison qu'elle est rarement utilisée pour la reconnaissance d'objets. En opposition à ce point de vue globalement accepté, des chercheurs ont récemment proposé des invariants de couleurs permettant la définition de caractéristiques invariantes aux changements d'éclairage. Pour dériver ces invariants il faut introduire le modèle général de couleur.

Le modèle général de couleur est classiquement composé de trois parties ($\mathbf{x}$ étant une position d'image bidimensionnelle et λ les longueurs d'ondes de la source d'éclairage) :

- $E(\mathbf{x}, \lambda)$ est la distribution de l'énergie spectrale de la source d'éclairage,
- $R(\mathbf{x}, \lambda)$ est la fonction de réflexion des surfaces de la scène,
- $r_k(\lambda)$ la fonction de sensibilité du capteur k (souvent $k \in \{Rouge, Vert, Bleu\}$).

Le signal $S_k(\mathbf{x})$ transmis par le capteur k peut être modélisé par l'équation suivante :

$$S_k(\mathbf{x}) = \int_\lambda E(\mathbf{x}, \lambda) R(\mathbf{x}, \lambda) r_k(\lambda) d\lambda \tag{3.38}$$

En utilisant trois signaux de capteur le but est la dérivation d'un descripteur qui dépend seulement de la fonction de réflexion $R(\mathbf{x}, \lambda)$ de l'objet. Ce descripteur doit être indépendant de la distribution de l'énergie spectrale $E(\mathbf{x}, \lambda)$ de la source d'éclairage et doit être indépendant des caractéristiques de capteurs $r_k(\lambda)$.

L'approche classique consiste à développer un "algorithme de la constance de couleurs" [3] qui extrait (soit exactement, soit approximativement) la fonction de réflexion $R(\mathbf{x}, \lambda)$ [Mal 85, Mal 86, For 90, Fun 91, Fin 95a, Fin 95b, Bar 96]. Comme cette extraction n'est en général pas possible, de nombreuses suppositions sont faites qui limitent l'applicabilité de ces algorithmes aux circonstances contraintes ou contrôlées.

Un petit nombre d'auteurs propose un calcul d'invariants de couleur dépendant surtout de la fonction de réflexion $R(\mathbf{x}, \lambda)$. Ces invariants ont été proposés par exemple par Nagao [Nag 95] ainsi que Funt et Finlayson [Fun 95].

Modèle de capteur à spectre étroit

Afin de simplifier le modèle général de couleur, plusieurs auteurs utilisent le modèle de capteur à spectre étroit :

$$r_k(\lambda) = \delta(\lambda - \lambda_k) \tag{3.39}$$

avec λ_k la longueur d'onde de la sensibilité maximale du capteur k. Le modèle général de couleur devient une multiplication simple :

$$S_k(\mathbf{x}) = R(\mathbf{x}, \lambda_k) E(\mathbf{x}, \lambda_k) \tag{3.40}$$

Malgré la simplification extrême de cette supposition, elle permet le calcul de plusieurs invariants qui sont robustes aux changements de la luminosité d'éclairage [Col 96]. Les sections suivantes introduisent les invariants proposés par Nagao et par Funt et Finlayson.

Invariants de couleurs proposés par Nagao

En supposant $E(\mathbf{x}, \lambda_i) = E(\lambda_i)$ de toute l'image (éclairage constant) la relation suivante peut être calculée pour des capteurs différents i et j (pour la même position d'image $\mathbf{x}$) :

$$\gamma_{ij} = \frac{S_i(\mathbf{x})}{S_j(\mathbf{x})} = \frac{R(\mathbf{x}, \lambda_i) E(\lambda_i)}{R(\mathbf{x}, \lambda_j) E(\lambda_j)} \tag{3.41}$$

La même relation γ'_{ij} peut être calculée pour une autre image I' (du même objet mais d'un éclairage différent) à la position correspondante $\mathbf{x}'$. La relation entre γ_{ij} et γ'_{ij} est constante pour toute l'image :

3. color constancy algorithm

$$\frac{\gamma_{ij}}{\gamma'_{ij}} = \epsilon_{ij} = \frac{E(\lambda_i)}{E(\lambda_j)} / \frac{E'(\lambda_i)}{E'(\lambda_j)} \tag{3.42}$$

Cette constante ne dépend que de la distribution de l'énergie de la source d'éclairage. Il est alors possible de normaliser l'image avec cette constante. L'image normalisée de couleurs est donc indépendante de l'éclairage.

Invariants de couleurs proposés par Funt et Finlayson

Funt et Finlayson proposent le calcul de la relation de réponses entre deux positions d'image voisines $\mathbf{x}$ et $\mathbf{y}$ du même capteur k :

$$\frac{S_k(\mathbf{x}, \lambda_k)}{S_k(\mathbf{y}, \lambda_k)} = \frac{R(\mathbf{x}, \lambda_k)E(\mathbf{x}, \lambda_k)}{R(\mathbf{y}, \lambda_k)E(\mathbf{y}, \lambda_k)} \tag{3.43}$$

En supposant $E(\mathbf{x}, \lambda_k) = E(\mathbf{y}, \lambda_k)$, c'est à dire que la distribution de l'énergie est localement constante, l'invariance suivante est définie :

$$\frac{S_k(\mathbf{x}, \lambda_k)}{S_k(\mathbf{y}, \lambda_k)} = \frac{R(\mathbf{x}, \lambda_k)}{R(\mathbf{y}, \lambda_k)} \tag{3.44}$$

Pour éviter l'opération de division, Funt et Finlayson proposent le calcul de différences de logarithmes :

$$\log S_k(\mathbf{x}, \lambda_k) - \log S_k(\mathbf{y}, \lambda_k) = \log R(\mathbf{x}, \lambda_k) - R(\mathbf{y}, \lambda_k) \tag{3.45}$$

Notre résumé d'expérimentations [Col 96] était que ces derniers invariants (équation 3.45) sont stables pour des changements importants de l'intensité de l'éclairage. En présence de changements mineurs de la couleur d'éclairage ces invariants se comportent bien. En présence de changements importants de la couleur d'éclairage les invariants ne se comportent plus très bien, probablement en raison de la simplification extrême du modèle à spectre étroit.

3.2 Normalisation de réponses de filtre

Les effets de variations de l'intensité du signal peuvent être supprimés par une normalisation de réponses de filtre. Cette normalisation doit être examinée selon deux points de vue. Le premier point de vue concerne le comportement de la normalisation en présence de bruit additif. Le deuxième point de vue concerne le résultat de la normalisation en présence de variations de l'intensité du signal causées par les changements de l'intensité de l'éclairage, de l'ouverture du diaphragme ou du gain de numérisateur. Dans cette section nous introduisons plusieurs techniques de normalisation.

Dans la section 3.2.4 nous examinons la robustesse des techniques de normalisation par rapport au bruit Gaussien additif. La section 5.3 décrit des expérimentations montrant une bonne robustesse par rapport aux changements de l'intensité de l'éclairage en utilisant la normalisation par énergie ou par variance. Dans la suite $I(x, y)$ signifie l'image et $M(i, j)$ le masque d'un filtre (comme par exemple le masque d'un filtre de Gabor ou d'une dérivée Gaussienne).

3.2.1 Normalisation par moyenne et variance

La normalisation par variance (ou par moyenne et variance) consiste en la soustraction de la moyenne de chaque voisinage et en la division par la variance du voisinage. La normalisation par variance est invariante par rapport à l'échelle dans le cas spécial d'images binaires. Cette invariance n'est pas valable pour des signaux arbitraires de niveaux de gris. Dans ces cas l'équivariance de filtres Gaussiens et de filtres de Gabor permet le choix explicite de l'échelle. L'échelle devient un paramètre libre et réglable pour la détection d'un objet (voir sections 3.1.1 et 3.1.2). La normalisation par variance est définie par :

$$I_{var}(x,y) = \frac{\sum_{i,j=-m,-n}^{m,n}(I(x+i,y+j) - \overline{I(x,y)})M(i,j)}{\sqrt{\sum_{i,j=-m,-n}^{m,n}(I(x+i,y+j) - \overline{I(x,y)})^2}\sqrt{\sum_{i,j=-m,-n}^{m,n} M(i,j)^2}} \tag{3.46}$$

avec

$$\overline{I(x,y)} \quad = \quad \frac{1}{(2m+1)(2n+1)} \sum_{i,j=-m,-n}^{m,n} I(x+i,y+j) \tag{3.47}$$

La moyenne du voisinage pourrait être remplacée par la moyenne de l'image entière en supposant un changement uniforme de l'intensité d'éclairage. Cette supposition n'est en général pas valable mais permet une simplification du calcul.

La normalisation par variance est relativement sensible au bruit Gaussien additif. Cette sensibilité peut être comprise en considérant les effets de la normalisation d'un signal constant. Par la soustraction de la moyenne, le signal est zéro et la variance est également zéro. Étant donné la quantification et l'échantillonage, la discrétisation d'une image introduit nécessairement du bruit Gaussien additif. Même si l'énergie du bruit est minimale, la normalisation de réponses de filtre répond uniquement au bruit. Cela rend la normalisation par variance inappropriée dans notre contexte.

3.2.2 Normalisation par énergie

L'énergie du signal est donné par la racine carrée de la somme des carrés de ses coefficients. La normalisation par énergie normalise l'énergie de chaque voisinage à l'unité. Le coût de la normalisation de chaque voisinage peut être minimisé par le calcul incrémental de la somme de carrés du voisinage pendant la convolution. La division par l'énergie du voisinage élimine les variations de l'intensité du signal qui proviennent par exemple de changements de l'intensité de l'éclairage. Cette division peut rendre les réponses de filtres invariantes par rapport aux changements de l'intensité de l'éclairage (voir l'expérimentation de la section 5.3) :

$$I_{ene}(x,y) = \frac{\sum_{i,j=-m,-n}^{m,n} I(x+i,y+j)M(i,j)}{\sqrt{\sum_{i,j=-m,-n}^{m,n} I(x+i,y+j)^2}\sqrt{\sum_{i,j=-m,-n}^{m,n} M(i,j)^2}} \tag{3.48}$$

3.2.3 Normalisation par max–min

La normalisation par max–min rend les réponses de filtres invariantes aux changements de l'intensité de l'éclairage. Pour assurer l'invariance le signal est supposé être à l'intérieur de la région linéaire de la gamme dynamique de la caméra [Bob 95].

Malheureusement, la normalisation par max–min est extrêmement sensible au bruit, en particulier pour des images presque binaires.

$$I_{maxmin}(x,y) = \frac{\sum_{i,j=-m,-n}^{m,n} I(x+i,y+j)M(i,j)}{\max_{i,j} I(x+i,y+j) - \min_{i,j} I(x+i,y+j)} \tag{3.49}$$

3.2.4 Robustesse de techniques de normalisation en présence de bruit Gaussien additif

Cette section examine des expérimentations pour déterminer la sensibilité de différentes techniques de normalisation par rapport au bruit Gaussien additif. Ces expérimentations emploient huit images artificielles. Les résultats de deux images sont résumés : la première image s'appelle *Sinus* et la deuxième *Grille*. L'image *Sinus* contient une courbe de sinus d'une longueur d'onde de 45 pixels. L'image *Grille* est une image binaire avec une grille de carrés de taille 30×30. Les deux images sont montrées par la figure 3.1.

FIG. 3.1 – *L'image Sinus et l'image Grille*

Les graphes de la figure 3.2 montrent principalement le comportement relatif des différentes techniques de normalisation. C'est à dire, la valeur absolue de χ^2_{qv} n'est pas vraiment intéressante car elle dépend de paramètres d'histogrammes. Ces expérimentations utilisent trois filtres de Gabor (voir section 3.1.2) :

- *G3*: filtres de Gabor avec longueur d'onde de 2.8 pixels (fenêtre de 7×7) dans les directions x et y,
- *G5*: filtres de Gabor avec longueur d'onde de 5.7 pixels (fenêtre de 15×15) dans les directions x et y,
- *G7*: filtres de Gabor avec longueur d'onde de 11.3 pixels (fenêtre de 30×30) dans les directions x et y,

La figure 3.2 montre les résultats pour l'image *Sinus*. Du bruit Gaussien additif était ajouté à l'image *Sinus* avec des déviations standards de $\sigma = 1, 2, 3, \ldots, 20$ (abscisse dans les graphes). L'histogramme bidimensionnel (*G3*, *G5* ou *G7*) de l'image initiale est stocké (équivalent à $\sigma = 0$). Cet histogramme est comparé (en utilisant χ^2_{qv} comme fonction de comparaison, voir section 5.1.3) aux histogrammes ayant du bruit Gaussien additif. L'ordonnée des graphes correspond à la fonction de comparaison.

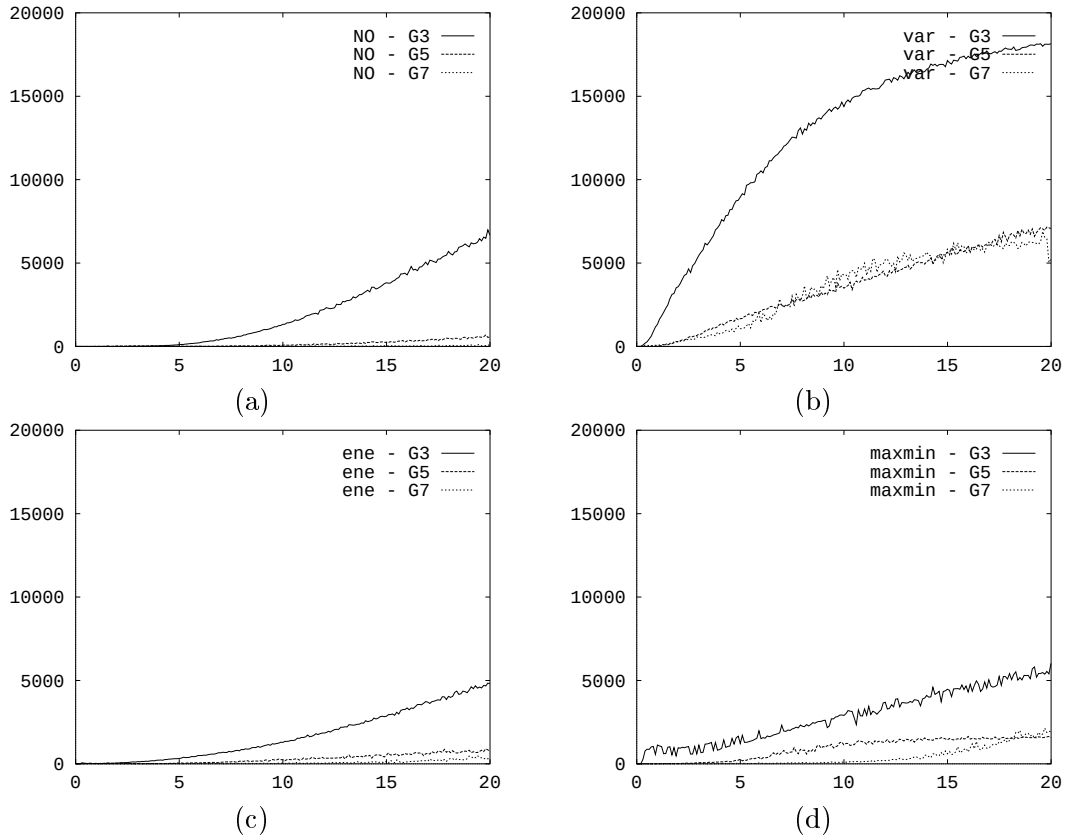


FIG. 3.2 – *L'image Sinus: (a) Robustesse sans normalisation. (b) Robustesse avec normalisation par variance. (c) Robustesse avec normalisation par énergie. (d) Robustesse par max–min*

Le premier résultat de l'image *Sinus* montre que les filtres de Gabor sont relativement robustes au bruit Gaussien additif (figure 3.2(a)). Cette robustesse est explicable par le lissage par la partie Gaussienne de la fonction de Gabor. La stabilité du filtre de Gabor augmente avec la taille du filtre de Gabor. C'est seulement dans le cas de normalisation par variance que les filtres de Gabor se comportent moins bien (voir figure 3.2(b)). Pour un deuxième résultat les différentes techniques de normalisation sont comparées : la robustesse *sans* normalisation est plutôt stable (figure 3.2(a)). La normalisation par *variance* par contre perturbe le bon comportement de filtres de Gabor. La normalisation par *max–min* donne des résultats moins stables mais encore bons (figure 3.2(d)). Les meilleurs résultats sont obtenus avec la normalisation par *énergie* (figure 3.2(c)).

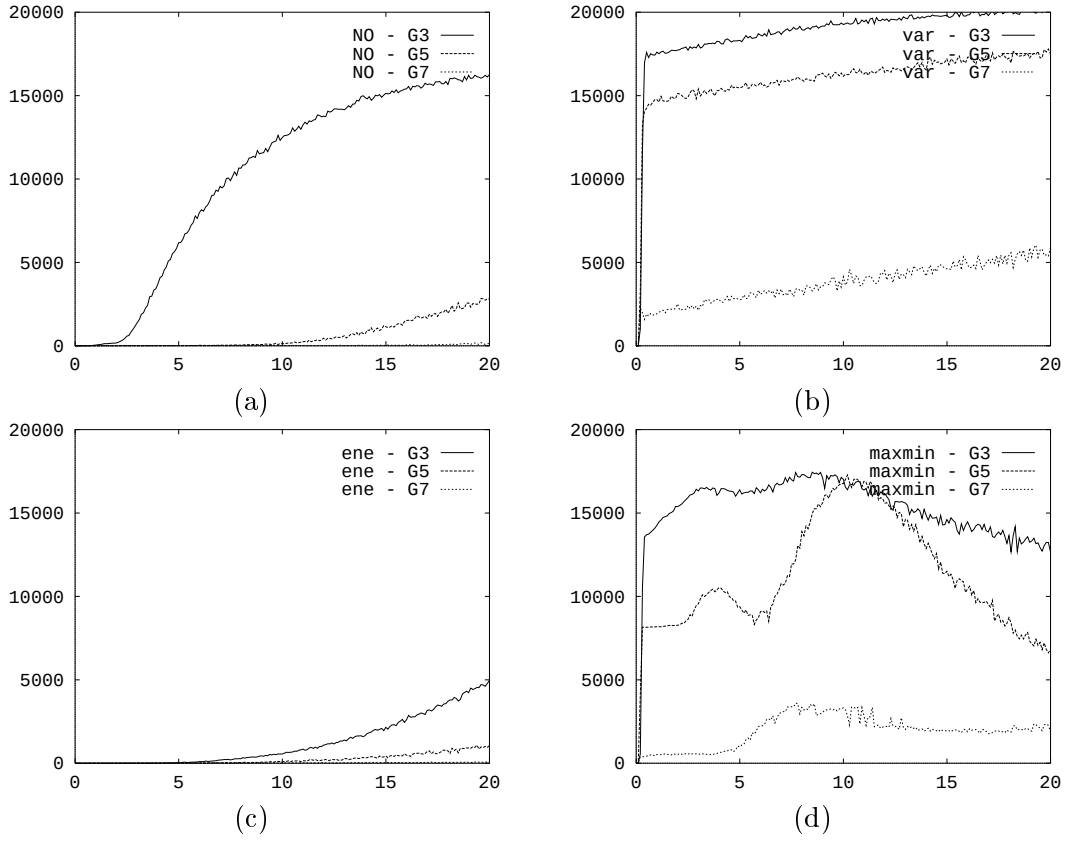


FIG. 3.3 – *L'image de Grille: (a) Robustesse sans normalisation. (b) Robustesse avec normalisation par variance. (c) Robustesse avec normalisation par énergie. (d) Robustesse par normalisation par max–min*

La figure 3.3 montre la robustesse des différentes techniques de normalisation pour l'image de *Grille*. La robustesse *sans* normalisation est moins bonne que dans le cas de l'image de *Sinus*. Les deux techniques de normalisation par *max–min* et par *variance* ne donnent pas de résultats satisfaisants. En particulier la normalisation par variance amplifie l'influence de bruit. Les meilleurs résultats sont obtenus par la normalisation par *énergie* qui est très stable.

Les chapitres suivants utilisent seulement la normalisation par *énergie* car elle semble appropriée pour rendre les réponses de filtres robustes aux bruits Gaussiens additifs. Les résultats

expérimentaux sont très satisfaisants en utilisant la normalisation par énergie.

3.3 Conclusion

Ce chapitre a décrit des descripteurs locaux fondés sur les dérivées Gaussiennes, les filtres de Gabor et l'information de couleurs. Les statistiques de vecteurs de tels descripteurs locaux sont utilisées pour la représentation d'objets. Cette thèse emploie souvent des dérivées Gaussiennes introduites dans la section 3.1.1. L'orientabilité de dérivées Gaussiennes à la rotation de l'image et l'équivariance par rapport à l'échelle ont été introduites. Ces deux propriétés rendent les dérivées Gaussiennes appropriées pour la représentation d'objets dans notre contexte. Les filtres de Gabor possèdent les mêmes propriétés mais l'implémentation récursive [Der 93] permet un calcul plus rapide de dérivées Gaussiennes. Les invariants de couleurs offrent la possibilité de descripteurs locaux et robustes en présence de changements de l'intensité et de la couleur de l'éclairage.

La section 3.2 a introduit les techniques de normalisation de réponses de filtres qui peuvent rendre les réponses robustes au bruit et au changement de l'intensité de l'éclairage. Une analyse de la robustesse par rapport au bruit Gaussien additif permet de conclure que la normalisation par énergie donne de meilleurs résultats. Étant donné ce résultat et étant donné les résultats d'expérimentations de la section 5.3 (étude de la robustesse par rapport au changement de l'intensité de l'éclairage) dans les expérimentations des chapitres suivants la normalisation par énergie est employée.

Pour généraliser l'approche des histogrammes de couleurs, le chapitre suivant développe la représentation statistique d'objets. Cette représentation statistique est fondée sur un ensemble arbitraire de vecteurs de descripteurs locaux introduits dans ce chapitre. En général, des combinaisons arbitraires de descripteurs sont imaginables si elles sont appropriées pour un ensemble particulier d'objets.

Chapitre 4

Représentation statistique d'objets

Le chapitre introduit l'utilisation d'histogrammes multidimensionnels de champs réceptifs pour la représentation statistique des apparences d'objets (section 4.1). Principalement, les objets 3D sont représentés par une densité probabiliste des caractéristiques locales. Ces caractéristiques sont bidimensionnelles et peuvent être extraites de façon robuste à partir des images d'objets. L'apprentissage du modèle d'objet, c'est à dire de la densité probabiliste, est alors automatique. Cette étude utilise des caractéristiques locales décrites dans le chapitre 3. La méthode peut et doit être appliquée en utilisant d'autres caractéristiques locales, comme par exemple, des moments de niveaux de gris ou des invariants géométriques.

La section 4.2 développe une analogie entre la théorie de l'information et la reconnaissance d'objets. Par conséquent, les concepts de la théorie de l'information s'appliquent dans le contexte de la reconnaissance d'objets. La transinformation, par exemple, peut servir pour l'évaluation quantitative d'un ensemble de descripteurs locaux (par rapport aux objets utilisés).

4.1 Représentation statistique d'objets

La suite développe une représentation statistique, fondée sur la densité probabiliste des caractéristiques locales d'un objet o_n. Nous supposons un ensemble M de mesures choisies et fixes. La densité probabiliste de l'ensemble M de mesures de l'objet o_n varie selon les changements de l'apparence de l'objet. Il est nécessaire de modéliser ces changements de l'apparence dans la densité probabiliste. Quatre catégories de changements peuvent être distinguées (voir figure

4.1) :

Transformation similaire dans le plan image : elle est décrite par trois degrés de liberté en translation (t_x, t_y et t_z) et un degré de liberté en rotation (r_z) (voir figure 4.1).

Transformation 3D d'un objet : il existe deux degrés de liberté supplémentaires en rotation (r_x et r_y) par rapport à la transformation similaire (voir figure 4.1).

Changements de la scène : ceux-ci incluent des occultations partielles et des changements du fond de la scène.

Conditions d'enregistrement : elles varient selon les modifications d'éclairage et selon les perturbations telles que les bruits de signal, les erreurs de discrétisation et le flou. Généralement, ces changements ne sont ni contrôlables ni prévisibles.

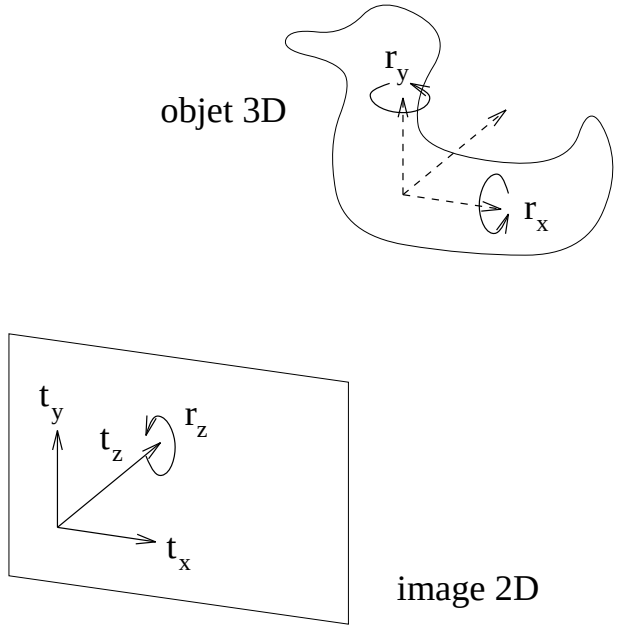


FIG. 4.1 – *Différentes composantes de la rotation et de la translation d'un objet 3D*

La densité probabiliste de l'objet o_n peut être écrite en fonction des paramètres suivants :

$$p(M|o_n, R, T, S, I) \tag{4.50}$$

où M est l'ensemble des caractéristiques locales m_k, o_n est la référence d'un objet (ou d'une classe d'objets), R correspond aux trois degrés de liberté de la rotation, T correspond aux trois

degrés de liberté de la translation, S décrit les changements de la scène et I décrit les conditions d'enregistrement.

L'estimation fiable de cette densité probabiliste multidimensionnelle est un problème délicat. La difficulté provient du besoin fondamental en données suffisantes requises pour l'estimation d'une densité probabiliste. En général, la quantité de données nécessaire à l'apprentissage croît de façon exponentielle avec le nombre de paramètres [Int 93]. La méthode la plus efficace pour réduire le nombre de dimensions consiste à utiliser des caractéristiques invariantes à certains changements. De nombreux chercheurs emploient les mesures invariantes avec succès dans des circonstances variées [Bur 92, Mun 92, Mun 93]. Malheureusement, ces invariants restreignent considérablement les types d'objets. Comme décrit dans le chapitre 3, l'applicabilité de notre approche ne doit pas être limitée par l'utilisation d'un type particulier d'invariants.

Le nombre de dimensions de la densité probabiliste (équation 4.50) peut être réduit également par l'utilisation des caractéristiques locales, qui sont robustes par rapport à certains changements de l'apparence. Des caractéristiques sont dites *robustes* lorsqu'elles sont peu sensibles aux changements. Typiquement, il existe un intervalle, dans lequel les caractéristiques sont pratiquement constantes. Dans ce contexte, les caractéristiques robustes peuvent être vues comme des quasi–invariants. Cette notion de robustesse présente de nombreux avantages car beaucoup de caractéristiques peuvent être calculées de façon robuste sans être invariantes en général.

Une catégorie de changements, les *conditions d'enregistrement*, contient des changements qui ne peuvent pas être contrôlés. Dans ce cas, il est nécessaire, pour le succès de notre approche, que les descripteurs locaux soient calculés de façon robuste par rapport à ces changements. Cela demande une analyse de la robustesse des caractéristiques locales en présence de ces changements : la section 3.2.4 examine la robustesse des caractéristiques locales et des différentes techniques de normalisation par rapport au bruit Gaussien additif. La section 5.3 décrit des expérimentations qui manifestent une robustesse considérable en présence de variations de l'intensité de l'éclairage.

La modélisation des changements de la scène, comme l'*occultation partielle* et le *changement du fond de la scène*, est difficile. Une possibilité de modélisation consiste à inclure ces changements dans le processus d'estimation de la densité probabiliste. Hornegger et Niemann [Hor 95] proposent la modélisation des occultations partielles par un objet particulier: le fond. L'introduction d'une probabilité – proportionnelle à la portion visible de l'objet – pour le fond permet le calcul de la probabilité de présence d'un objet. Le processus de reconnaissance n'estime pas seulement l'objet et sa pose, mais aussi le degré de son occultation partielle. La reconnaissance d'objets devient, alors, un processus d'optimisation. Bien que ce processus d'optimisation permette une formulation élégante, il demande un temps de calcul important. Contrairement à cette approche, le chapitre 7 propose une technique de reconnaissance probabiliste qui reconnaît des objets à partir d'une petite portion visible d'objet. Cet algorithme de reconnaissance est rapide et en même temps robuste par rapport aux occultations partielles. En conséquence, les occultations partielles ne sont pas modélisées dans la densité probabiliste. Dans ce contexte, les changements du fond sont considérés comme un cas particulier des occultations partielles.

Une transformation 3D d'un objet consiste en une *translation* (trois paramètres) et une rotation (trois paramètres). Les trois paramètres de la translation et le paramètre r_z de la rotation constituent une transformation similaire. Le paramètre r_z modélise une rotation dans le plan image (voir figure 4.1). Les deux derniers paramètres r_x et r_y de la rotation représentent

un changement de point de vue.

Une *transformation similaire* peut être traitée par une transformation du contenu de l'image. Cette transformation consiste en une translation bidimensionnelle dans le plan image (t_x et t_y), une rotation dans le plan image (r_z) et un changement d'échelle (t_z). Cette transformation du contenu de l'image inclut des interpolations et des échantillonnages. La résolution de l'image et la taille de l'objet dans l'image limitent la précision de cette transformation : une résolution plus grande permet le traitement d'un changement d'échelle t_z plus importante. Plus l'objet est grand dans l'image, plus l'erreur résultant de la transformation est petite.

En général, le problème de la mise en correspondance de modèles d'objets et d'une image test est difficile et requiert un temps de calcul important. Ce problème peut être contourné si les deux paramètres de la translation t_x et t_y ne sont pas représentés dans la densité probabiliste. Plusieurs avantages justifient ce choix : premièrement, le problème de la mise en correspondance en translation n'existe plus. Deuxièmement, le nombre de dimensions de la densité probabiliste est réduit de moitié. Troisièmement, le problème de l'estimation de la densité probabiliste devient faisable. L'estimation devient faisable grâce à la réduction du nombre de dimensions et aussi grâce à la quantité de données provenant de l'image d'un objet : une image de taille 512×512 contient approximativement $500^2 = 250000$ exemples d'apprentissage pour l'estimation de la densité probabiliste.

Comme expliqué plus haut, le troisième paramètre de la translation t_z peut être traité directement par une transformation de l'image. Pour le traitement de t_z, cette thèse exploite la propriété d'équivariance des descripteurs locaux par rapport au changement d'échelle. Une rotation r_z dans le plan image peut être traitée par des descripteurs invariants. Schmid [Sch 96h], par exemple, propose l'utilisation de tels descripteurs invariants pour la reconnaissance d'objets. L'inconvénient principal est que ces descripteurs ne préservent pas l'information d'orientation. Un autre inconvénient provient de la supposition que toutes les rotations d'un objet sont équiprobables. Cette thèse utilise des descripteurs invariants et variants par rapport à la rotation. Dans le contexte de descripteurs variants, une rotation dans le plan image est traitée par la propriété d'"orientabilité" [1] des descripteurs locaux (voir section 3.1.1).

Les deux paramètres r_x et r_y de la rotation représentent un *changement de point de vue* de l'observateur. Plusieurs auteurs [Bur 90, Cle 91] montrent, qu'en général, il n'existe pas de descripteurs invariants par rapport au changement de point de vue. Néanmoins, il existe des descripteurs intéressants pour des cas particuliers [Mun 92, Mun 93]. Comme argumenté plus haut, l'applicabilité de notre approche ne doit pas être limitée par des invariants spécialisés. Les deux paramètres r_x et r_y sont alors modélisés dans la densité probabiliste.

De la densité probabiliste originale (équation 4.50), il subsiste trois composantes de la rotation et une composante de la translation :

$$p(M|o_n, r_x, r_y, r_z, t_z) \tag{4.51}$$

En considérant un vecteur m_k des descripteurs locaux de dimension L, la représentation d'un objet o_n est donnée par une densité probabiliste de dimension $L + 4$. Dans le cas des descripteurs invariants par rapport à la rotation, la représentation est donnée par une densité de dimension $L + 3$.

1. terme anglais : steerability

4.1.1 Estimation et représentation par des histogrammes multidimensionnels de champs réceptifs

Il existe différentes possibilités pour estimer et représenter la densité probabiliste d'un objet (équation 4.51). Des schémas paramétriques et non–paramétriques sont distingués. Une estimation paramétrique suppose une distribution de données, comme par exemple, une distribution Gaussienne ou une distribution de Poisson. L'algorithme d'apprentissage estime alors les paramètres de la distribution supposée. Hornegger et Niemann [Hor 95] utilisent une mélange paramétrique de distributions Gaussiennes multi–variables. Leur modèle statistique considère le comportement statistique des caractéristiques, la mise en correspondance des caractéristiques, et la projection du modèle dans le plan de l'image. La supposition d'une mélange de distributions Gaussiennes est justifiée pour des caractéristiques comme les coordonnées de points. Par contre, cette supposition ne peut pas être appliquée pour des caractéristiques plus générales.

L'autre possibilité consiste en une estimation non–paramétrique de la densité probabiliste. Principalement deux méthodes s'appliquent pour les densités multi–dimensionnelles: le calcul des histogrammes et l'estimation fondée sur des fonctions de noyau [Pop 94]. L'avantage principal du calcul des histogrammes est la bonne représentation d'exemples d'apprentissage. Dans notre contexte, cette propriété est avantageuse, car toute information et, en particulier, celle discriminante, est préservée. Par opposition aux histogrammes, l'estimation fondée sur des fonctions de noyau permet la généralisation. La généralisation décrit la reconnaissance des exemples non utilisés pour l'apprentissage. Pour les histogrammes, la capacité de généralisation peut être obtenue par l'utilisation d'un nombre suffisant d'exemples d'apprentissage. Ce nombre est "suffisant" s'il est au moins du même ordre que le nombre de cellules d'histogrammes. La réduction du nombre de cellules facilite alors d'obtenir la capacité de généralisation. Comme notre approche ne modélise pas les translations (t_x et t_y) dans le plan image, le nombre d'exemples d'apprentissage est donné directement par la taille de l'objet dans l'image. Le nombre d'exemples – égal au nombre de mesures – est alors du même ordre que la taille de l'objet. En pratique, le nombre de mesures différentes est beaucoup plus petit que le nombre d'exemples d'apprentissage. Les histogrammes sont alors attendus de généraliser.

Ici on décrit le nombre de cellules d'histogrammes d'une expérimentation (pour les détails voir la section 7.3). Cette expérimentation exploite les plus grands histogrammes de cette thèse: les histogrammes à six dimensions d'une résolution de 32 cellules par axe d'histogramme. Le nombre de cellules d'un histogramme est alors de l'ordre de 10^9. Étant donné leur taille importante, ces histogrammes sont particulièrement difficiles à estimer. En pratique, seulement une partie de toutes les 10^9 cellules d'histogramme – c'est à dire les vecteurs de mesures à six dimensions – apparaissent dans les images des 103 objets utilisés. En moyenne, 5000 cellules différentes apparaissent au moins une fois par image.

Il est suffisant de représenter les mesures qui se trouvent au moins une fois dans un des histogrammes. Pour une estimation fiable et pour obtenir la capacité de généralisation, le nombre de ces mesures doit être petit. La figure 4.2 montre la distribution de mesures sur les cellules des 6367 histogrammes calculés (chaque histogramme correspond à un point de vue particulier, à une rotation dans le plan image et à une échelle, voir section 7.3). Le nombre de cellules d'histogrammes à représenter est de l'ordre de 10^6. Le nombre de cellules qui apparaissent au moins 100 fois est de l'ordre de 10^5. En utilisant une image de taille 512×512, le nombre de mesures par image est aussi de l'ordre de 10^5. En conclusion, les histogrammes sont attendus

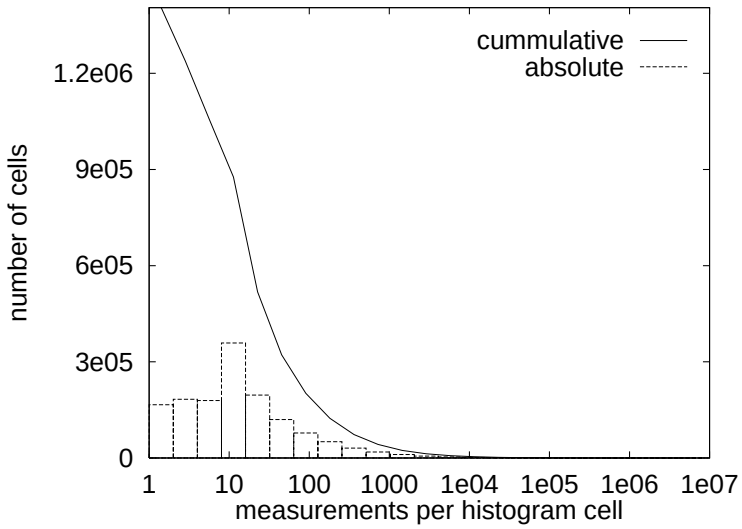


FIG. 4.2 – *Distribution des mesures en fonction de cellules d'histogramme. L'axe horizontal montre le nombre de mesures par cellule. L'axe vertical montre le nombre de cellules correspondantes*

de généraliser dans le contexte de notre travail. On doit noter que cet exemple correspond au cas d'estimation le plus difficile rencontré dans cette thèse. La capacité de généralisation peut être améliorée par la réduction du nombre de dimensions d'histogrammes ou par la réduction de la résolution des axes d'histogrammes.

En conséquence, la densité probabiliste d'un objet o_n est représentée par plusieurs histogrammes multidimensionnels. L'histogramme H d'un point de vue particulier (r_x r_y), d'une rotation dans le plan de l'image r_z et d'une certaine échelle t_z est donné par (avec M l'ensemble de mesures) :

$$H(M|o_n, r_x, r_y, r_z, t_z) \tag{4.52}$$

Ces histogrammes $H(M|o_n, r_x, r_y, r_z, t_z)$ sont calculés à partir de plusieurs images d'un objet. Le nombre d'images peut être réduit considérablement en utilisant deux propriétés des descripteurs locaux: l'"orientabilité" par rapport à la rotation dans le plan image et l'équivariance par rapport au changement d'échelle (l'"orientabilité" des dérivées Gaussiennes, par exemple, est décrite dans la section 3.1.1). Cela permet le calcul de plusieurs histogrammes qui correspondent aux différentes rotations r_z et aux différentes échelles t_z de l'objet, à partir d'une seule image par point de vue (r_x r_y).

Il est nécessaire d'estimer des histogrammes de différents points de vue à partir de différentes images. La section 6.5 analyse le nombre nécessaire d'histogrammes pour la représentation d'un objet 3D. Les expérimentations indiquent une stabilité considérable des histogrammes par rapport aux changements de point de vue. Par ce fait, un petit nombre d'histogrammes est suffisant pour obtenir des taux de reconnaissance élevés.

Dans cette thèse, la densité probabiliste d'un objet est toujours représentée (ou plutôt estimée) par les histogrammes multidimensionnels. Il faut noter l'existence de méthodes plus

efficaces pour sa représentation. Un exemple consiste en une représentation par des distributions paramétriques. Cette représentation est plus efficace car seulement un petit nombre de paramètres est mémorisé. Par contre, il existe un dilemme entre l'efficacité et la discrimination de la représentation. Un des principaux objectifs de la thèse est de montrer qu'une représentation d'objets par une densité probabiliste de caractéristiques locales contient des informations discriminantes et suffisantes pour la reconnaissance d'une variété d'objets. Dans ce contexte, la représentation choisie ne fait pas de compromis au niveau de la préservation des informations discriminantes, et l'estimation de la densité probabiliste n'est pas coûteuse en calcul. De plus, les histogrammes permettent la définition d'algorithmes de reconnaissance simples et rapides : un premier algorithme utilise la comparaison d'histogrammes (chapitres 5 et 6), et un deuxième algorithme consiste en une reconnaissance probabiliste (chapitre 7).

4.2 Application de la théorie de l'information à la reconnaissance d'objets

Cette section développe une analogie entre le processus de reconnaissance d'objets et la transmission d'information par un canal (bruité). Cette analogie s'applique le mieux à une représentation d'objets statistique comme celle de la section précédente. Cependant, l'analogie est appropriée pour une grande variété de processus de reconnaissance. La supposition nécessaire pour sa validité est que les messages (ou mesures) m_k reçus font partie d'un ensemble fini M.

La section 4.2.1 résume les concepts de base de la théorie de l'information (voir par exemple [Rez 95]) et donne une première interprétation. L'interprétation de ces concepts dans le contexte de la reconnaissance d'objets est discutée et plus détaillée par la section 4.2.2. La "transinformation" (ou information mutuelle) d'un processus de reconnaissance (section 4.2.3) peut servir pour évaluer un ensemble M de mesures. Fondées sur la transinformation, les performances relatives des processus de reconnaissance, utilisant des ensembles de mesures différentes, peuvent être prédites. La section 4.2.4 définit finalement les concepts de capacité, de redondance et d'efficacité dans le contexte de la reconnaissance d'objets.

4.2.1 Mesure de l'information

Soit un espace d'exemples Ω_X, réparti par un nombre fini d'événements x_n mutuellement exclusifs, et $p(x_n)$ les probabilités de ces événements. Les événements x_n forment une répartition complète dans le sens :

$$\bigcup_{n=1}^{N} x_n = \Omega_X \tag{4.53}$$

$$\sum_{n=1}^{N} p(x_n) = 1 \tag{4.54}$$

Un schéma probabiliste ayant ces propriétés s'appelle *schéma complet et fini de probabilité*. La théorie de l'information s'intéresse surtout à la définition d'une mesure de l'*incertitude* pour ce schéma de probabilité. Shannon et Wiener ont proposé l'utilisation de l'équation :

$$H(X) = -\sum_{n=1}^{N} p(x_n)\log(p(x_n)) \tag{4.55}$$

En utilisant la quantité $I(x_n) = -\log(p(x_n))$ comme mesure de l'*information propre* de l'événement x_n, l'entropie $H(X)$ peut être interprétée comme l'information moyenne de tous les événements x_n :

$$H(X) = \overline{I(x_n)} = \sum_{n=1}^{N} p(x_n)I(x_n) \tag{4.56}$$

Un ensemble d'objet $O = \Omega_O = \bigcup_{n=1}^{N} o_n$ avec les probabilités $p(o_n)$ forme un schéma complet et fini de probabilité, comme défini plus haut. En utilisant l'équation 4.55, la définition de l'information moyenne de tous les objets o_n devient :

$$H(O) = -\sum_{n=1}^{N} p(o_n)\log(p(o_n)) \tag{4.57}$$

La même analogie peut être appliquée à un ensemble de mesures $M = \Omega_M = \bigcup_{k=1}^{K} m_k$, qui forme aussi un schéma complet et fini de probabilité :

$$H(M) = -\sum_{k=1}^{K} p(m_k)\log(p(m_k)) \tag{4.58}$$

Dans le contexte de la transmission d'information par un canal, il est nécessaire de connaître la relation entre les symboles d'entrée et les symboles de sortie. La suite développe alors une mesure de l'information pour ce cas bidimensionnel.

Mesure de l'information pour le cas bidimensionnel :

La mesure $H(X)$ de l'*incertitude* ou de l'*information* peut être généralisée à un schéma bidimensionnel complet et fini de probabilité. Ce schéma est donné par deux espaces des exemples Ω_X et Ω_Y. Pour ces espaces, des ensembles d'événements complets dans le sens des équations 4.53 et 4.54 sont choisis. Chaque événement x_n de Ω_X peut apparaître avec chaque événement y_k de Ω_Y. Le produit d'espaces $\Omega_X \times \Omega_Y$ forme un ensemble complet et fini d'événements ayant la matrice de probabilité suivante :

$$P(X \wedge Y) = \begin{pmatrix} p(x_1 \wedge y_1) & p(x_1 \wedge y_2) & \dots & p(x_1 \wedge y_K) \\ p(x_2 \wedge y_1) & p(x_2 \wedge y_2) & \dots & p(x_2 \wedge y_K) \\ \vdots & & \ddots & \vdots \\ p(x_N \wedge y_1) & p(x_N \wedge y_2) & \dots & p(x_N \wedge y_K) \end{pmatrix} \tag{4.59}$$

Trois schémas complets de probabilité, plus précisément $P(X)$, $P(Y)$ et $P(X \wedge Y)$, sont donnés, ayant les trois entropies correspondantes :

$$H(X) = -\sum_{n=1}^{N} p(x_n) \log(p(x_n)) \tag{4.60}$$

$$H(Y) = -\sum_{k=1}^{K} p(y_k) \log(p(y_k)) \tag{4.61}$$

$$H(X \wedge Y) = -\sum_{n=1}^{N} \sum_{k=1}^{K} p(x_n \wedge y_k) \log(p(x_n \wedge y_k)) \tag{4.62}$$

avec

$$p(x_n) = \sum_{k=1}^{K} p(x_n \wedge y_k) \tag{4.63}$$

$$p(y_k) = \sum_{n=1}^{N} p(x_n \wedge y_k) \tag{4.64}$$

$H(X)$ est appelée entropie marginale de X, $H(Y)$ entropie marginale de Y et $H(X \wedge Y)$ entropie conjointe. Deux autres entropies peuvent être définies : l'entropie conditionnelle $H(X|Y)$ et l'entropie conditionnelle $H(Y|X)$. L'équation de $H(X|Y)$ se base sur l'information propre de l'événement $(x_n|y_k)$:

$$I(x_n|y_k) = -\log(p(x_n|y_k)) \tag{4.65}$$

L'information moyenne $H(X|y_k)$ d'un certain y_k est alors :

$$H(X|y_k) = \overline{I(x_n|y_k)} = -\sum_{n=1}^{N} p(x_n|y_k) \log(p(x_n|y_k)) \tag{4.66}$$

L'entropie conditionnelle $H(X|Y)$ est calculée par :

$$H(X|Y) = \overline{H(X|y_k)} = \sum_{k=1}^{K} p(y_k) H(X|y_k) \tag{4.67}$$

$$= -\sum_{k=1}^{K} p(y_k) \sum_{n=1}^{N} p(x_n|y_k) \log(p(x_n|y_k)) \tag{4.68}$$

$$= -\sum_{k=1}^{K} \sum_{n=1}^{N} p(x_n \wedge y_k) \log(p(x_n|y_k)) \tag{4.69}$$

L'équation de l'entropie conditionnelle $H(Y|X)$ peut être dérivée de manière analogue :

$$H(Y|X) = -\sum_{k=1}^{K} \sum_{n=1}^{N} p(x_n \wedge y_k) \log(p(y_y|x_n)) \tag{4.70}$$

4.2.2 Application de la théorie de l'information à la reconnaissance d'objets

La section précédente décrit cinq entropies d'un schéma bidimensionnel de probabilité. Dans cette section, ces entropies sont interprétées dans le contexte de la reconnaissance d'objets. Dans la théorie de l'information, le schéma bidimensionnel décrit classiquement un réseau de communication: les x_n correspondent aux entrées possibles (ou aux symboles de l'alphabet d'entrée) et les y_k correspondent aux sorties possibles du réseau. Chaque entrée x_k est "transformée" par le canal de communication en sorties possibles y_k. La matrice de probabilité (équation 4.59) décrit les caractéristiques du canal.

Dans le contexte de la reconnaissance d'objets, les "entrées" sont les objets o_n. Les "sorties" sont les mesures ou symboles m_k, qui sont extraits de l'image d'un objet. Le canal correspond à la transformation des objets vers l'espace de mesures. Le canal de communication correspond donc au processus de reconnaissance lui–même. Les caractéristiques du processus de reconnaissance sont décrites par la matrice de probabilité (équation 4.59). En utilisant cette analogie entre un réseau de communication et un processus de reconnaissance, les cinq entropies $H(O)$, $H(M)$, $H(O \wedge M)$, $H(O|M)$ et $H(M|O)$ peuvent être interprétés de façon suivante:

- $H(O)$ (équation 4.60) est l'information moyenne de chaque objet o_n,
- $H(M)$ (équation 4.61) est l'information moyenne de chaque mesure m_k,
- $H(O \wedge M)$ (équation 4.62) est l'incertitude du processus de reconnaissance,
- $H(M|O)$ (équation 4.70) donne une indication du "bruit" ou de l'erreur moyenne du processus de reconnaissance,
- $H(O|M)$ (équation 4.69) indique la qualité du processus de reconnaissance. Plus $H(M|O)$ est petit, mieux les objets peuvent être reconnus avec l'ensemble de mesures M.

En considérant les probabilités $p(o_n)$ fixes et connues pour chaque objet o_n, l'entropie $H(O)$ est constante. Dans ce manuscrit, tous les objets sont supposés équiprobables: $p(o_n) = 1/N$. L'entropie $H(O)$ est alors $H(O) = \log(N)$.

Les quatre autres entropies varient de façon significative, si l'ensemble de mesures M change. L'entropie $H(O|M)$ est très intéressante car elle indique l'erreur du processus de reconnaissance. Pour montrer l'influence du choix de l'ensemble M de mesures, l'entropie $H(O|M)$ (équation 4.69) est développée:

$$\begin{aligned}
H(O|M) &= -\sum_{k=1}^{K}\sum_{n=1}^{N} p(o_n \wedge m_k)\log(p(o_n|m_k)) \\
&= -\sum_{k=1}^{K}\sum_{n=1}^{N} p(o_n \wedge m_k)\log\frac{p(o_n \wedge m_k)}{p(m_k)} \\
&= -\sum_{k=1}^{K}\sum_{n=1}^{N} p(o_n \wedge m_k)\log(p(o_n \wedge m_k)) + \sum_{k=1}^{K}\sum_{n=1}^{N} p(o_n \wedge m_k)\log(p(m_k)) \\
&= H(O \wedge M) + \sum_{k=1}^{K} p(m_k)\log(p(m_k)) \\
&= H(O \wedge M) - H(M) \qquad (4.71)
\end{aligned}$$

En minimisant $H(O|M)$, l'erreur moyenne du processus de reconnaissance est réduite. En utilisant l'équation 4.71, cela peut être fait par la minimisation de l'entropie $H(O \wedge M)$ et par la maximisation de $H(M)$. $H(M)$ est maximale si toutes les mesures m_k sont équiprobables. $H(O \wedge M)$ est minimale si pour chaque objet o_n, il existe une et une seule mesure m_k telle que $p(o_n|m_k) = 1$ et $p(m_k|o_n) = 1$ (cela correspond au cas où chaque ligne de la matrice 4.59 contient à une et une seule position la valeur $1/N$). Autrement dit, plus $H(O \wedge M)$ est petit, plus l'ensemble de mesures est significatif en moyenne. L'équation 4.71 permet alors la comparaison numérique de différents ensembles de mesures m_k pour le même ensemble d'objets o_n. l'équation 4.71 peut être appliquée pour de nombreux processus de reconnaissance. La seule supposition est la possibilité du calcul ou de l'estimation de la matrice de probabilité 4.59.

4.2.3 La transinformation du processus de reconnaissance

En théorie de l'information, l'information mutuelle d'une paire d'événements (x_n, y_k) est la base du calcul de la *transinformation* [Rez 95]. L'analogie entre le réseau de communication et la reconnaissance d'objets permet le calcul de la transinformation d'une paire d'objet/mesure (o_n, m_k) :

$$T(o_n, m_k) = \log \frac{p(o_n \wedge m_k)}{p(o_n)p(m_k)} \tag{4.72}$$

La transinformation moyenne de toutes les paires est alors donnée par :

$$T(O, M) = \overline{T(o_n, m_k)} = \sum_{n=1}^{N} \sum_{k=1}^{K} p(o_n \wedge m_k) \log \frac{p(o_n \wedge m_k)}{p(o_n)p(m_k)} \tag{4.73}$$

Cette entropie indique l'information transmise par le canal (= processus de reconnaissance). Pour cette raison, elle est appelée la *transinformation* du canal. Par la définition (équation 4.73), on obtient :

$$\begin{aligned} T(O, M) &= H(O) + H(M) - H(O \wedge M) && (4.74)\\ &= H(O) - H(O|M) && (4.75)\\ &= H(M) - H(M|O) && (4.76) \end{aligned}$$

En utilisant l'équation 4.75, l'information transmise par le canal (c'est à dire le processus de reconnaissance) peut être maximisée par la minimisation de $H(O|M)$ (supposant $H(O)$ constante, comme annoncé plus haut). En utilisant l'équation 4.76, la maximisation de $H(M)$ et la minimisation de $H(M|O)$ maximise aussi la transinformation. Un résultat similaire a été obtenu par l'analyse de l'équation 4.71.

L'inégalité $H(O \wedge M) \geq \max(H(O), H(M))$ et l'équation 4.74 permettent la définition d'une borne supérieure pour la transinformation $T(O, M)$:

$$T(O, M) \leq \min(H(O), H(M)) \tag{4.77}$$

Comme $H(O)$ est considérée constante (tous les objets sont équiprobables), la borne supérieure $T(O, M)$ peut être augmentée par $H(M)$ (jusqu'à l'égalité de $H(M)$ et $H(O)$).

L'idée la plus intéressante est d'utiliser la transinformation pour comparer différents ensembles M de mesures pour un certain ensemble O d'objets. En appliquant la transinformation à un sous–ensemble de O, l'ensemble M de mesures le plus discriminant pour ce sous–ensemble peut être obtenu.

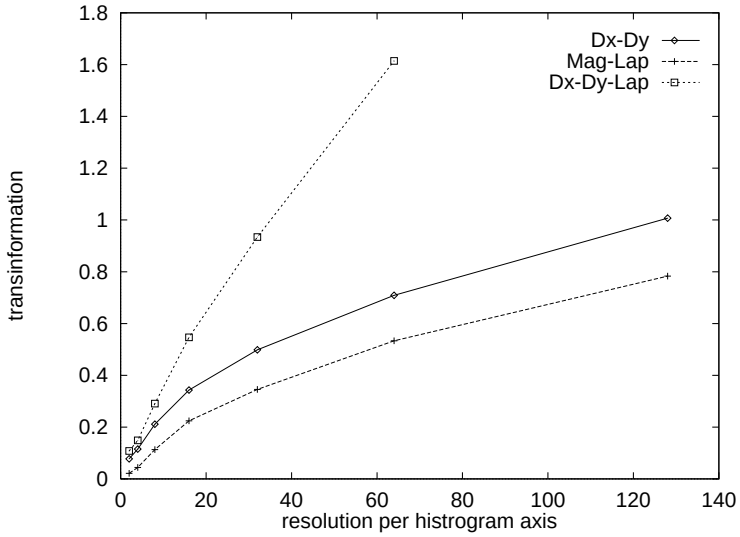


FIG. 4.3 – *La transinformation de 100 objets et de différents ensembles de mesures : Dx-Dy et Mag-Lap correspondent aux histogrammes bidimensionnels et Dx-Dy-Lap correspond au histogramme tridimensionnel. L'axe horizontal correspond aux différentes résolutions des histogrammes et l'axe vertical montre la transinformation pour une résolution particulière*

Ce paragraphe décrit un exemple d'illustration de l'utilisation de la transinformation pour l'évaluation de différents ensembles de mesures. La transinformation de 100 objets est calculée en fonction de différentes combinaisons de filtres et de différentes résolutions. La base d'images (celle de Columbia) contient 100 objets (voir figure A.6 de l'annexe). La figure 4.3 montre les résultats du calcul de la transinformation. Les trois graphes correspondent aux trois combinaisons de filtres : *Dx-Dy* est un histogramme bidimensionnel de premières dérivées Gaussiennes dans les directions x et y. *Mag-Lap* correspond aussi à un histogramme bidimensionnel, plus précisément de la norme de la première dérivée Gaussienne et de l'opérateur Laplacien. *Dx-Dy-Lap* correspond à un histogramme tridimensionnel de premières dérivées Gaussiennes dans les directions x et y et de l'opérateur Laplacien. Pour tous ces filtres, la définition des dérivées Gaussiennes de la section 3.1.1 est utilisée. Pour chaque combinaison de filtres, les histogrammes de différentes résolutions, plus précisément de résolutions de 2, 4, 8, 16, 32, 64 et 128, sont calculés.

La figure 4.3 montre que le choix de l'ensemble de mesures influence beaucoup la transinformation du processus de reconnaissance. Les graphes montrent une augmentation de la transinformation pour les ensembles tridimensionnels de mesures *Dx-Dy-Lap* par rapport aux ensembles bidimensionnels. Cette augmentation s'explique par l'information supplémentaire contenue dans la dimension indépendante. En ajoutant plus de dimensions indépendantes, l'augmentation résultant doit être encore plus importante. Il est intéressant de remarquer que les résultats de

Dx-Dy et *Mag-Lap* sont qualitativement similaires. Néanmoins, les résultats de *Dx-Dy* sont meilleurs que ceux de *Mag-Lap*, car l'information d'orientation est préservée par *Dx-Dy*. *Mag-Lap*, par opposition, est invariant par rapport aux rotations.

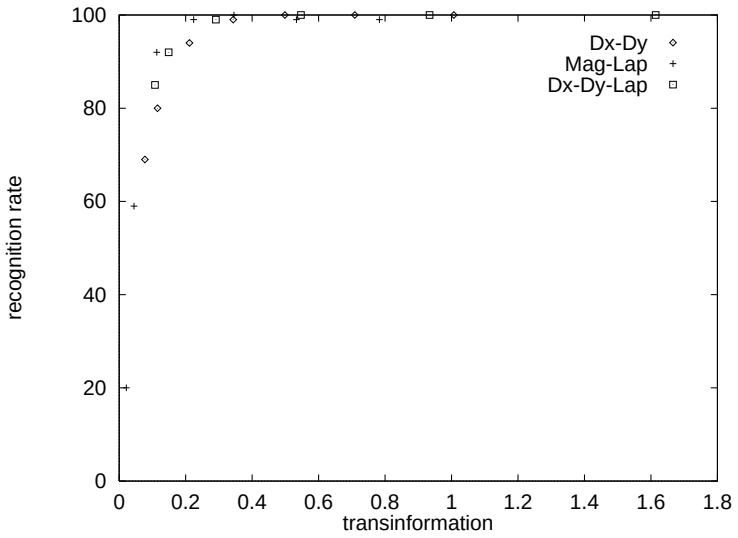


FIG. 4.4 – *La relation entre la transinformation de 100 objets (pour différents ensembles de mesures, voir figure 4.3) et le taux de reconnaissance de 100 images de test. Tous les points sont approximativement sur la même courbe. Cela montre la forte relation entre la transinformation et le taux de reconnaissance. Cela indique que la transinformation peut être utilisée pour prédire les taux de reconnaissance*

La figure 4.4 montre la relation entre la transinformation (figure 4.3) et le taux de reconnaissance de différents ensembles de mesures. Les ensembles de mesures et la base d'images utilisés sont les mêmes que ceux de la figure 4.3. 100 images d'objets avec un changement de point de vue de 5° par rapport à la base de données sont employées comme base de test. La figure 4.3 montre une forte relation entre la transinformation et le taux de reconnaissance. Cela indique l'applicabilité de la transinformation pour la prédiction de taux de reconnaissance.

4.2.4 Capacité, redondance et efficacité du processus de reconnaissance

Pour compléter l'application de la théorie de l'information à la reconnaissance d'objets, les concepts de capacité, de redondance et d'efficacité sont définis.

Dans la théorie de l'information la capacité du canal est définie comme le maximum de la transinformation :

$$\textit{Capacité} = \max_{P(O)} T(O,M) = \max_{P(O)} \left(H(M) - H(M|O)\right) \tag{4.78}$$

La maximisation est par rapport à tous les ensembles possibles de probabilités $P(O)$ d'objets o_n. La capacité est alors calculée pour le schéma $p(m_k|o_n)$ particulier de probabilités.

La redondance absolue est définie par la différence de la *Capacité* et de la transinformation :

$$Redondance_{absolue} = Capacité - T(O, M) \tag{4.79}$$

La redondance relative est définie comme relation entre la $Redondance_{absolue}$ et la $Capacité$:

$$Redondance_{relative} = \frac{Capacité - T(O, M)}{Capacité} \tag{4.80}$$

L'efficacité est définie par : $Efficacité = 1 - Redondance_{relative}$.

4.3 Conclusion

Ce chapitre introduit une formulation générale de la représentation d'objets par une densité probabiliste d'un ensemble de mesures. Cet ensemble de mesures est donné par des vecteurs de caractéristiques locales. Chaque degré de liberté de la reconnaissance d'objets, comme introduit dans la section 1.2, est pris en compte de manière appropriée. L'estimation et la représentation de la densité probabiliste par les histogrammes multidimensionnels de champs réceptifs sont justifiées et discutées.

La section 4.2 développe une analogie entre la théorie de l'information et la reconnaissance d'objets. Cette analogie permet l'utilisation de la transinformation pour la comparaison de différents ensembles de mesures. Les résultats indiquent que la qualité de reconnaissance peut être prédite par la transinformation.

Le schéma de la représentation statistique peut être interprété comme un cadre général de la représentation d'objets. L'utilisation des histogrammes multidimensionnels de champs réceptifs montre que ce modèle généralise directement la technique des histogrammes de couleurs. Les deux chapitres suivants emploient la comparaison d'histogrammes pour la reconnaissance d'objets. Le chapitre 7 étend l'application du modèle proposé à la reconnaissance probabiliste d'objets. Cet algorithme est capable de reconnaître les objets dans des scènes complexes avec des occultations partielles et avec plusieurs objets. Le chapitre 8 définit deux algorithmes de la reconnaissance active d'objets. Le chapitre 9 introduit un schéma pour la classification d'objets fondé sur le modèle de la représentation statistique d'objets.

Chapitre 5

Fonctions pour la comparaison d'histogrammes

Les chapitres précédents ont introduit les histogrammes multidimensionnels de champs réceptifs pour la représentation statistique d'objets. La comparaison d'histogrammes s'applique directement à la reconnaissance d'objets. Ce chapitre propose et analyse différentes fonctions de la comparaison d'histogrammes. La motivation principale de la comparaison d'histogrammes pour la reconnaissance d'objets est son faible volume de calcul. De plus, les informations statistiques de l'image entière sont utilisées, permettant la reconnaissance robuste en présence de changements importants d'apparence. Le chapitre suivant applique les fonctions proposées de la comparaison d'histogrammes à l'identification d'objets.

La section 5.1 décrit et discute différentes fonctions de comparaison dans le contexte des histogrammes multidimensionnels de champs réceptifs. La section 5.2 examine la stabilité des fonctions introduites en présence de bruit Gaussien additif, de flou et de rotation d'image. La section 5.3 est dédiée à l'analyse de la stabilité de la comparaison d'histogrammes par rapport aux changements d'intensité d'éclairage. La dernière section 5.4 résume les résultats de ce chapitre.

5.1 Fonctions de comparaison d'histogrammes

La section décrit des fonctions adaptées à la comparaison d'histogrammes. L'analyse de ces fonctions est importante car la fonction d'*intersection*, utilisée par Swain et Ballard [Swa 91]

dans le contexte des histogrammes de couleurs, possède des limitations pour le cas plus général d'histogrammes multidimensionnels de champs réceptifs. En particulier, la fonction d'intersection suppose l'équiprobabilité de toutes les mesures. Cette supposition n'est pas valable pour un ensemble $M = \cup_k m_k$ de mesures arbitraires.

La section 5.1.1 introduit la fonction d'intersection et deux extensions possibles. La section 5.1.2 décrit des distances quadratiques dans les cas spéciaux de la SSD, somme des carrés des distances, et la distance de Mahalanobis $maha$. La section 5.1.3 justifie l'utilisation des statistiques χ^2 pour la comparaison de deux histogrammes. L'indexation de grandes bases de données (comme par exemple pour l'accès aux bases d'images) requiert des stratégies efficaces de recherche. La 5.1.4 discute brièvement l'organisation d'une base d'histogrammes pour obtenir une stratégie efficace d'indexation.

Comme motivé dans la section 4.1, un objet o_n est représenté par une densité probabiliste sur un ensemble $M = \cup_k m_k$ de mesures. Chaque mesure correspond à un vecteur à L dimensions de descripteurs locaux. Plusieurs histogrammes multidimensionnels $H(M|o_n, r_x, r_y, r_z, t_z)$ sont utilisés comme approximation de la densité probabiliste d'un objet o_n. Chaque histogramme correspond à une apparence particulière de l'objet définie par une rotation (r_x, r_y, r_z) et une échelle t_z. Dans le contexte de la reconnaissance d'objet, les histogrammes de la base sont comparés à l'histogramme de l'image test. Les sections suivantes utilisent la notation $V = \cup_{\mathbf{i}} v_{\mathbf{i}}$ pour un histogramme de la base et $Q = \cup_{\mathbf{i}} q_{\mathbf{i}}$ pour l'histogramme de l'image test. $\mathbf{i}$ signifie l'indice d'histogrammes à L dimensions où L correspond au nombre de dimensions du vecteur m_k de mesures. C'est à dire que les histogrammes qui possèdent L dimensions. $v_{\mathbf{i}}$ (respectivement $q_{\mathbf{i}}$) correspondent à la valeur d'une cellule particulière de l'histogramme V (respectivement de Q).

5.1.1 Fonction d'intersection

Swain et Ballard ont introduit dans leur travail original [Swa 91] la fonction d'*intersection* pour la comparaison des histogrammes de couleurs. L'intersection de deux histogrammes V et Q est définie par :

$$\cap(Q, V) = \sum_{\mathbf{i}} \min(q_{\mathbf{i}}, v_{\mathbf{i}}) \tag{5.81}$$

La motivation intuitive de cette fonction est le calcul des parties communes (l'intersection) des deux histogrammes V et Q. Un avantage de cette fonction est qu'elle néglige explicitement les pixels du fond qui peuvent apparaître dans l'histogramme test Q mais pas dans l'histogramme de la base V. Cette fonction n'implique pas un lourd niveau de calcul car seulement deux opérations (minimum et addition) sont nécessaires par cellule d'histogramme. La complexité est linéaire par rapport au nombre C de cellules d'histogramme: $O(C)$.

Dans leur travail original, Swain et Ballard reportent la nécessité d'une distribution éparse[1] afin de discriminer des objets. Nos expérimentations ont vérifié ce résultat. Une telle distribution peut être obtenue par l'utilisation d'histogrammes à haute dimension. Dans ce cas il faut faire un compromis entre la possibilité de discriminer des objets et la stabilité par rapport aux perturbations [Cal 93]. Un deuxième inconvénient de la fonction d'intersection est que toutes les cellules d'histogrammes sont traitées de manière égale, impliquant l'équiprobabilité des cellules.

1. sparse distribution

Une répartition appropriée de chaque axe d'histogrammes peut être utilisée (par exemple en utilisant une répartition logarithmique) ce qui est souvent instable. C'est à dire que les cellules d'histogrammes ne peuvent pas être supposées de façon équiprobable dans le cas général des histogrammes multidimensionnels de champs réceptifs.

Pour résoudre le problème d'équiprobabilité de cellules, Ennesser et Medioni [Enn 93] proposent une pondération de la fonction d'intersection par des poids $w_\mathbf{i}$ pour chaque cellule. Ce poids $w_\mathbf{i}$ doit être choisi proportionnel à l'importance d'une couleur particulière (ou d'une cellule d'histogramme $\mathbf{i}$). Si l'histogramme moyen $A = \cup_\mathbf{i} a_\mathbf{i}$ est connu il peut être utilisé pour la définition des fonctions pondérées d'intersection :

$$\cap_{we}(Q,V) = \sum_\mathbf{i} \frac{1}{a_\mathbf{i}} \min(q_\mathbf{i}, v_\mathbf{i}) \tag{5.82}$$

$$\cap_{mo}(Q,V) = \sum_\mathbf{i} \frac{v_\mathbf{i}}{a_\mathbf{i}} \min(q_\mathbf{i}, v_\mathbf{i}) \tag{5.83}$$

Pour $\cap_{we}(Q,V)$ le poids $w_\mathbf{i} = \frac{1}{a_\mathbf{i}}$ est inversement proportionnel à l'occurrence de la cellule d'histogramme $\mathbf{i}$. Ce poids peut être vu comme indication de l'importance de la cellule $\mathbf{i}$. Pour le deuxième cas de $\cap_{mo}(Q,V)$ le poids $w_\mathbf{i} = \frac{m_\mathbf{i}}{a_\mathbf{i}}$ prend en compte l'importance de la cellule ainsi que le nombre d'occurrences dans l'histogramme modèle V. L'application de ces fonctions d'intersection pondérées peut améliorer la discrimination d'objets. Cependant ces fonctions sont moins stables pour les cellules $\mathbf{i}$ ayant un petit nombre de $a_\mathbf{i}$ (la stabilité est examinée dans la section 5.2).

Stricker [Str 92] propose une autre fonction de comparaison $m(Q,V)$, équivalente à la fonction d'intersection normalisée par la taille d'histogrammes. Stricker introduit la fonction suivante $m(Q,V)$:

$$m(Q,V) = 1 - \frac{d(Q,V)}{s(Q)+s(V)} \tag{5.84}$$

avec

$$d(Q,V) = \sum_\mathbf{i} ||q_\mathbf{i} - v_\mathbf{i}|| \tag{5.85}$$

$$s(Q) = d(Q,0) = \sum_\mathbf{i} q_\mathbf{i} \tag{5.86}$$

Il est possible de montrer que la fonction $m(Q,V)$ est équivalente à l'intersection normalisée par la somme des tailles d'histogrammes $s(Q)+s(V)$:

$$\begin{aligned} s(Q)+s(V) &= \sum_\mathbf{i} q_\mathbf{i} + \sum_\mathbf{i} v_\mathbf{i} \\ &= 2\sum_\mathbf{i} \min(q_\mathbf{i}, v_\mathbf{i}) + \sum_\mathbf{i} ||q_\mathbf{i} - v_\mathbf{i}|| \\ &= 2 \cap (Q,V) + d(Q,V) \\ \Longleftrightarrow s(Q)+s(V)-d(Q,V) &= 2 \cap (Q,V) \\ \Longleftrightarrow m(Q,V) &= \frac{2 \cap (Q,V)}{s(Q)+s(V)} \qquad \square \end{aligned}$$

5.1.2 Distances quadratiques

Les distances quadratiques entre deux histogrammes V et Q peuvent être écrites en utilisant une matrice carrée de poids W. La matrice W prend en compte les relations entre les différentes cellules d'histogrammes :

$$quad(Q,V) = (Q-V)^T W (Q-V) \tag{5.87}$$

Un cas spécial est constitué par la somme des carrés des distances SSD, souvent utilisées dans le domaine du traitement du signal (avec W la matrice d'identité) :

$$SSD(Q,V) = \sum_{\mathbf{i}} (q_{\mathbf{i}} - v_{\mathbf{i}})^2 \tag{5.88}$$

Un choix sensible de la matrice W est l'inverse de la matrice de covariance de cellules d'histogrammes. Cette matrice de covariance modélise l'importance de chaque cellule et les dépendances entre les différentes cellules. Ce cas spécial est appelé la distance de Mahalanobis. Si les différentes cellules d'histogrammes sont indépendantes les unes des autres, alors les éléments de la diagonale de la matrice W ne sont pas nuls. Cela correspond à un cas spécial de la distance de Mahalanobis appelée $maha(Q,V)$ par la suite :

$$maha(Q,V) = \sum_{\mathbf{i}} \frac{(q_{\mathbf{i}} - v_{\mathbf{i}})^2}{\eta_{\mathbf{i}}^2} \tag{5.89}$$

avec $\eta_{\mathbf{i}}^2$ la variance de la cellule d'histogramme $\mathbf{i}$. Comme mentionné plus haut, l'équation 5.87 permet l'introduction des relations entre les différentes cellules d'histogrammes. Cela peut être fait en utilisant l'inverse de la matrice de covariance entière comme matrice W. D'autres possibilités du choix de la matrice $W = (w_{\mathbf{ij}})$ incluent [Haf 95] :

$$w_{\mathbf{ij}} = 1 - \frac{d_{\mathbf{ij}}}{\max_{\mathbf{ij}}(d_{\mathbf{ij}})} \tag{5.90}$$

$$w_{\mathbf{ij}} = e^{-\sigma(\frac{d_{\mathbf{ij}}}{\max_{\mathbf{ij}}(d_{\mathbf{ij}})})^2} \tag{5.91}$$

avec $d_{\mathbf{ij}}$ la distance euclidienne (ou une autre distance) entre deux cellules d'histogramme $\mathbf{i}$ et $\mathbf{j}$. En considérant ces relations simples, nous pouvons obtenir des résultats stables en sacrifiant pour cette stabilité la diminution de la discrimination d'objets différents. Si la matrice de covariance peut être estimée, la distance de Mahalanobis est favorisée. Une autre idée intéressante est l'utilisation de la matrice d'information de Fisher comme matrice W.

La complexité des distances quadratiques $quad$ (équation 5.87) est généralement $O(C^2)$ avec C le nombre de cellules d'histogramme. La complexité exacte dépend de la matrice W. En particulier, la complexité de SSD (équation 5.88) et de la distance de Mahalanobis $maha$ (équation 5.89) est linéaire par rapport au nombre C de cellules : $O(C)$. Néanmoins, le calcul de ces deux fonctions est plus lourd que celui de la fonction d'intersection car SSD utilise trois opérations par cellule d'histogramme (soustraction, multiplication et addition) et la distance de Mahalanobis $maha$ utilise quatre opérations par cellule (soustraction, multiplication, division et addition).

5.1.3 Statistiques χ^2

La méthode formelle de statistiques pour déterminer si deux distributions sont différentes est le test χ^2. Ayant l'hypothèse nulle que deux ensembles de données (des histogrammes) sont tirés de la même population (par exemple des mesures du même objet) le but est de démontrer l'hypothèse nulle. Démontrer l'hypothèse nulle prouve que les histogrammes ont été tirés de différentes distributions. S'il n'est pas possible de démontrer l'hypothèse nulle, les histogrammes sont consistants et pourraient être tirés de la même population. Le test χ^2 est alors un test de consistance de deux histogrammes.

Dans le contexte d'histogrammes multidimensionnels de champs réceptifs, nous employons seulement les statistiques χ^2 sans utiliser la fonction de probabilité de χ^2. Les statistiques χ^2 sont utilisées pour calculer la "dissemblance" d'histogrammes. Deux différents calculs des statistiques χ^2 sont considérés dans la thèse [Pre 92]. Le premier – $\chi^2_v(Q,V)$ – suppose la connaissance exacte de l'histogramme modèle V :

$$\chi^2_v(Q,V) = \sum_{\mathbf{i}} \frac{(q_{\mathbf{i}} - v_{\mathbf{i}})^2}{v_{\mathbf{i}}} \tag{5.92}$$

Un deuxième calcul – $\chi^2_{qv}(Q,V)$ – compare deux histogrammes observés (aucun n'est dérivé théoriquement). Cette deuxième statistique χ^2 apparaît plus appropriée dans notre contexte car nous ne supposons par la connaissance exacte de l'histogramme modèle V. $\chi^2_{qv}(Q,V)$, défini par :

$$\chi^2_{qv}(Q,V) = \sum_{\mathbf{i}} \frac{(q_{\mathbf{i}} - v_{\mathbf{i}})^2}{q_{\mathbf{i}} + v_{\mathbf{i}}} \tag{5.93}$$

Les expérimentations montrent, pour la plupart des cas, une meilleure reconnaissance par ces statistiques χ^2 que par les autres fonctions de comparaison. Malheureusement, ces deux statistiques χ^2 ne sont pas métriques car l'inégalité triangulaire n'est pas satisfaite. Pour voir cela le cas dégénéré des histogrammes d'une cellule est considéré : $A = (a)$ et $C = (c)$. Pour tous les histogrammes d'une cellule $B = (b)$ avec $a < b < c$, l'inégalité suivante est vraie (qui correspond au contraire de l'inégalité triangulaire d'une métrique) :

$$\chi^2_{qv}(A,B) + \chi^2_{qv}(B,C) < \chi^2_{qv}(A,C) \tag{5.94}$$

Étant donné que les statistiques χ^2 donnent les meilleurs résultats de reconnaissance, nous introduisons une fonction modifiée qui est métrique. En supposant la connaissance d'un histogramme moyen $A = \cup_{\mathbf{i}} a_{\mathbf{i}}$ une fonction de comparaison peut être définie :

$$\chi^2_{av}(Q,V) = \sum_{\mathbf{i}} \frac{(q_{\mathbf{i}} - v_{\mathbf{i}})^2}{a_{\mathbf{i}}} \tag{5.95}$$

Cette fonction donne des taux de reconnaissance élevés (par rapport aux autres fonctions) mais pas la même qualité de résultats que les statistiques χ^2 originales.

La complexité des trois statistiques χ^2 est $O(C)$ avec C le nombre de cellules d'histogrammes. Le nombre d'opérations par cellule est cinq pour χ^2_{qv} et quatre pour χ^2_v et χ^2_{av}.

5.1.4 Indexation efficace d'histogrammes

Les fonctions introduites de comparaison possèdent une complexité pour la comparaison de deux histogrammes de $O(C)$, C le nombre de cellules (une complexité de $O(C^2)$ était obtenue pour les distances quadratiques *quad*, qui ne sont pas considérées dans la discussion suivante). La table 5.1 montre des fréquences typiques de cinq différentes fonctions pour la comparaison d'histogrammes de 4096 cellules. La table montre qu'entre 500 et 2000 histogrammes peuvent être comparés par seconde.

fonction de comparaison	$\cap$	χ^2_v	χ^2_{qv}	SSD	$maha$
fréquence [Hz]	1827	1241	845	1767	623

TAB. 5.1 – *Des fréquences typiques pour la comparaison d'histogrammes avec 4096 cellules. Les fréquences ont été obtenues pour les expérimentations de la section 6.3 sur une Silicon Indy 200 MHz*

A côté de la complexité de la comparaison de deux histogrammes, il est aussi important de discuter la complexité d'indexation d'une grande base de données. Cette discussion est importante, par exemple dans le contexte de l'accès aux bases d'images par leur contenu. L'accès aux bases d'images est motivé par le fait qu'il est impraticable (et coûteux en niveau du temps) de mémoriser des mots clés d'une manière manuelle pour chaque image de la base. Un algorithme couramment employé pour l'accès aux bases d'images est l'algorithme des histogrammes de couleurs de Swain et Ballard (la version originale ou une version améliorée).

En calculant pour chaque image de la base un histogramme multidimensionnel, l'accès à la base d'images est un problème d'indexation dans une base d'histogrammes. En supposant une complexité de $O(C)$ pour la comparaison de deux histogrammes avec C cellules et en appliquant une recherche linéaire sur la base de N histogrammes, la complexité d'"indexation" est $O(C \times N)$. Cette complexité n'est pas acceptable pour les grandes bases de données. Il est souhaitable d'obtenir une recherche logarithmique de la base ayant une complexité $O(C \log(N))$. En principe deux types de méthodes s'appliquent :

- Organisation de la base d'histogrammes dans une structure d'arbre (ou structure similaire).
- Indexation de la base d'histogrammes par une table de hachage. L'issue majeure est de trouver des indices appropriés.

La première méthode, l'organisation de la base d'histogrammes dans une structure d'arbre, s'appuie classiquement sur une métrique comme fonction de comparaison (les statistiques χ^2 ne sont pas métriques, voir section 5.1.3). En supposant une fonction métrique de comparaison, de nombreux algorithmes existent pour l'organisation d'une base de données dans une structure d'arbre.

La deuxième possibilité est plus intéressante car un indice d'histogrammes doit coder des informations importantes sans être instable. Un exemple est l'application des moments d'histogrammes comme la moyenne d'un histogramme [Haf 95]. Tous les algorithmes de réduction de dimension peuvent être appliqués pour l'extraction d'indices comme par exemple l'analyse de discriminants linéaires ou la transformation de Karhunen–Loeve. Comme le premier type de méthode, ce deuxième s'appuie sur le fait que la fonction de comparaison est métrique.

En utilisant une distance quadratique ou une fonction d'intersection, les deux types de méthodes sont utilisables. De nombreux algorithmes standards existent mais ne seront pas discutés plus en détail. Malheureusement, les statistiques χ^2 ne sont pas métriques (voir section 5.1.3) et rendront l'application d'un algorithme standard difficile. Un algorithme spécifique pour organiser la base d'histogrammes est alors proposé. L'idée est le calcul d'une pyramide d'histogrammes pour chaque histogramme de la base. Plus précisément, le premier histogramme de la pyramide est calculé directement à partir d'images avec la résolution maximale souhaitée. L'histogramme du niveau suivant est calculé sur la base du premier histogramme, en réduisant la résolution de chaque axe d'histogramme par un certain facteur (comme par exemple par un facteur de 2). Les pyramides d'histogrammes peuvent être calculées avant l'accès à la base. Pendant l'accès, un algorithme de branche-et-limite est appliqué à la base de la pyramide d'histogrammes. Pour l'application de cet algorithme il faut alors définir une limite inférieure et une limite supérieure. Le volume de calcul de l'algorithme dépend de la qualité de ces limites. Comme la qualité de limites ne peut être assurée pour une base particulière d'histogrammes, le calcul de la complexité moyenne de l'algorithme n'est pas possible. Néanmoins, la limite inférieure et la limite supérieure sont données et sont exactes, au sens où l'histogramme de statistique minimale χ^2_{qv} est trouvé. Les limites sont optimales car il n'existe ni une limite supérieure plus petite ni une limite inférieure plus grande.

Limite inférieure de χ^2_{qv} :

Pour l'application d'un algorithme de branche-et-limite une limite inférieure est nécessaire. Sans perte de généralité, deux histogrammes avec seulement deux cellules chacun sont considérés : $A = a_0 \cup a_1$ et $B = b_0 \cup b_1$. Pour chaque histogramme le niveau suivant de la pyramide d'histogramme est donné par : $A' = [a_0 + a_1]$ et $B' = [b_0 + b_1]$. Les deux histogrammes A' et B' contiennent une seule cellule. La question essentielle est d'obtenir une limite inférieure pour la comparaison de A et B par la comparaison de A' et B'. Nous allons montrer que la limite inférieure est donnée par :

$$\chi^2(A', B') \leq \chi^2(A, B) \tag{5.96}$$

Il faut montrer que :

$$\chi^2([a_0 + a_1], [b_0 + b_1]) \leq \chi^2(a_0, b_0) + \chi^2(a_1, b_1)$$

Preuve :

$$\begin{aligned} \frac{(a_0 + a_1 - b_0 - b_1)^2}{a_0 + a_1 + b_0 + b_1} &\leq \frac{(a_0 - b_0)^2}{a_0 + b_0} + \frac{(a_1 - b_1)^2}{a_1 + b_1} \\ \frac{((a_0 - b_0) + (a_1 - b_1))^2}{(a_0 + b_0) + (a_1 + b_1)} &\leq \frac{(a_0 - b_0)^2(a_1 + b_1) + (a_1 - b_1)^2(a_0 + b_0)}{(a_0 + b_0)(a_1 + b_1)} \end{aligned}$$

Avec les substitutions suivantes : $\alpha = a_0 - b_0, \beta = a_1 - b_1, \gamma = a_0 + b_0$ et $\delta = a_1 + b_1$, l'équation devient :

$$(\alpha + \beta)^2 \gamma\delta \leq (\alpha^2\delta + \beta^2\gamma)(\gamma + \delta)$$

$$\begin{aligned} 2\alpha\beta\gamma\delta &\leq \beta^2\gamma^2 + \alpha^2\delta^2 \\ 2(a_0^2 - b_0^2)(a_1^2 - b_1^2) &\leq (a_1 - b_1)^2(a_0 + b_0)^2 + (a_0 - b_0)^2(a_1 + b_1) \end{aligned}$$

Pour $a_i + b_i = 0$ pour $i = 1$ ou $i = 2$, le côté gauche est zéro. Pour $a_i + b_i > 0$ pour $i = 1, 2$ l'équation devient :

$$\begin{aligned} 8a_0a_1b_0b_1 &\leq 4a_1^2b_0^2 + 4a_0^2b_1^2 \\ 0 &\leq 4(a_1b_0 - a_0b_1)^2 \qquad \square \end{aligned}$$

Remarque : dans le cas de $a_0 = b_0 = 0$, l'égalité de l'équation 5.96 est satisfaite, indiquant qu'aucune limite inférieure plus grande n'existe.

Limite supérieure de χ^2_{qv} :

La limite supérieure est donnée par (utilisant $s(A) = \sum a_i$ comme taille de l'histogramme A) :

$$s(A') + s(B') \geq \chi^2_{qv}(A, B) \tag{5.97}$$

Il faut montrer que :

$$a_0 + a_1 + b_0 + b_1 \geq \chi^2(a_0, b_0) + \chi^2(a_1, b_1)$$

ce qui est directement donné par $(i = 1, 2)$:

$$\begin{aligned} (a_i + b_i)^2 &\geq (a_i - b_i)^2 \\ (a_i + b_i) &\geq \frac{(a_i - b_i)^2}{a_i + b_i} \qquad \square \end{aligned}$$

Remarque : en général aucune limite supérieure plus petite n'existe car l'égalité de l'équation 5.97 peut être obtenue dans le cas de $a_1 = b_0 = 0$.

5.2 Stabilité des fonctions de comparaison d'histogrammes

La section examine la stabilité des différentes fonctions de comparaison d'histogrammes introduites par la section précédente 5.1. La stabilité des fonctions par rapport aux changements d'histogrammes est une condition nécessaire à l'utilité de la comparaison d'histogrammes pour la reconnaissance d'objets. Ces changements d'histogrammes peuvent être causés par des variations d'apparence d'objets ou par des sources diverses de perturbation, comme le bruit d'image et le flou.

Dans les expérimentations huit différentes fonctions sont comparées :

- trois statistiques χ^2 (section 5.1.3) : χ^2_v, χ^2_{qv} et la fonction modifiée χ^2_{av},
- deux distances quadratiques (section 5.1.2) : SSD et $maha$ et

– trois fonctions d'intersection (section 5.1.1) : $\cap$, $\cap_{we}$ et $\cap_{mo}$.

Afin de comparer les différentes fonctions de comparaison, une mesure normalisée d_{norm} est calculée pour chaque fonction. La normalisation utilise la valeur optimale *opt* et la valeur moyenne *av* de chaque fonction. La valeur moyenne *av* est calculée à partir d'un grand nombre d'histogrammes (378 pour l'expérimentation de la rotation d'image et 500 pour les expérimentations du bruit et du flou). La valeur d_{raw} d'une fonction de comparaison est normalisée en utilisant *opt* et *av* par l'équation suivante :

$$d_{norm} = \frac{d_{raw} - opt}{av - opt} \times 0.5 \tag{5.98}$$

C'est à dire que $d_{norm} = 0.0$ correspond à la valeur optimale *opt* de la fonction de comparaison et $d_{norm} = 0.5$ correspond à la valeur moyenne *av*.

Pour les expérimentations de reconnaissance, une valeur de $d_{norm} \leq 0.1$ est typiquement suffisante pour l'identification d'objets. En présence de changements importants il est difficile d'atteindre cette valeur basse. Une valeur de $d_{norm} \leq 0.2$ est souvent appropriée pour obtenir l'objet correct dès les premières correspondances (par exemple dans les premiers 5% de la base). Pour l'analyse suivante de stabilité, ces deux valeurs sont utilisées comme références ($d_{norm} \leq 0.1$ et $d_{norm} \leq 0.2$). Les valeurs exactes de ces références n'influencent que légèrement l'analyse de stabilité car la stabilité relative entre les différentes fonctions de comparaison est examinée plutôt que la stabilité "absolue". Plus importante que la stabilité absolue est la possibilité de discriminer différents objets par la comparaison d'histogrammes. Le chapitre 6 se consacre à ce sujet.

Les valeurs de pixels (des niveaux de gris ou des couleurs) peuvent être interprétées comme les caractéristiques d'image les plus locales (et les plus instables). Comme la stabilité de différentes fonctions de comparaison est analysée, les *niveaux de gris* ont été choisis en tant que caractéristiques locales appropriées – c'est à dire instables. Ce choix est motivé par le souhait de séparer la stabilité liée à la fonction de comparaison de la stabilité liée à la robustesse des caractéristiques locales, comme par exemple la stabilité de dérivées Gaussiennes ou de filtres de Gabor. Les sections suivantes montrent les résultats pour deux différentes résolutions : résolution de 256 cellules d'histogramme (chaque cellule correspond à un niveau de gris, toujours montré à gauche dans les figures 5.2, 5.4 et 5.6) et résolution de 8 cellules d'histogramme (chaque cellule correspond à 32 niveaux de gris, toujours montré à droite dans les figures).

Les sections suivantes visualisent les résultats par des graphes. Pour chaque changement examiné (bruit additif Gaussien, flou et rotation d'image), six graphes sont montrés dans les figures 5.2, 5.4 et 5.6. Les colonnes de gauche montrent les résultats des histogrammes de niveaux de gris de résolution entière (256 cellules) et les colonnes de droite correspondent à la résolution réduite de 8 cellules. Les deux graphes de la même ligne de ces figures correspondent à un ensemble particulier de fonctions de comparaison. L'axe vertical d'un graphe montre la mesure normalisée d_{norm} dans l'intervalle $[0.0, 0.5]$. Comme référence, le graphe de "changements d'histogramme" est toujours inclu et correspond aux changements effectifs d'histogrammes (par exemple dans la figure 5.2 (a), le graphe de référence montre un changement absolu d'histogramme entre 0.0 (pas de changement) et 0.35 (35% de changements)).

La section suivante 5.2.1 examine la stabilité par rapport au bruit additif Gaussien et la

section 5.2.2 par rapport au flou. La section 5.2.3 examine la stabilité en présence de rotation d'image. Les notations suivantes sont employées plus loin :

- statistiques χ^2 : chsone pour χ^2_v, chstwo pour χ^2_{qv} et chsav pour χ^2_{av} (voir section 5.1.3),
- fonctions d'intersection : intersection pour $\cap$, inter_We pour $\cap_{we}$ et inter_Mo pour $\cap_{mo}$ (voir section 5.1.1),
- distances quadratiques : SSD pour SSD et maha pour $maha$ (voir section 5.1.2).

5.2.1 Stabilité par rapport au bruit Gaussien

Pour examiner la stabilité de différentes fonctions de comparaison par rapport au bruit additif Gaussien, 500 images ont été choisies arbitrairement dans notre base d'images. A chaque pixel de ces images, du bruit Gaussien de moyenne zéro et déviation standard de $\sigma = 0.5, 1, 1.5, \ldots 20$ est ajouté. La figure 5.1 montre une image avec du bruit additif Gaussien de $\sigma = 5, 10, 15$ et 20. Dans notre expérimentation, l'histogramme de l'image originale (sans bruit Gaussien) est comparé aux histogrammes des images bruitées. Dans les graphes de la figure 5.2, l'axe horizontal correspond à la déviation standard de σ du bruit additif Gaussien et l'axe vertical correspond à la mesure normalisée d_{norm} (voir équation 5.98). Chaque point de ces graphes est la moyenne des mesures normalisées, obtenues pour 500 images bruitées. Comme mentionné plus haut, le graphe de changements absolus d'histogrammes (moyenne de 500 histogrammes) est montré comme référence.

FIG. 5.1 – *Stabilité par rapport au bruit additif Gaussien : une de 500 images sans bruit Gaussien (à gauche) et avec du bruit Gaussien de $\sigma = 5, 10, 15, 20$ (de gauche à droite)*

Le premier résultat est que les histogrammes de résolution 8 sont plus stables que les histogrammes de résolution 256 indépendamment de la fonction de comparaison (voir figure 5.2). Ce résultat semble clair, mais il n'est pas seulement obtenu pour les valeurs brutes de fonctions de comparaison mais pour les mesures normalisées d_{norm}. Autrement dit, pour le cas de résolution de 8 cellules, les valeurs de fonctions de comparaison varient moins relativement à la valeur moyenne av, calculée séparément pour les résolutions de 8 et de 256. Ce résultat indique que les histogrammes de résolution plus faible sont plus stables au bruit additif Gaussien.

La première ligne de la figure 5.2 montre les résultats des statistiques χ^2. Les résultats obtenus sont stables pour l'ensemble des trois statistiques χ^2 car les mesures normalisées d_{norm} restent toujours en dessous de 0.2 et même en dessous de 0.1 pour des grandes valeurs de σ. En particulier, χ^2_v donne les meilleurs résultats en comparaison avec la totalité des huit mesures.

La deuxième ligne de la figure 5.2 montre les graphes de trois fonctions d'intersection. Les résultats sont stables (pour la résolution de 256 la mesure normalisée reste en dessous de 0.1

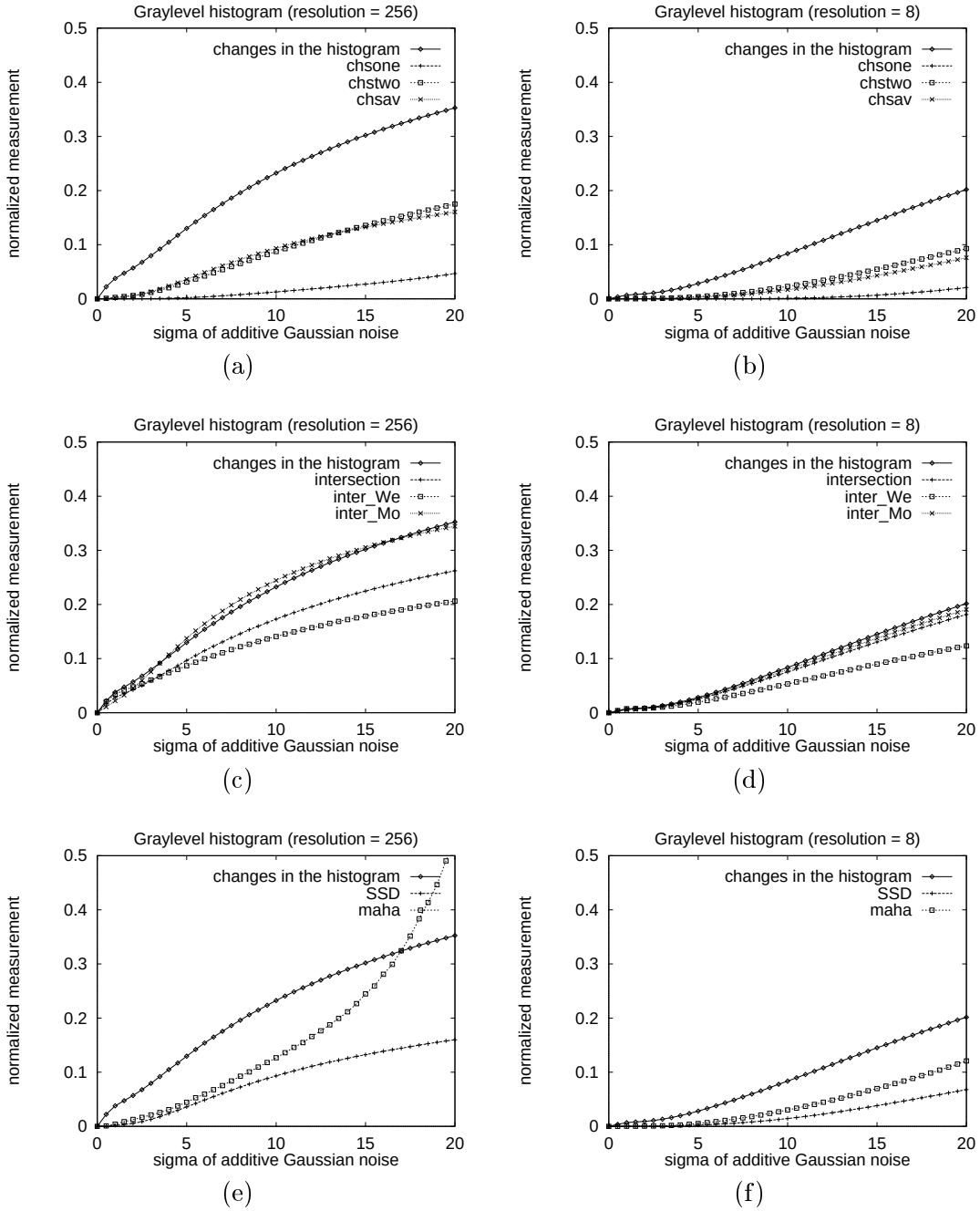


FIG. 5.2 – *Stabilité par rapport au bruit Gaussien. Commentaires, se reporter à la section 5.2.1*

pour $\sigma \leq 5$). Néanmoins, les résultats sont moins bons que pour les statistiques χ^2. $\cap_{we}$ obtient des résultats supérieurs aux autres fonctions d'intersection. $\cap$ est supérieur à $\cap_{mo}$.

La troisième ligne de la figure 5.2 montre les résultats de deux distances quadratiques (de la somme des carrés des distances SSD (équation 5.88) et la distance de la distance de Mahalanobis $maha$ (équation 5.89). SSD montre une bonne stabilité en présence de bruit Gaussien (comparable à χ^2_{qv} et χ^2_{av}). La distance de Mahalanobis montre aussi un comportement stable pour des petites valeurs de σ. Pour les grandes valeurs de σ, la distance de Mahalanobis devient instable en particulier pour la haute résolution de 256 (figure 5.2 (e)). Une bonne stabilité est obtenue par la résolution basse de 8. La dégradation observée pour la résolution de 256 est due au fait que l'estimation des variances $\eta^2_{\mathbf{i}}$ (nécessaire pour le calcul de la distance de Mahalanobis, voir équation 5.89) n'a pas considéré le bruit Gaussien. L'estimation de variances est toujours délicat pour l'application de la distance de Mahalanobis car le calcul de cette distance inclut la division par les variances estimées. La distance de Mahalanobis est typiquement moins stable que la somme des carrés des distances SSD.

S'appuyant sur cette première expérimentation, nous pouvons résumer que la statistique χ^2_v donne les résultats les plus stables, suivi par χ^2_{qv}, χ^2_{av} et SSD. La distance de Mahalanobis peut être stable si les variances sont estimées de manière robuste. Les trois fonctions d'intersection donnent des résultats acceptables mais moins stables que les autres fonctions de comparaison.

5.2.2 Stabilité en présence de flou

Une deuxième expérimentation examine la stabilité des fonctions de comparaison par rapport au filtrage répété par la moyenne[2], utilisant un masque de 3×3. Ce filtrage peut être utilisé comme fonction de lissage et peut être vu comme simulation de flou d'image, du à une mauvaise mise au point[3]. Afin d'augmenter l'effet de mauvaise mise au point, le filtrage est répété plusieurs fois pour la même image. La figure 5.3 montre une image et différentes versions filtrées de l'image. Comme dans l'expérimentation précédente, 500 images d'objets sont utilisées.

FIG. 5.3 – *Stabilité par rapport au flou : une de 500 images filtrées plusieurs fois par une masque de* 3×3 *:* $2, 4, 7$ *et* 10 *fois*

Comme introduit plus haut, la mesure normalisée d_{norm} est utilisée pour rendre les fonctions comparables. L'axe vertical des graphes de la figure 5.4 correspond à la mesure normalisée d_{norm}. L'axe horizontal de ces graphes montre le nombre d'applications du filtre (opération de flou).

Les graphes 5.4(a) et (b) montrent les résultats des statistiques χ^2. Comme dans la première expérimentation, χ^2_v donne les résultats les plus stables, relativement aux autres fonctions. Les

2. mean–filtering

3. defocusing

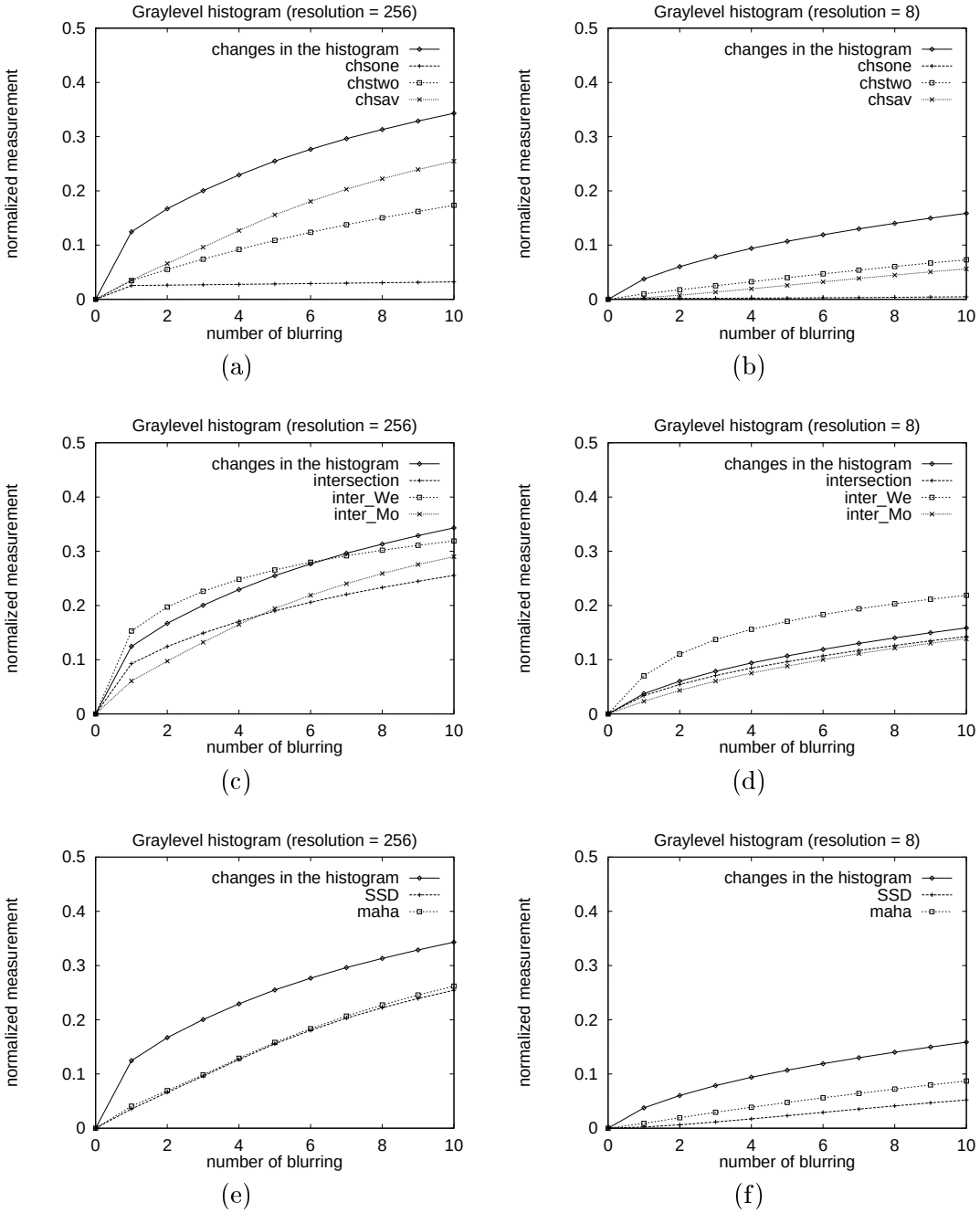


FIG. 5.4 – *Stabilité par rapport au flou. Commentaires, se reporter à la section 5.2.2*

résultats de χ^2_{qv} sont stables pour la résolution de 8 et raisonnablement stables pour la résolution de 256 (les deuxièmes meilleurs résultats). Cette fois χ^2_{av} produit des résultats moins stables que χ^2_{qv} pour la résolution de 256. Les résultats de χ^2_{av} pour la résolution de 8 sont légèrement meilleurs que ceux de χ^2_{qv}.

Les graphes 5.4(c) et (d) montrent les résultats des fonctions d'intersection. Dans le cas de résolution de 256 un seul filtrage cause une augmentation significative des valeurs de différentes fonctions d'intersection. Cela est dû aux changements d'histogrammes, montrés par le graphe de référence ("changes in the histogram") ainsi qu'à la dépendance directe des fonctions d'intersection de ces changements. Dans les deux graphes 5.4(c) (résolution = 256) et (d) (résolution = 8) les deux fonctions $\cap$ et $\cap_{mo}$ donnent des résultats très similaires. $\cap_{we}$ obtient des résultats moins bons.

Les graphes 5.4(e) et (f) montrent les résultats des distances quadratiques. Ces deux distances se comportent de façon très similaire. La stabilité obtenue est très haute pour la résolution de 8 et comparable aux résultats de χ^2_{qv} et χ^2_{av}. Dans le cas de la haute résolution de 256, la mesure normalisée d_{norm} reste en dessous de 0.1 jusqu'à trois filtrages et en dessous de 0.2 jusqu'à 7 filtrages. Ces résultats sont meilleurs que ceux des fonctions d'intersection, mais moins bons que ceux de χ^2_v.

En conclusion, les statistiques χ^2 donnent les résultats les plus stables. Plus précisément χ^2_v fournit les meilleurs résultats. Les deuxièmes meilleurs résultats sont obtenus par χ^2_{qv}, χ^2_{av} et *SSD*. *maha* donne des résultats légèrement inférieurs. Les résultats les moins bons sont donnés par les fonctions d'intersection avec $\cap$ et $\cap_{mo}$, typiquement supérieurs aux résultats de $\cap_{we}$.

5.2.3 Stabilité par rapport aux rotations d'image

Cette section présente une évaluation de la stabilité des histogrammes de niveaux de gris en présence de rotations d'image. Théoriquement, les histogrammes de niveaux de gris sont invariants par rapport aux rotations d'image (en utilisant un support circulaire pour le calcul d'histogrammes). Étant donné le bruit et la forme rectangulaire de pixels, les changements observés de ces histogrammes sont importants. Le graphe de changements d'histogrammes de la figure 5.6 (a) montre un changement (pour la résolution = 256) entre 13% et 20%. Pour la résolution de 8 (figure 5.6 (b)) ces changements sont moins importants (en dessous de 10%). Dans le cas de haute résolution, les changements correspondants à un bruit additif Gaussien d'approximativement $\sigma = 5$ et 8.

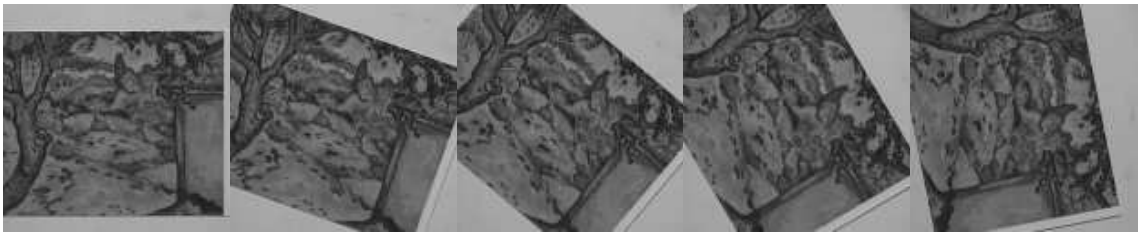


FIG. 5.5 – *5 rotations d'un de 21 objets*

Dans cette expérimentation, 18 images tournées de 21 objets sont utilisées. La figure 5.5 montre 5 rotations d'un objet. Les rotations sont d'une différence approximative de 20°. L'axe vertical des graphes de la figure 5.6 montre l'angle de rotation entre une image de référence et

une image test du même objet. Chacune des 18 images tournées d'un objet est utilisée comme image de référence. Chaque point d'un graphe correspond alors à la moyenne de $18 \times 21 = 378$ images.

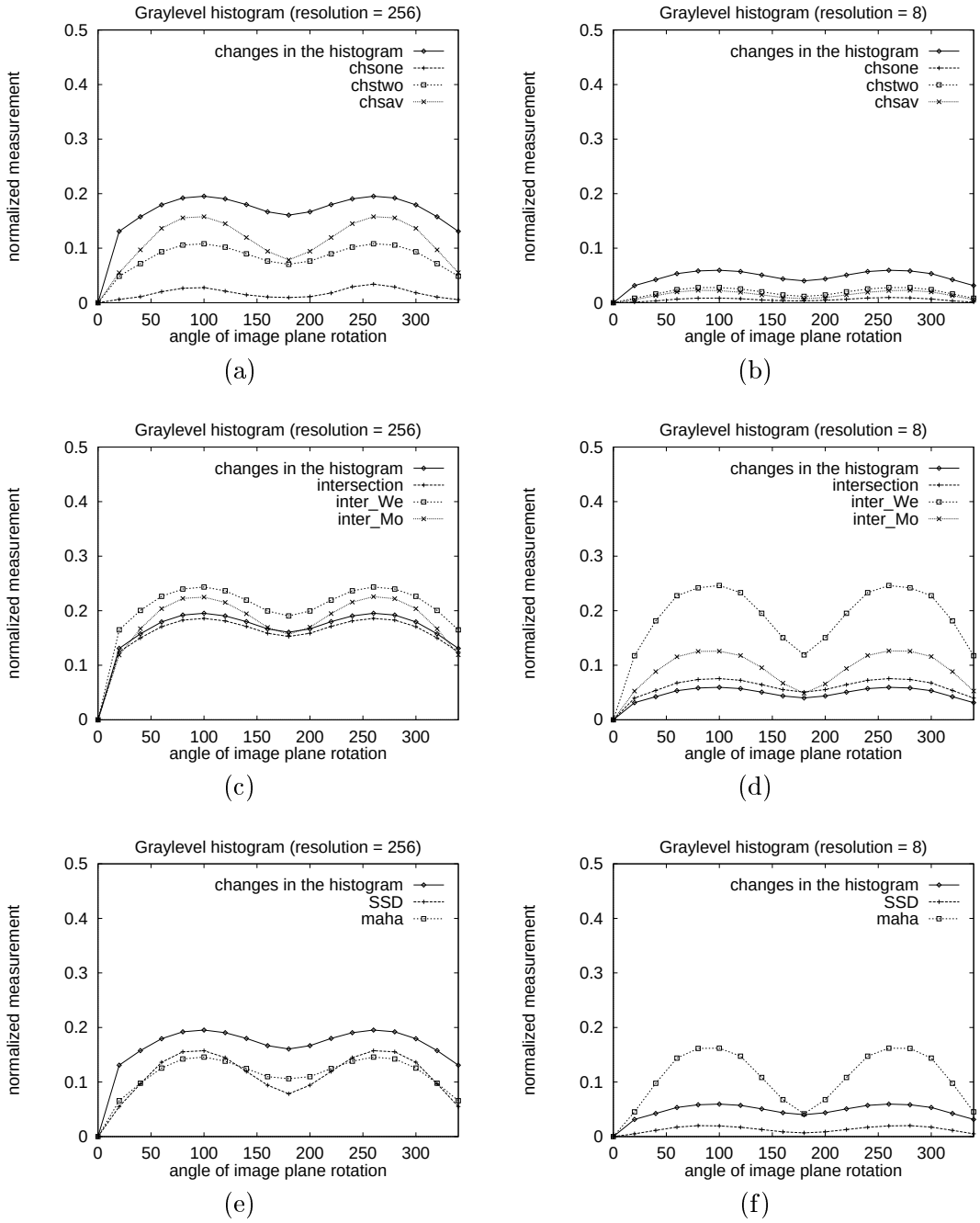


FIG. 5.6 – *Stabilité par rapport aux rotations d'image. Commentaires se reporter à la section 5.2.3*

La première ligne de la figure 5.6 montre les résultats obtenus par les statistiques χ^2. Dans

le cas de haute résolution, les résultats de χ^2_v sont meilleurs que ceux de χ^2_{qv}. Pour cette expérimentation, χ^2_{av} donne des résultats moins stables (néanmoins, les troisièmes meilleurs résultats avec SSD et $maha$). Dans le cas de basse résolution de 8, χ^2_v donne de meilleurs résultats et les deux autres fonctions donnent des résultats légèrement inférieurs.

La deuxième ligne de la figure 5.6 montre que les fonctions d'intersection donnent des résultats instables. La meilleure des fonctions d'intersection est l'intersection non–pondérée $\cap$. $\cap_{mo}$ donne des résultats moins stables et $\cap_{we}$ donne les résultats les plus instables de toutes les fonctions.

La troisième ligne de la figure 5.6 montre que les résultats de SSD et $maha$ sont comparables aux résultats de χ^2_{av} pour la haute résolution de 256. Dans le cas de basse résolution de 8 seulement SSD donne des résultats stables (comparables avec χ^2_{av} et χ^2_{qv}). Pour la basse résolution, la distance de Mahalanobis $maha$ fournit les résultats plus instables même si la mesure normalisée d_{norm} reste en dessous de 0.2.

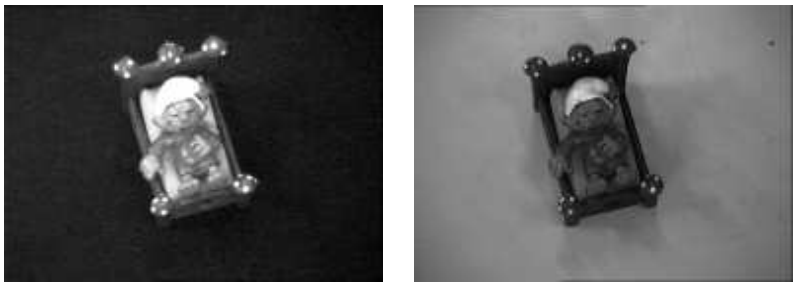
Le résumé de cette expérimentation est similaire aux résultats des deux premières expérimentations (sections 5.2.1 et 5.2.2) : χ^2_v donne les résultats les plus stables, suivi par SSD, χ^2_{av} et χ^2_{qv}. La distance de Mahalanobis est moins stable que les quatre premières fonctions de comparaison. Les fonctions d'intersection donnent les résultats les plus instables ($\cap$ meilleurs que $\cap_{mo}$ qui est supérieur à $\cap_{we}$).

5.3 Stabilité de la comparaison d'histogrammes par rapport aux changements de l'intensité d'éclairage

Cette section décrit une expérimentation de la stabilité de la comparaison d'histogrammes par rapport aux changements de l'intensité d'éclairage. Comme dans la section précédente 5.2, les différentes fonctions de comparaison doivent être évaluées. En opposition à la section précédente, il n'est pas judicieux d'utiliser des histogrammes de niveaux de gris car ces niveaux varient de façon dramatique en présence de changements de l'intensité d'éclairage. Étant donné ce fait, des dérivées Gaussiennes sont utilisées. De plus, différentes techniques de normalisation des réponses de filtres sont appliquées. En particulier, la normalisation par *énergie*, la normalisation par *variance-moyenne* et *absence de* normalisation sont examinées (voir la section 3.2 pour les détails de techniques de normalisation).

L'influence et le succès de la normalisation dépendent fortement des caractéristiques de la caméra employée. Pour la plupart de nos expérimentations nous avons utilisé la caméra Canon VCC1. Une analyse [Col 96] des caractéristiques de la caméra a montré que le contrôle automatique du gain de cette caméra maintient constantes la moyenne des niveaux de gris et aussi la variance de niveaux de gris. Des changements d'intensité de l'image entière ne sont alors pas mesurables comme changements de l'image. Si le changement d'intensité est constant sur l'image entière il n'est pas nécessaire de normaliser les réponses de filtres. Néanmoins, comme le contrôle du gain de la caméra est global sur l'image entière, des changements locaux d'intensité dans l'image peuvent être causés par exemple par le fond, illustré par la figure 5.7.

Afin d'obtenir une plage importante de changements d'intensité d'éclairage la caméra digital Indy-Cam, fournie avec une Silicon Graphics Indy, est utilisée. Typiquement, cette caméra n'est pas appropriée étant donné sa mauvaise qualité d'image. Par contre il est facile d'éteindre le contrôle automatique du gain pour cette expérimentation. Les caractéristiques de la caméra



(a) (b)

FIG. 5.7 – *Deux images du même objet devant deux différents fonds. Des changements importants de l'intensité locale peuvent être observés*

sont montrées par la figure 5.8. Ce graphe montre la relation entre la variance de niveaux de gris et la moyenne de niveaux de gris de trois séries d'images.

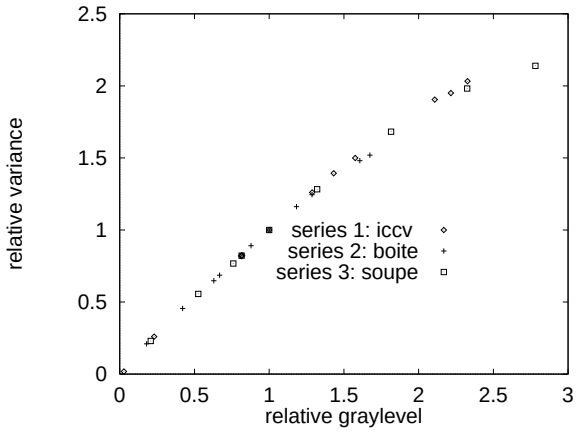


FIG. 5.8 – *Les dynamiques de la caméra Indy-Cam (caméra digital de Silicon Graphics)*

Les séries d'images employées de trois objets sont montrées dans la figure 5.9 et dans les figures B.1 et B.2 de l'appendice. Comme les résultats sont similaires pour l'ensemble de trois séries, les résultats d'une seule série sont décrits (voir figure 5.9). Pendant la prise de séries d'images, l'intensité d'éclairage d'une lampe halogène était diminuée de façon continue. Une image d'intensité moyenne d'éclairage est choisie comme image de référence. Les nombres en dessous des images de la figure 5.9 montrent le niveaux de gris moyen de l'image entière, relativement à l'image de référence

Comme mentionné plus haut, les histogrammes de niveaux de gris de la section précédente ne sont pas adaptés au contexte de variations de l'intensité d'éclairage. Cette section utilise alors des histogrammes bidimensionnels de la combinaison de filtres *Dx-Dy* (premières dérivées Gaussiennes dans les directions x et y avec $\sigma = 2.0$). La résolution est 32 cellules par axe d'histogramme. Ces histogrammes sont typiques pour les expérimentations décrites dans le chapitre 6.

La figure 5.10 montre l'application de plusieurs fonctions de comparaison d'histogrammes. L'axe horizontal montre le niveau de gris moyen des images de la figure 5.9. L'axe vertical montre

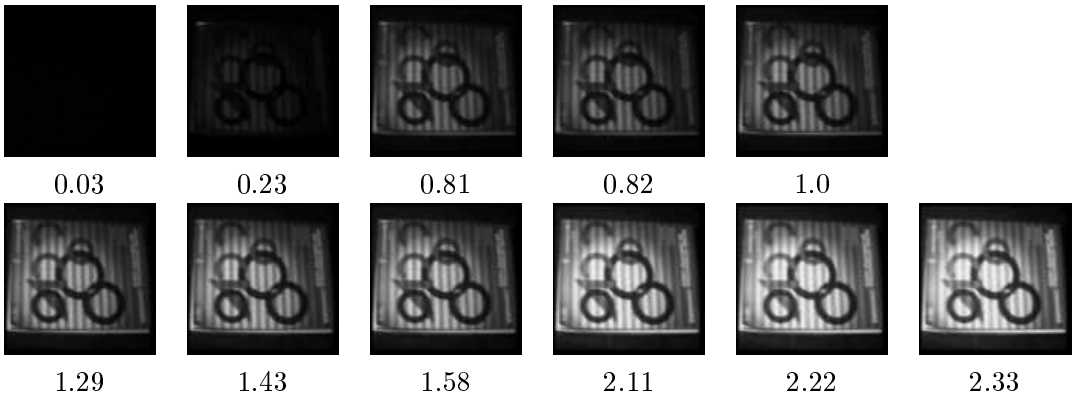


FIG. 5.9 – *Série d'images de changement d'intensité d'éclairage. Les nombres en dessous des images correspondent aux niveaux de gris moyens par rapport à l'image de référence*

la mesure normalisée comme définie par l'équation 5.98. Afin d'estimer la distance moyenne av, 500 images d'objets ont été utilisées. Plus précisément, la figure 5.10 (a) montre l'application de χ^2_q (chsone), (b) de χ^2_{qv} (chstwo), (c) de χ^2_{av} (chsav), (d) de $\cap$ (intersection), (e) de SSD (ssd), et (f) de *maha* (maha) (voir section 5.1 pour l'introduction des fonctions de comparaison).

Chaque graphe de la figure 5.10 montre l'application d'*absence de* (no) normalisation, de la normalisation par *énergie* (eg) et de la normalisation par *variance-moyenne* (vg). Pour *absence de* normalisation le meilleur résultat est obtenu pour la statistique χ^2_q, suivi par les résultats de χ^2_{qv}, χ^2_{av}, SSD et *maha*. Les résultats le moins bons sont obtenus pour la fonction d'intersection $\cap$. Les résultats de la normalisation par énergie et par variance-moyenne sont pratiquement identiques. Pour ces deux techniques de normalisation, les résultats de χ^2_q, χ^2_{qv}, χ^2_{av}, SSD et *maha* sont très similaires et les résultats de la fonction d'intersection $\cap$ sont moins bons.

Pour les six fonctions de comparaison la normalisation par énergie et par variance-moyenne stabilise les résultats de manière significative par rapport à une absence de normalisation de réponses de filtres. A part la fonction de comparaison d'intersection $\cap$, la distance maximale reste pratiquement toujours en dessous de 0.1, ce qui correspond à une bonne stabilité des techniques de normalisation par rapport aux changements de l'intensité d'éclairage.

Les résultats les moins bons sont obtenus pour les images d'intensité faible (intensité relative 0.03 et 0.23). Cela peut être expliqué par la relation faible de signal-bruit de ces images. Pour toutes les autres images la mesure normalisée est toujours proche de zéro (sauf la fonction d'intersection $\cap$ qui est légèrement plus grande).

Les résultats décrits sont typiques pour ce type d'expérimentations. Les figures B.1 et B.2 de l'appendice montrent des résultats similaires, obtenus pour deux autres séries d'images. La conclusion principale de cette expérimentation est que des résultats stables par rapport aux changements de l'intensité d'éclairage peuvent être obtenus par la normalisation des réponses de filtres par énergie ou par variance-moyenne. Les techniques de normalisation deviennent moins stables pour des images de faible luminosité. Une instabilité comparable est attendue pour des images saturées, ce qui n'était pas examiné par ces expérimentations. Les fonctions de comparaison fournissent toutes des résultats stables (sauf la fonction d'intersection) et elles montrent des comportements similaires.

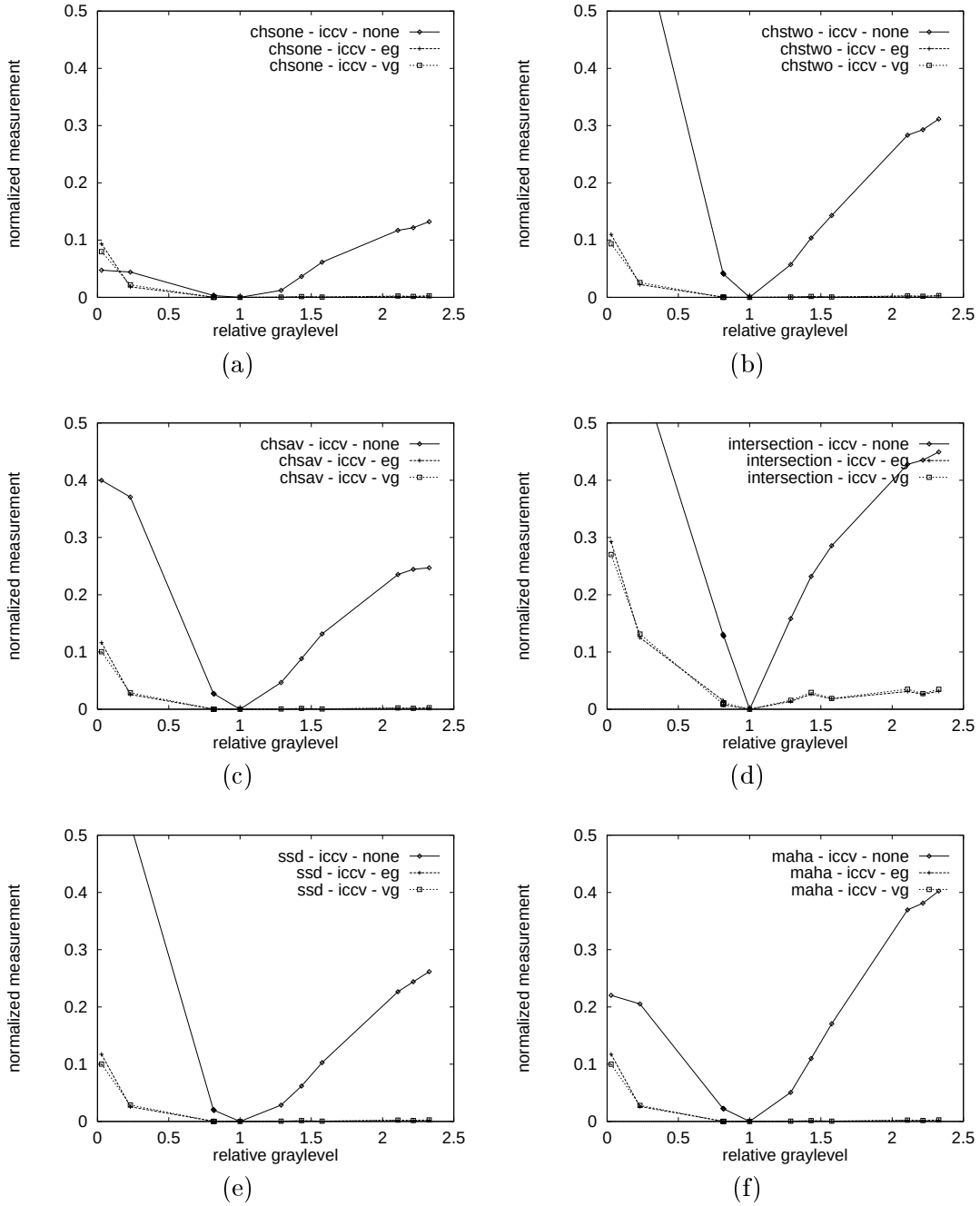


FIG. 5.10 – *Stabilité par rapport aux changements de l'intensité de la série d'images de la figure 5.9 : (a) chsone: χ^2_q, (b) chstwo : χ^2_{qv}, (c) chsav : χ^2_{av}, (d) intersection : $\cap$, (e) ssd, (f) maha : distance de Mahalanobis*

5.4 Conclusion

Ce chapitre a introduit des fonctions différentes de la comparaison d'histogrammes. L'analyse des ces fonctions est importante car la fonction d'intersection (proposée par Swain et Ballard pour la comparaison des histogrammes de couleurs [Swa 91]) possède des limitations théoriques pour l'application dans le contexte d'histogrammes multidimensionnels de champs réceptifs: afin d'obtenir une séparation optimale de différents histogrammes, les mesures m_k doivent être équiprobables et distribuées grossièrement. A côté de ces contraintes théoriques, nous avons observé un comportement instable de la fonction d'intersection en présence de bruit additif Gaussien, de flou et des rotations d'image. En présence de changements de l'intensité d'éclairage, la fonction d'intersection obtient les résultats les plus mauvais.

Pour réduire les limitations de la fonction d'intersection nous avons proposé des fonctions de comparaison de deux classes: les statistiques χ^2 et les distances quadratiques. En particulier, les statistiques χ^2 (la méthode proposée par la statistique pour la comparaison d'histogrammes) fournit des résultats stables par rapport à tous les changements examinés. L'utilisation des distances quadratiques permet l'introduction de relations entre les différentes cellules d'histogramme, particulièrement intéressant en présence de changements importants d'histogrammes. Deux versions de distances quadratiques ont été proposées: la somme des carrés des distances SSD et la distance de Mahalanobis $maha$, qui ne considèrent pas de relations entre les cellules. Les deux fonctions montrent un comportement plus stable que la fonction d'intersection.

Les résultats de la section 5.2 peuvent être résumés (analyse de stabilité par rapport au bruit additif Gaussien, au flou et aux rotations d'images):

- χ^2_v obtient les résultats les plus stables,
- χ^2_{qv}, χ^2_{av} et SSD donnent des résultats similaires et stables. Dans les cas particuliers, χ^2_{av} devient moins stable,
- la distance de Mahalanobis $maha$ donne des résultats similaires à SSD mais typiquement moins stables. En particulier dans le cas de rotations d'image les résultats ont été moins stables,
- les trois fonctions d'intersection donnent les résultats les moins stables.

La section 5.3 décrit l'analyse de stabilité des fonctions de comparaison par rapport aux changements de l'intensité d'éclairage. Les résultats soulignent la stabilité des statistiques χ^2 et des distances quadratiques. La fonction d'intersection $\cap$ fournit les résultats les moins stables. La section 5.3 n'examine pas seulement la stabilité des fonctions de comparaison mais aussi la stabilité de différentes techniques de normalisation de réponses de filtres. Les expérimentations montrent une stabilité considérable en présence de changements de l'intensité d'éclairage en utilisant la normalisation par énergie ou par variance-moyenne.

Ce chapitre a analysé la complexité et la stabilité de différentes fonctions de comparaison d'histogrammes. Encore plus importante est la possibilité de discriminer différents objets dans le contexte de la reconnaissance d'objet. Le chapitre suivant décrit des expérimentations de reconnaissance par comparaison d'histogrammes en présence de variations d'apparence d'objets. Les changements d'apparences incluent des changements d'échelle, des rotations dans le plan image et des changements de points de vue.

Chapitre 6

Reconnaissance d'objets par comparaison d'histogrammes

La comparaison d'histogrammes est une application directe des histogrammes multidimensionnels de champs réceptifs à la reconnaissance d'objets. Le chapitre 5 a introduit différentes fonctions de comparaison d'histogrammes et analysé leur stabilité par rapport à des changements divers. Ce chapitre décrit des expérimentations de l'identification d'objets par comparaison d'histogrammes, en présence de rotations d'image, de variations d'échelle, de changements de point de vue et d'occultations partielles.

Ce chapitre emploie les dérivées Gaussiennes suivantes (voir section 3.1.1) :

- *Dx* : première dérivée Gaussienne dans la direction x
- *Dy* : première dérivée Gaussienne dans la direction y
- *Lap* : opérateur Laplacien
- *G12* : filtre Gaussien invariant à la rotation d'image, fondé sur des dérivées Gaussiennes d'ordre un et deux

Différentes combinaisons de ces filtres sont employées : deux combinaisons de filtres variantes à la rotation (*Dx-Dy* et *Dx-Dy-Lap*) et deux combinaisons de filtres invariantes à la rotation (*Mag-Lap* et *Mag-G12-Lap*). Ce chapitre examine la performance de reconnaissance de différentes combinaisons de filtres ainsi que l'influence de résolutions diverses d'histogrammes.

La notation *Dx-Dy-32*, par exemple, correspond à l'histogramme de la combinaison de filtres *Dx-Dy* d'une résolution de 32 cellules par axe d'histogramme. L'histogramme bidimensionnel correspondant contient alors $32^2 = 1024$ cellules. Étant donné la discrétisation, chaque cellule correspond à un intervalle de réponses du filtre bidimensionnel *Dx-Dy*. Cet intervalle est appelé "vecteur de mesures" m_k.

Comme dans le chapitre précédent, nous utilisons les notations suivantes dans les figures pour les différentes fonctions de comparaison d'histogrammes :

- statistiques χ^2 : chsone pour χ^2_v, chstwo pour χ^2_{qv} et chsav pour χ^2_{av} (voir section 5.1.3),
- fonctions d'intersection : intersection pour $\cap$, inter_We pour $\cap_{we}$ et inter_Mo pour $\cap_{mo}$ (voir section 5.1.1),
- distances quadratiques : SSD pour SSD et maha pour $maha$ (voir section 5.1.2).

La section 6.1 présente un exemple de reconnaissance de 261 objets par comparaison d'histogrammes. La section 6.2 décrit la prise en compte des rotations d'image et la section 6.3 considère des changements d'échelle. Les expérimentations de la section 6.4 fournissent des résultats de reconnaissance en présence de rotations d'image et de variations d'échelle. La robustesse de la comparaison d'histogrammes aux changements de point de vue est examinée dans la section 6.5. Une expérimentation à l'intérieur d'une base de 103 objets, en présence de changements de point de vue, de rotations d'image et de variations d'échelle est décrite dans la section 6.6. Cette dernière section examine plus particulièrement la robustesse de la comparaison d'histogrammes dans le cadre des occultations partielles. La consommation de mémoire est discutée brièvement par la section 6.7. La section résume les résultats du chapitre.

6.1 Un exemple de l'identification d'objets par comparaison d'histogrammes

Afin de montrer l'applicabilité de la comparaison d'histogrammes pour l'identification d'objets, une expérimentation dans une base de 261 objets est proposée. Le calcul de la base d'histogrammes utilise une image par objet, correspondant à une apparence particulière d'un objet. Comme ensemble d'images tests nous utilisons différentes images des mêmes 261 objets, introduisant des changements d'apparence d'objets. En particulier, l'échelle et le point de vue ont été changés. La base de données se compose des objets suivants qui peuvent être trouvés dans l'appendice A :

- 61 objets de notre base d'images — voir figures A.1 et A.3
- 100 objets de la base d'images de Columbia — voir figure A.6
- 100 objets de la base d'images aériennes de Marseille — voir figures A.4 et A.5

La base de données contient une seule image par objet. Pour chaque image nous calculons un histogramme bidimensionnel *Dx-Dy-32* : premières dérivées Gaussiennes dans les directions x et y d'une résolution de 32 cellules par axe d'histogramme. Chaque histogramme contient alors $32^2 = 1024$ cellules. Pour la reconnaissance, ces histogrammes sont comparés aux histogrammes

d'images tests. Deux ensembles d'images tests ont été utilisés contenant chacun 261 images des mêmes objets. Les images de ces deux ensembles ont été prises dans des conditions différentes de celles de la base de données. Le premier ensemble contient les changements suivants : 10° de variation du point de vue pour les images de la base de Columbia, la deuxième série d'images aériennes de Marseille (correspondant à des variations de point de vue) et entre 10% et 15% de changements d'échelle pour les derniers 61 objets.

fonction de comparaison	$\cap$	$\cap_{we}$	$\cap_{mo}$	χ^2_v	χ^2_{qv}	χ^2_{av}	SSD	$maha$
reconnaissance	98.9	82.4	47.9	95.4	99.2	97.3	96.9	98.1
erreurs	3	46	136	12	2	7	8	5

TAB. 6.1 – *Résultats de reconnaissance d'une base de 261 objets et du premier ensemble d'images tests. Cet ensemble contient entre 10% et 15% de changements d'échelle et* 10° *de variations du point de vue*

La table 6.1 montre les résultats de reconnaissance du premier ensemble d'images tests par les 8 différentes fonctions de comparaison. La meilleure reconnaissance de 99% est obtenue par la statistique χ^2_{qv}. Une qualité de résultat quasiment identique est donnée par la fonction d'intersection $\cap$. Des résultats de reconnaissance au dessus de 95% sont fournis par les quatre fonctions *maha*, χ^2_{av}, SSD et χ^2_v. La fonction d'intersection pondérée $\cap_{we}$ obtient un taux de reconnaissance de 82%. $\cap_{mo}$ ne donne pas un résultat satisfaisant.

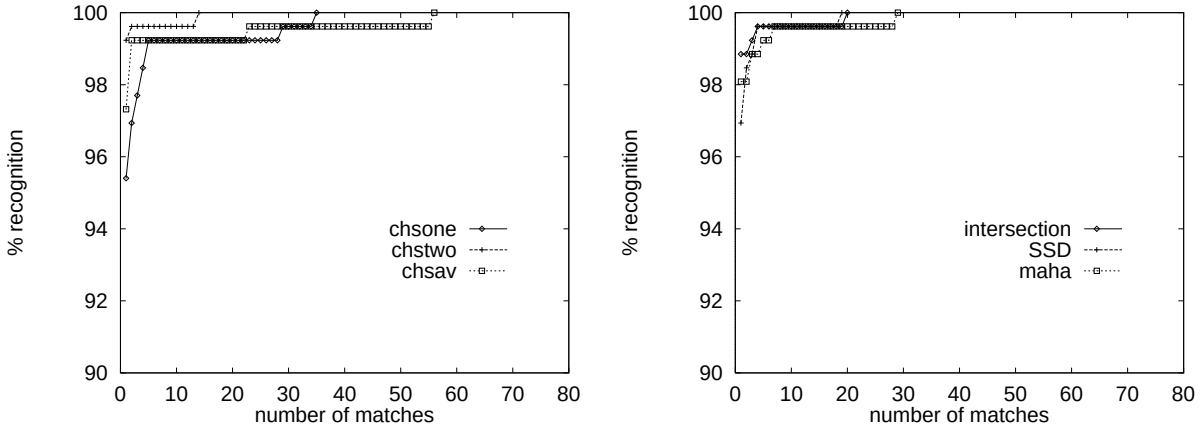


FIG. 6.1 – *Comparaison d'histogrammes pour le premier ensemble d'images tests de 261 objets. Le taux de reconnaissance est montré en fonction du nombre de correspondances inclues*

En n'utilisant pas uniquement la meilleure correspondance mais aussi les n premières correspondances, on obtient le graphe de la figure 6.1. L'axe horizontal montre le nombre de correspondances considérées pour le calcul du taux de reconnaissance. Le taux de reconnaissance est montré par l'axe vertical. Afin d'obtenir un taux de reconnaissance de 100%, χ^2_{qv} a besoin de 15 correspondances. SSD et $\cap$ ont besoin d'environ 20 correspondances et *maha* de 29 correspondances pour atteindre une reconnaissance de 100%. Pour ces quatre fonctions, le nombre de correspondances pour atteindre le taux de 100% est de l'ordre de 5% à 11% de la taille de la base de données. χ^2_v a besoin de 35 correspondances et χ^2_{av} déjà de 56 correspondances. Les

dernières fonctions $\cap_{we}$ et $\cap_{mo}$ (pas présentées dans la figure 6.1) ont besoin de plus de 100 correspondances.

Les taux de reconnaissance élevés de χ^2_{qv} et $\cap$ de cette première expérimentation montrent l'applicabilité de la comparaison d'histogrammes pour l'identification d'objets en présence de changements d'apparences d'objets. En particulier, l'objet correct était dans les premiers 5% de correspondances de χ^2_{qv}.

fonction de comparaison	$\cap$	$\cap_{we}$	$\cap_{mo}$	χ^2_v	χ^2_{qv}	χ^2_{av}	SSD	$maha$
reconnaissance	95.0	66.67	38.3	86.2	96.2	90.4	84.3	92.0
erreurs	13	87	161	36	10	25	41	21

TAB. 6.2 – *Résultats de reconnaissance d'une base de 261 objets pour le deuxième ensemble d'images tests. Cet ensemble contient entre 20% et 30% de changements d'échelle et* 20° *de variation du point de vue*

La table 6.2 montre les résultats de reconnaissance du deuxième ensemble de 261 images tests. Cet ensemble test contient plus de changements que le premier ensemble : 20° de variation de point de vue pour les 100 objets de la base d'images de Columbia, la troisième série d'images aériennes de Marseille (avec des variations plus importantes de point de vue) et entre 20% et 30% de changements d'échelle des derniers 61 objets. Naturellement, les taux de reconnaissance de toutes les fonctions de comparaison sont inférieurs à ceux du premier ensemble d'images tests. La table 6.2 montre un taux de reconnaissance de 96% pour χ^2_{qv} indiquant une robustesse de cette fonction aux changements importants d'échelle et de point de vue. Le deuxième meilleur résultat d'une reconnaissance de 95% est obtenu par $\cap$. Des taux de reconnaissance au dessus de 90% sont fournis par χ^2_{av} et *maha*. Un taux de reconnaissance au dessus de 85% est donné par χ^2_v et SSD. Comme pour le premier ensemble test, les fonctions d'intersection pondérée donnent les résultats les moins bons de toutes les fonctions.

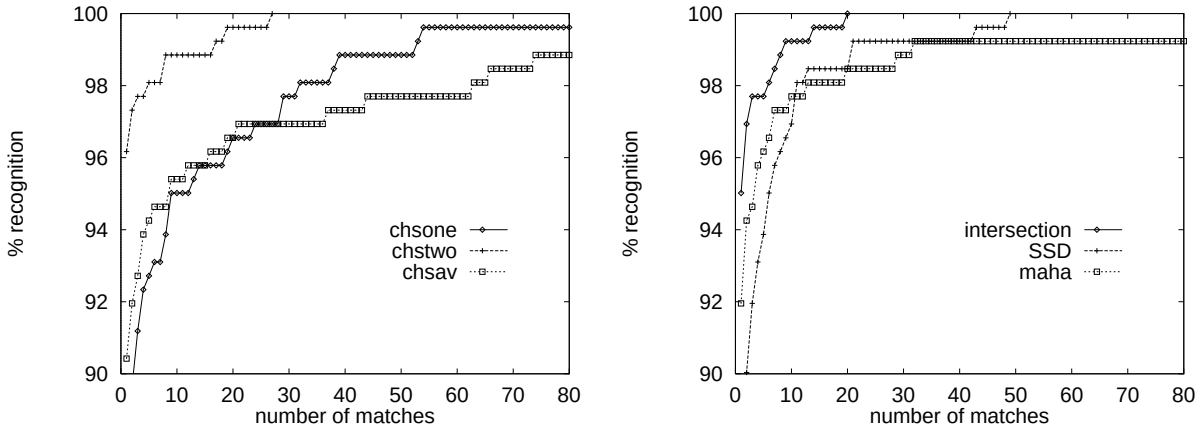


FIG. 6.2 – *Comparaison d'histogrammes pour le deuxième ensemble d'images tests de 261 objets. Le taux de reconnaissance est montré en fonction du nombre de correspondances inclues*

Comme pour le premier ensemble test, la figure 6.2 montre le nombre de correspondances nécessaires pour atteindre un taux de reconnaissance de 100%. Les fonctions d'intersection $\cap$ et

χ^2_{qv} requièrent les nombres les plus petits. Les nombres de correspondances pour ces deux fonctions sont inférieurs de 10% à la taille de la base de données. SSD — avec 49 correspondances — a besoin d'environ 20% de la taille de la base.

Ces deux expérimentations initiales indiquent que la comparaison d'histogrammes constitue un moyen fiable pour l'identification d'objets. La meilleure fonction de comparaison était χ^2_{qv}, suivie par la fonction d'intersection $\cap$. De bons résultats sont obtenus par la distance de Mahalanobis *maha*, par la somme des distances carrées SSD, par χ^2_{av} et par χ^2_v. Par contre, les résultats des fonctions d'intersection pondérée $\cap_{we}$ et $\cap_{mo}$ sont moins bons. Les sections suivantes vont montrer le traitement des changements d'apparence d'objet de façon explicite et appropriée.

6.2 Identification d'objets en présence de rotations d'image

La section décrit des expérimentations de l'identification d'objets en présence de rotations dans le plan image. Comme introduit dans la section 3.1.1, la réponse d'un filtre Gaussien d'une rotation arbitraire peut être calculée à partir d'un ensemble fini de filtres Gaussiens. En utilisant cette propriété — appelée orientabilité des dérivées Gaussiennes — nous calculons des histogrammes de dérivées Gaussiennes de rotations arbitraires d'image à partir d'une seule image d'objet. Cette propriété n'existe pas seulement pour les dérivées Gaussiennes mais par exemple aussi pour les filtres de Gabor (voir section 3.1.1 pour plus de détails).

Dans cette section, nous appliquons quatre combinaisons de filtres : *Dx-Dy*, *Dx-Dy-Lap*, *Mag-Lap* et *Mag-G12-Lap*. Les deux premières combinaisons sont variantes à la rotation et il faut les orienter aux rotations diverses de l'image. En utilisant l'orientabilité des dérivées, une seule image par objet est utilisée pour le calcul des réponses de filtres de rotations arbitraires. Les expérimentations ont montré qu'une différence entre les rotations calculées de $\Delta\alpha = 10°, 20°, 30°$ et $40°$ fournissent toutes des résultats très similaires pour l'identification des 22 objets employés. Comme ce résultat ne se généralise pas à un ensemble arbitraire d'objets nous ne citons pas l'analyse de $\Delta\alpha$ en détail. Pour la suite, nous supposons $\Delta\alpha = 20°$ et calculons les réponses de filtres de 18 différentes rotations, plus précisément des angles de $\alpha = 0°, 20°, 40°, \ldots, 340°$. Pour chacune des 18 réponses de filtres orientés d'une image, nous calculons un histogramme. Chaque objet est alors représenté par une collection de 18 histogrammes.

Les deux autres combinaisons de filtres (*Mag-Lap* et *Mag-G12-Lap*) sont invariantes à la rotation d'image. En opposition aux deux premières combinaisons de filtres nous pouvons utiliser un seul histogramme pour la représentation d'un objet. Par conséquent, le nombre d'histogrammes est réduit de manière significative. Néanmoins, nous nous attendons à une perte d'information qui peut être importante pour l'identification. Il est alors intéressant de comparer les résultats des histogrammes invariants à la rotation, à ceux des histogrammes variants.

La base de données contient 22 objets montrés dans la figure A.2 de l'appendice. Pour chaque objet, 18 différentes rotations ont été prises. Les objets ont été tournés devant la caméra de manière à ce que la différence entre chaque rotation soit approximativement 20°. Le support d'histogrammes de cette section est circulaire afin d'utiliser la même portion d'objet. La déviation standard σ de dérivées Gaussiennes était $\sigma = 2.0$ et le rayon du support circulaire de chaque histogramme était $110\sigma = 220$ pixels. Chaque histogramme est alors calculé à partir d'approximativement 150.000 positions. La taille des images est 500×500 pixels. Comme base de données nous avons choisi arbitrairement une image par objet. L'ensemble d'images tests

contient toutes les $22 \times 18 = 396$ images.

Dans une première expérimentation nous examinons les résultats de la reconnaissance de quatre combinaisons de filtres en utilisant 8 différentes fonctions de comparaison (voir section 5.1). La résolution d'axes d'histogrammes était 32 cellules. C'est à dire que les histogrammes bidimensionnels (*Dx-Dy-32* et *Mag-Lap-32*) contiennent $32^2 = 1024$ cellules, et les histogrammes tridimensionnels contiennent $32^3 = 32768$ cellules. Ces nombres de cellules correspondent au nombre maximum de cellules devant être mémorisées par histogramme. Comme le montre une analyse de la section 6.7, le nombre de cellules occupées est beaucoup plus petit et les histogrammes sont alors comprimés.

fonction de comparaison	$\cap$	$\cap_{we}$	$\cap_{mo}$	χ^2_v	χ^2_{qv}	χ^2_{av}	SSD	$maha$
Dx-Dy-32	99.5	29.0	65.7	95.7	100	97.5	88.9	99.0
Dx-Dy-Lap-32	100	56.6	16.7	97.5	100	98.7	95.5	99.7
Mag-Lap-32	99.8	43.4	30.6	91.9	99.5	99.5	96.7	92.7
Mag-G12-Lap-32	100	67.7	59.1	98.9	100	93.7	100	98.0

(a) reconnaissance par correspondance optimale

fonction de comparaison	$\cap$	$\cap_{we}$	$\cap_{mo}$	χ^2_v	χ^2_{qv}	χ^2_{av}	SSD	$maha$
Dx-Dy-32	2	17	4	14	1	6	6	1
Dx-Dy-Lap-32	1	10	14	7	1	3	3	4
Mag-Lap-32	2	21	9	10	2	2	2	7
Mag-G12-Lap-32	1	18	8	7	1	4	1	2

(b) nombre de correspondances nécessaires pour atteindre une reconnaissance de 100%

TAB. 6.3 – *Reconnaissance en présence de rotations d'image*

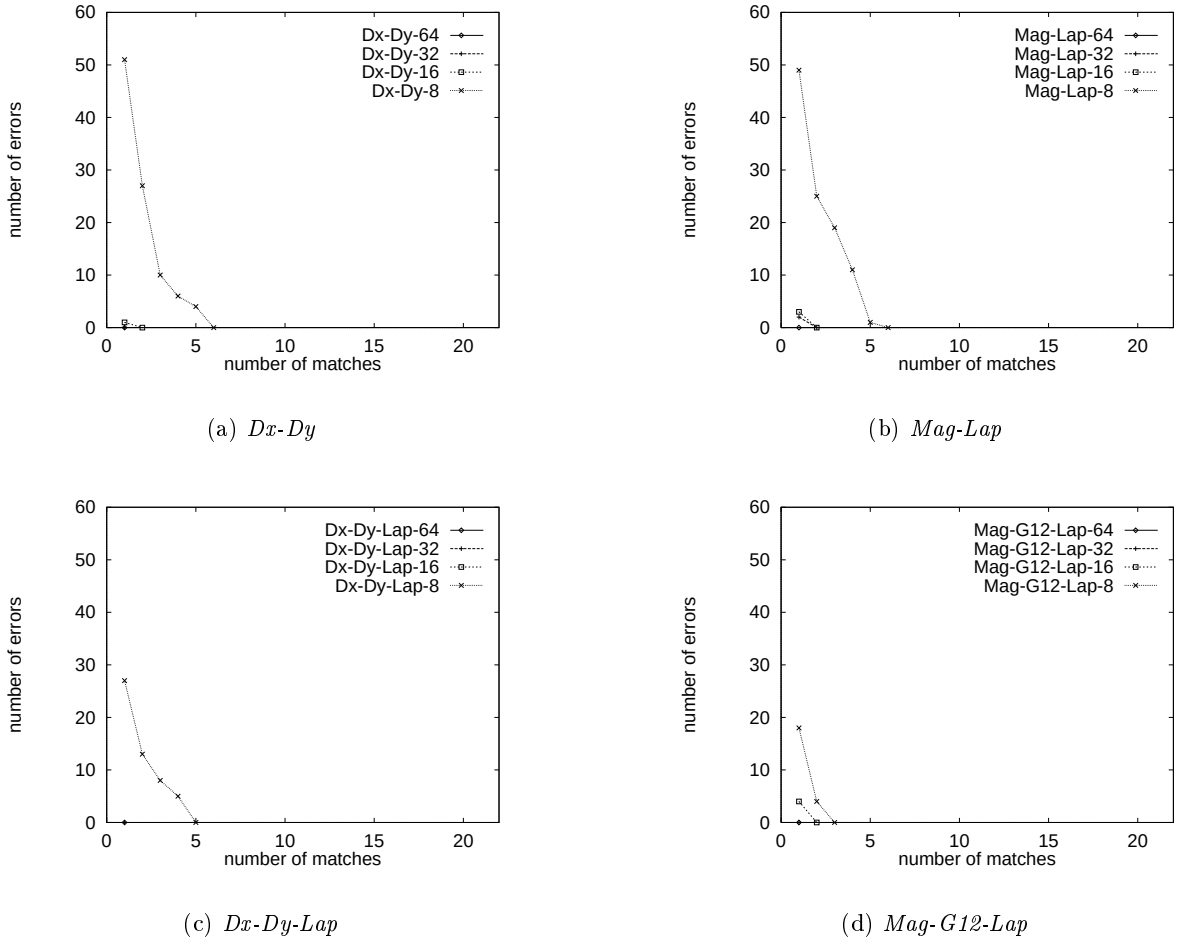
La table 6.3 montre les résultats obtenus pour les quatre combinaisons de filtres. La première table 6.3(a) montre le taux de reconnaissance avec la meilleure correspondance entre les histogrammes de la base et l'histogramme de l'image test. Les meilleurs résultats (presque toujours 100%) sont obtenus par la statistique χ^2_{qv}. La fonction d'intersection $\cap$ fournit une qualité de résultats légèrement inférieure. De très bons résultats sont donnés par la statistique χ^2_{av} et la distance de Mahalanobis *maha*. χ^2_v et SSD donnent des taux de reconnaissance autour de 90% ou plus élevés pour la plupart des cas. Les résultats des deux fonctions d'intersection $\cap_{we}$ et $\cap_{mo}$ ne sont pas acceptables.

Il est aussi intéressant d'examiner le nombre de correspondances nécessaires pour atteindre un taux de reconnaissance de 100%, ce qui est montré par la table 6.3(b). χ^2_{qv} et $\cap$ ont besoin deux correspondances au maximum pour la reconnaissance de la totalité des 396 images tests. SSD, *maha* et χ^2_{av} ont besoin d'un petit nombre de correspondances. Plus de correspondances sont nécessaires pour $\cap_{we}$, $\cap_{mo}$ et aussi χ^2_v. Malheureusement, χ^2_v a souvent besoin d'un grand nombre de correspondances pour atteindre une reconnaissance de 100% même si de bons résultats sont obtenus avec la première correspondance.

Dans la table 6.3 nous pouvons comparer les histogrammes bidimensionnels (*Dx-Dy* et *Mag-Lap*) aux histogrammes tridimensionnels (*Dx-Dy-Lap* et *Mag-G12-Lap*). Comme prévu, les résultats des histogrammes tridimensionnels sont meilleurs. Ceci s'explique par l'information sup-

plémentaire apportée par le troisième axe indépendant d'histogrammes.

Comme mentionné plus haut, il est aussi intéressant de comparer les histogrammes invariants à la rotation (*Mag-Lap* et *Mag-G12-Lap*) et les histogrammes variants à la rotation (*Dx-Dy* et *Dx-Dy-Lap*). Les résultats sont pratiquement identiques, indépendamment de l'invariance à la rotation. L'utilisation d'une fonction appropriée de comparaison d'histogrammes et (comme nous allons voir plus bas) la résolution des axes d'histogrammes sont plus importantes pour la reconnaissance. Malheureusement, cela n'est pas toujours vrai : dans le cas des images aériennes de Marseille (voir figures A.4 et A.5) les combinaisons de filtres variants à la rotation fournissent des résultats de reconnaissance nettement supérieurs à ceux de combinaisons invariantes. Dans ce cas la consommation supplémentaire de mémoire est justifiée. Le choix entre des filtres invariants et variants dépend fortement du contexte et de la base d'objets utilisée. Ce choix peut s'appuyer sur la transinformation comme introduit dans la section 4.2.3. La transinformation permet l'évaluation d'ensembles de mesures différentes et ainsi de différentes combinaisons de filtres dans le contexte de la reconnaissance d'objets.



(a) *Dx-Dy* (b) *Mag-Lap*

(c) *Dx-Dy-Lap* (d) *Mag-G12-Lap*

FIG. 6.3 – *Reconnaissance d'objets en présence de rotations d'image pour quatre combinaisons de filtres et quatre différentes résolutions*

Afin d'examiner l'influence de la résolution par axe d'histogramme nous avons utilisé un certain nombre de différentes résolutions. Pour chacune des quatre combinaisons de filtres, nous avons employé les résolutions de $8, 16, 32$ et 64 cellules par axe d'histogramme. La figure 6.3 montre le nombre d'erreurs — pour les mêmes 396 images tests — en fonction du nombre de correspondances. La statistique χ^2_{qv} est utilisée pour la comparaison d'histogrammes. Comme premier résultat, nous observons que pour les quatre combinaisons de filtres, la résolution 64 est suffisante pour atteindre une reconnaissance parfaite dès la première correspondance. Une résolution de 32 cellules est suffisante pour trois des quatre combinaisons (sauf pour *Mag-Lap*). La convergence de toutes les combinaisons est rapide. Utilisant une résolution de 16 cellules, les deux premières correspondances sont suffisantes pour reconnaître toutes les images tests. De plus, les convergences des histogrammes bidimensionnels sont très similaires. Aussi, les deux histogrammes tridimensionnels convergent de façon similaire. Néanmoins, les histogrammes tridimensionnels donnent des résultats de reconnaissance supérieurs et ils convergent plus rapidement.

6.3 Identification d'objets en présence de changements d'échelle

La section décrit des expérimentations d'identification d'objets en présence de changements d'échelle. Ces expérimentations montrent que l'approche peut traiter un facteur d'échelle de l'ordre de 2. Afin de traiter ces changements d'échelle, nous utilisons l'équivariance des dérivées Gaussiennes décrite dans la section 3.1.1. Des changements d'échelle supérieurs peuvent être traités par une pyramide d'images.

Afin de calculer des histogrammes de réponses de descripteurs locaux d'une échelle arbitraire, nous appliquons deux principes: premièrement, nous utilisons l'équivariance des dérivées Gaussiennes par rapport à l'échelle. Deuxièmement, nous adaptons le rayon de la région de support d'un histogramme en fonction de l'échelle. La propriété d'équivariance est décrite dans la section 3.1.1 et elle nous permet de calculer les réponses de dérivées Gaussiennes à une échelle arbitraire. Ayant donné une image $p(x, y)$, la dérivée Gaussienne de l'ordre n est définie par:

$$\frac{\partial^n}{\partial x^n} p(x, y) = G^{\sigma}_{x^n} \star p(x, y) \tag{6.99}$$

La dérivée de l'ordre n de l'image $f(x, y) = p(sx, sy)$ de l'échelle s peut être calculée à partir de l'image $p(x, y)$ en utilisant l'équation suivante (voir section 3.1.1 pour plus de détails):

$$\frac{\partial^n}{\partial x^n} f(x, y) \quad = \quad s^n G^{\sigma s}_{x^n}(x, y) \star p(x, y) \tag{6.100}$$

Il est alors suffisant d'adapter le calcul de la dérivée Gaussienne selon l'équation 6.100 pour obtenir la réponse correcte du filtre de l'image $f(x, y)$ fondé sur l'image $p(x, y)$ à une position d'image (x, y) et à une échelle arbitraire s. Afin de calculer l'histogramme des réponses de filtres Gaussiens d'un ensemble de positions d'image, il faut adapter les positions (x, y) dans l'image $p(x, y)$. Cela peut être fait par l'adaptation des distances entre les positions d'image, ce qui inclut des interpolations entre les pixels. Étant donné le volume de calcul élevé de l'interpolation, nous préférons utiliser les positions de pixels et adapter plutôt la région de support pour le

calcul d'histogrammes. L'équation $f(x, y) = p(sx, sy)$ indique qu'il faut multiplier le rayon de support par s. Cette adaptation de la région de support n'est pas lourde au point de vue calcul mais elle est moins précise (en particulier pour les petites $s < 1.0$). Les histogrammes correspondants de différentes échelles sont alors calculés à partir de différentes régions de support et ils contiennent des nombres différents d'entrées. Afin de rendre les histogrammes comparables, le nombre d'entrées de chaque histogramme doit être normalisé par une constante pré–définie.

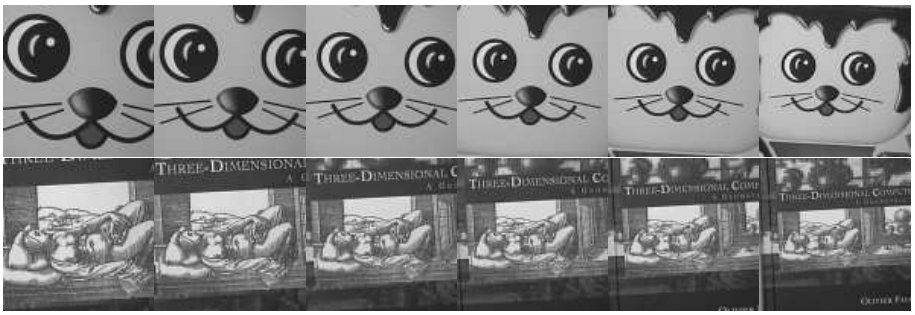


FIG. 6.4 – *Six différentes échelles de deux objets de la base de 30 objets*

La base de données contient 30 objets montrés dans la figure A.1. Pour chaque objet, six images de six différentes échelles ont été prises. Le facteur d'échelle entre la première et la dernière image est approximativement 2.1. Les six images de deux objets sont montrées dans la figure 6.4. Nous avons choisi une image par objet (le troisième à gauche de la figure 6.4) et nous avons calculé des histogrammes bidimensionnels *Dx-Dy-64* : premières dérivées Gaussiennes dans les directions x et y d'une résolution de 64 par axe d'histogramme. La déviation standard de dérivées Gaussiennes était $\sigma = 2.0$. L'ensemble d'images tests contient l'ensemble des autres images correspondant à 5 échelles différentes : $30 \times 5 = 150$ images. Pour chaque image test, nous avons calculé six différents histogrammes de réponses de filtres d'échelles différentes. Les six échelles sont données par $\sigma = 1.23, 1.45, 1.7, 2.0, 2.35$ et 2.76, ce qui correspond à un changement d'échelle de 15% entre chaque échelle. Le facteur d'échelle entre l'échelle la plus petite et la plus grande est alors 2.2. Le rayon de la région de support d'histogrammes était 40σ et varie alors entre 49.2 et 94 pixels.

fonction de comparaison	$\cap$	χ^2_v	χ^2_{qv}	χ^2_{av}	SSD	$maha$
reconnaissance	99.3	96.7	100	98.0	92.0	98.0
nombre de correspondances pour un taux de 100%	2	5	1	2	12	5

TAB. 6.4 – *Résultats de reconnaissance en présence de changements d'échelle pour 30 objets. Les résultats sont obtenus pour l'histogramme bidimensionnel Dx-Dy-64*

La table 6.4 montre les résultats expérimentaux. La statistique χ^2_{qv} obtient un taux de reconnaissance de 100% pour les 150 images tests. Le deuxième meilleur résultat est obtenu par la fonction d'intersection $\cap$, ayant une erreur. Cette erreur est reconnue par la deuxième correspondance. De très bons résultats sont fournis par χ^2_{av} et $maha$ d'un taux de reconnaissance de 98%. Un résultat légèrement inférieur est obtenu par χ^2_v. SSD n'obtient pas un résultat satisfaisant dans cette expérimentation.

6.4 Identification d'objets en présence de changements d'échelle et de rotations d'image

Les deux sections précédentes 6.2 et 6.3 ont décrit des expérimentations de l'identification d'objets en présence de changements d'échelle et de rotations d'image séparément. Cette section décrit deux expérimentations de l'identification d'objets en présence de changements d'échelle *et* de rotations d'image combinés. La première expérimentation utilise les 22 objets de l'expérimentation de rotations d'image de la section 6.2. La deuxième expérimentation emploie les bases de données des deux sections précédentes ensemble, c'est à dire que la base de données contient 52 objets.

La première expérimentation utilise les 22 objets de l'expérimentation de rotations d'image de la section 6.2 (voir figure A.2). De plus, aux 18 différentes rotations de chacun des 22 objets, nous avons pris deux images (par objet et rotation) d'une autre échelle. Le facteur d'échelle était approximativement ± 1.75. Pour chacun des 22 objets et pour chacune des 18 rotations par objet, la base d'images contient trois différentes échelles qui sont montrées dans la figure 6.5. Le nombre d'images est alors $22 \times 18 \times 3 = 1188$.

(a) petite échelle (b) image de la base de données (c) grande échelle

FIG. 6.5 – *Les trois échelles de la première expérimentation de l'identification d'objets en présence de changements d'échelle et de rotations d'image. Le changement d'échelle était de l'ordre de* ± 1.75. *Le facteur entre la petite et la grande échelle est approximativement 3.*

Pour le calcul de la base d'histogrammes, nous utilisons une image par objet d'échelle moyenne d'une rotation arbitraire d'objet. Utilisant cette image d'un objet nous calculons $18 \times 3 = 54$ histogrammes bidimensionnels correspondant à 18 différentes rotations (plus précisément $\alpha = 0°, 20°, \ldots, 340°$) à trois différentes échelles. Les différentes échelles ont été $\sigma = 1.7, 3.0$ et 5.2. La base d'histogrammes contient alors $22 \times 54 = 1188$ histogrammes. La combinaison de filtres est *Dx-Dy-64* (premières dérivées Gaussiennes dans les directions x et y d'une résolution de 64 cellules par axe d'histogrammes).

Comme ensemble d'images tests, nous utilisons toutes les images d'échelles différentes à la base de données: $22 \times 18 \times 2 = 792$ images. Pour chaque image test nous calculons un histogramme *Dx-Dy-64* avec $\sigma = 3.0$. Les résultats expérimentaux sont montrés dans la table 6.5. La table montre que χ^2_{qv} obtient un taux de reconnaissance de 95.7% en utilisant la première correspondance. En utilisant aussi la deuxième correspondance, un taux de reconnaissance de 100% est fourni. En appliquant $\cap$ comme fonction de comparaison, on obtient un taux de reconnaissance de 91.2% et un taux de 100% est atteint par l'inclusion des premières 6 correspondances.

Il est intéressant de noter la différence entre la reconnaissance d'images de petite échelle et d'images de grande échelle. Comme les changements d'échelle considérés sont importants (le facteur d'échelle est approximativement ± 1.75), la reconnaissance est difficile. Les images

fonction de comparaison	$\cap$	χ^2_{qv}
reconnaissance	91.2	95.7
nombre de correspondances pour un taux de 100%	6	2

TAB. 6.5 – *Résultats de reconnaissance en présence de changements d'échelle et de rotations d'image de 22 objets. Les résultats sont obtenus pour l'histogramme bidimensionnel Dx-Dy-64*

de grande échelle sont typiquement plus faciles à reconnaître que les images de petite échelle. Dans cette expérimentation particulière aucune erreur n'était obtenue pour les images de grande échelle. Cela peut être expliqué par la déviation standard σ plus petite qu'il faut utiliser pour les dérivées Gaussiennes. Ces petits σ augmentent la sensibilité au bruit de discrétisation et réduisent la région de support des histogrammes correspondants. En utilisant des σ plus larges la sensibilité peut être réduite. Par contre, des détails pouvant être importants pour la reconnaissance d'objets sont perdus.

fonction de comparaison	$\cap$	χ^2_{qv}
reconnaissance	98.1	99.1
nombre de correspondances pour un taux de 100%	8	2

TAB. 6.6 – *Résultats expérimentaux pour 52 objets en présence de changements d'échelle et de rotations d'image. Les résultats sont montrés pour l'histogramme bidimensionnel Dx-Dy-64*

La table 6.6 montre les résultats de la deuxième expérimentation de l'identification d'objets en présence de changements d'échelle et de rotations d'image. Dans cette expérimentation toutes les images d'expérimentations des sections 6.2 et 6.3 sont utilisées. La base d'images contient alors 52 objets avec un facteur d'échelle de 2 pour 30 objets et 18 rotations pour 22 objets. Pour le calcul de la base d'histogrammes nous utilisons une image pour chacun des 52 objets. Pour chacune des images nous calculons des histogrammes *Dx-Dy-64* correspondant à 18 différentes rotations d'image : $\alpha = 0°, 20°, 40°, \ldots, 340°$. Chaque objet est alors représenté par 18 histogrammes. La base d'histogrammes contient alors $52 \times 18 = 936$ histogrammes.

Comme images tests nous utilisons 18 différentes rotations d'image pour 22 objets et 6 différentes échelles pour 30 objets. Le nombre d'images tests est alors $22 \times 18 + 30 \times 6 = 576$. Pour chaque image test nous calculons six différents histogrammes *Dx-Dy-64* à différentes échelles. Comme dans la section 6.3, les déviations standards suivantes sont employées : $\sigma = 1.23, 1.45, 1.7, 2.0, 2.35$ et 2.76. La table 6.6 montre un taux de reconnaissance de 99.1% par χ^2_{qv}. En incluant la deuxième correspondance, l'algorithme reconnaît toutes les images tests correctement. En utilisant la fonction d'intersection $\cap$, un taux de reconnaissance de 98% est obtenu. 8 correspondances sont nécessaires pour atteindre une reconnaissance de 100%.

6.5 Identification d'objets en présence de changements de point de vue

Cette section décrit des expérimentations pour évaluer la robustesse des histogrammes multidimensionnels de champs réceptifs en présence de changements de point de vue en utilisant la base d'images de Columbia [Mur 95]. La base d'images contient 20 objets et 72 différents points

de vue par objet. La base d'images contient seulement des points de vue d'un *cercle de points de vue* plutôt que d'une *sphère de points de vue*. Néanmoins nous sommes convaincus que les résultats de cette section se généralisent au cas d'une sphère de points de vue. Dans les expérimentations, les performances de différentes combinaisons de filtres, de différentes fonctions de comparaison d'histogrammes et de paramètres de design des histogrammes multidimensionnels de champs réceptifs sont examinées.

FIG. 6.6 – *La base de 20 objets de Columbia*

La base d'images de Columbia contient 20 objets (figure 6.6) vus sous 72 différents angles de vue. La différence entre deux angles de vue voisins est 5° (figure 6.7). La base d'images contient alors $20 \times 72 = 1440$ images. Classiquement [Mur 95, Sch 96i], la moitié de ces images est utilisée comme base de données et l'autre moitié est employée comme ensemble test. Dans ce cas, l'angle entre les points de vue de la base de données est $\Delta\beta = 10°$.

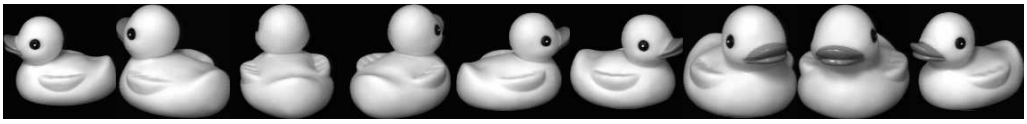


FIG. 6.7 – *9 de 72 rotations 3D d'un objet de la base d'images de Columbia*

Les expérimentations de cette section démontrent que la comparaison d'histogrammes est relativement robuste aux changements de points de vue (rotations 3D). Les expérimentations montrent aussi que nous pouvons utiliser une basse résolution pour chaque axe d'histogramme multidimensionnel et obtenir des taux de reconnaissance élevés. En particulier, la section donne une indication sur le nombre de points de vue nécessaires afin de reconnaître un objet d'un point de vue arbitraire. Un résultat de "graphes d'aspects" [1] est que le nombre de points de vue dépend fortement des objets utilisés. Les résultats obtenus dans cette section doivent être considérés alors comme une indication et non pas comme la solution générale à la modélisation d'objets 3D à partir d'images 2D. Ce problème est discuté dans le contexte de la classification d'objets où des points de vue similaires d'objets sont regroupés automatiquement (voir chapitre 9).

Les expérimentations examinent, en particulier, la dépendance de la robustesse aux rotations 3D et les paramètres de design des histogrammes. Les paramètres de design déterminent la discrimination d'objets (particulièrement le nombre de dimensions d'histogramme). Un paramètre important est la résolution d'histogrammes, c'est à dire le nombre d'intervalles discrets utilisé pour chaque dimension d'histogramme. La réduction de la résolution augmente la stabilité des

1. aspect–graphs

histogrammes mais diminue aussi la discrimination entre les objets.

La section décrit deux différentes séries d'expérimentations : la première série examine la relation entre le taux de reconnaissance et la fonction de comparaison d'histogrammes, la combinaison de filtres et la résolution d'histogrammes. Une deuxième série montre la dépendance entre la reconnaissance et le nombre de points de vue (et ainsi l'angle entre les différents points de vue) de la base de données.

Résultats de la première série d'expérimentations

La table 6.7 montre les résultats de la première série d'expérimentations avec les combinaisons de filtres suivantes : *Dx-Dy*, *Mag-Lap* et *Dx-Dy-Lap*. Chaque combinaison de filtres est calculée pour des résolutions différentes d'axes d'histogramme. La résolution varie entre 2 et 64 cellules par axe d'histogramme. Huit différentes fonctions de comparaison d'histogrammes ont été utilisées : trois fonctions d'intersection ($\cap$, $\cap_{we}$ et $\cap_{mo}$, voir section 5.1.1), trois statistiques χ^2 (χ^2_v, χ^2_{qv} et χ^2_{av}, voir section 5.1.3) et deux distances quadratiques (SSD et $maha$, voir section 5.1.2).

filtres	résolution	$\cap$	$\cap_{we}$	$\cap_{mo}$	χ^2_v	χ^2_{qv}	χ^2_{av}	SSD	$maha$
Dx-Dy	64	100	89.17	90.14	99.44	100	97.5	94.58	100
Dx-Dy	32	100	91.11	92.63	99.44	100	98.47	97.36	99.72
Dx-Dy	16	99.86	94.31	88.89	99.44	100	99.72	97.78	99.86
Dx-Dy	8	99.44	92.5	81.11	99.44	99.72	99.58	96.81	99.58
Dx-Dy	4	93.19	85.42	56.25	98.61	99.31	98.61	90.83	98.47
Dx-Dy	2	75.14	77.08	33.75	77.64	77.50	77.08	76.11	77.36
Mag-Lap	64	99.58	68.61	77.5	99.17	99.72	99.44	99.56	99.72
Mag-Lap	32	99.58	59.58	81.25	98.14	99.86	99.86	96.38	99.44
Mag-Lap	16	99.03	56.53	76.81	99.58	99.72	99.58	95.69	99.86
Mag-Lap	8	97.92	70.28	74.31	99.44	99.72	99.72	93.33	99.03
Mag-Lap	4	80.42	79.86	57.08	94.72	94.31	93.75	78.89	92.36
Mag-Lap	2	60.69	61.25	37.92	67.78	68.33	67.92	60.42	69.03
Dx-Dy-Lap	16	100	98.61	97.5	100	100	99.86	99.31	99.86
Dx-Dy-Lap	8	99.72	97.5	91.81	99.86	100	99.86	96.39	99.58
Dx-Dy-Lap	4	96.53	94.03	72.64	98.89	99.31	99.31	93.89	99.31
Dx-Dy-Lap	2	93.47	96.67	58.75	94.58	94.86	94.17	91.94	96.25

TAB. 6.7 – *Résultats de reconnaissance de la première série d'expérimentations*

La table 6.7 montre un taux de reconnaissance de 100% en utilisant, par exemple, l'histogramme bidimensionnel *Dx-Dy-16* et la fonction de comparaison d'histogrammes χ^2_{qv}. Un taux de 100% est aussi obtenu par l'histogramme tridimensionnel *Dx-Dy-Lap-8* en utilisant χ^2_{qv}.

Les résultats de la combinaison de filtres *Dx-Dy* sont légèrement meilleurs que les résultats de *Mag-Lap*. Cette différence est due au fait que *Dx-Dy* code des informations de rotations et *Mag-Lap* est invariant à la rotation d'image. La table 6.7 montre des résultats meilleurs pour l'histogramme tridimensionnel *Dx-Dy-Lap* relative aux histogrammes bidimensionnels *Dx-Dy* et *Mag-Lap* pour la plupart des cas. En particulier, dans les cas de basses résolutions les résultats de

Dx-Dy-Lap sont nettement meilleurs. Par exemple, un taux de reconnaissance de 96% est fourni par *Dx-Dy-Lap-2* (en utilisant *maha*), où chaque histogramme contient seulement $2^3 = 8$ cellules. Le gain de reconnaissance peut être expliqué par le fait que *Dx-Dy-Lap* possède trois dimensions indépendantes et code alors plus d'informations que les histogrammes bidimensionnels.

La table 6.7 montre aussi les résultats de huit différentes fonctions de comparaison d'histogrammes. Les meilleurs résultats ont été obtenus par la statistique χ^2_{qv}. Les résultats de cette fonction de comparaison sont meilleurs que ceux de toutes les autres fonctions, montrant ainsi la capacité de cette fonction pour la comparaison d'histogrammes. Les deuxièmes meilleurs résultats sont fournis par la distance de Mahalanobis *maha*, par χ^2_v et par χ^2_{av}. Les différences entre ces trois fonctions sont relativement petites. La fonction d'intersection $\cap$ obtient aussi de bons résultats. Il est intéressant de noter que cette fonction fournit des résultats relativement meilleurs — comparés avec χ^2_{qv} par exemple — pour des résolutions hautes. En fait, les résultats de $\cap$ sont presque identiques aux résultats de χ^2_{qv} pour les hautes résolutions. Dans ces cas, le besoin théorique en mesures distribuées de façon éparse est satisfait et permet l'application de cette fonction (voir section 5.1.1). Malheureusement, les résultats pour les basses résolutions sont nettement inférieurs comme par exemple pour la résolution de 4 cellules. Des résultats assez bons sont obtenus par *SSD*. Les deux fonctions d'intersection pondérée $\cap_{we}$ et $\cap_{mo}$ fournissent des résultats nettement inférieurs.

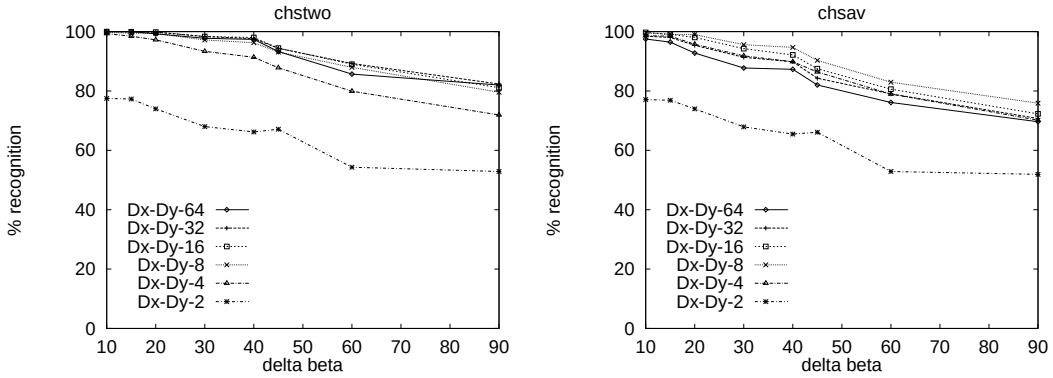
Un dernier paramètre, examiné dans ces expérimentations, est la résolution par axe d'histogrammes. Comme prévu, le taux de reconnaissance est augmenté pour les hautes résolutions. Comme la stabilité est diminuée par les résolutions plus hautes, il faut faire un compromis entre la possibilité de discriminer des objets et la stabilité. La table 6.7 montre que l'augmentation de la résolution à 16 cellules donne toujours un gain du taux de reconnaissance. Une augmentation de la résolution à 32 ou même à 64 diminue les taux de reconnaissance dans certains cas (par exemple pour *Dx-Dy* en utilisant χ^2_{av} ou *SSD*).

Robustesse par rapport aux changements de point de vue

Dans une deuxième série d'expérimentations nous avons varié l'angle entre les points de vue de la base de données. En particulier, nous avons utilisé $\Delta\beta = 10°, 15°, 20°, 30°, 40°, 45°, 60°$ et $90°$. Cette variation nous permet d'examiner la robustesse de la comparaison d'histogrammes par rapport aux changements de point de vue. A côté de $\Delta\beta$, nous avons aussi changé les paramètres suivants de la technique :

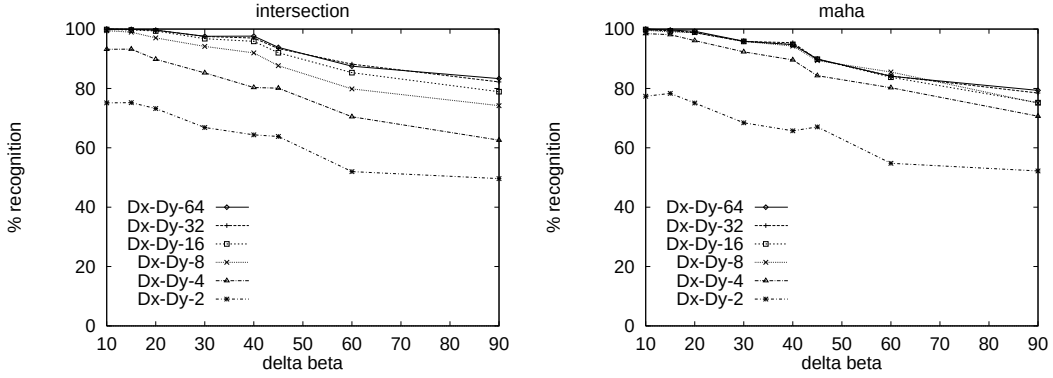
- la résolution de chaque axe d'histogramme variait entre 2 et 64 cellules
- deux combinaisons de filtres ont été utilisées : *Dx-Dy-Lap* et *Dx-Dy*
- quatre fonctions de comparaison d'histogrammes ont été appliquées : χ^2_{qv}, χ^2_{av}, $\cap$ et *maha*.

Les figures 6.8 et 6.9 montrent les taux de reconnaissance de différents $\Delta\beta$, de différentes résolutions, ainsi que de différentes fonctions de comparaison et de différentes combinaisons de filtres. La performance des histogrammes bidimensionnels *Dx-Dy* est moins bonne que celle des histogrammes tridimensionnels *Dx-Dy-Lap*. Même pour les grands $\Delta\beta$ et pour les basses résolutions, la combinaison de filtres *Dx-Dy-Lap* fournit des taux de reconnaissance élevés. Ce résultat indique que des taux de reconnaissance élevés peuvent être obtenus en ajoutant des



(a) χ^2_{qv}

(b) χ^2_{av}



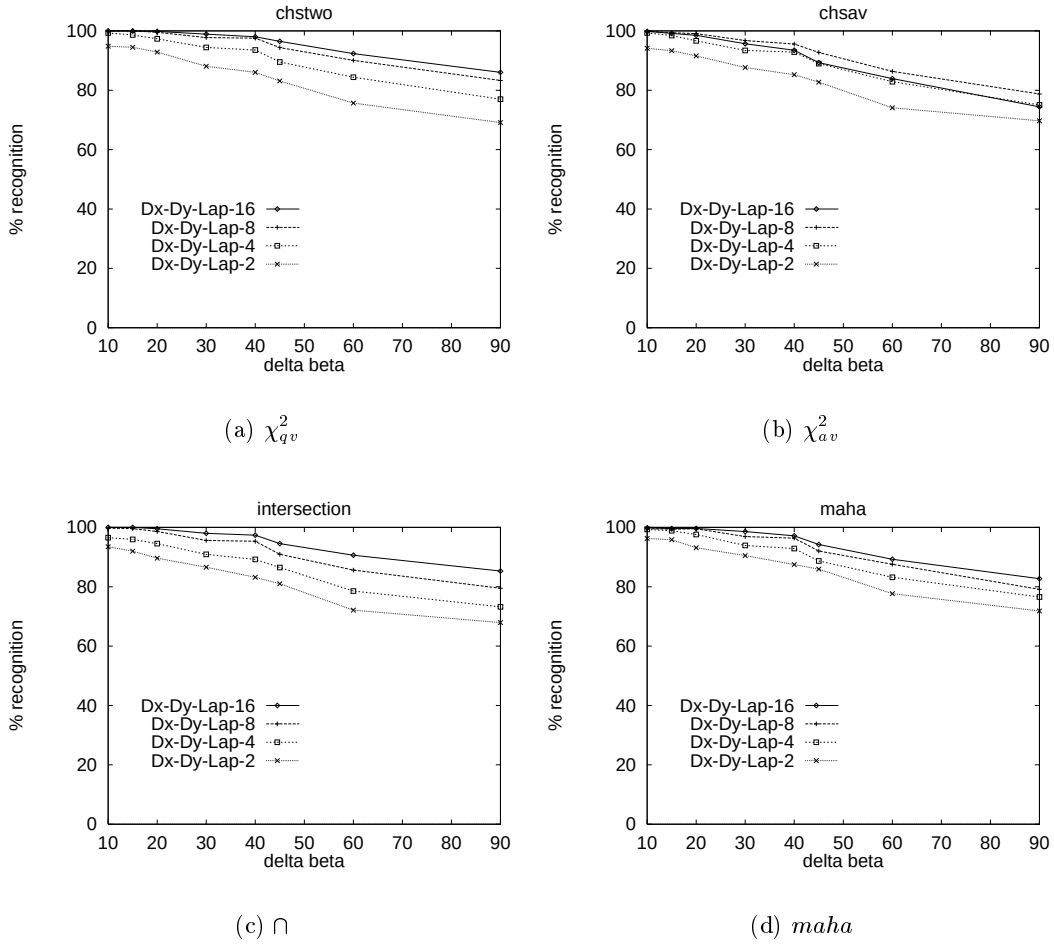
(c) $\cap$

(d) *maha*

FIG. 6.8 – *Base d'images de Columbia : histogrammes bidimensionnels Dx-Dy. Relation entre les taux de reconnaissance et $\Delta\beta$ (voir texte).*

dimensions supplémentaires aux histogrammes. L'augmentation de la consommation de mémoire peut être compensée en partie par une diminution de la résolution (voir section 6.7).

Les figures 6.8 et 6.9 montrent les taux de reconnaissance en fonction de l'angle $\Delta\beta$ entre les différents points de vue de la base de données. Les deux combinaisons de filtres, toutes les résolutions ainsi que l'ensemble de quatre fonctions de comparaison d'histogrammes montrent une dégradation lente des taux de reconnaissance avec l'augmentation de $\Delta\beta$. Cela indique la robustesse désirée de l'approche en présence de changements de points de vue. La dégradation la plus lente est donnée par χ^2_{qv}, suivie par la fonction d'intersection $\cap$. *maha* donne des résultats légèrement supérieurs aux résultats de χ^2_{av}.



(a) χ^2_{qv} (b) χ^2_{av}

(c) $\cap$ (d) *maha*

FIG. 6.9 – *Base d'images de Columbia : histogrammes tridimensionnels Dx-Dy-Lap. Relation entre les taux de reconnaissance et $\Delta\beta$ (voir texte).*

6.6 Identification d'objets en présence d'occultations partielles

La section décrit une expérimentation sur 103 objets. Plus précisément, nous utilisons les objets suivants :

- 20 objets de la base d'images de Columbia (figure 6.6)
- 30 objets de l'expérimentation en présence de changements d'échelle de la section 6.3 (figure A.1)
- 22 objets de l'expérimentation en présence de rotations d'image de la section 6.2 (figure A.2)
- 31 objets supplémentaires (figure A.3)

Dans les expérimentations des sections précédentes nous avons utilisé des histogrammes bidimensionnels et tridimensionnels de champs réceptifs. Dans le cas de 52 objets (section 6.4) nous avons employé, par exemple, des histogrammes bidimensionnels *Dx-Dy-64* qui ont fourni un taux de reconnaissance de 95.7%. Les résultats de la section précédente indiquent que les taux de reconnaissance peuvent être augmentés en ajoutant des dimensions indépendantes aux histogrammes. Cette section utilise des histogrammes à six dimensions de la combinaison de filtres *Dx-Dy-32* , à trois différentes échelles : $\sigma_1 = \sigma$, $\sigma_2 = 2\sigma$ et $\sigma_3 = 4\sigma$. Les résultats, décrits plus bas, montrent que ces histogrammes contiennent des informations suffisantes pour reconnaître toutes les images tests des 103 différents objets.

Pour la base d'images de Columbia nous avons calculé $20 \times 36 = 720$ histogrammes, correspondant à 36 différents points de vue de 20 objets ($\sigma = 2.0$). Les points de vue sont pris tous les 10°. Pour les 52 objets (utilisés dans l'expérimentation en présence de changements d'échelle et de rotations d'image) nous avons calculé $52 \times 18 \times 6 = 5616$ histogrammes correspondant à 18 différentes rotations d'image et à 6 différentes échelles. Nous avons utilisé les rotations d'image de $\alpha = 0°, 20°, 40°, \ldots, 340°$. Les différentes échelles de dérivées Gaussiennes ont été $\sigma = 1.48, 1.7, 2.0, 2.26, 2.62$ et 3.0. Pour chacun des derniers 31 objets, nous avons calculé un histogramme ($\sigma = 2.0$). La base d'histogrammes contient alors $720 + 5616 + 31 = 6367$ histogrammes.

L'ensemble d'images tests se compose des images suivantes : la deuxième moitié des 720 images de la base de Columbia, 6 différentes échelles pour 30 objets, 18 différentes rotations d'image de 22 objets et 31 images différentes pour les derniers 31 objets présentant des changements mineurs par rapport aux images de la base de données. L'ensemble d'images tests contient alors $720 + 30 \times 6 + 22 \times 18 + 31 = 1327$ images. Pour chaque image test nous calculons un seul histogramme à six dimensions avec $\sigma = 2.0$.

La figure 6.10 montre les taux de reconnaissance obtenus par deux fonctions de comparaison : χ^2_{qv} et $\cap$. Les taux de reconnaissance sont montrés en fonction de la partie visible de l'objet. Comme les objets sont centrés à l'intérieur des images, nous calculons les histogrammes d'une région de support au centre d'image. Ce choix correspond au cas idéal où la position de l'objet est connue approximativement.

La figure 6.10 montre les taux de reconnaissance de 100% obtenus pour les deux fonctions de comparaison en utilisant l'objet entier comme région de support du calcul d'histogrammes. En utilisant seulement 62% de l'objet, la fonction d'intersection obtient encore un taux de 100%.

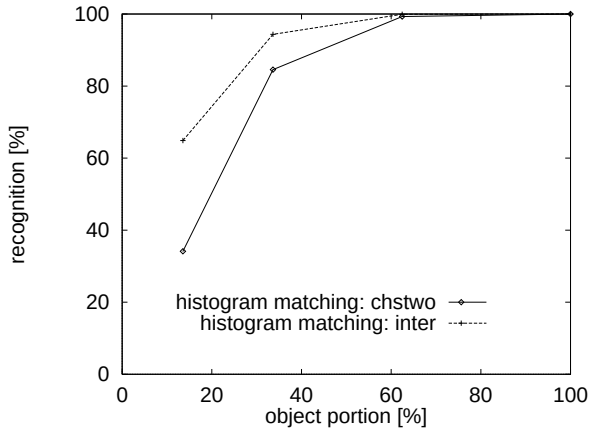


FIG. 6.10 – *Identification d'objets en présence d'occultations partielles pour 1327 images tests de 103 différents objets*

Dans ce cas, χ^2_{qv} fournit un taux de 99.3%. Dans le cas d'une visibilité de 33% de l'objet, $\cap$ obtient encore une reconnaissance de 94%. χ^2_{qv} fournit un taux de 84% dans ce cas. Cette expérimentation montre, en particulier, une robustesse attendue de la fonction d'intersection $\cap$ par rapport aux occultations partielles.

Même si une portion visible d'objet de 62% est suffisante pour la reconnaissance de 103 objets, l'expérimentation correspond au cas idéal: un seul objet est considéré par image et les histogrammes ont été centrés sur les objets. Afin de reconnaître les objets dans des scènes complexes de plusieurs objets, nous pouvons calculer les histogrammes des sous–fenêtres de l'image qui correspondent à certaines portions de l'objet. En général, cette approche exige de lourds calculs et peut être inadaptée pour des objets de forme arbitraire. Le chapitre 7 développe une approche de reconnaissance d'objets utilisant seulement un petit nombre de vecteurs de mesures, choisis arbitrairement dans l'image. Les résultats de reconnaissance de la figure 6.10 vont servir à la comparaison de la reconnaissance par comparaison d'histogrammes et de l'algorithme de reconnaissance probabiliste proposé dans le chapitre 7.

6.7 Consommation de mémoire pour des histogrammes multidimensionnels de champs réceptifs

Dans les sections précédentes, nous avons établi que l'augmentation du nombre de dimensions d'histogrammes multidimensionnels de champs réceptifs donne typiquement de meilleurs résultats de reconnaissance. Cette section discute brièvement une limite supérieure de la consommation de mémoire par les histogrammes multidimensionnels. Par ailleurs, nous décrivons la consommation de mémoire pour deux cas: pour les 20 objets de la base d'images de Columbia (voir figure 6.6) et pour l'expérimentation de la section 6.6.

La consommation maximale de mémoire des histogrammes multidimensionnels est donnée par la résolution R par axe d'histogramme et par le nombre de dimensions L de l'histogramme: R^L. Le nombre maximal varie de façon exponentielle avec le nombre de dimensions et implique que l'on se penche plus en détail sur la question de la consommation de mémoire.

Une limite supérieure est donnée par la taille de la région de support d'un histogramme. Dans le cas des images de taille 128^2 pixels de la base de Columbia, la limite supérieure est par exemple de l'ordre de $128^2 = 16384$. Cela est vrai simplement parce que chaque position de pixels correspond à un seul vecteur de mesures. En supposant que chaque vecteur de mesures apparaisse exactement une seule fois dans l'image, nous obtenons 16384 différents vecteurs de mesures qui doivent être représentés. La consommation de mémoire possède alors la taille d'image comme limite supérieure.

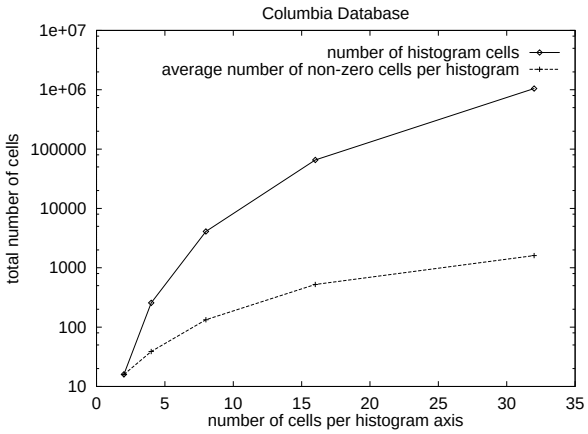


FIG. 6.11 – *Nombre de cellules non-nulles des histogrammes à quatre dimensions*

En réalité, les vecteurs de mesures apparaissent plus d'une fois dans une image. Le graphe en haut de la figure 6.11, par exemple, montre le nombre théorique de cellules R^L. Le graphe en bas de cette figure 6.11 montre le nombre moyen de vecteurs différents de mesures (= nombre de cellules d'histogrammes non-nulles) en fonction de la résolution R. La base d'images employée est la base de Columbia de $20 \times 72 = 1440$ images. La combinaison de filtres était *Dx-Dy* à deux différentes échelles : $\sigma = 2.0$ et 4.0. Les histogrammes correspondants sont alors à quatre dimensions. La figure 6.11 montre que le nombre moyen de différentes cellules d'histogrammes est nettement plus petit que la taille théorique des histogrammes. Nous utilisons ce fait pour comprimer la représentation d'histogrammes par un facteur significatif.

Dans la section 4.1 nous avons analysé la consommation de mémoire pour l'expérimentation décrite dans la section 6.6. Le nombre moyen de cellules était approximativement 5000. La figure 4.2 montre le nombre de cellules d'histogrammes en fonction de leur occurrence dans l'image. Les histogrammes à six dimensions employées dans les expérimentations de la section 6.6 sont les histogrammes les plus coûteux en niveau mémoire de cette thèse. Typiquement, le nombre moyen de cellules d'histogrammes non-nulles est nettement plus petit montré par la figure 6.11.

Ces deux exemples démontrent que la consommation de mémoire des histogrammes multidimensionnels de champs réceptifs est de l'ordre de quelques kilo-octets. Si plusieurs rotations d'image, plusieurs échelles et plusieurs points de vue doivent être mémorisés par objet, la consommation de mémoire par objet est de l'ordre d'un 1 mega-octet. C'est à dire que des algorithmes de réduction automatique de la représentation sont souhaitables. De tels algorithmes sont proposés dans le chapitre 9, dans le contexte de la classification d'objets.

6.8 Conclusion

Les expérimentations de ce chapitre ont démontré l'applicabilité de la comparaison d'histogrammes pour l'identification d'objets en présence de changements d'échelle, de rotations d'image, de variations de point de vue et d'occultations partielles.

A partir de ces expérimentations, nous pouvons conclure les points suivants :

- l'augmentation du nombre de caractéristiques locales et ainsi l'augmentation du nombre de dimensions des histogrammes multidimensionnels de champs réceptifs permet l'augmentation des taux de reconnaissance
- une bonne reconnaissance peut être obtenue même dans le cas de basse résolution d'axes d'histogrammes. Néanmoins, l'augmentation de la résolution donne typiquement une meilleure reconnaissance. Une résolution de 32 était suffisante dans la plupart des cas
- la meilleure fonction de comparaison était χ^2_{qv}. Cette fonction donne des taux de reconnaissance élevés et aussi le nombre le plus petit permettant d'atteindre un taux de reconnaissance de 100%
- de bons résultats sont aussi fournis par la fonction d'intersection $\cap$. En particulier en présence d'occultations partielles, cette fonction donne des résultats meilleurs que χ^2_{qv}
- de bons résultats sont donnés par χ^2_{av} et *maha*
- la comparaison d'histogrammes est robuste par rapport aux changements de point de vue et une dégradation lente des taux de reconnaissance est reportée dans la section 6.5
- le nombre de cellules d'histogrammes occupées est de l'ordre de 10^3 dans le cas d'histogrammes à hautes dimensions de champs réceptifs.

Dans le cas d'occultations partielles, la fonction d'intersection $\cap$ fournit des résultats meilleurs que χ^2_{qv}. Un taux de reconnaissance de 100% est obtenu pour 1327 images tests de 103 objets en utilisant une portion visible de 62% (voir section 6.6). Néanmoins, il est souhaitable d'obtenir une robustesse plus haute en présence d'occultations partielles. Ce souhait justifie le développement d'un algorithme de reconnaissance probabiliste dans le chapitre 7 suivant. Cet algorithme est capable de calculer la probabilité de chaque objet à partir d'un petit nombre de vecteurs de mesures choisis arbitrairement dans l'image.

Chapitre 7

Reconnaissance probabiliste d'objets

Les chapitres précédents ont montré que les histogrammes multidimensionnels de champs réceptifs constituent un moyen fiable pour la représentation statistique d'objets. En particulier, les *fonctions de comparaison* d'histogrammes ont été appliquées pour l'identification d'objets en présence de changements d'échelle et d'orientation. La robustesse de la comparaison d'histogrammes par rapport aux variations du point de vue a été démontrée. Dans les chapitres précédents, la technique n'a pas pris en compte les *occultations partielles* d'une façon spécifique. Néanmoins, en utilisant les histogrammes multidimensionnels, une certaine robustesse aux occultations partielles a été observée (voir section 6.6).

Pour le traitement approprié des occultations partielles, ce chapitre développe une méthode de reconnaissance fondée sur des vecteurs de mesures locales sélectionnés arbitrairement dans l'image. Ces vecteurs permettent le calcul des probabilités de présence de chaque objet de la base. Une propriété remarquable de cet algorithme est son indépendance vis à vis de la mise en correspondance de l'image test et de la base d'objets.

La section 7.1 introduit une méthode pour le calcul de probabilités d'objets utilisant un certain nombre de vecteurs de mesures locales. Une expérimentation sur une base de 52 objets en présence de changements en rotation et en échelle est décrite (section 7.2). Les bases de données et de tests utilisées dans la section 6.6 sont employées par la section 7.3. Les 1327 images tests de 103 objets incluent des variations d'échelle, des rotations dans le plan image et des changements du point de vue. L'utilisation des histogrammes à six dimensions de la section 6.6 permet de comparer la *reconnaissance par comparaison d'histogrammes* à la *reconnaissance probabiliste* proposée dans ce chapitre. Les résultats sont donnés en fonction de la partie visible

d'objet pour montrer la robustesse de l'algorithme probabiliste par rapport aux occultations partielles. La section 7.4 propose une extension de la méthode pour la reconnaissance de plusieurs objets dans des scènes complexes.

7.1 Reconnaissance d'objets sans correspondance

La reconnaissance probabiliste consiste à calculer la probabilité d'un objet o_n à partir d'une région R de l'image : $p(o_n|R)$. Dans notre contexte, la région la plus petite possible contient un seul vecteur m_k de mesures locales. La probabilité $p(o_n|m_k)$ peut être calculée par la règle de Bayes :

$$p(o_n|m_k) = \frac{p(m_k|o_n)p(o_n)}{p(m_k)} = \frac{p(m_k|o_n)p(o_n)}{\sum_i p(m_k|o_i)p(o_i)} \tag{7.101}$$

avec

- $p(o_n)$ la probabilité *a priori* de l'objet o_n,
- $p(m_k)$ la probabilité *a priori* du vecteur m_k de mesures locales (= combinaison de sorties de filtres),
- $p(m_k|o_n)$ la densité probabiliste de l'objet o_n. Cette densité peut être estimée en normalisant les histogrammes multidimensionnels de champs réceptifs de l'objet o_n par leurs tailles.

Dans la plupart de cas, un seul vecteur de mesures n'est pas suffisant pour la reconnaissance d'objets. En employant deux vecteurs m_k et m_j du même objet o_n, la probabilité de l'objet o_n peut être calculée :

$$p(o_n|m_k \wedge m_j) = \frac{p(m_k \wedge m_j|o_n)p(o_n)}{\sum_i p(m_k \wedge m_j|o_i)p(o_i)} \tag{7.102}$$

En supposant l'*indépendance* de m_k et m_j on obtient :

$$p(o_n|m_k \wedge m_j) = \frac{p(m_k|o_n)p(m_j|o_n)p(o_n)}{\sum_i p(m_k|o_i)p(m_j|o_i)p(o_i)} \tag{7.103}$$

K vecteurs *indépendants* $m_1, m_2, \ldots, m_K$ permettent de calculer la probabilité de chaque objet o_n :

$$p(o_n|\bigwedge_k m_k) = \frac{p(\bigwedge_k m_k|o_n)p(o_n)}{\sum_i p(\bigwedge_k m_k|o_i)p(o_i)} \tag{7.104}$$

$$= \frac{\prod_k p(m_k|o_n)p(o_n)}{\sum_i \prod_k p(m_k|o_i)p(o_i)} \tag{7.105}$$

Dans notre contexte, un vecteur de mesures correspond à un vecteur multidimensionnel de champs réceptifs (par exemple, le vecteur bidimensionnel des premières dérivées Gaussiennes dans les directions x et y). Les K vecteurs m_k correspondent alors aux K vecteurs de champs réceptifs sélectionnés dans la même région R de l'image. Pour la validité de l'équation 7.105 tous les K vecteurs doivent provenir du même objet. Dans les expérimentations décrites plus bas un petit nombre de mesures – c'est à dire une petite portion visible d'objet – génère des hypothèse fiables. Cela permet de supposer en général que tous les K vecteurs de mesures proviennent réellement du même objet. Une segmentation objet–fond serait certainement utile, mais les résultats indiquent l'applicabilité de l'algorithme sans ce pré–traitement.

Les probabilités a priori $p(o_n)$ de l'occurrence de chaque objet o_n ne peuvent pas être déterminées à partir des histogrammes multidimensionnels de champs réceptifs. Ces probabilités dépendent du contexte et de l'environnement. Classiquement, elles sont constantes pour un certain contexte et un certain environnement. Pour les expérimentations de ce chapitre (sections 7.2 et 7.3) et des chapitres suivants, les objets sont supposés équiprobables. Les probabilités a priori sont alors données par $p(o_n) = \frac{1}{N}$, avec N le nombre d'objets. Cette supposition simplifie l'équation 7.105 :

$$p(o_n | \bigwedge_k m_k) \quad = \quad \frac{\prod_k p(m_k|o_n)}{\sum_i \prod_k p(m_k|o_i)} \tag{7.106}$$

Comme mentionné plus haut, la densité probabiliste $p(m_k|o_n)$ d'un objet o_n est donnée directement par les histogrammes multidimensionnels de champs réceptifs de l'objet o_n. Autrement dit, l'équation 7.106 calcule la probabilité de chaque objet o_n entièrement fondée sur les histogrammes multidimensionnels de N objets.

Le choix des positions de vecteurs dans l'image est arbitraire. La technique est alors rapide (seulement un certain nombre de vecteurs est calculé) et robuste aux occultations partielles (l'approche est strictement locale). De plus, la méthode ne dépend pas de la mise en correspondance de l'image test et de la base d'objets.

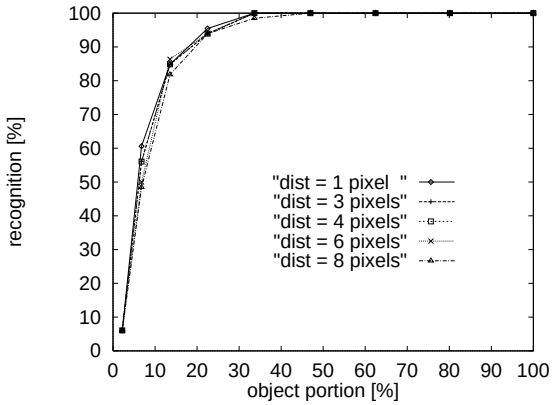
Les prochaines deux sections décrivent les expérimentations de l'application de l'équation 7.106 : la section 7.2 examine la reconnaissance de 52 objets en présence de rotations dans le plan image et de changements d'échelle. L'expérimentation de la section 7.3 augmente le nombre d'objets à 103 objets et considère des variations du point de vue pour 20 objets.

Évaluation expérimentale de paramètres d'implémentation

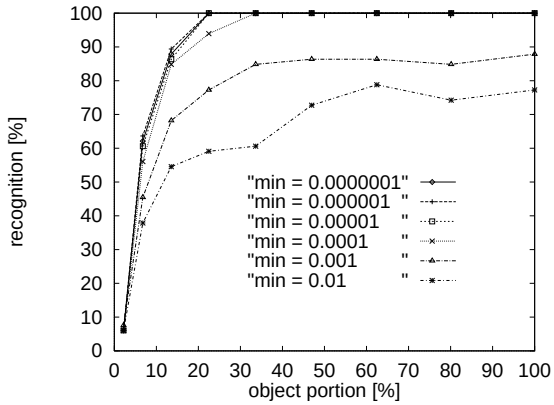
Une supposition importante de l'équation 7.106 est l'indépendance de K différents vecteurs m_k de mesures locales. Pour la satisfaire, une distance minimale $d(m_k, m_j)$ entre différents vecteurs doit être respectée. D'un point de vue du traitement du signal et étant donné l'enveloppe Gaussienne des dérivées Gaussiennes, la distance de $d(m_k, m_j) \geq 3\sigma$ est suffisante pour l'indépendance mutuelle de deux vecteurs m_k et m_j. Une distance de $d(m_k, m_j) \geq 2\sigma$ correspond à une dépendance faible des vecteurs de mesures.

Dans le contexte de la reconnaissance probabiliste, l'intérêt majeur est le calcul de probabilités d'une région locale d'image utilisant l'équation 7.106. Cette région doit être la plus petite possible. Il existe deux influences complémentaires : d'un côté, la supposition de l'indépendance requiert une certaine distance $d(m_k, m_j)$. D'un autre côté, l'augmentation de cette distance

diminue le nombre K de vecteurs de mesures pour une région locale choisie. Ce paragraphe examine l'influence de la distance $d(m_k, m_j)$ pour une région fixée.



(a) *différentes distances* $d(m_k, m_j)$, $p_{min} = 0.0001$ *constant*



(b) $d(m_k, m_j) = 4$ *pixels constant, différentes probabilités minimales* p_{min}

FIG. 7.1 – *Évaluation empirique de deux paramètres de l'approche : (a)* $d(m_k, m_j)$ *et (b)* p_{min}. *La base de données consiste en 22 objets et l'ensemble de test en 66 images*

La figure 7.1(a) montre l'influence de la distance $d(m_k, m_j)$ sur la reconnaissance de 22 objets. Pour chaque objet, plusieurs histogrammes à six dimensions sont calculés. Ces histogrammes correspondent à 18 rotations différentes et à 6 échelles différentes (pour les détails, voir section 7.2). La base d'histogrammes contient alors $22 \times 18 \times 6 = 2376$ histogrammes. Trois images de chaque objet d'une rotation dans le plan image ont été choisies arbitrairement comme base de test. La figure 7.1(a) montre le taux de reconnaissance en fonction de la portion visible de l'objet. Les différents graphes sont obtenus pour différentes distances $d(m_k, m_j)$ entre les vecteurs de mesures. Comme mentionné plus haut, le nombre total K de vecteurs d'une région fixée d'objet augmente avec la diminution de la distance $d(m_k, m_j)$. Les figures montrent les résultats des distances $d(m_k, m_j)$ de $1, 3, 4, 6$ et 8 pixels entre les vecteurs de mesures. Les distances différentes obtiennent des résultats pratiquement identiques. Cela peut être expliqué par un effet de compensation entre l'indépendance et le nombre K de vecteurs. Seul le résultat de la distance de 8 pixels est légèrement inférieur. Pour les expérimentations suivantes, la distance $d(m_k, m_j)$ minimale entre deux vecteurs de mesures m_k et m_j est fixée à 4 pixels (correspondant à 2σ).

L'influence d'un autre paramètre d'implémentation est montrée par la figure 7.1(b) : la probabilité minimale $p_{min} \leq p(m_k|o_n)$ utilisée dans l'équation 7.106. Cette probabilité minimale est nécessaire car l'estimation de probabilités $p(m_k|o_n)$ est sensible aux perturbations du signal pour les probabilités $p(m_k|o_n)$ petites. Ces perturbations sont liées aux variations de l'apparence d'objets. Pour rendre la technique robuste à ces perturbations, une probabilité minimale p_{min} est associée aux histogrammes multidimensionnels. La figure 7.1(b) montre le taux de reconnaissance en fonction de la portion visible pour les 66 images tests. Les différents graphes pour p_{min} entre 10^{-2} et 10^{-7} sont montrés. Les graphes justifient le choix de $p_{min} = 10^{-5}$ car des valeurs de p_{min} inférieures n'améliorent plus la reconnaissance. Par contre, les valeurs de p_{min}

plus petites augmentent la sensibilité aux perturbations du signal, qui n'est pas examinée dans la figure.

7.2 Identification d'objets en présence de changements d'échelle et de rotation

Une expérimentation décrit l'identification de 52 objets en présence de changements d'échelle et de rotations dans le plan image. La base d'objet et les images tests sont celles de la deuxième expérimentation de la section 6.4 : la figure A.2 montre 22 objets et la figure A.1 les 30 autres objets. La comparaison d'histogrammes de la section 6.4 est remplacée par l'application de l'équation 7.106 pour la reconnaissance probabiliste.

Cette section et la section suivante emploient la combinaison des filtres *Dx-Dy* à trois échelles : $\sigma_1 = 2.0, \sigma_2 = 4.0$ et $\sigma_3 = 8.0$. Chaque vecteur m_k de mesures et son histogramme possèdent six dimensions. Comme pour la section 6.4, l'orientabilité de dérivées Gaussiennes est utilisée pour calculer les réponses de filtres de rotations arbitraires. La combinaison de filtres est calculée pour 18 différentes rotations d'une distance de $\Delta\alpha = 20°$.

En utilisant la propriété de l'équivariance, les réponses de dérivées Gaussiennes d'échelles arbitraires peuvent être calculées à partir d'une seule image. Comme les images tests contiennent un changement d'échelle de facteur approximatif de 2.2, six différentes échelles sont calculées par objet (utilisant les σ de la section 6.4 : $\sigma_1 = 1.23, 1.45, 1.7, 2.0, 2.35$ et 2.76, et $\sigma_2 = 2\sigma_1$ et $\sigma_3 = 4\sigma_1$).

Pour les histogrammes correspondant aux différentes rotations, la région de support doit être circulaire. En opposition à une région circulaire, une région carrée – utilisant le rayon du cercle comme demi–longueur du carré – contient approximativement 20% en plus de vecteurs. Cela est avantageux pour les petits rayons utilisés plus bas. Heureusement, les imprécisions dues à l'utilisant de régions carrées n'interviennent que pour le bord d'objets. Dans cette expérimentation, les régions sont carrées, petites et centrées dans l'image. La taille de la région de support est limitée par la taille des images. Comme les histogrammes sont calculés pour différentes échelles d'objets, le rayon maximal possible de la région de support est $40\sigma_1$. Cela correspond à un rayon de 49 pixels (pour $\sigma_1 = 1.23$) et 110 pixels (pour $\sigma_1 = 2.76$). La région de support d'histogrammes varie d'un facteur de $5 \approx \frac{110^2}{49^2}$. Le choix consistant à centrer la région de support peut être vu comme une segmentation objet–fond pour l'apprentissage du modèle.

La base d'histogrammes est calculée à partir d'une seule image par objet et contient $52 \times 18 \times 6 = 5616$ histogrammes. Ces histogrammes correspondent à 18 différentes rotations et 6 différentes échelles de chacun des 52 objets.

Pour appliquer l'algorithme de reconnaissance probabiliste (équation 7.106), il faut choisir K vecteur m_k de mesures. Comme il l'est dit plus haut, l'équation dépend de deux suppositions : premièrement, tous les vecteurs doivent correspondre au même objet et deuxièmement, les K vecteurs sont supposés indépendants. La deuxième supposition est satisfaite par la distance fixe entre deux vecteurs de $2\sigma_1$. La première supposition est remplie par l'utilisation d'images tests qui contiennent un seul objet et par le choix de vecteurs de mesures d'une région centrale de l'image. Les résultats de cette expérimentation correspondent alors au cas idéal où tous les K vecteurs de mesures proviennent du même objet. En général, il n'existe pas de moyen trivial pour satisfaire la première supposition. Néanmoins, les résultats expérimentaux décrits plus

bas indiquent qu'une bonne hypothèse d'objet est obtenue à partir d'une petite portion visible d'objet. La section 7.4 définit une extension de l'algorithme pour reconnaître plusieurs objets dans les scènes complexes.

L'ensemble d'images tests est composé de 18 rotations différentes de 22 objets (figure A.2) et de 6 échelles différentes de 30 objets (figure A.1). L'ensemble test contient alors $22 \times 18 + 30 \times 6 = 576$ images. Comme introduit plus haut, les vecteurs de mesures sont choisis dans une région centrale et carrée de l'image. Les rayons (demi–longueur du carré) varient entre $1\sigma_1, 5\sigma_1, 10\sigma_1, 15\sigma_1, \ldots, 50\sigma_1$ pixels correspondant à 1, 25, 100, 225, ..., 2500 vecteurs de mesures. Le rayon maximal de $50\sigma_1$ couvre une région plus large que la région de support d'histogrammes de la base (qui ont été calculés avec un rayon de $40\sigma_1$). L'influence du fond – qui n'est pas modélisée d'une manière explicite – peut être analysée. La combinaison de filtres à six dimensions est donnée par *Dx-Dy* à trois échelles calculée pour $\sigma_1 = 2.0, \sigma_2 = 4.0$ et $\sigma_3 = 8.0$.

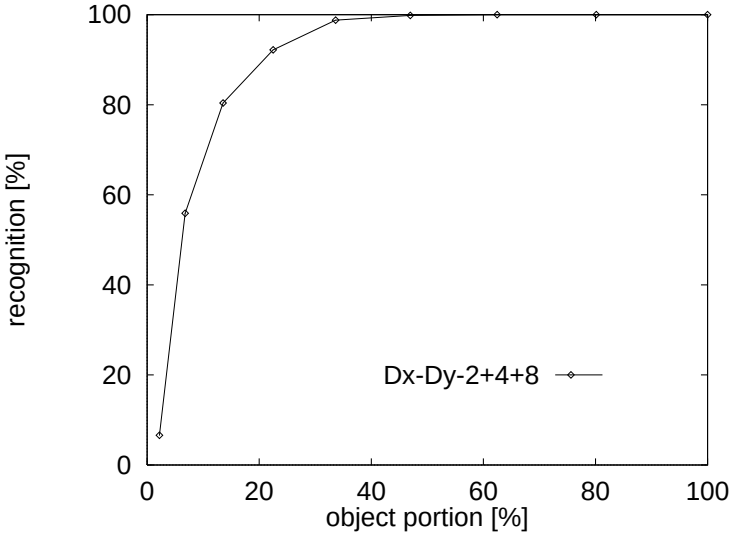


FIG. 7.2 – *Résultats expérimentaux pour 52 objets*

La figure 7.2 et la table 7.1 résument les taux de reconnaissance pour les 576 images tests. Le premier résultat est qu'une portion visible de 62% est suffisante pour la reconnaissance complète. Pour une portion visible de 34%, il n'y a que 7 erreurs correspondant à un taux de reconnaissance de presque 99%. Une portion visible de 13.5% (plus de 85% d'occultation) permet la reconnaissance de plus de 80% des images tests.

rayon [σ_1]	1	5	10	15	20	25	30	35	40	45	50
portion d'objet [%]	2.2	6.8	13.5	22.5	33.6	47.0	62.5	80.1	100.0	122.1	146.3
reconnaissance [%]	6.7	56.0	80.4	92.2	98.8	99.8	100	100	100	100	100
erreurs	538	254	113	45	7	1	0	0	0	0	0

TAB. 7.1 – *Résultats expérimentaux utilisant 52 objets*

La table 7.1 montre un taux de reconnaissance d'également 100% aussi pour les rayons de $45\sigma_1$ et $50\sigma_1$. Ces rayons correspondent à 122% et 146% de la région de support des histo-

grammes de la base. Dans ces cas, l'influence du fond est importante. Les taux de reconnaissance de 100% indiquent que l'algorithme peut reconnaître les objets même sans segmentation parfaite de l'objet du fond et sans modélisation explicite du fond.

En conclusion, les résultats expérimentaux démontrent la capacité de l'approche à reconnaître des objets en présence d'occultations partielles importantes. De plus, une petite portion de l'objet est suffisante pour obtenir une bonne hypothèse de l'objet.

7.3 Identification d'objets en présence d'occultations partielles

La section précédente a décrit une expérimentation de 52 objets en présence de rotations dans le plan image et de changements d'échelle. Pour considérer des variations du point de vue, cette section ajoute 20 objets de la base d'images de Columbia (voir figure 6.6). Pour chacun des 20 objets, 36 points de vue sont utilisés, avec une différence entre eux de 10°. La base d'objets entière contient 103 objets différents et correspond à la base de la section 6.6.

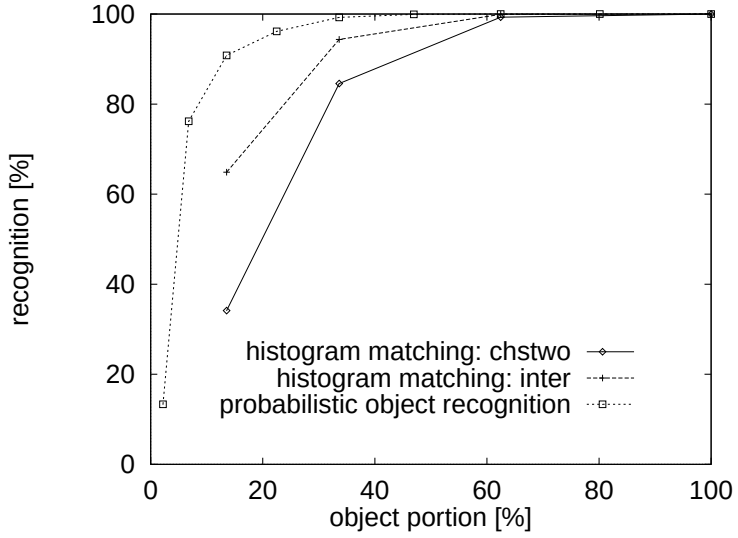


FIG. 7.3 – *Résultats expérimentaux pour 103 objets. Comparaison de la reconnaissance probabiliste à la reconnaissance par comparaison d'histogrammes : χ^2_{qv} (chstwo) et $\cap$ (inter)*

Les détails techniques (la combinaison de filtres, les σ_i, les rayons de la région de support d'histogrammes) de cette expérimentation sont ceux de la section 7.2 précédente. $52 \times 18 \times 6 = 5616$ histogrammes sont calculés pour les 52 objets utilisés précédemment. Pour les 36 points de vue d'objets de Columbia, $20 \times 36 = 720$ histogrammes sont calculés. Un histogramme pour chacun des 31 derniers objets est ajouté. Au total, la base d'histogrammes contient $5616 + 720 + 31 = 6367$ histogrammes.

Pour les 52 objets, 576 images tests sont employées (voir la section précédente). La base de Columbia contient 36 autres points de vue par objet ajoutant 720 images tests. 31 images tests pour les 31 derniers objets sont utilisées correspondant aux changements mineurs (intensité de luminosité, rotation et échelle). L'ensemble test consiste alors en 1327 images déjà utilisées par la section 6.6.

La figure 7.3 et la table 7.2 résument les résultats de reconnaissance. Malgré un doublement du nombre d'objets par rapport à la section précédente, les résultats sont pratiquement les mêmes. Étant donné la simplicité des images de la base de Columbia, les *taux* de reconnaissance sont même meilleurs qu'auparavant. Une portion visible d'objet de 62% est suffisante pour la reconnaissance des 1327 images tests (le même résultat était obtenu par la comparaison d'histogrammes de la section 6.6). Avec une visibilité de 33.6%, le taux de reconnaissance est encore supérieure à 99% (10 erreurs au total). L'utilisation d'une portion visible de 13.5% permet une reconnaissance supérieur à 90%. Encore plus remarquable, un taux de reconnaissance de 76% est obtenu pour une visibilité d'objet de 6.8% seulement. Ce taux s'explique par l'information discriminante portée par chaque vecteur, confirmée par le taux de reconnaissance de 13% utilisant un seul vecteur de mesures.

rayon [σ_1]	1	5	10	15	20	25	30	35	40
portion d'objet [%]	2.2	6.8	13.5	22.5	33.6	47.0	62.5	80.1	100.0
reconnaissance [%]	13.3	76.2	90.8	96.2	99.3	99.9	100	100	100
erreurs pour les 52 objets	547	255	115	48	9	1	0	0	0
erreurs pour la base de Columbia	573	42	0	0	0	0	0	0	0
erreurs pour les 31 derniers objets	30	19	7	3	1	0	0	0	0

TAB. 7.2 – *Résultats expérimentaux pour 103 objets*

La section 6.6 utilise exactement les mêmes histogrammes à six dimensions, la même base d'images et les mêmes 1327 images tests. La figure 7.3 compare alors les résultats de reconnaissance de l'algorithme probabiliste à celui de comparaison d'histogrammes. En particulier, les résultats des fonctions de la χ^2–statistique χ^2_{qv} et de l'intersection $\cap$ de la section 6.6 sont montrés. La robustesse par rapport aux occultations partielles est considérablement augmentée par l'algorithme probabiliste développé dans ce chapitre.

L'algorithme proposé de reconnaissance probabiliste est alors capable de discriminer 103 objets en présence de changements importants en échelles, en rotations dans le plan image et en points de vue. De plus, la technique est robuste par rapport aux occultations partielles car une petite portion visible d'objet est suffisante pour une hypothèse fiable d'objets. Comme mentionné plus haut, les résultats de reconnaissance ont été obtenus sans correspondance des images tests et de la base d'objets.

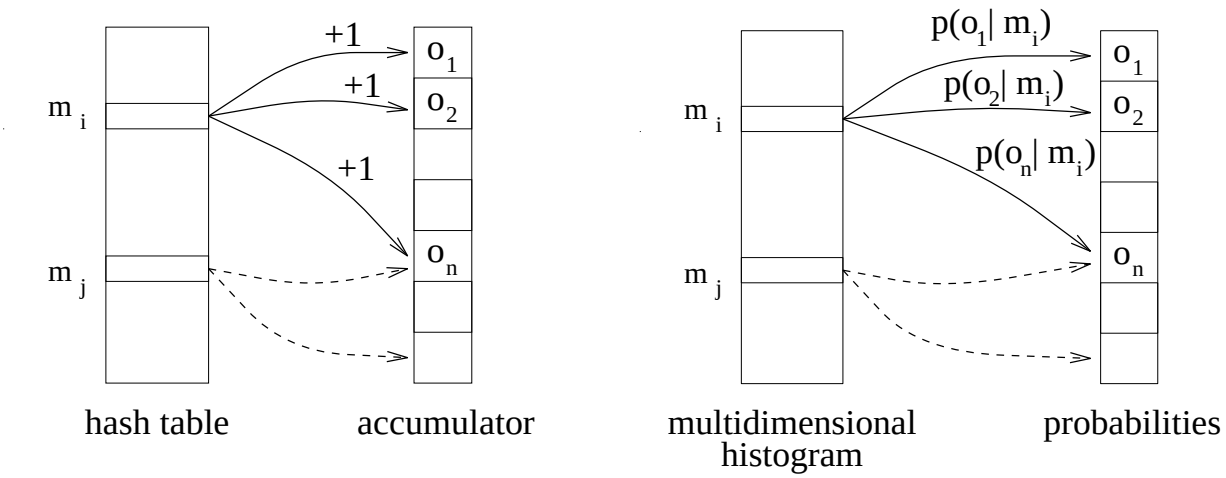
7.4 Reconnaissance probabiliste dans des scènes complexes

La section 7.1 a défini un algorithme de reconnaissance probabiliste calculant la probabilité $p(o_n|R) = p(o_n| \wedge_k m_k)$ pour tous les objets o_n à partir d'une région R de l'image. Les expérimentations des sections 7.2 et 7.3 ont démontré la capacité de l'algorithme à reconnaître 103 objets en présence de changements d'échelle, de rotation et de point de vue. Néanmoins, les résultats expérimentaux ont été obtenus pour le cas d'un seul objet par image. Cette section discute donc une extension de l'algorithme pour la reconnaissance de plusieurs objets dans des scènes complexes.

Les résultats des sections précédentes justifient la définition d'un algorithme de reconnais-

sance pour des scènes complexes. Cet algorithme peut employer, par exemple, une stratégie hypothèse–test présentée au chapitre 8. Une autre possibilité utilise les probabilités calculées à partir des régions de l'image pour voter pour les objets. Ce dernier algorithme ressemble à un algorithme de reconnaissance par une table de hachage utilisant les régions d'image comme caractéristiques d'objets. L'avantage majeur d'une table de hachage fondée sur notre algorithme réside dans le fait que la taille et la forme des régions d'image (= caractéristiques) peuvent être choisies dynamiquement et même pendant l'*exécution de l'algorithme*. Cette section définit un tel algorithme comme extension de notre algorithme probabiliste de la section 7.1.

Un algorithme standard de la reconnaissance d'objets dans des scènes complexes se base sur une table de hachage et sur un accumulateur pour chaque objet. Schmid et Mohr ont démontré récemment la capacité d'un tel algorithme fondé sur des vecteurs discriminants de caractéristiques locales (voir la section 2.2.3). Comme notre algorithme probabiliste de la section 7.1 possède des similarités structurelles à un algorithme d'une table de hachage, les deux techniques sont discutées. Ensuite une extension de notre algorithme est définie comme synthèse des deux méthodes.



(a) *Algorithme fondé sur une table de hachage*

(b) *Reconnaissance probabiliste d'objets*

FIG. 7.4 – *La structure d'un algorithme fondé sur une table de hachage, et celle de notre algorithme probabiliste*

La figure 7.4(a) montre le schéma général d'un algorithme fondé sur une table de hachage. Cet algorithme calcule des indices m_k à partir d'une image test. En utilisant un indice m_i comme indice de la table de hachage, un ensemble d'objets est obtenu pouvant produire l'indice m_i. Souvent, l'accumulateur de chaque objet de l'ensemble est incrémenté d'un (ou d'un facteur approprié de certitude). Cette partie de l'algorithme est le vote. L'algorithme répète le calcul d'indices m_k et vote pour les ensembles d'objets obtenus. L'algorithme s'arrête si un objet possède un nombre suffisant de votes.

La figure 7.4(b) montre le schéma de notre algorithme probabiliste. Cet algorithme utilise un indice m_i pour le calcul d'une probabilité $p(o_n|m_i)$ (en utilisant l'équation 7.101) pour tous

les objets o_n. L'algorithme répète le calcul d'indices m_k et le calcul de probabilités $p(o_n|R) = p(o_n| \wedge_k m_k)$ (voir équation 7.106) pour tous les objets et une région R. L'algorithme peut s'arrêter si un objet a obtenu une probabilité suffisante.

Même si les structures de ces deux algorithmes sont similaires, des différences significatives existent. Leurs avantages et inconvénients sont discutés à partir des critères suivants:

- souplesse de décision
- test de consistance d'information provenant de différentes régions de l'image
- caractéristiques utilisées

L'algorithme d'une table de hachage accumule les votes pour les objets dans une table d'accumulateurs. C'est à dire que les décisions sont "dures", car l'algorithme décide si un objet peut produire un indice ou non. Comme les votes de différents indices sont indépendants, toutes les combinaisons possibles d'indices sont considérées. Cela produit de nombreuses réponses *fausses positives* motivant un test de consistance entre différents indices. Ces tests sont capables de diminuer les réponses fausses positives, mais il est nécessaire de coder ces tests dans la table de hachage avant la reconnaissance.

L'algorithme probabiliste défini dans la section 7.1 utilise la souplesse de décision en calculant les probabilités de chaque objet. Une des différences majeure par rapport à une table de hachage consiste à supposer que dans l'équation 7.106, tous les indices m_k proviennent du même objet. Le test de consistance est fait alors pour chaque objet sans considération des combinaisons de différents objets. Cette supposition est injustifiée pour des scènes complexes de plusieurs objets. Comme une table de hachage permet l'incorporation de tests de consistance pour plusieurs objets, une extension de notre algorithme est fondée sur les capacités de ces deux techniques.

Les caractéristiques utilisées forment un dernier critère de comparaison. D'une part, les caractéristiques d'une table de hachage sont fixes et choisies a priori. D'autre part, dans notre approche, la région d'image R du calcul de probabilités $p(o_n|R)$ peut être choisie dynamiquement et sans recalcul des histogrammes multidimensionnels. La région d'image R peut être interprétée comme indice dynamique d'une table de hachage.

La figure 7.5 présente l'idée de la reconnaissance d'objets par une table de hachage utilisant des régions R_k comme indices dynamiquement adaptables. Une table de hachage "dynamique" est implémentée par notre algorithme (dans la figure montrée par des histogrammes multidimensionnels et par une table d'accumulateurs). Si un critère est satisfait (par exemple si la probabilité d'un objet a dépassé un seuil où la région d'image couvre une taille fixée) l'algorithme vote pour un ensemble d'objets et incrémente l'accumulateur de ces objets. Pour l'application de cet algorithme il faut définir des *critères* appropriés. En principe, deux types de critères sont utilisables: les critères concernant les régions d'image R_k (comme par exemple la taille, la quantité d'informations discriminantes), et/ou les critères des probabilités $p(o_n|R_k)$ (par exemple au dessus d'un seuil). Pour ces deux cas, les critères peuvent être changés dynamiquement pendant l'exécution de l'algorithme sans recalcul de la "table de hachage des indices dynamiques R_k".

Comme illustration de l'algorithme de reconnaissance proposé, fondé sur une "table de hachage dynamiquement indexée", une expérimentation d'une base de 100 objets est décrite. La tâche est de reconnaître plusieurs objets dans les scènes complexes comme celles de figures 7.6(a), (c) et (e). Pour chacun des 100 objets, une image (256×256 pixels) est stockée et un histogramme

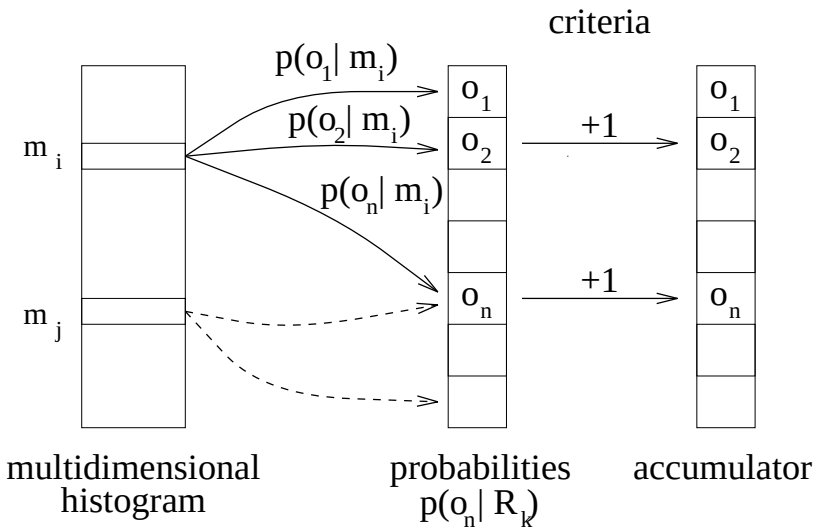


FIG. 7.5 – *Reconnaissance probabiliste d'objets dans les scènes complexes*

à six dimensions est calculé : *Mag-Lap-32* (norme de la première dérivée et opérateur Laplacien, résolution de 32 par axe d'histogramme) à trois échelles différentes ($\sigma_1 = 2.0, \sigma_2 = 4.0$ et $\sigma_3 = 8.0$). Les régions d'image sont les carrés fixes d'une taille de 64^2 pixels. $6 \times 6 = 36$ régions ont été choisies en utilisant un recouvrement de 50% des régions voisines. Pour chacune de 36 régions, l'algorithme probabiliste calcule les probabilités qui sont ajoutées aux accumulateurs d'objets. Les objets qui recouvrent plusieurs régions d'image R_k accumulent alors les probabilités de ces régions. L'accumulateur d'un objet augmente avec le nombre des régions recouvertes par un objet. A la fin les plus grands accumulateurs correspondent aux objets de la scène.

La figure 7.6(a) montre une scène complexe et la figure 7.6(b) montre les quatre objets ayant les plus grands accumulateurs de 100 objets. Ces quatre objets correspondent aux objets de la scène. Une autre image test est montrée par la figure 7.6(c) et les quatre meilleurs objets sont montrés par la figure 7.6(d). Comme pour la première image test, les quatre objets sont reconnus. Une troisième scène est donnée par la figure 7.6(e). Trois objets sont dans la scène, correspondant aux trois meilleurs objets (voir figure 7.6(f)) de l'algorithme. Le quatrième objet retiré correspond à une autre partie du premier objet retiré. C'est à dire que les objets similaires (la similarité est "simulée" par différentes parties du même objet physique) rivalisent et l'objet correct est reconnu à la fin.

En conclusion, l'algorithme proposé combine les avantages de notre algorithme de reconnaissance probabiliste et les avantages d'une table de hachage. En adaptant les régions d'image R_k et en combinant les résultats de manière appropriée, un algorithme de reconnaissance puissant peut être défini applicable dans des circonstances variées. Les résultats indiquent que notre algorithme probabiliste est applicable pour des scènes complexes en utilisant une table de hachage "dynamiquement indexée".

(a) *Image test 1*

(b) *Premières images trouvées pour image test 1*

(c) *Image test 2*

(d) *Premières images trouvées pour image test 2*

(e) *Image test 3*

(f) *Premières images trouvées pour image test 3*

FIG. 7.6 – *Résultats expérimentaux pour trois images tests de scènes complexes*

7.5 Conclusion

Ce chapitre a défini un *algorithme probabiliste de reconnaissance d'objets sans correspondance* comme extension de la *comparaison* d'histogrammes multidimensionnels. La technique calcule la probabilité de la présence de chaque objet en se fondant entièrement sur les histogrammes multidimensionnels de champs réceptifs. Des résultats prometteurs ont été obtenus pour une base de 103 objets en présence de changements d'échelle, de rotation dans le plan image et de point de vue.

La dernière section du chapitre a introduit une extension de l'algorithme probabiliste à une *table de hachage dynamiquement indexée*. Cet algorithme permet la reconnaissance de plusieurs objets dans des scènes complexes en profitant d'avantages d'une table de hachage. Les régions d'image sont employées comme "indices". Pour chacune de ces régions, l'algorithme probabiliste est appliqué pour déterminer des objets qui puissent correspondre à la région d'image. L'avantage majeur de la méthode proposée est que les régions d'image – c'est à dire les indices de la table de hachage – peuvent être adaptées dynamiquement pendant l'exécution. Les résultats prometteurs on été obtenus pour une base de 100 objets.

Chapitre 8

Reconnaissance active d'objets

Le chapitre 7 a proposé un algorithme probabiliste de reconnaissance d'objets en utilisant les histogrammes multidimensionnels de champs réceptifs. Cet algorithme probabiliste permet le calcul d'hypothèses fiables sur l'identité et sur la pose d'objets à partir de petites régions d'image.

L'algorithme probabiliste calcule des probabilités pour chacun des histogrammes multidimensionnels de champs réceptifs, correspondant aux identités différentes o_n et aux poses différentes S_j de ces objets (voir section 4.1). Il fournit alors des hypothèses sur l'identité et sur la pose d'objets. Les résultats des expérimentations nous ont permis de définir (section 7.4) une extension de l'algorithme pour la reconnaissance de plusieurs objets dans des scènes complexes, en utilisant une table de hachage dynamique.

Ce chapitre discute une autre extension possible de la technique en introduisant une *reconnaissance active d'objets*. Cette reconnaissance s'appuie sur des hypothèses fiables sur l'identité et sur la pose d'objets provenant de l'algorithme probabiliste de reconnaissance. Des algorithmes de reconnaissance active en 2D et en 3D sont proposés comme extension de la reconnaissance par histogrammes de champs réceptifs. Le cadre des algorithmes actifs est donné par :

1. La génération d'une hypothèse : l'identité et la pose d'objet sont estimées en utilisant la comparaison d'histogrammes ou l'algorithme probabiliste de reconnaissance :

$$\begin{aligned} \hat{o}_n(t-1) &\equiv \text{hypothèse sur l'identité de l'objet à l'instant } t-1 \\ \hat{S}_j(t-1) &\equiv \text{hypothèse sur la pose de l'objet à l'instant } t-1 \end{aligned}$$

2. Le déplacement de la caméra vers une région discriminante de l'image en 2D ou vers un point de vue discriminant en 3D (la région discriminante et le point de vue discriminant sont choisis à l'avance) :

$$\Delta S(t-1) \equiv \text{paramètre du mouvement de la caméra}$$

3. La vérification de l'hypothèse : la différence entre la prévision et l'observation est employée pour vérifier l'hypothèse :

$$\begin{aligned} \hat{o}_n(t) &= \hat{o}_n(t-1) \\ \hat{S}_j(t) &= \hat{S}_j(t-1) + \Delta S(t-1) \pm \epsilon \end{aligned}$$

L'hypothèse se compose de l'identité $\hat{o}_n$ et de la pose $\hat{S}_j$ de l'objet. Cette hypothèse est donnée, par exemple, par la probabilité maximale $p(o_n|R, S_j)$ (équation 7.106) calculée pour une région R de l'image :

$$(\hat{o}_n, \hat{S}_j) : \max_{n,j} p(o_n|R, S_j) \tag{8.107}$$

L'histogramme multidimensionnel $h(M|o_n, S_j) = h(M|o_n, t_z, r_x, r_y, r_z)$, correspondant à la probabilité maximale, fournit une hypothèse sur l'identité de l'objet $\hat{o}_n$ et sur sa pose $\hat{S}_j = (t_z, r_x, r_y, r_z)$.

Une autre possibilité pour obtenir une hypothèse est la comparaison d'histogrammes par la statistique χ^2. L'hypothèse sur l'identité et sur la pose est alors donnée par le minimum de la statistique χ^2 de l'histogramme $H(M|I)$ de l'image test I et de la base d'histogrammes :

$$(\hat{o}_n, \hat{S}_j) : \min_{n,j} \chi^2_{qv}(H(M|I), H(M|o_n, S_j)) \tag{8.108}$$

Pour vérifier l'hypothèse, la caméra doit être déplacée vers une position discriminante choisie à l'avance (en 2D ou en 3D). Cette position discriminante[1] est spécifique à l'objet o_n. La section 8.1 propose le calcul d'un réseau de régions discriminantes pour la détermination de la prochaine position de la caméra en 2D. Dans ce contexte, l'algorithme actif de reconnaissance peut être vu comme un algorithme de contrôle de fixation pour la reconnaissance. La section 8.2 propose la sélection de points de vue discriminants fondée sur la théorie de l'information. Les points de vue discriminants servent pour la définition d'un algorithme actif de la planification de points de vue. Des résultats expérimentaux sont décrits pour la base d'images de Columbia de 100 objets et 72 points de vue par objet. Il est remarquable que les résultats de l'algorithme soient cohérents en 3D. Cette cohérence est remarquable car l'algorithme se base entièrement sur l'apparence des objets sans incorporation explicite d'informations 3D. Néanmoins, les résultats des sections 8.1 et 8.2 sont préliminaires et doivent être vus comme tels.

Une question importante pour le contrôle de fixation dans un système de vision active se pose comme telle : *"Où regarder après?"*[2] [Swa 93b]. Les différentes approches de ce problème se

1. salient position

2. *"where to look next?"*

divisent en trois catégories de mécanismes d'attention : ceux dirigés par la tache, par le contexte et par les caractéristiques. Dans ce chapitre, une méthode statistique est proposée pour répondre à la question *"où regarder après?"*. Notre approche est fondée sur les histogrammes multidimensionnels de champs réceptifs et sur un réseau de points discriminants. Les histogrammes multidimensionnels servent à générer des hypothèses de la présence d'objets. Le réseau de points discriminants (respectivement le réseau de points de vue discriminants) d'objets permet de déplacer la caméra vers des régions d'intérêt (respectivement vers un point de vue discriminant). La technique dirige alors la fixation afin de vérifier les hypothèses générées. Comme l'algorithme possède une prévision de l'observation à la nouvelle position de la caméra, la méthode répond aussi à la question *"que chercher?"*[3].

Fondée sur le chapitre 7, la section 8.1 développe une technique de calcul d'un réseau de points discriminants et utilise ce réseau pour le contrôle de fixation dans le contexte de la reconnaissance active. Quelques exemples de calcul de points discriminants et d'une expérimentation de leur robustesse par rapport aux changements d'échelle sont décrits. La section 8.2 propose un algorithme de reconnaissance active de la planification de points de vue. Les points de vue discriminants sont sélectionnés par la maximisation de la transinformation de différents points de vue d'un objet. Des résultats expérimentaux sont donnés pour la base d'images de Columbia de 100 objets.

8.1 Reconnaissance active dans une seule image

Le chapitre précédent a proposé un algorithme probabiliste comme extension de l'application d'histogrammes multidimensionnels de champs réceptifs. Dans cette section, le problème de la reconnaissance d'objets est renversé en demandant *"où regarder après?"* pour la vérification de la présence d'un objet, pour le suivi d'un objet ou pour l'exploration active d'une scène. Cette section décrit une méthode pour dériver un réseau de points discriminants d'un objet, qui est caractéristique pour cet objet. Ces points discriminants correspondent à un seul objet ou à un petit nombre d'objets. Ces points maximisent alors la discrimination entre les objets. Un réseau de points discriminants peut être utilisé pour le contrôle de fixation dans le contexte de la reconnaissance active d'objets.

8.1.1 Réseau de points discriminants

La première partie de la section introduit un concept pour déterminer les points les plus significatifs d'une image et/ou d'un objet. Ce concept peut être utilisé pour définir des réseaux de points discriminants d'un ensemble d'objets. Ces réseaux peuvent être employés dans le contexte du contrôle de fixation pour la reconnaissance d'objets (voir section 8.1.2). Il est intéressant de mentionner que le concept se généralise à un *détecteur de points d'intérêt*. Ces points d'intérêt d'un objet ou d'une image peuvent être utilisés directement dans une approche de hachage géométrique [Wol 90] (voir section 2.2.1). Le détecteur de points d'intérêt peut aussi être employé pour trouver les points d'un objet appropriés pour le suivi de l'objet, c'est à dire qui sont discriminants et faciles à détecter.

Comme décrit dans la section 7.1, le calcul de la probabilité d'un objet o_n, étant donné un vecteur m_k de mesures locales, est :

3. *"what to look for?"*

$$p(o_n|m_k) = \frac{p(m_k|o_n)p(o_n)}{p(m_k)} \tag{8.109}$$

avec

- $p(o_n)$ la probabilité a priori de l'objet o_n,
- $p(m_k)$ la probabilité a priori du vecteur m_k de mesures locales (= combinaison de réponses de filtres),
- $p(m_k|o_n)$ la densité probabiliste de l'objet o_n. Cette densité peut être estimée en normalisant les histogrammes multidimensionnels de champs réceptifs de l'objet o_n par leurs tailles.

Cette section s'intéresse aux points les plus discriminants d'un objet o_n (ou d'une image). Ces points sont obtenus par la maximisation de $p(o_n|m_k)$ par rapport à toutes les réponses de filtres m_k d'un objet o_n. La relation entre $p(m_k|o_n)$ et $p(m_k)$ est alors maximisée (comme décrit plus bas, la probabilité $p(o_n)$ peut être négligée). Les maxima correspondent aux combinaisons de filtres extrêmement discriminantes et, comme les exemples vont le montrer, faciles à détecter.

Les applications de ces points discriminants (ou points d'intérêt) sont nombreuses. Comme les points sont les plus discriminants, ils sont utilisables dans le contexte de la vision active pour le suivi d'objets ou pour le contrôle de fixation (section 8.1.2). Pour la reconnaissance d'objets, les premiers K maxima sont utilisés afin de minimiser le nombre de points nécessaires pour déterminer l'identité d'un objet. Ces K maxima définissent un réseau de points discriminants utilisé par la section suivante dans le contexte d'un système de reconnaissance active.

Pour l'utilisation de l'équation 8.109, il faut déterminer la probabilité a priori $p(m_k)$ de chaque réponse de filtres m_k. Cette distribution probabiliste ne dépend pas seulement de la combinaison de filtres employée mais aussi du contexte de la tache de vision, c'est à dire que l'estimation des probabilités $p(m_k)$ nécessite une discussion plus profonde.

Dans le chapitre 7 nous avons utilisé l'équation $p(m_k) = \sum_i p(m_k|o_i)p(o_i)$ pour le calcul de la probabilité a priori de vecteurs de mesures m_k. Cette équation est approprié dans le contexte de la reconnaissance probabiliste d'objets mais elle dépend de la base de données utilisée. Cette dépendance peut être justifiée si l'environnement d'une tache visuelle est connue à l'avance. Cela est souvent vrai dans le cas du suivi d'objets utilisant une caméra stationnaire. Cette section veut proposer une méthode plus générale où certains objets/fonds ne sont pas connus à l'avance. Nous voulons estimer alors la *vraie* probabilité a priori $p(m_k)$. Pour cela, le calcul des histogrammes moyens d'un large ensemble d'histogrammes représentatifs est proposé. La figure 8.1 montre trois histogrammes bidimensionnels des combinaisons de filtres *Dx-Dy*, *Mag-Dir* et *Mag-Lap* qui ont été calculés sur une base de 832 images. Ces histogrammes moyens dépendent encore de la base de données employée, mais ils apparaissent plus appropriés qu'une approximation purement analytique.

En utilisant un histogramme moyen $p(m_k|moyen)$ pour l'estimation de la probabilité a priori $p(m_k)$, l'équation 8.109 peut être réécrite (dans cette équation la probabilité $p(o_n|m_k)$ dépend seulement de l'objet considéré et de l'histogramme moyen) :

$$p(o_n|m_k) = \frac{p(m_k|o_n)p(o_n)}{p(m_k|moyen)} \tag{8.110}$$

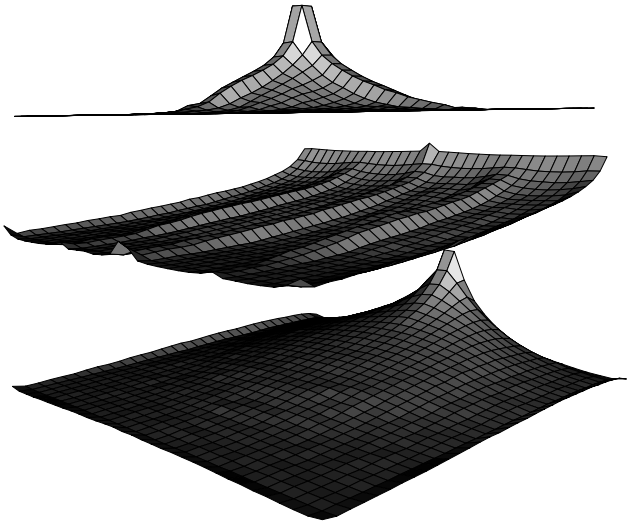


FIG. 8.1 – *Histogrammes moyens de différentes combinaisons de filtres : en haut Dx-Dy, en milieu Mag-Dir et en bas Mag-Lap ($\sigma = 1.0$, la résolution par axe d'histogramme est 32)*

La probabilité $p(o_n)$ peut être éliminée car ces valeurs n'affectent pas l'ordre des maxima de $p(o_n|m_k)$. Pour la stabilisation de l'équation aux "mesures bruitées" [4], l'histogramme $h(m_k|o_n)$ est ajouté à l'histogramme moyen $h(m_k|moyen)$. L'équation suivante est alors employée pour déterminer les points les plus discriminants d'un objet o_n :

$$p^\star(o_n|m_k) = \frac{h(m_k|o_n)}{h(m_k|o_n) + h(m_k|moyen)} \tag{8.111}$$

Dans cette équation, $h(m_k|o_n)$ correspond à la valeur de l'histogramme de vecteurs m_k de mesures de l'objet o_n, et $h(m_k|moyen)$ correspond à la valeur de l'histogramme moyen.

En utilisant les K premiers maxima de l'équation 8.111, un réseau de points discriminants est obtenu ainsi qu'une mesure quantitative de l'importance de chaque point (en particulier $p^\star(o_n|m_k)$). Cette mesure peut être utilisée pour déterminer si un objet possède les points discriminants ou des régions discriminantes. Plus la valeur $p^\star(o_n|m_k)$ d'un point est grande, plus le point est discriminant. La section 8.1.3 utilise ces valeurs pour déterminer des objets d'une base qui contiennent les points les plus discriminants et les objets qui contiennent les points les moins discriminants.

La figure 8.2 montre un exemple des premiers points discriminants d'un objet (marqués par des cercles). La table 8.1 montre les valeurs de réponses de filtres correspondantes. Un histogramme *Dx-Dy-Lap* avec $\sigma = 1.0$ est employé. La taille des cercles est choisie arbitrairement et ne correspond pas au support de filtres Gaussiens. Les maxima de la table 8.1 correspondent

4. outliers

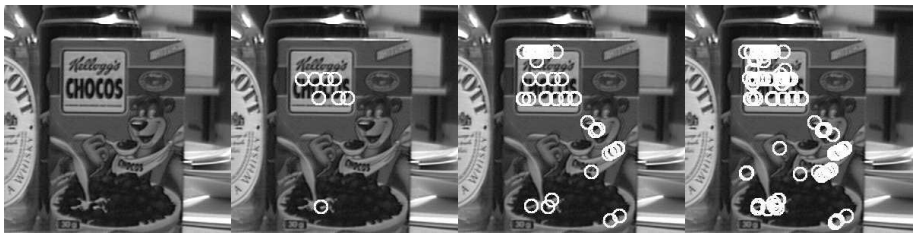


FIG. 8.2 – *Un objet et ses premiers 8, 50 et 103 points discriminants*

aux grandes valeurs de réponses des filtres, c'est à dire qu'ils sont discriminants et faciles à détecter dans l'image. Ces maxima correspondent aux petites valeurs de $h(m_k|moyen)$ (le premier maximum par exemple apparaît seulement 0.04 fois par image). Ces petites valeurs produisent de très grandes valeurs de $p^\star(o_n|m_k)$.

Nombre de maxima	Dx	Dy	Lap	$h(m_k\|o_n)$	$h(m_k\|moyen)$	$p^\star(o_n\|m_k)$
1	34.2	49.0	-99.0	1	0.0445	0.9574
2	-19.4	-50.5	56.8	4	0.2067	0.9509
3	-27.3	-49.0	63.7			
4	-26.6	-48.7	54.4			
5	-25.5	-53.1	62.7			
6	36.0	50.8	6.2	3	0.1635	0.9483
7	35.8	50.7	7.8			
8	40.1	48.2	3.3			
			...			
103	16.1	43.8	16.0	19	1.2848	0.9367

TAB. 8.1 – *Les réponses de filtres – projetées dans l'intervalle [-128,+128] – des premiers points discriminants de l'objet de la figure 8.2, les valeurs de l'histogramme de l'objet $h(m_k|o_n)$, les valeurs de l'histogramme moyen $h(m_k|moyen)$ et la valeur calculée par $p^\star(o_n|m_k)$*

Dans de nombreux cas, des regroupements de points discriminants sont possibles (comme par exemple dans la figure 8.2). Il semble alors raisonnable de définir un réseau de régions discriminantes plutôt qu'un réseau de points discriminants. Le concept du contrôle de fixation, décrit par la suite, peut être adapté facilement pour un réseau de régions discriminantes.

8.1.2 Contrôle de fixation pour la reconnaissance active d'objets

Le schéma général proposé pour la reconnaissance active d'objets se compose de deux étapes majeures :

Génération d'une hypothèse : fondée sur l'algorithme probabiliste de reconnaissance, l'hypothèse de la présence d'un objet à une certaine rotation et à une certaine échelle peut être générée. Après une recherche locale d'un point discriminant des meilleures hypothèses, l'hypothèse de la position d'objet dans l'image est générée.

Vérification : pour vérifier l'hypothèse générée, le *réseau de points discriminants* détermine

une région d'intérêt (*"où regarder après?"*). Le réseau de points discriminants permet aussi de répondre à la question *"que chercher?"*.

La *génération d'hypothèse* s'appuie sur l'algorithme probabiliste de reconnaissance. Cet algorithme calcule les probabilités pour chaque objet pour une région arbitraire d'image, c'est à dire pour une région initiale d'intérêt arbitraire. Comme mentionné plus haut, l'algorithme calcule des probabilités pour des histogrammes multidimensionnels qui correspondent à des échelles et des rotations différentes des objets. Des hypothèses candidates sont obtenues par les maxima des probabilités calculées. Ces hypothèses consistent en une identité, en une échelle et en une rotation d'objet.

Afin de compléter une des hypothèses sur la pose d'objet, il faut estimer la position d'un des objets dans l'image. Nous proposons de rechercher la région actuelle d'intérêt pour trouver un des points discriminants des objets d'hypothèses candidates. Les caractéristiques locales calculées pendant cette recherche peuvent être utilisées pour valider les hypothèses candidates. Après avoir trouvé un point discriminant d'un des objets, la position de ce point dans le réseau de points discriminants détermine la position approximative de l'objet dans l'image. L'hypothèse sur l'identité de cet objet et sur sa pose complète est alors générée pour la suite.

Pour la *vérification* de l'hypothèse, une nouvelle région d'intérêt est choisie dans l'image. La région la plus appropriée est donnée par le point le plus significatif du réseau de points discriminants de l'objet de l'hypothèse générée. Le choix de cette région répond à la question *"où regarder après?"*. Comme le vecteur de caractéristiques locales de ce point discriminant est connu, l'algorithme possède une prévision de l'observation dans la nouvelle région d'intérêt (*"que chercher?"*).

Si le point discriminant de la nouvelle région d'intérêt peut être trouvé, l'évidence de l'hypothèse est augmentée. L'étape de vérification peut être répétée jusqu'à ce que l'évidence de l'hypothèse soit suffisante pour *confirmer* l'hypothèse. Si le point discriminant n'est pas trouvé, les vecteurs calculés de caractéristiques locales dans la nouvelle région d'intérêt sont utilisés pour calculer et/ou modifier les probabilités de chaque objet (*récupérer*).

Dans le contexte de cette thèse, étant donné des contraintes de temps, cet algorithme n'a pas été implémenté entièrement. Néanmoins, le détecteur de points d'intérêt est implémenté. La section 8.1.3 décrit des exemples d'illustration de l'application du détecteur de points discriminants et donne une indication de la robustesse du détecteur aux changements d'échelle. La prochaine étape de la réalisation de l'algorithme du contrôle de fixation est l'analyse de la répétabilité et de la stabilité du détecteur de points discriminants par rapport aux changements d'échelle, de rotation et d'éclairage. De plus, il faut définir une représentation du réseau de points discriminants et trouver un algorithme efficace pour la recherche locale de points discriminants. En particulier, la répétabilité et la stabilité du détecteur de points discriminants peuvent limiter l'applicabilité de l'approche proposée.

8.1.3 Exemples d'illustration du détecteur de points discriminants

Dans cette section, le détecteur de points discriminants est appliqué sur la base des 30 objets montrés par la figure A.1. Des histogrammes à six dimensions sont utilisés : la combinaison de filtres *Dx-Dy-Lap* à deux échelles ($\sigma = 1.0$ et 2.0). En utilisant les premiers 30 maxima de $p^\star(o_n|m_k)$ (voir équation 8.111), un ensemble de points discriminants est obtenu pour chaque

objet. La première partie de la section détermine les objets contenant les points les plus discriminants et contenant les points les moins discriminants. La deuxième partie examine la robustesse du détecteur de points discriminants en présence de changements d'échelle.

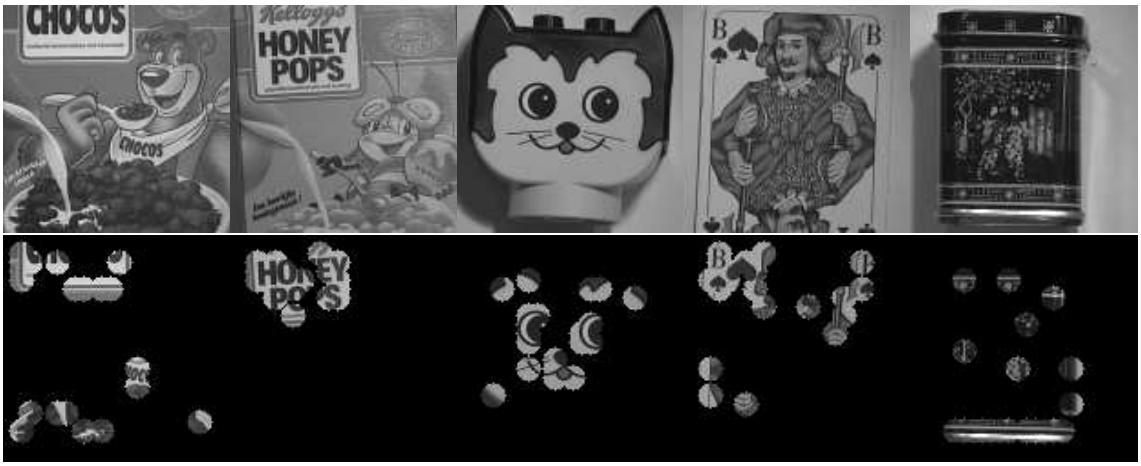


FIG. 8.3 – *La première ligne montre les 5 objets contenant les points les plus discriminants. La deuxième ligne montre les 20 points les plus discriminants de ces objets*

FIG. 8.4 – *Les 5 objets contenant les points les moins discriminants. La deuxième ligne montre les 20 points les plus discriminants de ces objets*

Comme mentionné dans la section 8.1.1, la valeur de $p^\star(o_n|m_k)$ détermine les objets de la base qui contient les points les plus discriminants. La première ligne de la figure 8.3 montre les 5 objets contenant les points les plus discriminants. La deuxième ligne de cette figure montre les 20 points les plus discriminants de chacun de ces objets. La figure 8.4 montre les 5 objets qui contiennent les points les moins discriminants correspondant au cas le plus délicat d'application de la technique. La deuxième ligne de cette figure montre les 20 points les plus discriminants de ces objets.

Pour examiner la robustesse par rapport aux changements d'échelle, nous avons employé un

FIG. 8.5 – *Robustesse de détecteur de points d'intérêt par rapport aux changements d'échelle. Le changement d'échelle est approximativement 20%*

objet de la figure 8.3 et un objet de la figure 8.4. La figure 8.5 montre les résultats du même détecteur de points d'intérêt pour trois échelles différentes. Le changement d'échelle est approximativement de 20% entre chaque image. Les points extraits sont pratiquement identiques, indiquant une certaine robustesse du détecteur de points d'intérêt.

8.2 Reconnaissance active d'objets comme planification de points de vue

La section précédente a défini un algorithme de reconnaissance active dans le contexte du contrôle de fixation d'une caméra. Les mouvements de la caméra ont été en 2D, c'est à dire dans le plan de l'image. Cette section propose le calcul de points de vue discriminants fondé sur la transinformation de points de vue. Ce calcul permet la définition d'un algorithme de reconnaissance active comme planification de points de vue.

8.2.1 Transinformation d'un point de vue singulier

La section 4.2 a montré l'application des concepts de la théorie de l'information au processus de reconnaissance d'objets. Cette section donne une application des concepts développés dans le contexte de la planification de points de vue pour la reconnaissance d'objets. Plus précisément, nous calculons la transinformation de chaque point de vue d'un objet afin de choisir le point de de vue le plus discriminant dans un algorithme de reconnaissance active.

Dans la section 4.2.3, nous avons défini la transinformation du processus de reconnaissance d'objets fondée sur une paire d'événements (o_n, m_k). Cette section développe une équation pour le choix du point de vue le plus discriminant d'un objet. En utilisant l'équation initiale 4.73, la transinformation peut être réécrite :

$$T(O,M) = \sum_{n=1}^{N}\sum_{k=1}^{K} p(o_n \wedge m_k) \log \frac{p(o_n \wedge m_k)}{p(o_n)p(m_k)} \tag{8.112}$$

$$= \sum_{n=1}^{N} p(o_n) \sum_{k=1}^{K} p(m_k|o_n) \log \frac{p(m_k|o_n)p(o_n)}{p(o_n)p(m_k)} \tag{8.113}$$

$$= \sum_{n=1}^{N} p(o_n) \sum_{k=1}^{K} p(m_k|o_n) \log \frac{p(m_k|o_n)}{\sum_i p(m_k|o_i)p(o_i)} \tag{8.114}$$

La transinformation peut être interprétée comme la transinformation moyenne de l'objet o_n qui peut être définie par :

$$T(o_n, M) = \sum_{k=1}^{K} p(m_k|o_n) \log \frac{p(m_k|o_n)}{\sum_i p(m_k|o_i)p(o_i)} \tag{8.115}$$

La probabilité $p(m_k|o_n)$ correspond à la densité probabiliste d'un objet entier, introduite et utilisée dans les chapitres précédents. Comme nous voulons calculer la transinformation de chaque point de vue d'un objet séparément, nous définissons la transinformation d'un objet o_n pour une pose particulière $S_j = (t_z r_x r_y r_z)$ par :

$$T(o_n, S_j, M) = \sum_{k=1}^{K} p(m_k|o_n, S_j) \log \frac{p(m_k|o_n, S_j)}{\sum_{i,l} p(m_k|o_i, S_l)p(o_i, S_l)} \tag{8.116}$$

8.2.2 Planification de points de vue

Cette section décrit un algorithme de reconnaissance active. L'idée de base de l'algorithme est la génération d'hypothèse sur l'identité et sur la pose d'objet à partir d'une image test (reconnaissance à partir d'un seul point de vue). Fondé sur cette hypothèse, l'algorithme déplace la caméra vers le point de vue le plus discriminant de l'objet de l'hypothèse. L'information du point de vue nouveau est utilisée pour la vérification de l'hypothèse, c'est à dire pour vérifier l'identité et la pose de l'objet.

Les résultats expérimentaux (section 8.2.3) montrent que l'algorithme peut résoudre toutes les ambiguïtés (sauf une) inhérentes à la reconnaissance, à partir d'un seul point de vue. Il est intéressant de noter que l'algorithme intègre des informations 3D de la représentation statistique d'objets. Les dernières erreurs (utilisant une étape de vérification) sont alors cohérentes en 3D. En appliquant plusieurs étapes de vérification, l'algorithme intègre plus d'informations 3D permettant d'atteindre une reconnaissance presque parfaite.

Description de l'algorithme de reconnaissance active

Comme mentionné dans l'introduction de ce chapitre, l'algorithme de reconnaissance active contient trois étapes – en utilisant la comparaison d'histogrammes pour la génération d'hypothèse :

Génération d'hypothèse : l'hypothèse sur l'identité et sur la pose d'objet est générée par la comparaison de l'histogramme $H(M|I)$ de l'image test I avec la base d'histogrammes. L'hypothèse est donnée par l'histogramme multidimensionnel de champs réceptifs qui correspond à la statistique χ^2_{qv} minimale: l'hypothèse consiste en l'identité d'objet $\hat{o_n}$ et en sa pose $\hat{S}_j$ (voir équation 8.108).

Mouvement de la caméra : la caméra est déplacée vers le point de vue le plus discriminant (ou le deuxième le plus discriminant) de l'objet $\hat{o_n}$. Ce point de vue est calculé avant l'application de l'algorithme par la maximisation de l'équation 8.116. Le mouvement de la caméra est calculé en s'appuyant sur la différence ΔS entre la pose estimée $\hat{S_j}$ et la pose du point de vue le plus discriminant.

Vérification de l'hypothèse : à la nouvelle position de la caméra, une autre hypothèse sur l'identité d'objet et sur sa pose est obtenue (voir équation 8.108). Plusieurs contraintes peuvent être utilisées pour la vérification. Premièrement, l'identité d'objet doit être la même pour les deux hypothèses, c'est a dire avant (au temps $(t-1)$) et après (au temps (t)) le déplacement de la caméra :

$$\hat{o_n}(t) = \hat{o_n}(t-1) \tag{8.117}$$

Deuxièmement, la connaissance du mouvement de la caméra peut être employée. Ce mouvement est calculée sur la base de la différence ΔS. En utilisant cette différence, l'observation à la nouvelle position de la caméra peut être prédite :

$$\hat{S_j}(t) \quad = \quad \hat{S_j}(t-1) + \Delta S(t-1) \pm \epsilon \tag{8.118}$$

où ϵ correspond à l'erreur permise pour l'estimation.

L'algorithme proposé de reconnaissance active déplace la caméra vers le point de vue le plus discriminant de l'objet de l'hypothèse. Si l'objet de l'hypothèse est présent dans la scène, l'étape de vérification doit vérifier la présence de cet objet. Par contre, l'algorithme doit être capable de rejeter l'hypothèse si l'objet de celle–ci n'est pas présent. Nous pouvons l'espérer car le point de vue le plus discriminant a été choisi par rapport à la base d'objets. Les ambiguïtés 2D de la base, difficiles à résoudre à partir d'un seul point de vue, doivent être résolues par l'algorithme proposé.

La propriété la plus intéressante de l'algorithme proposé est l'incorporation des informations 2D (une seule image) ainsi que des informations 3D de la densité probabiliste d'objets. En particulier, les informations 3D sont utilisées pour la vérification d'hypothèses. Les erreurs de l'algorithme proposé doivent correspondre aux objets qui sont "similaires" en 3D (comme par exemple des cubes). Les expérimentations soulignent cette propriété de l'algorithme.

Afin d'obtenir un nombre minimal de réponses fausses positives et fausses négatives, il est possible d'introduire plusieurs étapes de vérifications. En utilisant L étapes de vérifications, les contraintes suivantes peuvent être utilisées :

$$\forall t = 2, ..., L : \hat{o_n}(t) \quad = \quad \hat{o_n}(t-1) \tag{8.119}$$

$$\hat{S_j}(t) \quad = \quad \hat{S_j}(t-1) + \Delta S(t-1) \pm \epsilon \tag{8.120}$$

8.2.3 Résultats expérimentaux

Cette section décrit une expérimentation pour montrer l'applicabilité de l'algorithme proposé à la reconnaissance active. La base d'images de Columbia [Nen 96] est employée, celle–ci contient 100 objets et 72 vues par objets, c'est à dire 7200 images. Les images de cette base sont en couleur et ont été converties en images de niveaux de gris. Ces images sont prises sous des conditions contrôlées d'éclairage et devant un fond noir. La figure A.6 montre les 100 objets. Les 72 points de vue d'un objet sont pris à une position fixe de la caméra et en tournant les objets en intervalles de 5° sur une table tournante.

Malheureusement, la base d'image ne contient qu'un seul degré de liberté de rotation. Par conséquent, trois paramètres sur l'estimation de la pose ne sont pas considérés dans cette expérimentation. La raison de l'utilisation de cette base est la possibilité de simuler les mouvements d'une caméra en "tournant" les objets devant l'objectif. L'applicabilité de l'algorithme peut être montrée sans dépendance de la précision du mouvement et de la calibration de la caméra (afin de déplacer la caméra relativement à l'objet, la caméra doit être calibrée au moins de façon approximative). La moitié des images d'objets (tous les 10°, $100 \times 36 = 3600$ images) est employée comme base de données. Pour chaque image, nous calculons son histogramme des premières dérivées Gaussiennes dans les directions x et y. Pour cette expérimentation une résolution de 16 cellules par axe d'histogramme est employée. Chacune de 3600 images est alors représentée par $16^2 = 256$ nombres. Pour chacun des 100 objets, nous calculons le point de vue le plus discriminant.

étapes de vérification	reconnaissance [%]	nombre d'erreurs
0	98.36	59
1	99.14	31
2	99.97	1

TAB. 8.2 – *Résultats expérimentaux utilisant la base de Columbia de 100 objets 3D*

La table 8.2 montre les résultats obtenus par l'application de l'algorithme de reconnaissance active introduit plus haut. La première colonne montre le nombre d'étapes de vérification utilisées pour accepter une hypothèse d'objet. Sans l'utilisation de vérification (première ligne de la table) 59 erreurs sans obtenues pour les 3600 images. En utilisant une seule étape de vérification, ce nombre est réduit à 31 erreurs. En appliquant deux étapes de vérification toutes les images, sauf une, sont correctement reconnues. Ces résultats montrent l'applicabilité de l'algorithme proposé pour la reconnaissance active. (Une erreur ϵ jusqu'à 5° est acceptée pendant une étape de vérification, voir équation 8.120).

FIG. 8.6 – *Les 5 objets confondus par l'algorithme de reconnaissance avec une seule étape de vérification*

Il est intéressant de noter que toutes les erreurs obtenues avec une seule étape de vérification

correspondent aux objets de la même forme géométrique 3D. La figure 8.6 montre les 5 objets qui sont confondus. La plupart des erreurs (20 erreurs sur 31) sont obtenues pour les deux derniers objets (qui peuvent être distingués dans la base originale par leur couleur). Les cinq objets possèdent la même forme géométrique, ce sont des cuboïdes. Les erreurs sont alors plutôt "systématiques". Ces objets peuvent être distingués par l'information de couleur ou par un autre vecteur de mesures appropriées (qui peut être évalué par la transinformation comme introduit dans la section 4.2.3). La seule erreur obtenue avec deux étapes de vérification est la confusion des deux derniers objets de la figure 8.6.

8.3 Conclusion

Ce chapitre a proposé l'application des histogrammes multidimensionnels de champs réceptifs pour la reconnaissance active d'objets. Deux algorithmes ont été proposés : un algorithme actif de contrôle de la fixation et un algorithme actif de la planification de points de vue.

Dans le cas du premier algorithme de contrôle de la fixation, l'utilisation des histogrammes multidimensionnels de champs réceptifs est proposée pour déterminer les points les plus discriminants d'un objet. Un réseau de ces points discriminants peut être utilisé pour déterminer *"où regarder après?"* et *"que chercher?"* dans le contexte de la reconnaissance active d'objets. En particulier, un détecteur de points d'intérêt peut être défini, qui est utilisable, par exemple, dans le contexte de hachage géométrique.

Les points de vue discriminants peuvent être choisis par la maximisation de la transinformation de tous les points de vue d'un objet. Cette sélection du point de vue le plus discriminant s'appuie sur l'analogie entre la théorie de l'information et la reconnaissance d'objets (voir section 4.2). Ces points de vue discriminants peuvent être utilisés dans le contexte de la reconnaissance active d'objets en déplaçant la caméra vers le point de vue le plus discriminant d'un objet de l'hypothèse. Les premiers résultats sur la base d'images de Columbia indiquent qu'une reconnaissance cohérente d'objets en 3D est obtenue par l'algorithme proposé. Cette cohérence en 3D est remarquable car l'algorithme de reconnaissance, employé à chaque point de vue, n'utilise pas de mesures 3D de manière explicite.

Les résultats de ce chapitre et ceux des deux chapitres précédents sont encourageants (chapitre 6 : reconnaissance fiable par comparaison d'histogrammes, chapitre 7 : reconnaissance robuste par un algorithme probabiliste et ce chapitre : reconnaissance d'objets en 3D) et permettent la proposition de la *classification d'objets* fondée sur les histogrammes multidimensionnels de champs réceptifs. Ceci est présenté dans le chapitre 9 suivant. Celui-ci propose le concept de classes visuelles comme cadre général de classification d'objets.

Chapitre 9

Classification d'objets

Les chapitres précédents ont démontré que les histogrammes multidimensionnels de champs réceptifs constituent un moyen fiable pour la reconnaissance d'objets 3D. Trois algorithmes de reconnaissance ont été proposés: la comparaison d'histogrammes, la reconnaissance probabiliste, et dans le chapitre précédent la reconnaissance active. Les résultats de ces chapitres indiquent que des objets visuellement similaires possèdent des histogrammes multidimensionnels de champs réceptifs similaires. Cette propriété permet d'espérer trouver des objets visuellement similaires à partir d'histogrammes multidimensionnels similaires. Dans ce contexte, la reconnaissance active d'objets 3D par l'incorporation de plusieurs points de vue est particulièrement remarquable. Les résultats des chapitres précédents nous permettent l'introduction du concept de classes visuelles comme cadre de la classification d'objets fondée sur l'apparence. La classification d'objets constitue l'un des défis principaux d'une technique de reconnaissance utilisant seulement des mesures 2D.

La première section du chapitre introduit le concept de *classes visuelles* comme cadre de la classification d'objets. Les classes visuelles associent des apparences visuellement similaires à un ensemble de mesures d'image. Comme défini dans ce chapitre, les classes visuelles sont implicites à un grand nombre de schémas de représentation d'objets (comme les modèles géométriques et les modèles fondés sur l'apparence). Notre argument est que l'identification de classes visuelles constitue un outil performant pour la classification d'objets. L'identification de classes visuelles est la première étape de la classification d'objet. La classification ne dépend pas seulement du contenu de l'image mais aussi d'autres informations dont l'observateur dispose, telles que les dépendances du contexte et les relations spatiales et temporelles entre les objets.

En utilisant la représentation statistique par les histogrammes multidimensionnels de champs réceptifs, la section 9.2 propose une technique selon le maximum de vraisemblance pour la reconnaissance de classes visuelles. La section 9.3 décrit des expérimentations permettant d'accéder à des images visuellement similaires d'une base de 200 images de paysages. La question importante liée à l'extraction de classes visuelles d'une base d'histogrammes est discutée et un algorithme de regroupement est proposé (section 9.4).

9.1 Le concept de classes visuelles

Cette section introduit le concept de *classes visuelles* pour la classification d'objets. Les classes visuelles sont définies à partir de similitudes d'apparence en 2D et/ou en 3D. Les classes visuelles dépendent, alors, seulement du contenu d'images. L'argument est que l'identification de classes visuelles constitue un outil performant pour la classification d'objets. Les classes visuelles – définies par l'information intrinsèque de l'image – représentent la première étape de la classification d'objet. La classification ne dépend pas seulement du contenu d'image mais aussi du contexte et des relations spatiales et temporelles entre les objets. Il est permis de penser que les classes visuelles peuvent être généralisées en dehors de la base de données considérée. Cette généralisation permet d'identifier les classes visuelles (et ainsi les classes d'objets possibles) d'un objet inconnu.

Les chapitres précédents ont développé et appliqué la représentation statistique pour les objets 3D, fondée sur les histogrammes multidimensionnels de champs réceptifs. Cette représentation est compatible avec de nombreux schémas de représentation d'objets. La représentation peut être utilisée pour l'identification de similitudes d'objets en 2D et/ou en 3D. Le chapitre précédent a montré la reconnaissance d'objets en 3D. Le concept de classes visuelles est définie par l'utilisation de similitudes visuelles entre les objets.

Les classes visuelles associent des apparences similaires à un ensemble de mesures d'image. Les classes visuelles doivent être fondées entièrement sur des mesures d'image. Dans la plupart des cas, cela implique que les classes visuelles *ne coïncident pas* avec les classes d'objets (d'un certain contexte). Un exemple connu est la classe d'objets *chaise* qui possède de nombreuses apparences différentes. L'approche classique est la recherche d'une représentation unique pour toutes les apparences (ou au moins la plupart des apparences) de la classe *chaise* (souvent appelée représentation invariante). Notre argument est qu'une telle représentation n'existe pas en général. Par contre, il est possible de définir des classes visuelles de différentes apparences de la classe d'objet *chaise*. D'une part, une classe d'objets est alors définie par plusieurs classes visuelles. D'autre part, il est possible qu'une classe visuelle fasse partie de plusieurs classes d'objet.

S'appuyant sur notre argument, l'identification de classes visuelles ,comme définies dans ce chapitre, peut être vue comme sous–problème de la classification d'objets. En utilisant les classes visuelles, les classes d'objets peuvent être définies comme un ensemble de classes visuelles. En général, une classe d'objets est définie par la combinaison de différentes classes visuelles. Les types de combinaisons concernent les relations (spatiale et temporelle) entre les objets, le contexte et les autres connaissances disponibles. Il est difficile d'extraire directement les classes d'objets à partir des images. De ce fait, l'extraction de classes visuelles est proposée comme étape préliminaire de la classification d'objets. Le concept de classes visuelles peut combler le vide qui existe entre les modèles fondés sur l'apparence et un classificateur général d'objets.

La représentation statistique introduite par la section 4.1 est interprétée comme une formulation générale de la représentation d'objets car elle est compatible avec de nombreuses représentations différentes d'objets. Comme mentionné plus haut, les classes visuelles sont définies par des similitudes (en 2D et en 3D) par rapport à un ensemble de mesures d'image. Ces similitudes peuvent être extraites par les schémas de réduction de dimensions comme la transformation de Karhunen–Loeve, l'analyse de discriminantes et les algorithmes de regroupement. La section 9.2 introduit une représentation et un algorithme de reconnaissance selon le maximum de vraisemblance de classes visuelles. Les premières expérimentations démontrent la validité du concept proposé de classes visuelles.

9.2 Reconnaissance de classes visuelles

Cette section décrit la reconnaissance de classes visuelles selon le maximum de vraisemblance. L'idée principale est d'adapter un algorithme de Moghaddam et Pentland [Mog 95] qui ont proposé un algorithme selon le maximum de vraisemblance pour la recherche visuelle ainsi que la détection de visages et de mains. Leur technique est fondée sur l'estimation de la densité probabiliste d'un vecteur $\mathbf{x}$ donné une classe Ω: $p(\mathbf{x}|\Omega)$. Dans leur cas le vecteur $\mathbf{x}$ est donné par une image en niveaux de gris et la classe Ω est la classe de visage, la classe d'une partie de visage (l'oeil, la bouche ou le nez) ou la classe de silhouettes de main.

Leur approche peut être adaptée pour la reconnaissance de *classes visuelles* Ω en représentant les histogrammes multidimensionnels par les vecteurs $\mathbf{x}$. Soient donnés K histogrammes $H_1, H_2, \ldots, H_K$ d'une base d'histogrammes représentant une classe visuelle Ω. Chaque histogramme correspond à l'apparence particulière d'un objet 3D. La représentation vectorielle d'un histogramme est donnée par la transformation t: $\mathbf{x}_i = t(H_i)$. Chaque cellule des histogrammes H_i correspond à une dimension du vecteur $\mathbf{x}_i$.

Pour estimer la densité probabiliste $p(\mathbf{x}|\Omega)$, des estimations fiables du vecteur moyen $\overline{\mathbf{x}}$ et de la matrice de covariance $\Phi = \sum_{i=1}^{K}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T$ sont réalisées à partir de la distribution d'exemples $\mathbf{x}_i = t(H_i)$. Supposons de plus une distribution Gaussienne de $p(\mathbf{x}|\Omega)$:

$$p(\mathbf{x}|\Omega) = \frac{e^{-\frac{d(\mathbf{x})}{2}}}{\sqrt{(2\pi)^N ||\Phi||}} \tag{9.121}$$

avec $d(\mathbf{x})$ la distance de Mahalanobis :

$$d(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}})^T \Phi^{-1} (\mathbf{x} - \overline{\mathbf{x}}) \tag{9.122}$$

Par le calcul de l'espace propre d'exemples $\mathbf{x}_i$, la distance de Mahalanobis peut être réexprimée. En projetant tous les K histogrammes $H_1, H_2, \ldots, H_K$ sur les vecteurs $\mathbf{x}_i = t(H_i), i = 1, 2, \ldots, K$ et en utilisant le vecteur moyen $\overline{\mathbf{x}}$, la matrice de covariance est donnée par :

$$\Phi = \sum_{i=1}^{K} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T \tag{9.123}$$

Cette matrice de covariance fournit une série de valeurs propres λ_i et de vecteur propres e_i (voir pour les détails de calcul [Pre 92, Tur 91a]) :

$$[e_1 e_2 \dots e_N] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_N \end{bmatrix} = E\Lambda = \Phi E \tag{9.124}$$

L'utilisation de tous les N vecteurs propres permet la définition d'une base orthogonale $E = [e_1 e_2 \dots e_N]$. Chaque histogramme H_i peut être projeté dans cet espace propre. Cette projection $\mathbf{y}_i$ est définie par :

$$\mathbf{y}_i = E^T(\mathbf{x}_i - \overline{\mathbf{x}}) \tag{9.125}$$

Les vecteurs propres de la base E et la matrice de valeurs Λ permettent la réécriture de la distance de Mahalanobis :

$$\begin{aligned} d(\mathbf{x}) &= (\mathbf{x} - \overline{\mathbf{x}})^T \Phi^{-1} (\mathbf{x} - \overline{\mathbf{x}}) && (9.126)\\ &= (\mathbf{x} - \overline{\mathbf{x}})^T E \Lambda^{-1} E^T (\mathbf{x} - \overline{\mathbf{x}}) && (9.127)\\ &= \mathbf{y}^T \Lambda^{-1} \mathbf{y} && (9.128)\\ &= \sum_{i=1}^{N} \frac{y_i^2}{\lambda_i} && (9.129) \end{aligned}$$

avec $\mathbf{y} = E^T(\mathbf{x} - \overline{\mathbf{x}})$ la projection du vecteur $\mathbf{x}$ sur l'espace propre. En calculant seulement les M premières projections Moghaddam et Pentland proposent l'estimation suivante de la distance de Mahalanobis $d(\mathbf{x})$:

$$\begin{aligned} \hat{d}(\mathbf{x}) &= \sum_{i=1}^{M} \frac{y_i^2}{\lambda_i} + \frac{1}{\rho} \sum_{i=M+1}^{N} y_i^2 && (9.130)\\ &= \sum_{i=1}^{M} \frac{y_i^2}{\lambda_i} + \frac{1}{\rho} \quad \epsilon^2(\mathbf{x}) && (9.131) \end{aligned}$$

L'espace couvert par les M premiers vecteurs propres est appelé *espace caractéristique* F dans la suite. Le complément $\overline{F}$ de cet espace est couvert par les $N - M$ autres vecteurs propres. Utilisant la distance estimée $\hat{d}(\mathbf{x})$, la probabilité $p(\mathbf{x}|\Omega)$ peut être estimée comme le produit de deux densités Gaussiennes mutuellement indépendantes. La première densité $p_F(\mathbf{x}|\Omega)$ est celle de l'espace caractéristique F et la deuxième densité $\hat{p}_{\overline{F}}(\mathbf{x}|\Omega)$ est celle de l'espace complémentaire $\overline{F}$:

$$\begin{aligned} \hat{p}(\mathbf{x}|\Omega) &= p_F(\mathbf{x}|\Omega) \quad \times \quad \hat{p}_{\overline{F}}(\mathbf{x}|\Omega) && (9.132)\\ &= \frac{e^{-\frac{1}{2}\sum_{i=1}^{M} \frac{y_i^2}{\lambda_i}}}{\sqrt{(2\pi)^M \prod_{i=1}^{M} \lambda}} \times \frac{e^{-\frac{\epsilon^2(\mathbf{x})}{2\rho}}}{\sqrt{(2\pi)^{(N-M)} \rho^{(N-M)}}} \end{aligned}$$

La première partie de cette distance estimée est appelée la *distance dans l'espace caractéristique* F car elle est à l'intérieur de l'espace F, couvert par les M vecteurs propres. La distance $\epsilon^2(\mathbf{x})$ est appelée la *distance à l'espace caractéristique* et elle est calculée directement à partir de la projection sur l'espace propre :

$$\epsilon^2(\mathbf{x}) = \sum_{i=M+1}^{N} y_i^2 = ||(\mathbf{x} - \overline{\mathbf{x}})||^2 - \sum_{i=1}^{M} y_i^2 \tag{9.133}$$

Une solution optimale du paramètre ρ par rapport à l'entropie relative – comme fonction de coût – entre la densité réelle $p(\mathbf{x}|\Omega)$ et la densité estimée $\hat{p}(\mathbf{x}|\Omega)$ est :

$$\rho^\star = \frac{1}{N-M} \sum_{i=M+1}^{N} \lambda_i \tag{9.134}$$

Une extension par Moghaddam et Pentland de l'approche utilise les densités multimodales pour l'espace F. La supposition de l'indépendance de composantes de l'espace $\overline{F}$ de celles de l'espace F permet l'utilisation de la formulation séparée de la densité estimée $\hat{p}(\mathbf{x}|\Omega)$ de l'équation 9.132. Dans ce contexte, $p_F(\mathbf{x}|\Omega)$ est donnée par une densité arbitraire de composantes principales de $\mathbf{y}$. Il est possible d'utiliser une mélange de densités Gaussiennes, estimée par l'algorithme de maximisation de l'espérance (voir la section 2.3.2).

La combinaison proposée de la distance dans l'espace caractéristique et de la distance à l'espace caractéristique constitue une méthode de reconnaissance de classes visuelles selon le maximum de vraisemblance. La section suivante applique, en particulier, la *distance à l'espace caractéristique* $\epsilon^2(\mathbf{x})$ pour la classification d'images.

9.3 Exemples de classification

La section précédente a introduit la *distance à l'espace caractéristique* et la *distance dans l'espace caractéristique* pour la reconnaissance de classes visuelles. Cette section montre des exemples de reconnaissance de la classe visuelle FLEUR et de la classe visuelle CÔTE. Trois images d'une classe visuelle sont détaillées et l'espace bidimensionnel propre, couvert par les trois histogrammes correspondants de ces images, est calculé. Pour la recherche d'images visuellement similaires, la distance à l'espace caractéristique $\epsilon^2(\mathbf{x})$ est calculée pour 200 images [1] par rapport à l'espace propre bidimensionnel. La base d'images contient des images de paysages telles que des montagnes, des fleurs, des paysages et des forêts. Pour chacune des 200 images un histogramme est calculé. La *distance à l'espace caractéristique* $\epsilon^2(\mathbf{x})$ de ces histogrammes est employée pour extraire les images les plus similaires à une classe visuelle.

La combinaison de filtres est donnée par *Dx-Dy-16* (premières dérivées Gaussiennes dans les directions x et y ayant une résolution de 16 cellules par axe d'histogramme) à trois différentes échelles : $\sigma_1 = 1.0, \sigma_2 = 2.0$ et $\sigma_3 = 4.0$. Les histogrammes multidimensionnels de champs réceptifs sont alors à six dimensions. Pour chacune des 200 images un histogramme est calculé. Trois histogrammes sont utilisés pour le calcul de l'espace bidimensionnel d'une classe visuelle comme par exemple pour la classe visuelle FLEUR.

(a) *Prototype* (b) *Prototype* (c) *Prototype*

(d) *Première correspondance* (e) *Deuxième correspondance* (f) *Troisième correspondance*

(g) *Quatrième correspondance* (h) *Cinquième correspondance*

FIG. 9.1 – *Reconnaissance de fleurs. Les trois premières images (a), (b) et (c) sont utilisées pour la définition de la classe visuelle* FLEUR*; les autres images correspondent aux cinq images ayant les plus petites distances aux autres 197 images*

Les figures 9.1(a), (b) et (c) montrent trois images choisies manuellement de la base d'images pour la représentation de la classe visuelle FLEUR. En utilisant la distance à l'espace caractéristique $\epsilon^2(\mathbf{x})$, les autres 197 images peuvent être mises en ordre. Les figures 9.1(d) à (h) montrent les premières cinq images. Les quatre premières images sont toutes des images de fleurs montrant l'applicabilité de la distance à l'espace caractéristique pour la reconnaissance de la classe visuelle FLEUR. La cinquième image est visuellement similaire car la texture des herbes fait partie des trois prototypes de la classe. Il faut noter que l'on n'obtient pas la même qualité de résultats avec chacune des autres combinaisons de trois images de fleurs. Un algorithme approprié doit être défini pour l'extraction des images représentant une classe visuelle.

La figure 9.2 détaille un deuxième exemple : la reconnaissance de la classe visuelle CÔTE. Comme pour la première fois, les trois images de côtes sont choisies manuellement (voir figure 9.2 (a), (b) et (c)) et sont utilisées pour le calcul de l'espace propre bidimensionnel de la classe CÔTE. Les cinq images ayant les plus petites distances sont toutes visuellement similaires et montrent des scènes de côte.

L'appendice montre trois autres exemples de l'application de la distance à l'espace caractéristique pour la reconnaissance de classes visuelles. Ces exemples utilisent les mêmes histogrammes et la même base de 200 images. La reconnaissance de la classe visuelle de FORÊT (voir figure B.3), de la classe visuelle de GLACIER (voir figure B.4) et de la classe de CHUTES D'EAU (voir figure B.5) est présentée plus en détail. Pour ces trois cas les cinq images de distances minimales correspondent aux images visuellement similaires. Pour les images de forêts (respectivement de glaciers), les cinq images sont des scènes de forêts (respectivement de glaciers). Dans le cas de la reconnaissance de chutes d'eau, l'ensemble des trois images de chutes d'eau des 197 autres images est inclu dans les quatre premières images.

Ces résultats, dans le contexte de l'accès à une base d'images, permettent de conclure que la technique proposée, fondée sur les histogrammes multidimensionnels de champs réceptifs et sur la distance à l'espace caractéristique $\epsilon^2(\mathbf{x})$, constitue un moyen de classification d'images. La même méthode, utilisant les histogrammes d'apparences d'objets ou même les histogrammes représentant des objets 3D entiers, peut constituer un cadre puissant pour la classification d'objets. Une question importante reste l'extraction automatique des histogrammes représentatifs d'une classe visuelle. Ce sujet est discuté brièvement dans la section suivante.

9.4 L'extraction de classes visuelles

La section précédente a montré des exemples de la reconnaissance de classes visuelles en appliquant la distance à l'espace caractéristique (voir aussi la section 9.2). Les résultats soulignent la validité de l'algorithme proposé pour la reconnaissance de classes visuelles.

Pour les cas de la section précédente, les images d'exemples d'une classe visuelle ont été choisies manuellement. Ce choix manuel n'est ni souhaitable, ni faisable en général. L'extraction d'histogrammes d'une classe visuelle est maintenant discutée. Il existe au moins deux besoins complémentaires à l'extraction d'histogrammes H_i d'une classe visuelle :

– les histogrammes H_i doivent correspondre aux apparences d'objet visuellement similaires

1. La base d'images utilisée pour cette expérimentation est fournie par Jutta Kreyß de l'Université de Bremen, Allemagne

(a) *Prototype* (b) *Prototype* (c) *Prototype*

(d) *Première correspondance* (e) *Deuxième correspondance* (f) *Troisième correspondance*

(g) *Quatrième correspondance* (h) *Cinquième correspondance*

FIG. 9.2 – *Reconnaissance de côte. Les trois premières images (a), (b) et (c) sont utilisées pour la définition de la classe visuelle* CÔTE*; les autres images correspondent aux cinq images ayant les plus petites distances aux autres 197 images*

– les histogrammes H_i doivent être représentatifs de la classe visuelle pour le calcul de la représentation de l'espace propre et l'estimation de la densité probabiliste $\hat{p}(\mathbf{x}|\Omega)$

Comme illustration, la reconnaissance de la classe visuelle FLEUR peut être examinée de plus près. Les trois images de fleurs choisies manuellement correspondent aux objets visuellement similaires, car les trois images contiennent des fleurs. Ces trois images correspondent aussi à la distance maximale entre les 7 images de fleurs des figures 9.1(a) à (g). La distance maximale peut être retenue comme un critère pour l'extraction des exemples représentatifs de la classe fleur. Au total il existe 9 images de la base de 200 images qui peuvent être classifiées "fleur" par un oeil humain. Si l'une des 2 images de fleur est ajoutée (par exemple celle présentant la distance minimale dans l'espace caractéristique assurant qu'une dimension indépendante est ajoutée), toutes les images de fleurs de la base constituent les meilleures correspondances.

Cette section définit un algorithme d'extraction des histogrammes correspondant aux objets visuellement similaires. Pour obtenir des exemples représentatifs, il est nécessaire d'avoir plus d'images que le total disponible. Supposant un nombre suffisant d'exemples, la densité probabiliste $\hat{p}(\mathbf{x}|\Omega)$ peut être estimée de façon robuste. Dans ce cas, l'algorithme de reconnaissance défini par la section 9.2 peut constituer un moyen fiable de la reconnaissance de classes visuelles. Étant donné le nombre insuffisant d'exemples pour la section précédente, les images des classes visuelles ont été choisies manuellement. Ce choix manuel est typique dans le contexte de l'accès aux bases d'images [Pen 96]. L'extraction des exemples *représentatifs* est le sujet de la recherche future.

Le prochain paragraphe définit un algorithme de regroupement pour l'extraction d'histogrammes d'une base correspondant aux objets visuellement similaires.

Regroupement fondé sur l'intersection

Un algorithme de regroupement est proposé dans le contexte d'une base d'histogrammes. L'algorithme de regroupement se compose, comme la plupart, d'algorithmes de regroupement, de deux étapes :

– recherche de la paire d'histogrammes les plus similaires de la base d'histogrammes

– fusion des représentations des deux histogrammes

Des fonctions différentes peuvent être appliquées pour la détermination de la paire d'histogrammes les plus similaires. Principalement, toutes les fonctions proposées dans la section 5.1 sont utilisables. La fonction la plus performante de comparaison était la χ^2–statistique χ^2_{qv}. Malheureusement, cette fonction n'est pas métrique, rendant cette mesure inappropriée dans le contexte du regroupement. L'intersection $\cap$ d'histogrammes est proposée, car cette fonction de comparaison a aussi donné des très bons résultats. En particulier, l'intersection est robuste par rapport aux occultations partielles.

L'application de l'intersection $\cap$ pour la comparaison d'histogrammes permet la fusion de deux histogrammes, fondée sur une opération d'intersection. La fusion de deux histogrammes $Q = \cup_{\mathbf{i}} q_{\mathbf{i}}$ et $V = \cup_{\mathbf{i}} v_{\mathbf{i}}$ est définie par :

$$merge(Q, V) = \cup_{\mathbf{i}} \cap (q_{\mathbf{i}}, v_{\mathbf{i}}) = \cup_{\mathbf{i}} \min(q_{\mathbf{i}}, v_{\mathbf{i}}) \tag{9.135}$$

Cette fusion permet l'extraction des parties communes de deux histogrammes. Ces parties communes correspondent aux parties visuellement similaires des images. Pour illustrer l'impact de cette remarque, l'algorithme de regroupement est appliqué à la base des 200 images, employée dans la section précédente.

La figure 9.3 montre un exemple de regroupement obtenu par l'algorithme décrit. Les carrés blancs indiqueront les histogrammes regroupés et fusionnés par l'opération d'intersection définie plus haut. Les images regroupées correspondent aux images visuellement similaires. Ceci autorise la conclusion que l'algorithme de regroupement proposé permet l'extraction des histogrammes de classes visuelles. Comme mentionné plus haut, l'extraction des histogrammes *représentatifs* est très importante et constitue le sujet de la recherche future.

9.5 Conclusion

Le chapitre a introduit le concept de classes visuelles en tant que cadre général de la classification d'objets. Centré autour de la représentation statistique d'objets, le chapitre a développé l'extraction et la représentation de classes visuelles. Les classes visuelles sont définies par des similitudes en 2D et/ou en 3D des apparences d'objets. Les similitudes en 2D correspondent aux apparences visuellement proches. Les similitudes en 3D sont plus restrictives car la forme 3D et la texture de toutes les surfaces de deux objets doivent être similaires. Comme les classes visuelles sont fondées seulement sur des mesures d'images, elles peuvent être extraites de façon fiable. De plus les classes visuelles ne dépendent pas du contexte. Contrairement aux classes visuelles, les classes d'objets dépendent, en général, du contexte (spatial et temporel). Notre argument est alors que le concept de classes visuelles constitue un cadre puissant pour la classification d'objets. Ce chapitre donne une extension de l'application des histogrammes multidimensionnels de champs réceptifs à la reconnaissance de classes visuelles comme sous–problème de la classification d'objets.

Le chapitre a proposé un algorithme de reconnaissance de classes visuelles selon le maximum de vraisemblance. Une application de l'algorithme dans le contexte de l'accès aux bases d'images démontre la validité du concept. Des cas particuliers sont décrits, par exemple, pour l'accès aux images de côte d'une base de 200 images. La question importante de l'extraction d'histogrammes correspondant à une classe visuelle est discutée. Un algorithme de regroupement est défini et appliqué. Néanmoins, la question de l'extraction d'histogrammes représentatifs pour une classe visuelle est le sujet de la recherche future.

D'un point de vue plus général, ce chapitre a proposé le concept de classes visuelles pour la *classification d'objets fondée sur l'apparence*. Comme énoncé dans l'introduction de la thèse, la classification d'objets constitue l'un des principaux défis pour les modèles fondés sur l'apparence. Le concept de classes visuelles proposé dans ce chapitre peut combler le vide qui existe entre les modèles fondés sur l'apparence et un classificateur général d'objets.

FIG. 9.3 – *Images regroupées. Les images de côtes sont montrées regroupées automatiquement, dans la base de 200 images de paysages*

Chapitre 10

Conclusions et perspectives

La fiabilité et de la rapidité de l'approche de Swain et Ballard, s'appuyant sur l'utilisation d'histogrammes de couleurs, ont guidé la présente étude. Le but de ce travail était de généraliser leur technique. La méthode généralisée utilise les statistiques de vecteurs de champs réceptifs. Cette généralisation permet une formulation globale de la représentation statistique d'objets. La représentation fondée sur les histogrammes multidimensionnels de champs réceptifs est justifiée, en tant que schéma particulier. Différents algorithmes de l'identification et de la classification d'objets sont proposés et évalués expérimentalement. Les algorithmes de reconnaissance par comparaison d'histogrammes ainsi que de reconnaissance probabiliste et de reconnaissance active, sont analysés en détail. Pour la classification d'objets, cette recherche propose un algorithme selon le maximum de vraisemblance.

Pour guider notre étude, deux références ont été choisies : d'une part, les *degrés de liberté de la reconnaissance d'objets* définis par la section 1.2 ont guidé le développement de la représentation statistique d'objets et l'évaluation des algorithmes de reconnaissance, d'autre part, les *défis des modèles fondés sur l'apparence*, résumés par la section 1.4, ont été la base des extensions de l'approche.

10.1 Résultats principaux

Cette section se consacre aux résultats et conclusions de cette thèse. Les trois investigations et résultats principaux de la thèse sont les suivants :

– Les vecteurs de caractéristiques locales et, en particulier, les statistiques de ces vecteurs

constituent un moyen fiable pour la représentation et la reconnaissance d'objets.

- L'algorithme probabiliste de reconnaissance, défini dans le chapitre 7, est capable de reconnaître les objets indépendamment de la mise en correspondance.
- Le schéma de la classification d'objets, fondé sur l'apparence, était proposé et appliqué dans le contexte d'accès à une base d'images par leur contenu.

L'ensemble de ces points est de l'intérêt général pour le domaine de la vision par ordinateur. Le premier point souligne le fait que la représentation d'objets, fondée sur les vecteurs de champs réceptifs, est extrêmement discriminante et fiable. Ce résultat peut être utilisé directement par d'autres chercheurs. Le deuxième point est intéressant, car la plupart des algorithmes de reconnaissance utilisent la mise en correspondance entre l'image test et les objets de la base. Comme cette mise en correspondance est, en général, difficile et requiert de lourds calculs, l'algorithme proposé permet le calcul d'hypothèse d'objets. Le troisième point constitue une critique fondamentale des modèles fondés sur l'apparence. Le chapitre 9 définit une formulation de la classification fondée sur l'apparence et fournit des résultats indiquant l'applicabilité de la formulation pour la classification d'objets. Une des perspectives de la thèse est alors cette extension de l'approche à la classification d'objets fondée sur l'apparence.

A côté de ces trois points, les points suivants doivent être cités comme résultats de la thèse :

- Le développement d'une analogie entre la reconnaissance d'objets et la théorie de l'information. Cette analogie permet l'évaluation d'un ensemble d'opérateurs locaux dans le contexte d'une tâche particulière de reconnaissance.
- L'algorithme actif de reconnaissance, défini dans le chapitre 8, permet la reconnaissance d'objets 3D en utilisant les informations bidimensionnelles de différents points de vue. C'est à dire que les informations 3D contenues dans l'ensemble des histogrammes multidimensionnels de champs réceptifs d'un objet peuvent être employées d'une manière implicite, sans l'extraction explicite des informations 3D.
- La technique originale de la comparaison d'histogrammes de couleur, a été généralisée à la comparaison d'histogrammes multidimensionnels de champs réceptifs. Plusieurs fonctions de comparaison ont été proposées, analysées et appliquées pour la reconnaissance de 103 objets en présence de rotation dans le plan image, de changements d'échelle et de variations du point de vue.

10.2 Perspectives

Les chapitres 8 et 9 exposent deux extensions de l'approche. Ces deux extensions font partie des perspectives les plus intéressantes de la thèse :

La reconnaissance active d'objets Le chapitre 8 introduit les concepts et décrit les résultats préliminaires de la reconnaissance active d'objets. Notre équipe ne possédant pas de tête active de caméra appropriée, nous n'avons pas pu continuer les expérimentations. Il est intéressant d'appliquer les principes développés à l'exploration active par un robot, ou à la reconnaissance et le suivi d'objets dans le contexte de l'interaction homme–machine.

La classification d'objets fondée sur l'apparence Les applications d'une classification d'objets fondée sur l'apparence sont nombreuses. De ces applications on peut citer l'accès aux bases d'images par leur contenu, la modélisation automatique d'objets et la généralisation à partir d'exemples. Les propositions du chapitre 9 doivent être considérées comme un premier essai de la classification d'objet fondée sur l'apparence. Les techniques d'apprentissage diverses doivent être examinées afin d'obtenir un algorithme général pour l'extraction des classes visuelles à partir de la représentation statistique de vecteurs de caractéristiques locales.

L'extension à d'autres caractéristiques locales La plupart des expérimentations décrites dans la thèse se basent sur les vecteurs de dérivées Gaussiennes, employés comme caractéristiques locales d'objets. Une extension directe de l'approche applique les descripteurs locaux fondés sur des invariances de couleurs introduites par la section 3.1.3. Néanmoins, ces vecteurs sont inadaptés à la représentation d'objets polyédriques ayant des surfaces lisses. Il est alors logique d'exploiter la technique en utilisant différents types de caractéristiques locales comme des moments, des descripteurs invariants et des descripteurs de contour.

L'extension à d'autres caractéristiques permet le calcul d'hypothèses d'objets venant de différents types et de différents ensembles de caractéristiques. La formulation probabiliste permet l'incorporation d'informations provenant de différentes sources d'une manière facile et élégante. On peut imaginer aussi une hiérarchie de caractéristiques où chaque niveau de la hiérarchie calcule une hypothèse d'objet fondée sur un ensemble de caractéristiques appropriées.

Le regroupement de représentations Le regroupement des histogrammes multidimensionnels de champs réceptifs permet la réduction de la consommation de mémoire par objet. Cette réduction est particulièrement intéressante pour les grandes bases d'objets. Dans le contexte de l'accès aux bases d'images par leur contenu, ce regroupement permet un accès hiérarchique réduisant la mémoire et le temps d'exécution. Le chapitre 9 emploie la transformation de Karhunen–Loeve pour la représentation compacte d'histogrammes. Ce dernier ainsi que d'autres schémas de la réduction de dimensions peuvent être appliqués pour réduire la consommation de mémoire.

La reconnaissance d'objets utilisant plusieurs capteurs Comme mentionné dans l'introduction, la formulation statistique permet l'incorporation d'informations provenant de différentes sources. Ces informations sont issues des connaissances a priori, des différents ensembles de caractéristiques locales et des différents capteurs, comme des caméras ou des télémètres à laser. La formulation statistique permet l'incorporation de toutes ces informations de différentes sources d'une façon cohérente et élégante.

Appendix A

Les bases d'images

Les bases d'images suivantes ont été utilisées dans les différents chapitres :

1. 30 objets à six échelles : figure A.1
2. 22 objets à 18 rotations d'image et à trois échelles : figure A.2
3. 31 objets supplémentaires : figure A.3
4. 100 images aériennes de Marseille, fournies par l'entreprise française ISTAR. Nous voulons aussi remercier Roger Mohr que a rendu possible l'utilisation de ces images : figures A.4 et A.5
5. base d'images de Columbia de 100 objets : figure A.6

FIG. A.1 – *30 × 6 images d'échelles : pour chaque objet nous avons pris six images d'échelles différentes. Le facteur d'échelle entre la première et la dernière image est approximativement 2.1. La figure 6.4 montre six images de deux objets*

FIG. A.2 – 22×18×3 *images de rotations d'image et images d'échelles : pour 22 objets nous avons pris 18 images correspondant à différentes rotations d'image. Pour chacune de ces images nous avons pris trois images correspondant à différentes échelles. Le facteur d'échelle entre les trois images est approximativement* ±1.75. *Les trois échelles d'un objet d'une rotation particulière sont montrées dans la figure 6.5*

FIG. A.3 – *31 × 2 images de 31 objets. Les deux images par objets diffèrent légèrement en niveau d'illumination, de mise au point, d'orientation et d'échelle*

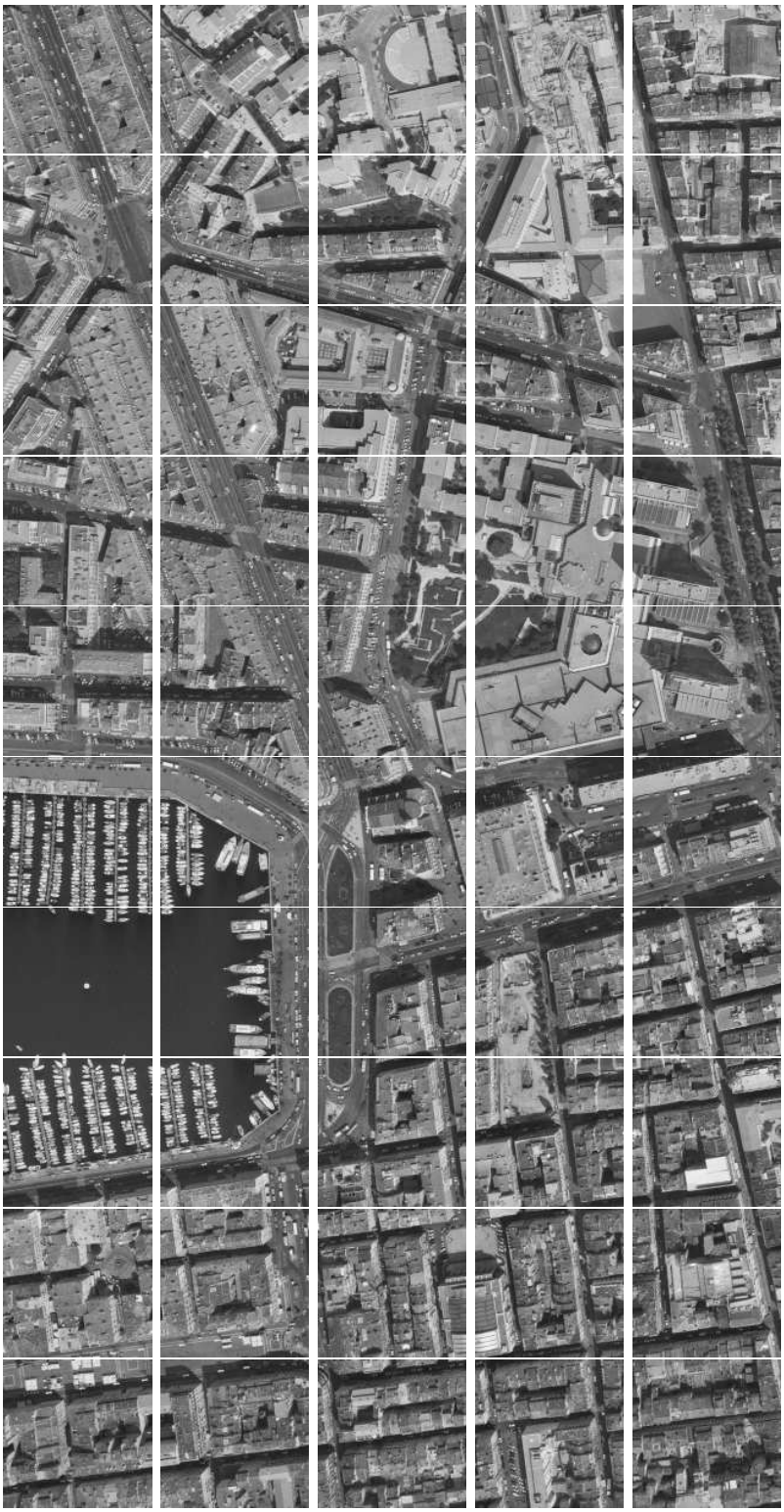


FIG. A.4 – *La première moitié des 100 images aériennes de Marseille, fournies par ISTAR, France. Il existe trois autres séries de 100 images de la même région de Marseille, correspondant aux changements de point de vue causés par le mouvement de l'avion*

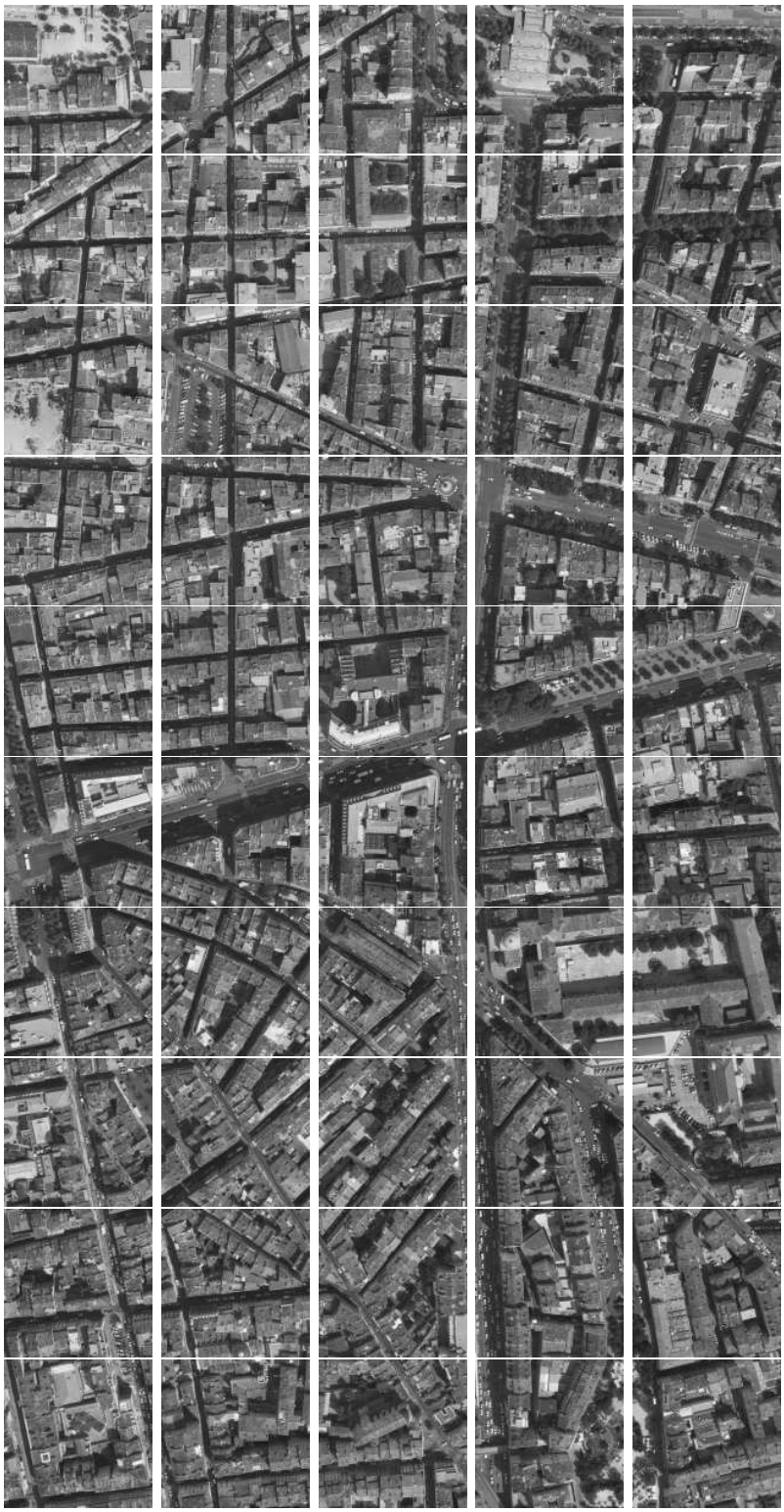


FIG. A.5 – *La deuxième moitié de 100 image aériennes de Marseille, fournies par ISTAR, France. Voir pour des commentaires la figure A.4*

FIG. A.6 – *100 × 72 images de 100 objets de la base d'images de Columbia. Les 72 images par objet correspondent aux différents points de vue d'une différence de* 5° *entre chaque*

Appendix B

Collection de résultats supplémentaires

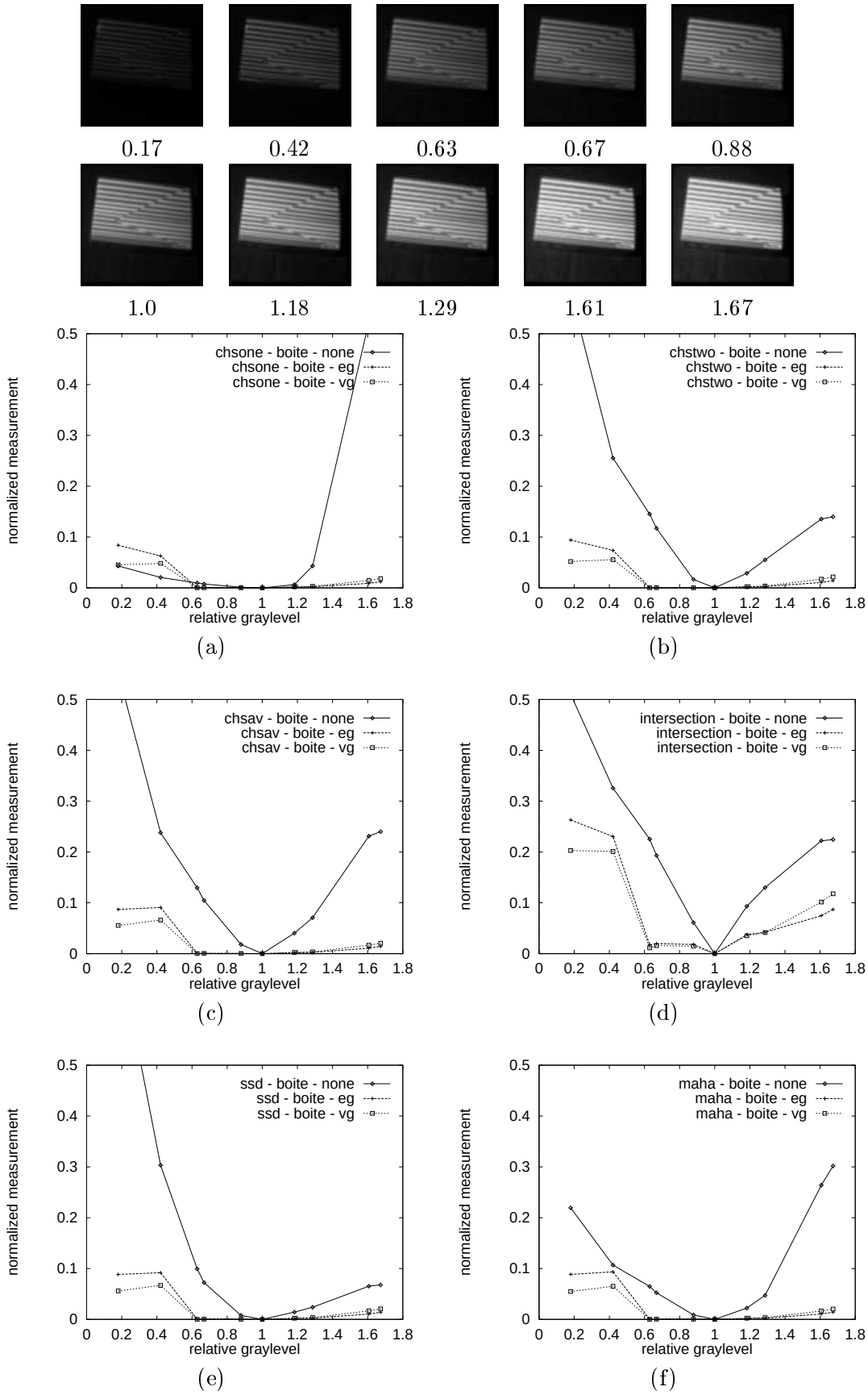


FIG. B.1 – *Changements de l'intensité d'éclairage de la série d'images "boite" et la stabilité de différentes techniques de normalisation et de différentes fonctions de comparaison d'histogrammes*

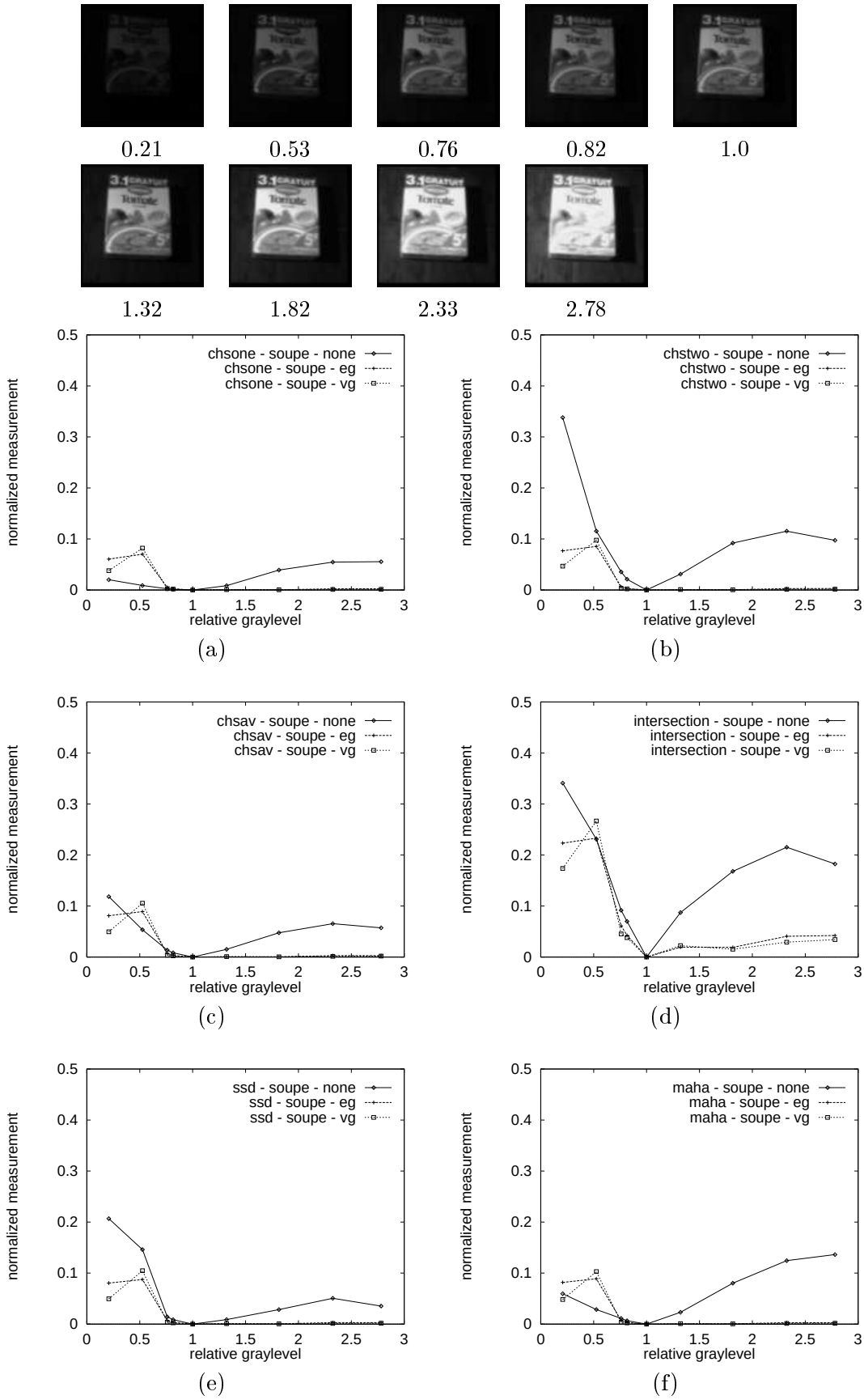


FIG. B.2 – *Changements de l'intensité d'éclairage de la série d'images "soupe" et la stabilité de différentes techniques de normalisation et de différentes fonctions de comparaison d'histogrammes*

(a) *Prototype* (b) *Prototype* (c) *Prototype*

(d) *Première correspondance* (e) *Deuxième correspondance* (f) *Troisième correspondance*

(g) *Quatrième correspondance*

(h) *Cinquième correspondance*

FIG. B.3 – *Reconnaissance de forêt*

(a) *Prototype*

(b) *Prototype*

(c) *Prototype*

(d) *Première correspondance*

(e) *Deuxième correspondance*

(f) *Troisième correspondance*

(g) *Quatrième correspondance*

(h) *Cinquième correspondance*

FIG. B.4 – *Reconnaissance de glacier*

(a) *Prototype*

(b) *Prototype*

(c) *Prototype*

(d) *Première correspondance*

(e) *Deuxième correspondance*

(f) *Troisième correspondance*

(g) *Quatrième correspondance*

(h) *Cinquième correspondance*

FIG. B.5 – *Reconnaissance de chutes d'eau*

Bibliographie

[Bal 82] D.H. Ballard and C.M. Brown. *Computer Vision*. Prentice Hall, 1982.

[Bal 93] D.H. Ballard and L.E. Wixson. Object recognition using steerable filters at multiple scales. In *IEEE Workshop on Qualitative Vision*, pages 2–10, June 1993.

[Bal 94] D.H. Ballard and R.P.N. Rao. Seeing behind occlusions. In *ECCV'94 Third European Conference on Computer Vision, Volume I*, pages 274–285, 1994.

[Bar 96] K. Barnard, G. Finlayson, and B. Funt. Colour constancy for scences with varying illumination. In *ECCV'96 Fourth European Conference on Computer Vision, Volume II*, pages 3–15, April 1996.

[Beb 95] G. Bebis, M. Goergiopoulos, and N. da Vitoria Lobo. Learning geometric hashing functions for model–based object recognition. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 543–548, 1995.

[Bei 94] J.S. Beis and D.G. Lowe. Learning indexing functions for 3-d model based object recognition. In *International Conference on Computer Vision and Pattern Recognition*, pages 275–280, 1994.

[Bel 89] Z.W. Bell. A bayesian/monte carlo segmentation method for images dominated by gaussian noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(9): 985–990, September 1989.

[Bel 96] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegmann. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In *ECCV'96 Fourth European Conference on Computer Vision, Volume I*, pages 45–58, 1996.

[Bel 97] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Recognition of images in large databases using color and texture. submitted to *International Conference on Computer Vision and Pattern Recognition*, 1997.

[Bla 96] M.J. Black and A.D. Jepson. Eigentracking: robust matching and tracking of articulated objects using a view–based representation. In *ECCV'96 Fourth European Conference on Computer Vision, Volume I*, pages 329–342, 1996.

[Bob 95] P. Bobet. *Tête stéréoscopique, Réflexes oculaires et Vision.* PhD thesis, INPG France, 1995.

[Bre 93] T.M. Breuel. Higher–order statistics in object recognition. In *International Conference on Computer Vision and Pattern Recognition*, pages 707–708, 1993.

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

[Bur 92] H. Burkhardt and A. Zisserman, editors. *Invariants for Recognition*, 1992. ESPRIT–Basic-Research-Workshop, ECCV'92.

[Cal 93] A. Califano and R. Mohan. Systematic design of indexing strategies for object recognition. In *International Conference on Computer Vision and Pattern Recognition*, pages 709–710, 1993.

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

[Cle 91] D.T. Clemens and D.W. Jacobs. Space and time bounds of indexing 3-d models from 2-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10): 1007–1017, 1991.

[Col 96] V. Colin de Verdière. Reconnaissance d'Objets par leurs Statistiques de Couleurs. Master's thesis, GRAVIR, I.N.P. Grenoble, 1996. DEA d'Imagerie, Vision et Robotique, in French.

[Cro 96] J.L. Crowley. Multi–modal tracking for video compression. In *Proceedings of the 4th International Symposium in Intelligent Robotic Systems*, pages 317–324, July 1996.

[Dau 93] J.G. Daugman. High confidence visual recognition of persons by test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11): 1148–1161, November 1993.

[Der 93] R. Deriche. Recursively implementing the gaussian and its derivatives. Technical Report 1893, INRIA–Sophia Antipolis, April 1993.

[Ede 93] S. Edelman. On learning to recognize 3-d objects from examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(8): 833–837, 1993.

[Egg 93] D.W. Eggert, K.W. Bowyer, C.R. Dyer, H.I. Christensen, and D.B. Goldgof. The scale space aspect graph. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11): 1114–1130, 1993.

[Enn 93] F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. In *International Conference on Computer Vision and Pattern Recognition*, pages 711–712, 1993.

[Enn 95] F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8): 805–809, 1995.

[Fau 92] O.D. Faugeras, J.L. Mundy, N. Ahuja, C.R. Dyer, A.P. Pentland, R. Jain, K. Ikeuchi, and K.W. Bowyer. Why aspect graphs are not (yet) practical for computer vision. *CVGIP*, 55(2): 212–218, 1992.

[Fin 93] G.D. Finlayson, M.S. Drew, and B.V. Funt. Diagonal transforms suffice for color constancy. In *ICCV'93 Fourth International Conference on Computer Vision*, pages 164–171, 1993.

[Fin 95a] G.D. Finlayson. Color constancy in diagonal chromatic space. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 218–223, 1995.

[Fin 95b] G.D. Finlayson, B.V. Funt, and K. Barnard. Color constancy under varying illumination. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 720–725, 1995.

[Fin 96] G.D. Finlayson, S.C. Chatterjee, and B.V. Funt. Color angular indexing. In *ECCV'96 Fourth European Conference on Computer Vision, Volume II*, pages 16–27, 1996.

[FL 95] G. Funka-Lea and R. Bajcsy. Combining color and geometry for the active, visual recognition of shadows. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 203–209, 1995.

[Fli 95] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Juang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, pages 23–32, September 1995.

[Flo 91] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever. General intensity transformations and second order invariants. In *Proceedings of the 7th Scandinavian Conference on Image Analysis*, pages 338–345, 1991.

[For 90] D.A. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1): 5–36, 1990.

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

[Fuk 90] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Computer Science and Scientific Computing. Academic Press, New York, 2nd edition, 1990.

[Fun 91] B.V. Funt, M.S. Drew, and J. Ho. Color constancy from mutual reflection. *International Journal of Computer Vision*, 6(1): 5–24, 1991.

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

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

[Gor 94] M.M. Gorkani and R.W. Picard. Texture orientation for sorting photos "at a glance". In *ICPR'94 Proceedings of the International Conference on Pattern Recognition*, pages 259–464, October 1994.

[Gri 90] W.E.L. Grimson and D.P. Huttenlocher. On the sensitivity of the hough transform for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3): 255–274, March 1990.

[Gri 91] W.E.L. Grimson and D. Huttenlocher, editors. Special issue on interpretation of 3-d scenes – part i. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10), October 1991.

[Gri 92] W.E.L. Grimson and D. Huttenlocher, editors. Special issue on interpretation of 3-d scenes – part ii. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), February 1992.

[Gui 96] L.J. Guibas and C. Tomasi. Image retrieval and robot vision research at stanford. In *ARPA Image Understanding Workshop*, 1996.

[Haf 95] J. Hafner, H.S. Sawhney, W. Equitz, M. Flickner, and W. Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7): 729–736, July 1995.

[Hea 94] G. Healey and D. Slater. Using illumination invariant color histogram descriptors for recognition. In *International Conference on Computer Vision and Pattern Recognition*, pages 355–360, 1994.

[Hor 95] J. Hornegger and H. Niemann. Statistical learning, localization and identification of objects. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 914–919, 1995.

[Hor 96] J. Hornegger. *Statistische Modellierung, Klassifikation und Lokalisierung von Objekten.* Shaker Verlag GmbH, 1996. PhD thesis, in German, see also [Hor 95].

[Hun 94] M. Hunke. Locating and tracking of human faces with neural networks. Technical Report CMU–CS–94–155, Carnegie Mellon University, August 1994.

[Hut 87] D.P. Huttenlocher and S. Ullman. Object recognition using alignment. In *ICCV'87 First International Conference on Computer Vision*, 1987.

[Hut 95] D. Huttenlocher and E.W. Jaquith. Computing visual correspondence: Incorporating the probability of a false match. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 515–520, 1995.

[Int 93] N. Intrator and J.I. Gold. Three–dimensional object recognition using an usupervised bcm network: The usefulness of distinguishing features. *Neural Computation*, 5: 61–74, 1993.

[Jon 92] D.G. Jones and J. Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *ECCV'92 Second European Conference on Computer Vision*, pages 395–410, 1992.

[Kan 78] T. Kanade. Region segmentation: Signal vs. semantics. In *Proceedings of the 4th International Conference on Pattern Recognition*, pages 95–105, 1978.

[Kir 90] M. Kirby and L. Sirovich. Application of the karhunen–loève procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1): 103–108, January 1990.

[Koe 79] J.J. Koenderink and A.J. 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 image. *Biological Cybernetics*, 50: 363–396, 1984.

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

[Koe 91] J.J. Koenderink. Local image structure. In *Proceedings of the 7th Scandinavian Conference on Image Analysis*, pages 1–7, 1991.

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

[Lam 88b] Y. Lamdan and H.J. Wolfson. Geometric hashing: A general and efficient model based recognition scheme. In *ICCV'88 Second International Conference on Computer Vision*, pages 238–249, 1988.

[Lam 88c] Y. Lamdan and H.J. Wolfson. On recognition of 3-d objects from 2-d images. In *International Conference on Robotics and Automation*, pages 1407–1413, 1988.

[Lam 96] B. Lamiroy and P. Gros. Rapid object indexing and recognition using enhanced geometric hashing. In *ECCV'96 Fourth European Conference on Computer Vision, Volume I*, pages 59–70, 1996.

[Mal 85] L.T. Maloney. *Computional approaches to color constancy.* PhD thesis, Stanford Unversity, 1985.

[Mal 86] L.T. Maloney and B.A. Wandell. Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*, 3(1): 29–33, January 1986. Interest: Color Constancy, CV, Co.

[Mal 89] J. Malik and P. Perona. A computational model of texture segmentation. In *International Conference on Computer Vision and Pattern Recognition*, pages 326–332, 1989.

[Mar 78] D. Marr. Representing visual information – a computational approach. In A.R. Hanson and E.M. Riseman, editors, *Computer Vision Systems*, pages 61–80. Academic Press New York, 1978.

[Mar 97] J. Martin and J.L. Crowley. An appearance-based approach to gesture-recognition. to appear in ICIAP'97, Florence, Italy, September 1997.

[Mat 95] J. Matas, R. Marik, and J. Kittler. On representation and matching of multi–colored objects. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 726–732, 1995.

[Mel 96] B.W. Mel. SEEMORE: Combining color shape, and texture histogramming in a neurally-inspired approach to visual object recognition. Technical report, Department of Biomedical Engineering, University of Southern California, MC 1451, 1996. appeared in *Proceedings of International Conference of Pattern Recognition 1996.*

[Mog 95] B. Moghaddam and A. Pentland. Maximum likelihood detection of faces and hands. In *International Workshop on Automatic Face– and Gesture–Recognition*, pages 122–128, 1995.

[Moh 97] R. Mohr, S. Picard, and C. Schmid. Bayesian decision versus voting for image retrieval. submitted to CAIP'97, Kiel, Germany, 1997.

[Mun 92] J. L. Mundy and Andrew Zisserman, editors. *Geometric Invariance in Computer Vision.* MIT Press, 1992.

[Mun 93] J. L. Mundy, A. Zisserman, and D. Forsyth, editors. *Application of Invariance in Computer Vision.* Number 825 in Lecture Notes in Computer Science. Springer Verlag, 1993.

[Mur 93] H. Murase and S. K. Nayar. Learning and recognition of 3-d objects form brightness images. In *AAAI Fall Symposium, Machine Learning in Computer Vision*, pages 25–29, October 1993.

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

[Nag 92] H.-H. Nagel. Reflections on active (machine) vision. In A.K. Sood and H. Wechsler, editors, *Active Perception and Robot Vision*, pages 23–42. Springer Verlag, 1992.

[Nag 95] K. Nagao. Recognizing 3d objects using photometric invariants. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 480–487, 1995.

[Nel 95] R.C. Nelson. 3-d recognition via 2-stage associative memory. Technical Report TR 565, Department of Computer Science, University of Rochester, January 1995.

[Nel 96] R.C. Nelson. Memory-based recognition for 3-d objects. In *ARPA, Image Understanding Workshop*, pages 1305–1310, February 1996.

[Nen 96] S.A. Nene, S.K. Nayar, and H. Murase. Columbia object image library (coil-100). Technical Report CUCS-006-96, Department of Computer Science, Columbia University, 1996.

[Obj 96] International workshop on object representation for computer vision. Cambridge, England, April 1996.

[Ohb 96] K. Ohba and K. Ikeuchi. Recognition of the multi specularity objects for bin–picking task. In *IROS'96 Intelligent Robots and Systems*, pages 1440–1447, Osaka, Japan, 1996.

[Pen 94] A. Pentland, B. Moghaddam, and T. Starner. View–based and modular eigenspaces for face recognition. In *International Conference on Computer Vision and Pattern Recognition*, 1994. see also M.I.T. Media Laboratory TR 245.

[Pen 96] A. Pentland, R.W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3): 233–254, 1996.

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

[Pic 96] R.W. Picard. A society of models for video and image libraries. Technical Report 360, M.I.T. Media Laboratory, 1996. submitted to IBM Systems Journal.

[Pog 90] T. Poggio and S. Edelman. A network that learns to recognize 3d objects. *Nature*, 343, January 1990.

[Pop 93] A.R. Pope and D.G. Lowe. Learning object recognition models from images. In *ICCV'93 Fourth International Conference on Computer Vision*, pages 296–301, 1993.

[Pop 94] K. Popat and R.W. Picard. Cluster–based probability model applied to image restoration and compression. In *IEEE Conference on Acoustics, Speech and Signal Processing*, Adeline, Australia, April 1994. also M.I.T. Media Laboratory TR 253.

[Pop 95] A.R. Pope. *Learning to Recognize Objects in Images: Acquiring and Using Probabilistic Models of Appearance*. PhD thesis, Department of Computer Science, University of British Columbia, 1995.

[Pop 96] A.R. Pope and D.G. Lowe. Learning appearance models for object recognition. In *International Workshop on Object Representation for Computer Vision*, Cambridge, England, April 1996.

[Pre 92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[Rao 95a] R.P.N. Rao and D.H. Ballard. An active vision architecture based on iconic representations. *Artificial Intelligence (Special Issue on Vision)*, 78: 461–505, 1995.

[Rao 95b] R.P.N. Rao and D.H. Ballard. Object indexing using an iconic sparse distributed memory. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 24–31, 1995.

[Rez 95] F.M. Reza. *An Introduction to Information Theory.* Dover Publications, New York, 1995.

[Rig 93] I. Rigoutsos and R. Hummel. Distributed bayesian object recognition. In *International Conference on Computer Vision and Pattern Recognition*, pages 180–186, 1993.

[Rom 91] B.M. ter Haar Romeny, L.K.J. Florack, J.J. Koenderink, and M.A. Viergever. Invariant third order properties of isophotes: T-junction detection. In *Proceedings of the 7th Scandinavian Conference on Image Analysis*, pages 346–353, 1991.

[Ros 84] A. Rosenfeld. Image analysis: Problems progress and prospects. *Pattern Recognition*, 17(1): 3–12, 1984.

[Sch 95a] B. Schiele and J.L. Crowley. Object recognition using multidimensional receptive field histograms and its robustness to view point changes. Presented at *1995 Rosenon Workshop on Computational Vision*, July 1995.

[Sch 95b] B. Schiele and A. Waibel. Estimation of the head orientation based on a face–color–intensifier. In *3rd International Symposium on Intelligent Robotic Systems '95*, pages 339–346, 10–14 July 1995.

[Sch 95c] B. Schiele and A. Waibel. Gaze–tracking based on face–color. In *IWAFGR 95, International Workshop on Automatic Face–and Gesture–Recognition*, pages 344–349, June 1995.

[Sch 96a] B. Schiele and J.L. Crowley. Object recognition using multidimensional receptive field histograms. In *ECCV'96 Fourth European Conference on Computer Vision, Volume I*, pages 610–619, 14–16 April 1996.

[Sch 96b] B. Schiele and J.L. Crowley. Probabilistic object recognition using multidimensional receptive field histograms. In *ICPR'96 Proceedings of the 13th International Conference on Pattern Recognition, Volume B*, pages 50–54, August 1996.

[Sch 96c] B. Schiele and J.L. Crowley. The robustness of object recognition to rotation using multidimensional receptive field histograms. submitted. available via www[1], 1996.

[Sch 96d] B. Schiele and J.L. Crowley. Where to look next and what to look for. In *4th International Symposium on Intelligent Robotic Systems'96*, pages 139–146, July 1996.

[Sch 96e] B. Schiele and J.L. Crowley. Where to look next and what to look for. In *IROS'96 Intelligent Robots and Systems*, pages 1249–1255, Osaka, Japan, November 1996.

[Sch 96f] B. Schiele and J.L. Crowley. The robustness of object recognition to view point changes using multidimensional receptive field histograms. Presented at *ECIS-VAP meeting, Object Recognition Day.* Available via WWW[2], March 1996.

1. http://pandora.imag.fr/Prima/schiele/

2. http://pandora.imag.fr/Prima/schiele/

[Sch 96g] B. Schiele and J.L. Crowley. Reconnaissance des objets utilisant des histogrammes multidimensionnels des champs réceptive. In *Journée Orasis'96, Clermont–Ferrand, France*, pages 7–12, May 1996.

[Sch 96h] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris*. PhD thesis, I.N.P.Grenoble, 1996.

[Sch 96i] C. Schmid and R. Mohr. Combining grayvalue invariants with local constraints for object recognition. In *International Conference on Computer Vision and Pattern Recognition*, 1996.

[Sch 97a] B. Schiele and J.L. Crowley. The concept of visual classes for object classification. In *SCIA'97, Proceedings of the Scandinavian Conference on Image Analysis*, pages 43–50, June 1997.

[Sch 97b] B. Schiele and J.L. Crowley. Transinformation of object recognition and its application to viewpoint planning. *Robotics and Autonomous Systems*, 1997. To appear.

[Sch 98] B. Schiele and J.L. Crowley. Transinformation for active object recognition. In *ICCV'98 Sixth International Conference on Computer Vision*, January 1998. accepted.

[Shv 90] H. Shvaytser. Learnable and nonlearnable visual concepts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5): 459–466, 1990.

[Sir 87] L. Sirovich and M. Kirby. Low–dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, 4(3): 519–524, March 1987.

[Sla 95] D. Slater and G. Healey. Combining color and geometric information for the illumination invariant recognition of 3d objects. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 563–568, 1995.

[Str 92] M.A. Stricker. Color and geometry as cues for indexing. Technical Report CS 92–22, University of Chicago, November 1992.

[Str 94] M. Stricker and M. Swain. The capacity of color histogram indexing. In *International Conference on Computer Vision and Pattern Recognition*, pages 704–708, 1994.

[Swa 90] M.J. Swain. Color indexing. Technical Report 360, University of Rochester Computer Science, November 1990.

[Swa 91] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1): 11–32, 1991.

[Swa 93a] M.J. Swain. Interactive indexing into image databases. In *SPIE, Storage and Retrieval for Image and Video Databases*, 1993.

[Swa 93b] M.J. Swain and M.A. Stricker. Promising directions in active vision. *International Journal of Computer Vision*, 11(2): 109–126, 1993.

[Tom 94] C. Tomasi and L.J. Guibas. Image descriptors for browsing and retrieval. In *ARPA Image Understanding Workshop*, November 1994.

[Tur 91a] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1): 71–86, 1991.

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

[Wei 94] C.F.R. Weiman. Efficient discrete gabor functions for robot vision. In *SPIE Vol. 2242 Wavelet Applications*, pages 148–160, April 1994.

[Wes 92] C.-J. Westelius. *Preattentive Gaze Control for Robot Vision.* PhD thesis, Department of Electrical Engineering, Linköping University, 1992.

[Wol 90] H.J. Wolfson. Model-based object recognition by geometric hashing. In *ECCV'90 First European Conference on Computer Vision*, pages 526–536, 1990.

[You 86] R.A. Young. Simulation of human retinal function with the gaussian derivative model. In *International Conference on Computer Vision and Pattern Recognition*, pages 564–569, 1986.