

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université El Hadj Lakhdar – BATNA

Faculté des sciences de
l'ingénieur



Département
d'informatique

N° d'ordre :

Série :

Mémoire présenté en vue de l'obtention du diplôme de :
Magister en Informatique

Option : Informatique Industrielle

Classification et Reconnaissance d'Objets en Utilisant les Points d'Intérêt

Présenté le : /.... /2009

Par : LAKEHAL Elkhamssa

Composition du jury :

Dr. BELAATAR Brahim	Président	(Maître de Conférence à l'Université de Batna).
Dr. YAHIAOUI Itheri	Rapporteur	(Maître de Conférence à l'Université de Reims).
Dr. CHAOUI Allaoua	Examineur	(Maître de Conférence à l'Université de Constantine).
Dr. KAZAR Okba	Examineur	(Maître de Conférence à l'Université de Biskra).

Année Universitaire : 2008/2009

Résumé

Ce mémoire, inscrit dans le domaine de la reconnaissance d'objets et plus particulièrement les descripteurs locaux, présente un nouveau détecteur des points d'intérêt, que nous appellerons points de focus. La conclusion à laquelle ont abouti plusieurs études psychologiques est que la symétrie attire l'attention. En conséquence, la détection des centres de symétrie donne des points ayant une grande importance par rapport au système visuel humain et sont donc plus appropriés à la reconnaissance d'objets. L'objectif de beaucoup de travaux est la recherche des points d'intérêt dans les fortes variations du signal. Malheureusement, en l'absence de ces variations les approches s'avèrent inefficaces. Par contre, l'utilisation des centres de symétrie se présente comme une solution intéressante. Le but de l'exploitation de ce nouveau détecteur est l'extraction de primitives de focus qui ont une signification visuelle. En prenant comme point de départ les travaux cités dans la littérature, nous apportons une amélioration consistant en l'utilisation de la couleur pour enrichir la détection et la robustesse des points détectés.

Abstract

This thesis, subscribed in the image retrieval domain and particularly the local descriptor one, presents a new points of interest' detector, which we call points of focus. Many psychological works state that symmetry attracts the attention of human visual system. This fact lead to the choice of symmetry centers as points having a high visual interest and then suitable for object recognition. The objective of many works is the search for points of interest in high signal variations. Unfortunately, with images presenting very limited signal variations these approaches seem to be useless. In contrast, the use of symmetry centers is a promising solution. The goal of this new detector is the extraction of primitives of focus that have a visual signification according to the human visual system. Taken as a starting point works cited in the literature, we provide an enhancement which consist of the color use to improve the detection process and the robustness of the detected points.

Remerciements

Ce mémoire présente le fruit de mes travaux de recherche, qui se sont déroulés entre juin 2007 et janvier 2009. Avant d'entrer dans le vif du sujet, j'aimerais profiter de cet espace pour remercier les personnes qui ont contribué à ce projet.

Pour commencer, je remercie avant tout le bon Dieu le tout puissant de m'avoir dotée d'une patience et d'une volonté pour achever mon projet.

Je dois un grand merci à ma directrice de recherche, Itheri Yahiaoui. D'abord, pour m'avoir introduit à cet intéressant domaine de recherche qu'est la reconnaissance d'objets. Aussi, je la remercie pour son aide inestimable et pour son support continu.

Merci ensuite aux gens qui m'ont doté leur temps pour lire attentivement ce mémoire et qui ont contribué de près ou de loin au bon déroulement de ce travail. Les gens concernés sont nombreux, j'espère qu'ils se reconnaîtront, car j'en oublierais certainement si j'entreprenais de les nommer individuellement. La somme des contributions de chacun a représenté une aide précieuse à l'avancement de mon projet.

Remerciements aux examinateurs, Belaatar B., Kazar O. et Chaoui A., pour avoir pris le temps de réviser ce mémoire.

Table des matières

Résumé	1
Abstract	2
Remerciements.....	3
Table des matières	4
Liste des figures	6
Introduction et Motivations.....	10
Introduction	10
Motivations	11
Approche proposée	12
Contributions	12
Plan du mémoire	13
Chapitre I Les Principes de la Recherche par le Contenu Visuel.....	14
Introduction	14
1. Les différents paradigmes de la recherche	15
1.1. Recherche par mots clés	15
1.2. Recherche par le contenu Visuel	16
1.2.1. Requêtes globales	19
1.2.2. Requêtes partielles	34
1.2.3. Requêtes locales par points d'intérêt	38
Conclusion	41
Chapitre II Les Points d'intérêt.....	42
1. Définition des points d'intérêt	42
2. Les différents types de points d'intérêt	43
3. Phases du processus d'indexation par point d'intérêt	43
3.1. Détection de primitives	43
3.2. Caractérisation des points d'intérêt	57

3.3. Appariement d'images	70
Conclusion	80
Chapitre III Les points de focus couleur.....	82
1. Approches de détection de points de focus	82
1.1. Transformée généralisée de Reisfeld	82
1.2. Transformée généralisée de symétrie en couleur	84
1.3. Transformée de Loy et Zelinsky	87
2. Les points d'intérêt et la couleur	88
3. Extension de l'approche d'Alexis et Rebai	89
3.1. Approche d'Alexis et Rebai	89
4. Approche proposée	91
5. Calcul automatique de l'angle θ	96
6. Résultats expérimentaux	98
6.1. Intérêt de la symétrie radiale en couleur	98
6.2. Relation Temps de calcul /Rayon maximal	98
6.3. Écartement des scores de symétrie	100
6.4. Evaluation face aux transformations de l'image	101
Conclusion	105
Conclusion et Perspectives	106
Conclusion	106
Perspectives	106
Bibliographie.....	108

Liste des figures

Figure I-1. Exemple de recherche d'images par les requêtes textuelles « Picsearch ».....	15
Figure I-2. Ambigüité de recherche par le mot clé « Avocat ».....	15
Figure I-3. Ambigüité de recherche par le mot clé « Religieuse ».....	16
Figure I-4. Ambigüité de recherche par la requête « Port du voile ».....	16
Figure I-5. Recherche par une image requête.....	16
Figure I-6. Recherche par l'ébauche graphique.	17
Figure I-7. Recherche par composition d'icônes.	17
Figure I-8. Processus d'une recherche multimodale où deux paradigmes sont mixés.....	18
Figure I-9. Les étapes de processus général de recherche d'image.....	18
Figure I-10. L'espace couleur RGB.....	20
Figure I-11. Le diagramme de chromaticité CIE.	21
Figure I-12. Limitation d'histogrammes couleur.....	24
Figure I-13. Histogramme de bloc.....	25
Figure I-14. Histogramme de blob.....	26
Figure I-15. Exemples d'images texturées. [Castelli, 2002].....	27
Figure I-16. Exemple de recherche d'image par le contenu à base de requêtes partielles.	34
Figure I-17. Processus de recherche d'image par les requêtes partielles.....	35
Figure I-18. Représentation en utilisant les descripteurs basés couleur de région. [Castelli, 2002]	37
Figure I-19. Approche d'appariement d'images intégrant les primitives d'espace et de couleur par la comparaison des arrangements spatiaux des régions. [Castelli, 2002]	37
Figure I-20. Exemple de recherche d'image en utilisant les points d'intérêt.....	39
Figure I-21. Processus de recherche d'image par les points d'intérêt.....	39
Figure II-1. Types de points d'intérêt : coins, jonction en Y et points de fortes variations de texture. [Boyer, 2002]	43
Figure II-2. Etapes du processus de mise en correspondance.	43
Figure II-3. Exemple de points détectés par le détecteur de HARRIS	46
Figure II-4. Variation du signal dans le cas de contour et de coin.....	48

Figure II-5. Quatre masques circulaires avec coloration similaire; l'USANs sont présentés par les parties blanches des masques. [Smith, 1997]	49
Figure II-6. Coins résultants de l'application du détecteur SUSAN sur (a) une image de test générique (synthétique) (b) sur une image réelle capturée par une caméra vidéo. [Smith, 1997]	50
Figure II-7. Exemple d'utilisation d'un cadre multi-échelle. Il est clair que le taux de répétabilité est inversement proportionnel au facteur d'échelle. [Schmid, 1996]	50
Figure II-8. Points d'intérêt extraits avec les détecteurs Harris, IPGP1, IPGP2. [Trujillo, 2006b]	53
Figure II-9. Des points d'intérêt détectés par l'opérateur Harris couleur. [Gouet, 2000]	54
Figure II-10. Homographie entre Deux images planes.	55
Figure II-11. Exemple d'images transformées par une rotation par 35° à gauche et par 325° à droite. [Nene, 1996]	61
Figure II-12. Exemple de caractérisation non paramétrique	63
Figure II-13. Représentation d'un changement d'échelle.	64
Figure II-14. Exemple d'images soumises à des changements de luminosité. [Geusebroek, 2005]	65
Figure II-15. Exemple d'images prises sous différents points de vue.	66
Figure II-16. Exemples de changements internes de luminosité [Gouet, 2000].	69
Figure II-17. Exemples de changements externes de luminosité [Gouet, 2000].	70
Figure II-18. Exemples de changements complexes de luminosité [Gouet, 2000].	70
Figure II-19. La corrélation de signal entre deux points.	72
Figure II-20. Distance des points de l'espace au centre d'inertie d'un nuage de points, en distance euclidienne (à gauche) et en distance de Mahalanobis (à droite).	75
Figure II-21. La mise en correspondance par le principe d'appariement croisé.	76
Figure II-22. Exemple de contrainte de voisinage.	78
Figure III-1. Symétrie réflexive de Reifeld [Reifeld, 1995]	83
Figure III-2. Apport de la couleur dans la symétrie réflexive [Heidemann, 2004]	85
Figure III-3. Gradient multi spectral pour deux points en symetrie reflexive	86
Figure III-4. Valeurs de la fonction PWF suivant l'angle d'orientation des vecteurs gradients	86
Figure III-5. Points affectés positivement et négativement	87
Figure III-6. Illustration du principe de vote ; dans ce cas le point k va avoir une valeur de symétrie égale à 3 parce que 3 vecteurs gradients votent pour lui comme un centre de symétrie.	89

Figure III-7. Un des objets symétriques qui ne sont pas circulaires et dont les détecteurs classiques ne peuvent pas détecter le centre de symétrie.	91
Figure III-8. Les votes en utilisant le gradient multi-spectral ; le point k aura plus de votes (7 dans ce cas) ce que renforce sa discrimination.	92
Figure III-9. (a) image original, (b) matrice de symétrie avec des gradients CMYK, (c) matrice de symétrie avec des gradients RGB ($R_{max}=40, \theta=0$).	93
Figure III-10. (a) image d'orientation couleur, (b) image d'orientation en niveau de gris, avec les points détectés marqués par des cercles bleu sur l'image originale.	94
Figure III-11. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=9$, Nombre maximum de points de focus = 15, $\theta=0$)	94
Figure III-12. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=8$, Nombre maximum de points de focus = 40, $\theta=0$)	94
Figure III-13. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=13$, Nombre maximum de points de focus = 13, $\theta=0$)	95
Figure III-14. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=70$, Nombre maximum de points de focus = 3, $\theta=0$)	95
Figure III-15. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=35$, Nombre maximum de points de focus = 2, $\theta=0$)	95
Figure III-16. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=30$, Nombre maximum de points de focus = 2, $\theta=0$)	95
Figure III-17. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=20$, Nombre maximum de points de focus = 6, $\theta=\pi/45$)	96
Figure III-18. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=40$, Nombre maximum de points de focus 1, $\theta=\pi/2$)	96
Figure III-19. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=30$, Nombre maximum de points de focus = 1, $\theta=\pi/16$)	96
Figure III-20. Relation angulaire entre le gradient et l'angle θ	97
Figure III-21. Résultats des votes en niveaux de gris et en couleur pour des objets qui ne se diffère pas de leur arrière plan en niveaux de gris. (Les tests pour (a, d, g) sont réalisés avec un $R_{max}=15, \theta=0$)	98
Figure III-22. Variation de temps de calcul suivant différents valeurs du rayon maximal (R_{max}).	99
Figure III-23. Variation de nombre de points détectés entre la valeur maximale et la valeur maximale-la distance D	101
Figure III-24. Exemples d'images sous différentes conditions d'illumination (a) séquence d'images de la scène « rose », (b) séquence d'images de la scène « ballons »	102

Figure III-25. Variation de taux de répétabilité suivant le changement d'illumination	102
Figure III-26. Exemples d'images sous différents taux de bruit. (a) sequence d'images de la scene « rose », (b) sequence d'images de la scene « ballons ».	103
Figure III-27. Variation de taux de répétabilité face au bruit.....	104
Figure III-28. Exemples d'images occultées. (a) sequence d'images de la scene « rose », (b) sequence d'images de la scene « ballons ».	104
Figure III-29. Variation de taux de répétabilité face à l'occultation.....	105

Introduction et Motivations

Introduction

« Une image vaut mille mots », c'est la raison pour laquelle l'utilisation de l'image est devenue une constante pour la quasi-totalité des gens. Avec l'apparition des caméras numériques, ces images se voient de plus en plus liées à l'ordinateur pour être traitées, imprimées ou stockées. Face à cette situation, la communauté des informaticiens se trouve forcée de mettre à la disposition des utilisateurs une gamme d'outils qui leur permettent de stocker, d'indexer et de rechercher les images. L'image est parfois très cruciale surtout dans des périodes critiques comme celles de crises et de guerres. Si on prend l'exemple de la guerre sur Ghaza, on peut clairement ressentir le vrai besoin des rapporteurs aux images dans les plus courts délais. Ces rapporteurs doivent donc disposer des outils qui leur permettent de retrouver facilement voire rapidement les images dont ils ont besoin. Ceci est une nécessité pour le rapporteur. De l'autre côté, l'implémentation d'un processus d'indexation qui soit le plus rapide et le plus efficace est primordiale. Cela exige une description fidèle du contenu de l'image car la simple annotation textuelle est souvent subjective.

Du déficit énorme en outils fiables de recherche d'image par le contenu, vient le recours intuitif à la classification et la reconnaissance d'objets. C'est le domaine de base du travail présenté dans ce mémoire, on s'intéresse particulièrement aux descripteurs locaux ou détecteurs de points d'intérêt. Ceci est une phase très intéressante dans le processus de reconnaissance d'objets. La compréhension de l'image et de son contenu est une étape clé dans le processus de reconnaissance d'objets à partir d'images. Toutefois, nous ne devons retenir de l'image que les points les plus intéressants et non pas la totalité des pixels qui la forment. Dans cette intention, l'utilisation de points d'intérêt pour représenter le contenu de l'image s'avère une alternative très fructueuse. Les travaux réalisés dans ce domaine sont nombreux, surtout ceux qui se basent sur l'information du signal pour définir les points d'intérêt. Cependant, le travail proposé dans cette thèse vise l'intérêt visuel que porte un pixel. Nous exploitons particulièrement l'intérêt visuel de la symétrie. A partir des approches déjà existantes, nous proposons une nouvelle approche qui exploite la richesse de l'information contenue dans les images couleur et ceci dans le but de développer des détecteurs de points d'intérêt plus robustes et plus performants.

Ce chapitre introductif est organisé comme suit : dans une première section les motivations qui nous ont amené à tirer profit de la symétrie dans la détection des points d'intérêt sont présentées. Ensuite une présentation de l'approche proposée vient clarifier ce choix de symétrie et de l'information couleur pour représenter fidèlement le contenu de l'image. Nous donnons ensuite nos contributions dans ce domaine de recherche, et nous finissons par un plan du mémoire.

Motivations

L'émergence de larges collections d'images, a mis la communauté de la vision par ordinateur face à plusieurs difficultés parmi lesquelles l'indexation, la recherche et la reconnaissance d'objets. Plusieurs techniques ont été essayées pour représenter le contenu d'une image de la manière la plus réduite et la plus fiable. Les différentes approches proposées sont classées en deux catégories principales : les approches globales (texture, couleur, forme) et les approches locales (régions d'intérêt, points d'intérêt). Ces dernières prouvent leur puissance face aux transformations d'images comme la rotation, le changement d'illumination et le changement de points de vue. De par leur capacité de représenter fidèlement le contenu d'une image, les points d'intérêt sont jugés les plus intéressants parmi les méthodes locales. Ils sont largement utilisés dans plusieurs applications de la vision par ordinateur comme le CBIR (Content-Based Image Retrieval) qui est une application permettant l'indexation automatique d'images par la récapitulation de leurs contenus visuels à travers l'extraction automatique de primitives de l'image ou de descripteurs comme la couleur, la texture, la forme ou encore des points d'intérêt. La reconnaissance d'objets est une autre application qui exploite aussi la richesse de ces points. Toutefois l'extraction de ces points à partir d'images n'est pas une tâche triviale car ils doivent être les plus discriminants possible tout en étant robustes et fiables.

Les points d'intérêt les plus répandus dans la littérature sont ceux déduits à partir du signal. Cependant, d'autres sources d'intérêt peuvent être exploitées. Un exemple d'intérêt visuel dominant dans les scènes aussi bien naturelles qu'artificielles est la symétrie. Cette information a permis l'apparition de nouveaux points d'intérêt qui sont les points de symétrie ou les points de focus. Cette symétrie peut être radiale, réfléchive ou encore axiale, mais qui donne toujours des points avec un intérêt visuel important. Plusieurs chercheurs utilisent la symétrie afin de réaliser des détecteurs de points de focus. Cependant, dans la majorité des travaux la couleur n'est pas exploitée. Les approches se limitent à l'information en niveau de gris pour calculer les centres et les axes de symétrie. Cette situation empêche les détecteurs existants de profiter de la richesse de l'information couleur en faveur de la robustesse et la précision des détecteurs. En fait, plusieurs auteurs utilisent la couleur pour améliorer les méthodes existantes et profiter du contenu informatif de celle-ci. Après avoir passé en revue un état de l'art des différentes approches et détecteurs de points d'intérêt, nous avons observé que le travail de Gouet [Gouet, 2000] est une extension à la couleur du fameux détecteur de points d'intérêt de Harris [Harris, 1988]. Le travail de Heidemann [Heidemann, 2004] est une extension à la couleur de l'approche proposée par Reisfeld dans [Reisfeld, 1995]. Sur le même axe, nous avons opté pour l'extension à la couleur de l'approche proposée par Rebai et al [Rebai, 2006b].

Approche proposée

Dans ce travail, nous introduisons un nouveau détecteur de points de symétrie ou points de focus qui permet de s'affranchir des limitations des détecteurs existants. Il profite de la richesse de couple symétrie-couleur pour une meilleure discrimination et robustesse des points détectés.

Nous choisissons les centres de symétrie radiale comme des points d'intérêt ou encore points de focus. Pour être choisi comme un centre de symétrie, un pixel de l'image doit être sélectionné par tous les pixels de l'image. Ce pixel est celui qui a un score de symétrie important par rapport à celui de ses pixels voisins.

Pour calculer un score de symétrie d'un pixel donné, il faut procéder comme suit :

- Réaliser l'élection suivant le principe d'élection approprié. Nous avons proposé dans ce travail l'utilisation du principe d'élection généralisé de Rebai [Rebai, 2006b] étendu à la couleur. A l'issue de cette étape, un score de symétrie est accordé au pixel en question.
- Un seuillage des scores de symétrie est ensuite envisagé pour déterminer les pixels élus.

Les pixels élus constituent les pixels centres de symétrie ou points de focus.

Contributions

Le travail proposé introduit deux principales contributions : la première contribution est le développement d'un détecteur de points de focus qui exploite l'information couleur dans le processus de vote afin de détecter des points plus discriminants et plus robustes face aux transformations usuelles de l'image comme les translations, les rotations, les changements d'illumination, les changements d'échelle, les changements de points de vue, les occultations et le bruit de la caméra. Nous avons étendu le principe d'élection pour prendre en compte l'information couleur. Avec notre approche, les 3 canaux couleur font le vote. Après avoir effectué le vote, chaque canal génère sa propre matrice d'orientation suivant les gradients calculés pour ce canal. La somme de ces matrices d'orientation nous donne une matrice d'orientation couleur qui soit plus discriminante et plus robuste face aux transformations de l'image.

La deuxième contribution est la proposition d'une approche de calcul automatique de l'angle d'élection θ utilisé dans les travaux en relation pour faire des votes non plus dans la direction du gradient mais dans différentes directions. L'approche proposée calcule les valeurs que peut prendre l'angle θ pour chaque couple de pixels dans l'image au lieu de prendre des valeurs arbitraires. Cet angle est aussi utilisé pour caractériser les points détectés. Par conséquent, le calcul automatique de cet angle facilite et améliore le processus de caractérisation.

Plan du mémoire

Les objectifs de ce mémoire sont d'une part la détection des points d'intérêt comme étant une phase importante dans le processus de reconnaissance d'objets et de l'appariement d'images, et d'autre part la considération des points de symétrie comme faisant partie des points d'intérêt. Ces buts imposent une organisation en trois chapitres.

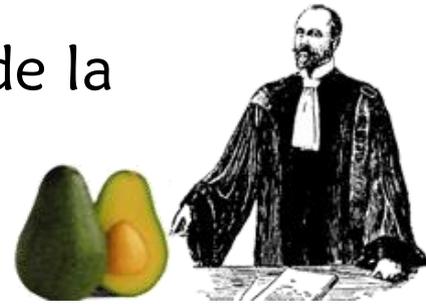
Le premier chapitre, traitant des principes de la recherche par le contenu visuel, passe en revue les méthodes de descriptions globales pour mieux situer ces approches par rapport aux approches de descriptions locales utilisant entre autres les points d'intérêt.

Le deuxième chapitre présente les trois étapes du processus d'appariement d'images par points d'intérêt. Pour chacune de ces étapes, nous avons présenté un état de l'art pour constituer une idée plus détaillée sur la mise en correspondance d'images.

Le troisième chapitre constitue l'approche que nous proposons. Il détaille aussi bien les contributions que les expérimentations réalisées. Il présente également les résultats expérimentaux et leur discussion.

Les conclusions et les perspectives sont une synthèse et une conclusion de ce travail.

Chapitre I Les Principes de la Recherche par le Contenu Visuel



Introduction

Ce chapitre présente une introduction à la recherche d'images par le contenu Visuel. C'est une application du domaine de la vision par ordinateur qui est en plein essor à cause des grandes quantités de données multimédia du monde actuel.

Les progrès dans les ordinateurs modernes et les technologies de télécommunication sont à l'origine des gigantesques archives de données multimédia. Cela se retrouve dans divers domaines d'application comme la médecine, le loisir, l'éducation et les services d'information en ligne. Le développement des données multimédia est l'équivalent du développement des données alphanumériques durant les premiers jours de l'informatique. A l'époque cette quantité de données a nécessité le développement de système de gestion de bases de données. Toutefois, les systèmes de bases de données ne peuvent pas permettre le stockage, l'exploration, l'indexation et la recherche d'images d'une façon efficace. Les différents types de données multimédia peuvent requérir des outils et des méthodologies d'indexation et de recherche spécifiques.

Depuis les années 70, la recherche d'images est un domaine de recherche très actif. Deux communautés de chercheurs, celle de la gestion des bases de données et celle de la vision par ordinateur s'y intéressent et s'investissent dans de nouvelles approches. Ces deux communautés étudient la recherche d'images suivant deux perspectives assez différentes : La première est principalement basée sur le texte, alors que la deuxième se base sur les propriétés visuelles des données.

La recherche d'images basée sur le texte a vu le jour avant la fin des années 1970. Par cette approche, les images sont annotées par des mots clés qui sont stockés comme des clés de recherche dans les bases de données traditionnelles. Cette approche souffre de deux problèmes majeurs rendant l'annotation inappropriée. Le premier est la difficulté et l'énorme effort impliqué dans la phase d'annotation. Ceci se traduit par un grand investissement financier. Le deuxième problème est dû à la subjectivité de la perception

humaine et à la gêne de représenter le contenu visuel d'une image par un petit nombre de mots clés.

Afin de surmonter les problèmes susnommés de la recherche d'images par le texte, les chercheurs ont proposé, au début des années 90, une recherche d'images par le contenu. Ce nouvel axe de recherche nommé CBIR (Content Based Image Retrieval) porte sur l'indexation automatique d'images par la caractérisation du contenu visuel.

1. Les différents paradigmes de la recherche

1.1. Recherche par mots clés

Il existe plusieurs paradigmes de recherche d'information visuelle. L'annotation texte constitue l'un de ces paradigmes. Les exemples de systèmes de recherche par le texte sont les moteurs de recherche : Google, Exalead, Picsearch...etc (Figure I-1). Toutefois, ce paradigme présente certains problèmes comme l'ambiguïté d'utilisation des mots clés (Figure I-2, Figure I-3, Figure I-4).

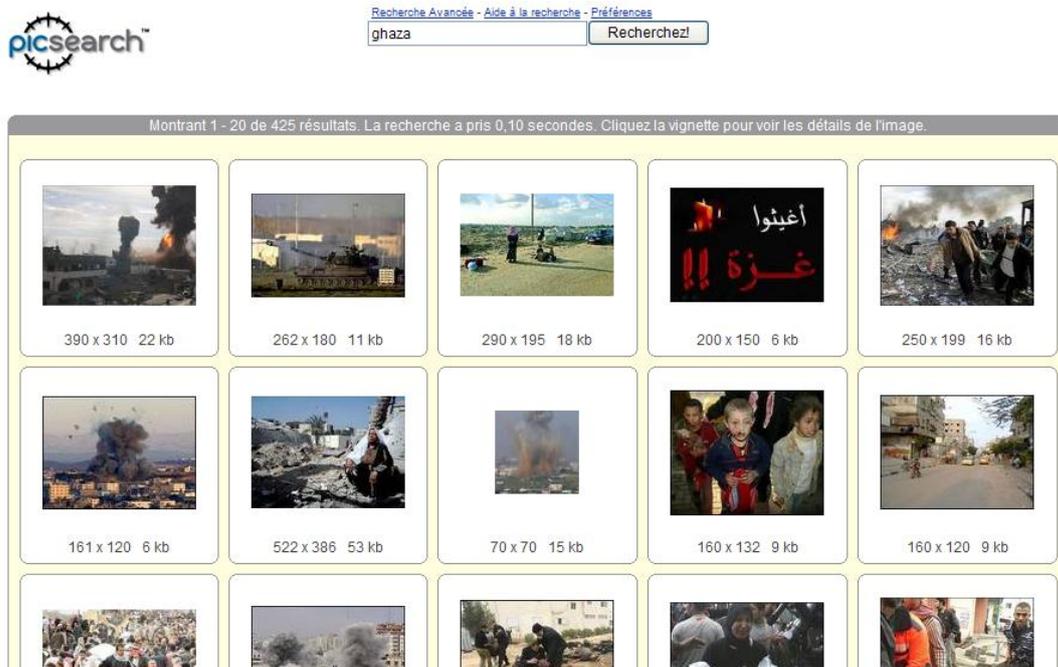


Figure I-1. Exemple de recherche d'images par les requêtes textuelles « Picsearch ».

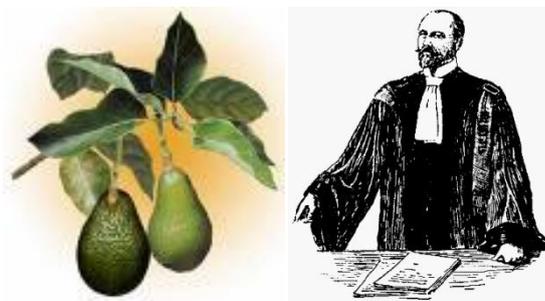


Figure I-2. Ambiguïté de recherche par le mot clé « Avocat ».



Figure I-3. Ambiguïté de recherche par le mot clé « Religieuse ».



Figure I-4. Ambiguïté de recherche par la requête « Port du voile ».

1.2. Recherche par le contenu Visuel

En l'absence ou l'incomplétude de l'annotation texte, la recherche par le contenu peut la remplacer ou la compléter. Dans les méthodes de recherche par le contenu, la recherche est réalisée par des primitives extraites directement des médias visuels comme la couleur, la texture et la forme. Les paradigmes de recherche de l'information visuelle VIR (Visual Information Research) les plus utilisés sont les suivants [Lew, 2001]:

Image requête

Dans le paradigme de requêtes d'images similaires, l'utilisateur sélectionne une image et le système répond par une liste d'images similaires à celle donnée par l'utilisateur Figure I-5.

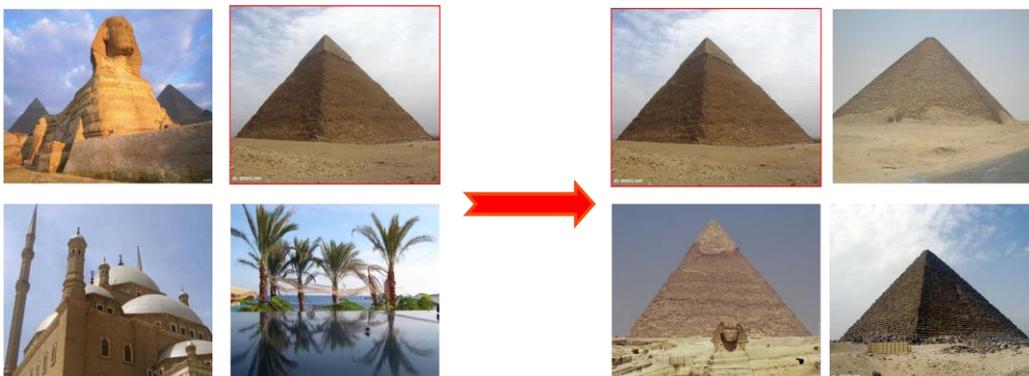


Figure I-5. Recherche par une image requête.

Ebauche graphique (sketch)

Dans le paradigme de recherche par esquisse "sketch", l'utilisateur dessine un sketch de forme, de texture ou de couleurs et le système répond par une liste d'images qui contient des primitives similaires au sketch dessiné par l'utilisateur (Figure I-6).



Figure I-6. Recherche par l'ébauche graphique.

Composition d'icônes

Dans le paradigme de requêtes iconiques, l'utilisateur place des icônes symboliques là où il faut avoir des primitives visuelles et le système répond par une liste d'images qui contient des primitives similaires aux icônes placées par l'utilisateur (Figure I-7).



Figure I-7. Recherche par composition d'icônes.

Recherche Mentale

Avec la recherche d'images par l'exemple visuel, l'image requête se trouve dans la base d'images et elle sera comparée par le système d'appariement avec les autres images. Si on suppose que l'image requête est externe et l'appariement est mental, par exemple, une image qui réside dans l'esprit de l'utilisateur, l'utilisateur peut donc chercher des versions de cette même image, par exemple, le même visage, ou des images appartenant à la même classe et répond aux requêtes générées par la machine pour accélérer la recherche. L'utilisateur ensuite déclare laquelle des images vues est la plus proche à sa requête. Ces décisions de similarité sont entièrement subjectives et dépendantes de l'utilisateur. Cette approche, proposée récemment par [Geman, 2006], discute un système de recherche interactif qui se base sur la théorie de l'information et l'inférence statistique. L'algorithme proposé entraîne un modèle bayésien de rendement de pertinence et un critère d'optimalité basé sur l'entropie conditionnelle.

Combinaison de modalités (visuelle -textuelle)

Chacun des paradigmes discutés auparavant a ses avantages et ses inconvénients. Toutefois, aucun des paradigmes de recherche cités n'est complet. Mais ils peuvent être combinés pour une meilleure recherche voir (Figure I-8).

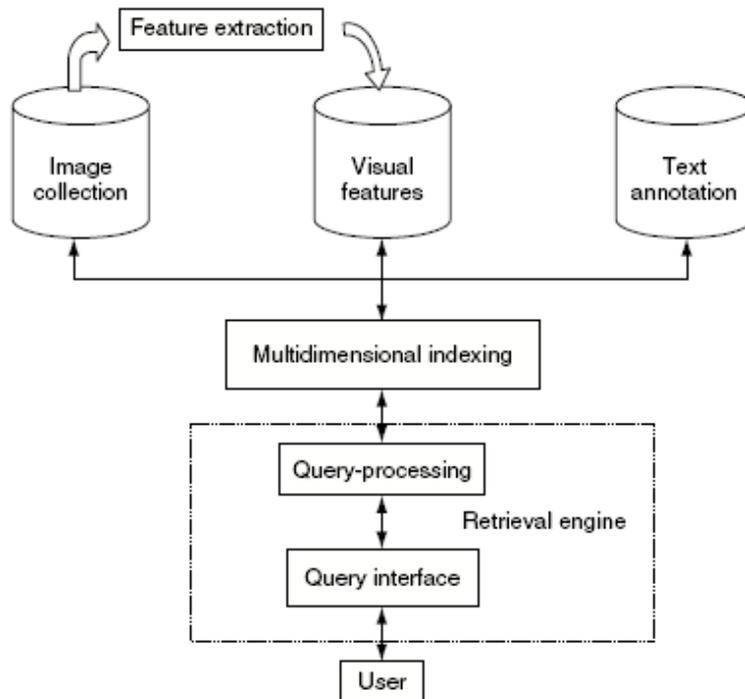


Figure I-8. Processus d'une recherche multimodale où deux paradigmes sont mixés.

Lors de la conception d'un système VIR où la recherche est basée sur des primitives visuelles, il est fondamental de bien choisir des primitives appropriées. Dans les systèmes actuels, ces primitives sont typiquement reliées à la couleur, la texture et la forme. Le processus du choix automatique ou interactif des primitives adéquates est nommé sélection de primitives. Après avoir sélectionné les primitives, une mesure (distance) de similarité est choisie pour trier les images candidates. Les différentes étapes d'un système VIR sont données dans le schéma de la Figure I-9.

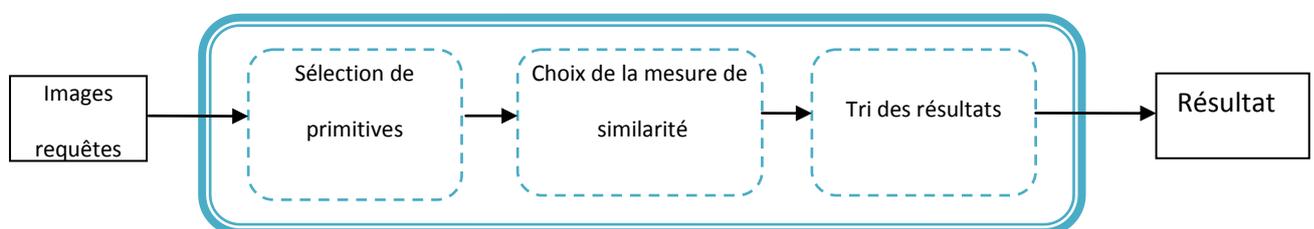


Figure I-9. Les étapes de processus général de recherche d'image.

La sélection de primitives : à partir de l'ensemble des primitives de bas niveau (couleur, texture et forme) et de haut niveau (primitives sémantiques), il faut choisir un sous-ensemble de primitives qui a une bonne discrimination par rapport à l'application sous-jacente.

La sélection de mesure de similarité : il s'agit de choisir une mesure ou une distance de similarité. Cette mesure de similarité permet de donner les distances entre l'image requête et les images résultats.

Le tri des résultats : il s'agit de trier les résultats obtenus en utilisant la mesure de similarité choisie.

La question la plus intuitive est donc : de quels outils disposent les chercheurs pour assurer la description automatique d'images? Trois approches se distinguent, suivant que l'on cherche à décrire l'image dans sa totalité, certaines de ses parties ou seulement des points précis de celle-ci.

1.2.1. Requêtes globales

L'utilisation des requêtes globales dans les bases d'images est adoptée dans le projet d'IBM Query By Image Content (QBIC). Ce projet explore les méthodes de recherche d'images basées sur la similarité de descripteurs globaux de couleur, de texture et de forme. Le projet MIT Photobook a étendu certaines des premières méthodes de requêtes basées sur le contenu par le développement de descripteurs qui fournissent un appariement effectif tout en ayant l'alternative de reconstruire les images et leurs caractéristiques à partir de ces descripteurs [Castelli, 2002]. L'architecture générale d'un système à base de requêtes globales est donnée par la Figure I-9. On peut classifier les approches globales en trois catégories selon la primitive utilisée : la couleur, la texture ou la forme.

1.2.1.1. Recherche basée sur la couleur

L'objectif de la recherche basée sur la couleur consiste à retourner les images qui présentent les descripteurs couleur les plus similaires aux descripteurs couleur de l'image requête. Swain et Ballard [Swain, 1991] ont étudié l'utilisation des descripteurs d'histogramme couleur pour la recherche d'objets colorés dans les images cibles. D'autres travaux proposent l'utilisation des descripteurs de moments couleur. Dans le projet SaFe [Smith, 1997] une autre approche propose l'utilisation des descripteurs 166-bin couleurs dans l'espace couleur HSV. Chacune des méthodes proposées nécessite la conception de descripteurs couleurs qui incluent la sélection de l'espace couleur et une métrique de distance pour mesurer la similarité des descripteurs couleurs.

La couleur est une dimension importante de la perception visuelle humaine qui permet la discrimination et la reconnaissance d'information visuelle. De ce fait, les primitives couleurs sont jugées fiables pour l'indexation et la recherche d'images couleur dans les bases d'images. Les descripteurs couleurs sont relativement simples à extraire et à comparer, ils sont alors très appropriés aux requêtes basées sur le contenu.

Typiquement, la spécification d'un descripteur couleur nécessite la fixation d'un espace de couleur. L'utilisation de la couleur pour l'extraction de primitives passe par trois principales étapes qui sont : le choix de l'espace couleur, la sélection d'un quantificateur et finalement le calcul des descripteurs couleurs.

a) Les espaces couleur

Un espace couleur, comme défini dans [Nieman, 1990], est l'espace multidimensionnel dans lequel les différentes dimensions représentent les différentes composantes de couleur. La couleur ou l'illumination couleur notée par la fonction $F(\lambda)$, est supposée comme étant la radiation électromagnétique dans l'intervalle d'ondes visibles ($\lambda \in \{380 \text{ nm} \dots 780 \text{ nm}\}$). Il était démontré expérimentalement que la couleur est perçue à travers trois récepteurs couleurs indépendants qui dominent la réponse à approximativement la longueur d'onde Rouge (r), Vert (g), Bleu (b): $\lambda_r = 700 \text{ nm}$, $\lambda_g = 546.1 \text{ nm}$, $\lambda_b = 435.8 \text{ nm}$, respectivement.

Par conséquent, toute couleur peut être représentée par une combinaison linéaire de trois couleurs primaires (R, G, B). L'espace construit par les valeurs R, G et B décrit complètement les couleurs visibles, qui sont représentées comme des vecteurs dans l'espace couleur 3D RGB. De ce fait, l'espace couleur RGB constitue un bon point de départ pour la représentation des primitives couleurs d'images. Toutefois, l'espace couleur RGB n'est pas perceptiblement uniforme. En général, les couleurs RGB peuvent être "mappées" vers différents espaces couleurs moyennant une transformation couleur. Une transformation linéaire produit un nombre important d'espaces couleur comme YIQ/YUV. D'autres espaces couleur comme HSV, CIE sont générés par des transformations non linéaires de l'espace couleur RGB. Ils sont plus convenables pour le traitement d'images car très proches de la perception humaine.

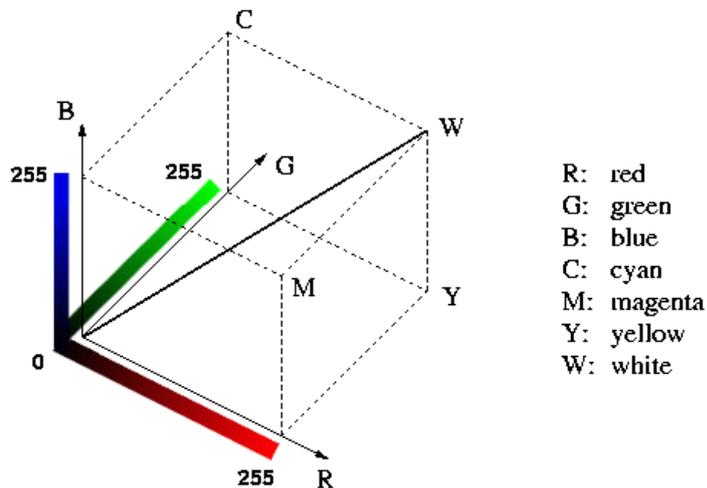


Figure I-10. L'espace couleur RGB.

Comme l'illustre la Figure I-10 l'espace couleur RGB présente une progression additive du noir vers blanc. La combinaison du Rouge et du Vert donne le Jaune, le Vert plus le Bleu donne le Cyan, et le Bleu plus le Rouge donne le Magenta. Les niveaux de gris s'allongent sur la diagonale du cube. Cyan, Yellow et Magenta sont des couleurs primaires soustractives utilisées pour l'impression, qui si soustraient du Blanc, il reste les couleurs Rouge, Bleu et Vert, respectivement.

Le diagramme de chromaticité CIE (Commission Internationale de l'Éclairage) est le plus ancien parmi les espaces couleur. C'est une couleur définie sur un plan 2D comme présenté dans la Figure I-11. Le troisième axe est la luminance.

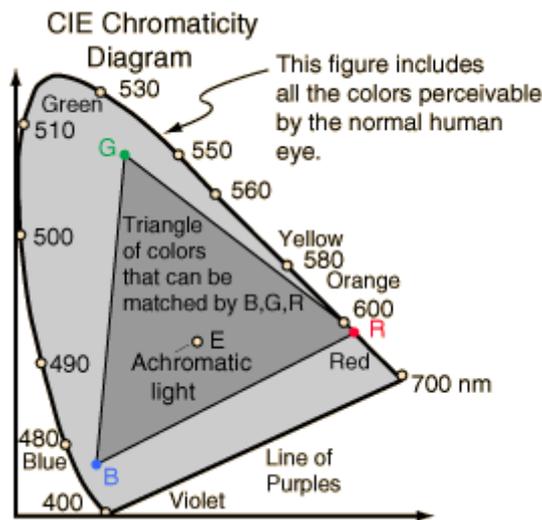


Figure I-11. Le diagramme de chromaticité CIE.

Dans la Figure I-11 le contour foncé contient les couleurs visibles, qui sont totalement saturées le long des contours. Les valeurs mentionnées donnent les longueurs d'onde de la lumière en nanomètres. Le triangle inclut donc les couleurs qui peuvent être produites par les CRTs couleurs typiques par le mixage du Rouge, du Vert et du Bleu.

Le diagramme CIE offre un outil pour la définition des couleurs, mais il ne correspond ni à l'opération du hardware ni directement à la vision humaine. Une approche qui fait ceci est incorporée dans les systèmes HSV (hue, saturation, et value), HIS (hue, saturation, et intensity) et HLS (hue, lightness, et saturation) ils sont très proches des concepts artistiques de teinte, ombre et conversation. Dans ce système, hue est la couleur comme décrite par la longueur d'onde, par exemple la distinction entre Rouge et Jaune. Saturation est le degré de la couleur qui est présent par exemple la distinction entre Rouge et Rose. Le troisième axe (nommé lightness, intensity ou value) est le degré de lumière, par exemple la distinction entre un Rouge foncé et un Rouge clair [Russ, 1998].

b) Quantificateur couleur

L'objectif des méthodes de quantification des couleurs est de réduire le nombre de couleurs de l'image originale par la création d'une distorsion minimale entre l'image quantifiée et l'image originale.

Donc, le partitionnement de l'espace couleur est une phase très importante dans la construction des descripteurs couleur. Généralement, le processus de quantification des couleurs image peut être fractionné en deux étapes [Kanjanawanishkul, 2003]: le "design" de la palette et le "mapping" des pixels. Dans la première étape, le "design" de la palette exige le choix d'un sous ensemble de couleurs à partir de l'espace couleur réel. Dans la deuxième étape, les pixels de l'image originale sont "mappés" aux couleurs de la palette.

i. Le "design" de la palette

Il existe deux classes générales de méthodes de quantification : les méthodes à palette fixe ou universelle et celles à palette adaptative ou dédiée. La quantification fixe est indépendante du contenu de l'image, alors que dans la quantification adaptative, la palette est conçue d'une manière adaptative au contenu de l'image. Les approches adaptatives existantes peuvent être classées en trois catégories :

1. Schéma Hiérarchique ou pré-groupement

La plupart des algorithmes dans cette catégorie sont basés sur l'analyse statistique de la distribution couleur de l'image dans l'espace couleur. Des exemples d'algorithmes de ce schéma sont the Popularity, Median Cut, Variance Minimization, Octree et Principal Analysis Algorithm.

2. Schéma Itératif ou post-groupement

Il consiste à sélectionner initialement une palette qui sera raffinée itérativement par l'utilisation de l'algorithme de K-Means et cela pour minimiser l'erreur quadratique minimale. Des exemples de ce schéma sont : fuzzy C-mean qui est une extension de l'algorithme K-mean, the Hierarchy Competitive Learning (HCL), l'algorithme Genetic C-means et NeuQant qui exploite la carte d'auto-organisation de Kohonen.

3. La quantification scalaire améliorée

Il y a un groupe d'algorithmes très simples qui peuvent être classés comme étant basés sur la quantification scalaire améliorée plutôt que sur la quantification vectorielle. La quantification scalaire séquentielle est un exemple de ce schéma.

ii. Le "mapping" des pixels

Après l'obtention de la palette couleur, on "mappe" chaque pixel de l'image originale à une couleur dans la palette. Plusieurs techniques sont utilisées pour cette fin. Citons :

1. L'algorithme de force brute

Il s'agit du chemin le plus simple pour trouver le bon "mapping".

2. K-d Trees

C'est la méthode classique k-d trees pour trouver le plus proche voisin en utilisant le partitionnement orthogonal.

3. La recherche triée localement

Elle présente des avantages par rapport à celle de force brute quand le nombre de couleurs représentatives est grand et quand les couleurs dans l'image originale ont une large distribution.

4. Le "mapping" centroïde

C'est la solution la plus rapide ; elle consiste à utiliser la partition obtenue pendant le processus de "design" de la palette couleur et à "mapper" les couleurs image au représentatif qui est le centroïde du "cluster" qui contient la couleur.

c) Les descripteurs couleur

Un descripteur couleur est défini comme étant la quantité numérique qui décrit une primitive couleur d'une image [Castelli, 2002]. Il existe plusieurs descripteurs couleur dans la littérature, les paragraphes suivants présentent quelques-uns.

i. Histogrammes des niveaux de gris

L'histogramme d'intensité présente comment les niveaux de luminance sont occupés dans une image. Le contraste image est mesuré par l'intervalle des niveaux de luminance. L'histogramme donne le nombre de pixels avec un niveau de luminosité particulier face au niveau de luminance. Pour les pixels 8-bit, la luminance va de zéro (noir) à 255 (blanc) [Nixon, 2002].

ii. Histogrammes couleur

La majorité des descripteurs couleur sont des histogrammes couleur ou dérivés de ceux-ci. Ils capturent la distribution des couleurs dans une image ou une région d'image. Il s'agit donc de "mapper" l'image à un espace couleur approprié, de quantifier l'image "mappée", et de dénombrer le nombre d'apparitions de chaque couleur quantifiée. Quand il s'agit d'observation des distributions continues ou qui peuvent contenir un grand nombre de valeurs possibles, un histogramme est construit en associant chaque "cluster" à un ensemble de valeurs observées. Chaque "cluster" de l'histogramme contient le nombre d'observations appartenant à l'ensemble associé. La couleur appartient à cette catégorie de valeurs aléatoires : par exemple l'espace couleur d'images 24-bit contient 2^{24} couleurs distinctes.

L'histogramme couleur H d'une image I contenant N pixels de n couleurs est un vecteur $(h_{c1}, h_{c2}, \dots, h_{cn})$ où chaque élément h_{ci} représente le nombre de pixels de couleur c_i dans I . L'avantage majeur de ce descripteur est qu'il est invariant à la translation, la rotation et le changement d'échelle (après normalisation). Parmi les limitations de l'histogramme couleur est le fait qu'il est une distribution du premier ordre. La deuxième limitation est la distribution globale des histogrammes couleurs Figure I-12.

1. Sensibilité aux changements d'illumination

Funt et Finlayson y apportent une solution appelée méthode CCCI « Color Constant Color Indexing » sous l'hypothèse du modèle d'illumination diagonal. Elle consiste à calculer des histogrammes de rapports de couleur : l'indexation n'est plus réalisée sur la couleur elle-même mais sur les dérivées du logarithme des couleurs. Cette méthode est toutefois moins bonne lorsque l'éclairage reste constant.

2. Comment tenir compte de l'information spatiale ?

L'inconvénient majeur des histogrammes est la perte de la distribution spatiale des niveaux de gris. Pour résoudre ce problème, quelques travaux ont été réalisés et qui utilisent le découpage de l'image en régions pour lesquelles un histogramme couleur est calculé.

Un autre type d'approche consiste à enrichir l'histogramme d'une certaine information structurelle (corrélogramme couleur). Une autre variante utilise les histogrammes en y introduisant une information de voisinage. Elle est basée sur la méthode de calcul des invariants par généralisation et contrainte. Cette méthode a connu plusieurs améliorations qui malheureusement augmentent la complexité et le temps de calcul.

Il est clair que l'indexation à partir des histogrammes est en général très coûteuse en temps de calcul ; ce qui a fait orienter les travaux vers des critères stochastiques moins 'gourmands' comme les moments.

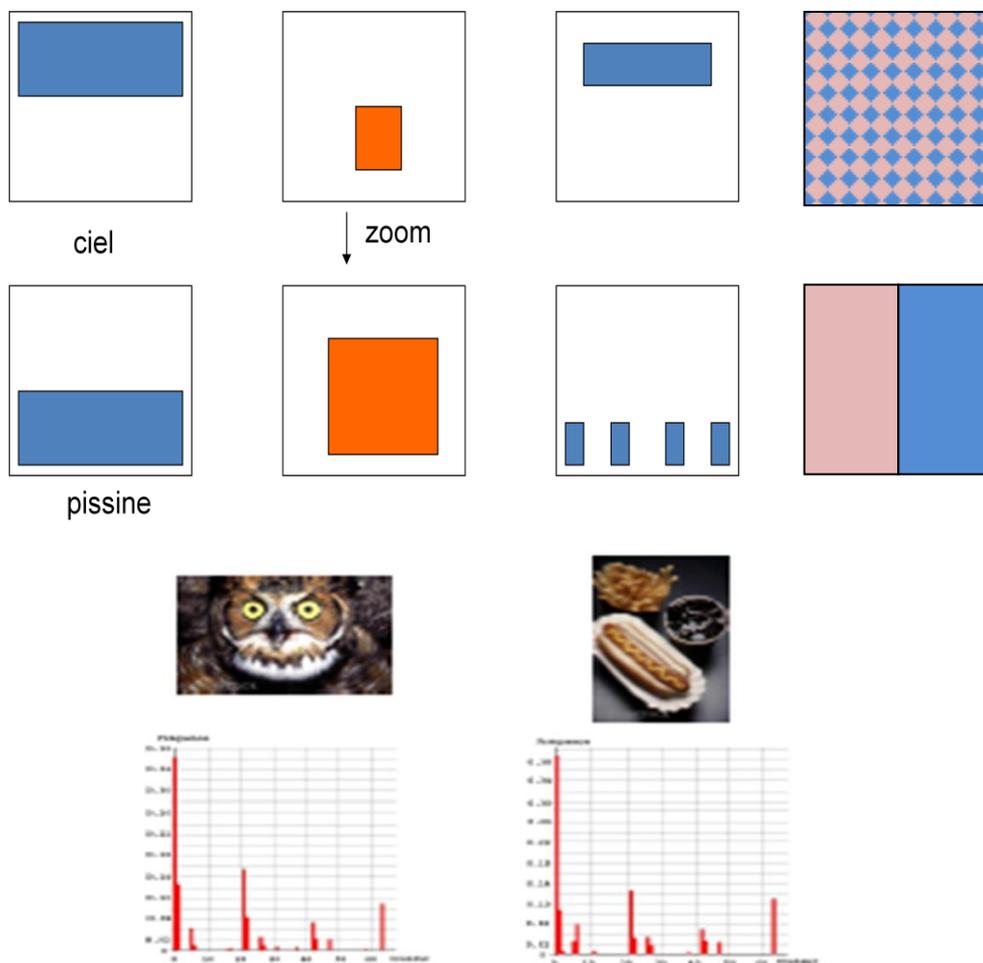


Figure I-12. Limitation d'histogrammes couleur.

iii. Histogrammes de bloc

C'est une approche qui consiste à diviser l'image en bloc et calculer pour chacun des blocs son histogramme. L'histogramme de l'image sera la concaténation des histogrammes

calculés pour chaque bloc. Ce type d'histogramme permet de voir la distribution de couleurs dans l'image.

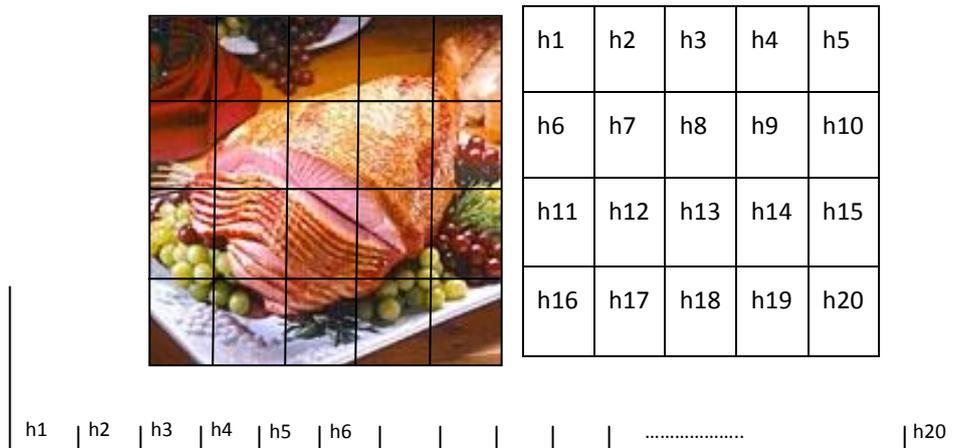


Figure I-13. Histogramme de bloc.

iv. Histogrammes de blob

Ce type d'histogramme [Qian, 2000] est capable de distinguer entre des images qui présentent la même distribution couleur mais qui contiennent des objets de différentes tailles ou formes. Ce type d'histogramme capture la taille relative et la fréquence d'information d'un groupe local de pixel avec des attributs uniformes par exemple la couleur. Ils sont capables de distinguer entre les images que les histogrammes conventionnels ne peuvent pas distinguer. Comme les histogrammes conventionnels, les histogrammes de blob sont faciles à calculer, invariants à la rotation et la translation et insensibles au bruit. Ils peuvent être utilisés avec la couleur, la texture ou d'autres types de primitives.

Cet histogramme est calculé en glissant une fenêtre (élément structurel) au-dessus de l'image et en identifiant les groupes de pixels avec une couleur uniforme à l'intérieur de la fenêtre glissante. Ces groupes de pixels qui ont une couleur uniforme sont nommés les blobs couleur. Ce qui fait que la notion de pixel couleur est étendue à la notion de blob couleur.

Par la collection de tailles relatives et de distributions fréquentielles des blobs couleur à différentes échelles, on peut créer une description image avec une information spatiale incorporée Figure I-14.

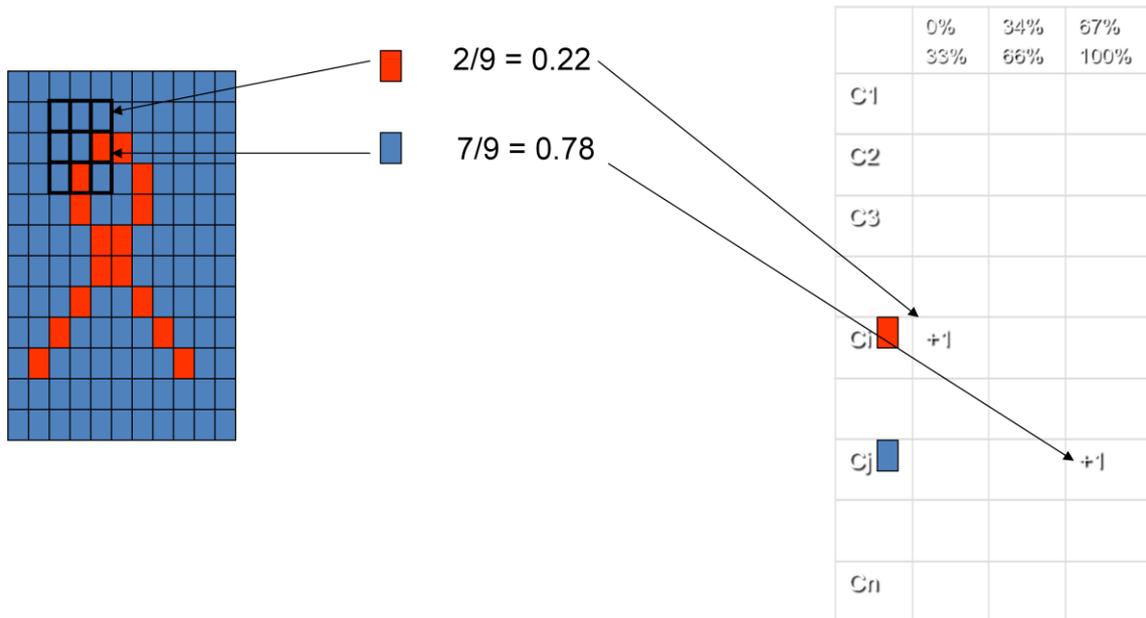


Figure I-14. Histogramme de blob.

v. Les moments

L'utilisation des moments pour l'indexation des images est une idée très naturelle partant du principe qui considère l'image comme la distribution aléatoire des couleurs.

Plusieurs approches ont été mises en place, celles qui utilisent les moments jusqu'au troisième ordre et celles qui combinent l'utilisation des moments avec les modèles linéaires d'illumination. En 1990, une approche définit la notion de moments couleurs généralisés à vu le jour. Ensuite une autre qui utilise les moments du quatrième ordre est apparue.

Comme toutes les descriptions basées sur la distribution des couleurs, ces approches ont le principal inconvénient de ne pas tenir compte de l'information spatiale.

vi. Autres approches

Il existe trois autres approches. La première applique à la couleur, la classique mesure de corrélation. La deuxième est une méthode basée sur des mesures d'angles. La troisième est une méthode où la distribution est envisagée sous forme de mélange gaussien des couleurs résistant aux changements de luminosité.

1.2.1.2. Recherche basée sur la texture

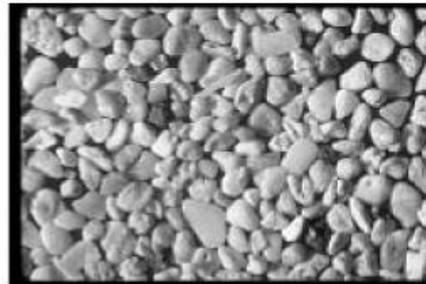
Les images de l'eau, de l'herbe, ou de motifs sur un tissu sont des exemples de texture image. On peut distinguer plusieurs objets que ce soit naturels ou artificiels à partir de leurs textures. Toutefois, il est difficile de décrire en termes qualitatifs ces objets texturés. Il est indispensable de créer des descriptions quantitatives nécessaires à l'analyse de la machine.

La texture dépend généralement des conditions de l'éclairage, des points de vue et de la distance. Ces conditions peuvent être changées durant une période de temps comme avec les images de paysage. La texture est une propriété de l'image mais sans définition formelle universelle malgré qu'il est simple de visualiser ce que signifie une texture [Castelli, 2002].

On peut penser à la texture comme étant constituée de quelques primitives de base (texels ou micro-motifs) dont la distribution spatiale dans l'image crée une apparence de texture. Ces texels facilement identifiables dans la plupart des objets artificiels ont une distribution spatiale régulière (périodique) ou "random". Dans la Figure I-15(a) la brique est un micro-motif à partir duquel une distribution particulière constitue un motif structuré (le mur). Les primitives ne doivent pas être de la même taille et forme, comme c'est illustré par les textures de briques et de cailloux Figure I-15(b). Les micro-motifs bien définis peuvent ne pas exister dans plusieurs cas comme les images du sable sur la plage, l'eau et les nuages. Des exemples d'images texturées sont donnés dans la Figure I-15.



(a) Mur de brique



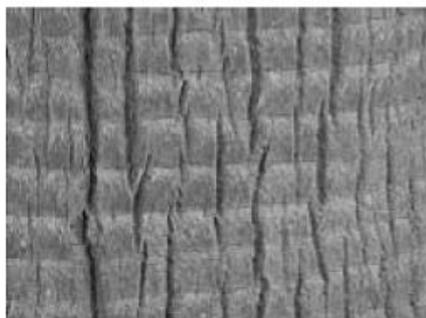
(b) Cailloux



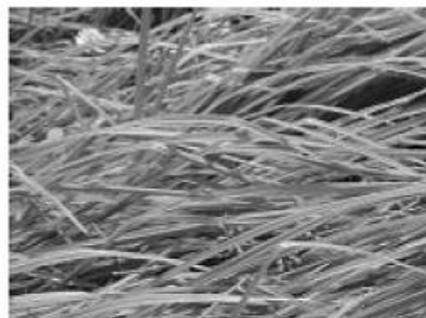
(c) Sable



(d) Eau



(e) Ecorce d'arbre



(f) Herbe

Figure I-15. Exemples d'images texturées. [Castelli, 2002]

La texture, étant une primitive image, est très utile pour la recherche de similarité. On doit calculer des descripteurs texture qui nous permettront de faire des comparaisons entre des images texturées de la base d'images. Notant que dans la classification de texture on calcul un label pour une image texturée donnée. Ce label peut avoir une sémantique associée : par exemple texture eau, texture nuage...etc. Si les textures dans la base d'images sont classées similairement, leurs labels peuvent être utilisés pour rechercher d'autres images contenant la texture eau ou nuage.

Mais cette solution présente des inconvénients. Premièrement, il se peut qu'il ne soit pas possible de créer un dictionnaire exhaustif de labels de classes. Deuxièmement, même si cette information de label de classe est disponible, il se peut que le but soit de trouver les N top appariements à l'intérieur de cette classe qui sont visuellement similaires au motif donné. Dans ce cas, la base d'images doit stocker les détails de description texture pour permettre la recherche de motifs texturés similaires.

a) La primitive texture

Une primitive est définie comme une caractéristique distincte d'une image représentée par un descripteur. Un descripteur définit la syntaxe et la sémantique de la représentation des primitives. Alors une primitive texture capture un attribut spécifique d'une image comme la rugosité. Ensuite un descripteur de rugosité est utilisé pour représenter cette primitive. Dans la littérature de traitement d'images et de la vision par ordinateur les termes: primitive et descripteur (d'une primitive) sont souvent utilisés comme synonymes. Les premiers travaux sur la discrimination de la texture utilisent différents types de statistiques de textures image. On peut par exemple considérer l'histogramme en niveau de gris car il représente une distribution de premier ordre d'intensités de pixels. La moyenne et la déviation calculées de ces histogrammes peuvent être utilisées comme des descripteurs de texture pour la discrimination de différentes textures. Les statistiques du premier ordre traitent les valeurs d'intensité de pixels comme des variables aléatoires. Par conséquent, elles ignorent les dépendances avec les intensités des pixels voisins et ne capturent pas efficacement plusieurs propriétés de texture. Les statistiques du deuxième ordre peuvent être utilisées pour développer des descripteurs plus efficaces.

b) La perception humaine de la texture

La texture, étant l'une des primitives visuelles de base, a été étudiée largement par les psycho-physiciens à travers les trois décades passées. Julesz et ses collègues [Julesz, 1983] ont étudié la texture dans le contexte de la vision pré-attentive. Julesz a utilisé le mot *textons* pour décrire de telles primitives qui incluent par conséquent la couleur, l'orientation, la longueur et la largeur, les lignes de terminaison, et les intersections de segments de droites. Les travaux de Julesz s'intéressent à la caractérisation de la texture de bas niveau en utilisant les *textons*, alors que d'autres auteurs s'intéressent à la perception de la texture par des primitives de haut niveau. A la différence de la perception pré-attentive, les primitives de haut niveau sont concernées par l'analyse attentive. Trois

primitives importantes dans la perception humaine de la texture : répétition, orientation et complexité. La répétition est liée aux motifs périodiques et généralement associée à la régularité. Par exemple le mur de briques est un motif répétitif alors que l'image de l'eau de l'océan est non répétitive. L'orientation est reliée à la présence ou l'absence de textures directionnelles comme les vagues et le bois. La complexité est liée à la complication de description de textures, elle est liée à la primitive de rugosité de Tamura.

c) Les descripteurs de texture

i. Descripteurs basés sur l'analyse dans le domaine spatial

1. Les matrices de cooccurrence

La texture se manifeste comme une variation d'intensité de l'image dans une région. En se basant sur les travaux de Julesz qui reposent sur les textons et son hypothèse disant que la discrimination humaine de la texture est basée sur les statistiques d'intensités d'image de deuxième ordre, on accorde beaucoup plus d'attention à la caractérisation de la distribution spatiale d'intensité de la texture. Ce qui a conduit à l'émergence du descripteur très populaire *la matrice de cooccurrence* [Bello, 1996]. Les matrices de cooccurrence sont basées sur les statistiques de deuxième ordre des paires de valeurs d'intensités des pixels d'une image. Une matrice de cooccurrence calcule l'expression qui montre comment apparaissent les paires de niveaux de gris des pixels séparées par une certaine distance et qui se prolongent sur une certaine orientation, dans une image. Soit $(x, y) \in \{1, \dots, N\}$ les valeurs d'intensité d'un pixel de l'image à la position (x, y) . Soit $[(x_1 - x_2)^2 + (y_1 - y_2)^2]^{1/2} = d$ la distance qui sépare les deux pixels aux positions (x_1, y_1) et (x_2, y_2) , et qui ont les intensités i et j , respectivement. Les matrices de cooccurrence pour une distance d sont données par :

$$c(d) = [c(i, j)], i, j \in \{1, \dots, N\}$$

où : $c(i, j)$ est le nombre de paires de pixels séparés par la distance d qui satisfait $I(x_1, y_1) = i$ et $I(x_2, y_2) = j$. Notant que l'orientation entre les paires de pixels peut être utilisée pour distinguer les matrices de cooccurrence pour une distance d . Malgré tous les travaux sur ces descripteurs de primitives, il est apparu que cette caractérisation de texture n'est pas très efficace pour la classification et la recherche. Elle est très coûteuse en temps de calcul et par conséquent rarement utilisée dans les applications de base de données.

2. Les primitives de Tamura

Le travail de Tamura [Tamura, 1978], est l'un des travaux influents sur les primitives de texture. Il caractérise la texture image à travers les dimensions de rugosité, contraste, orientation, linéarité, et régularité.

Rugosité :

La rugosité correspond à l'échelle de l'image. Prenant l'exemple de deux images aériennes d'une ville prises à différentes altitudes : celle prise de la distance la plus grande est dite moins brute que celle prise d'une distance plus petite. Alors une image avec une résolution plus fine a une texture plus brute. Plusieurs approches pour le calcul de cette propriété de texture sont proposées dans la littérature. En général, ces approches essayent de mesurer le niveau du taux de changement spatial dans les intensités de l'image et alors définir le niveau de rugosité.

Contraste :

Le contraste mesure le taux de variation d'intensité locale présente dans l'image. Il est aussi lié à la qualité de l'image. Une image avec un grand contraste est considérée comme étant de meilleure qualité qu'une autre avec un moindre contraste. Le contraste est défini par :

$$F_{con} = \sigma / (\alpha 4) n$$

Où n est un nombre positif, σ est la déviation standard de la distribution probabiliste des niveaux de gris et $\alpha 4$ est une mesure de polarisation entre les régions noires et les régions blanches dans l'image.

Orientation :

L'orientation est une propriété globale de la texture, elle est due à la forme des éléments de la texture et aux règles de placement utilisées pour créer la texture. Le degré d'orientation mesuré sur une échelle de 0 à 1 peut être utilisé comme un descripteur.

ii. Descripteurs random ou autorégressifs

La texture peut être considérée comme une matrice de nombres aléatoires. Alors, l'intensité de chaque pixel est une variable aléatoire.

On peut modéliser une image par une fonction $f(r, \omega)$, où r est le vecteur de position qui représente la localisation de pixel dans l'espace 2D et ω est un paramètre aléatoire. Pour une valeur donnée de r , $f(r, \omega)$ est une variable aléatoire. Dès qu'on sélectionne une texture spécifique ω , $f(r, \omega)$ soit une image –une fonction à travers la matrice 2D indexée par r -. $f(r, \omega)$ est nommée un champ aléatoire.

Les modèles de champs aléatoires imposent des hypothèses sur la distribution de l'intensité. Ils sont très utilisés pour la synthèse de texture. Il y a une littérature très riche sur les modèles de champs aléatoires pour l'analyse de texture. Citons : le champ aléatoire Markovien (MRF) [Besag, 1973] et le modèle autorégressif simultané.

iii. Descripteurs fréquentiels

Au lieu de calculer les primitives de texture dans le domaine spatial, une alternative est d'utiliser les primitives de domaine de transformée. La transformée discrète de Fourier

(DFT) et la transformé cosine discrète (DCT) ont été largement utilisées pour la classification de texture dans le passé. D'autres travaux récents utilisent la transformée de Gabor [Dunn, 1994] et la transformée en ondelette (DWT) [Daubechies, 1990]. Les primitives qui en découlent caractérisent la distribution des contours orientés dans l'image et cela à différentes échelles.

1.2.1.3. Recherche basée sur la forme

L'utilisation de la forme est moins développée que celle de la couleur ou de la texture. Ceci est dû principalement à la complexité inhérente de sa représentation [Castelli, 2002]. C'est le cas de système ARTISAN qui utilise la forme pour la recherche d'image [Eakins, 1996]. La question centrale est comment représenter l'information sur la forme -quel soit complète ou partielle- de sorte qu'elle soit facile à extraire, à apparier et à rechercher ?

Cinq questions de base doivent être résolues :

L'image et la préparation de la requête. Comment extraire les formes de l'image ?

Beaucoup de techniques d'extraction de la forme ont vu le jour, comme celles qui segmentent progressivement l'image pour extraire les primitives des niveaux inférieurs associés comme par exemple les contours qui apportent une représentation partielle de la forme.

La formulation de la requête et l'extraction de la forme sont fortement liées. Le mécanisme de spécification de la requête fourni par l'interface utilisateur (requête par l'exemple, requête par mots clés, dessin d'esquisse..) doit s'apparier strictement au processus d'extraction de la forme et mettre l'accent particulièrement sur la représentation de la forme durant la recherche.

La représentation de la forme. Comment est représentée la forme? Y a-t-il une invariance à une classe de transformations ? La représentation est elle basée contour ou basée région ? Est ce qu'elle est basée sur des primitives locales ou des attributs globaux ? Est ce que la relation spatiale entre les différentes parties ou primitives est représentée explicitement ? La représentation est elle multi-échelle ?

L'appariement de la forme. Comment sont apparés les requêtes et les éléments de la base de données? L'appariement est basé sur le hachage géométrique, appariement de graphe, minimisation d'énergie, formulation probabiliste ? Comment est représentée la similarité entre deux objets représentés ?

Indexation et recherche. Comment est organisée la base de données ? Utilisent-elles des catégories ? Le processus de recherche est il guidé par des modèles ?

Validation. Les résultats des approches utilisées sont ils justes et précis ? Quel est le taux d'efficacité de la recherche.

a) Préparation de l'image et formulation de la requête

Les deux questions sur la manière dont les images sont préparées avant leur stockage dans la base d'images et comment les requêtes peuvent être formulées sont intimement liées à l'utilisation de la forme comme mécanisme d'indexation. Une forme 2D est représentée complètement par son contour. Un contour est une courbe continue dans le plan, et peut être définie par un grand nombre de points [Castelli, 2002]. Cette représentation volumineuse n'est pas efficace pour la recherche d'images ce qui nécessite l'utilisation de représentations partielles.

Tandis qu'une description complète de la forme est utilisée dans l'indexation, l'image doit être segmentée et les formes entières doivent être stockées. Ce processus apparaît simple lorsque l'image contient des formes binaires comme les marques commerciales, les logos, signatures, clipart, design, graphiques,...etc. Toutefois, la tâche de ségrégation de l'arrière plan est redoutable. Dans certains domaines, il est possible d'utiliser la segmentation automatique, par exemple l'intersection d'histogrammes couleurs est utilisée pour éliminer les objets de la base d'images qui diffèrent en couleur des objets recherchés. A cause de la difficulté de ségrégation d'arrière plan, des représentations partielles peuvent être utilisées quand les besoins de l'application les permettront. Les méthodes les plus communes se basent sur le contenu du gradient, qui indique les frontières d'objet.

b) La représentation de la forme

Dans le cadre de la recherche d'image, seulement les représentations approximatives sont pratiquement utilisées. Il y a une relation entre la complexité de la représentation et sa puissance à capturer différents aspects de la forme. Toutefois, la nature ambiguë des formes fait qu'il est la plupart des temps impossible d'analyser cette relation formellement. Par conséquent, la forme est représentée par une variété de descripteurs comme les moments, les descripteurs de Fourier (FD), les invariants géométriques et algébriques, les polygones, les polynomiales, les splines,... pour la reconnaissance d'objet et l'indexation des bases d'images.

La comparaison de forme est un problème très difficile, il est connu que ni les descriptions mathématiques ni les métriques formelles pour la comparaison de forme, ne capturent entièrement les aspects importants de la forme. L'observation clé est que la forme est multi-face. Les méthodes de représentation existantes peuvent être catégorisées selon les faces ciblées par la représentation. On distingue entre méthodes qui décrivent les contours et méthodes qui décrivent l'intérieur, entre représentations globales et celles locales. On distingue aussi entre les approches basées composition et celles basées déformation.

i. Contour Vs Intérieur

Les représentations de contour se basent sur la courbe fermée qui entoure la forme. Cette courbe a été décrite par plusieurs modèles, entre autres, les chaînes de code, les polygones, les arcs circulaires, les splines, les polynômes explicites et implicites, et les

descripteurs de contour de Fourier. Un contour peut aussi être décrit par ses primitives, par exemple, les maximums des courbures ou les points d'inflexion.

A la différence des approches basées contour, les descripteurs de l'intérieur de la forme mettent l'accent sur la « matière » à l'intérieur de contour fermé. Beaucoup de modèles ont été utilisés pour représenter cet intérieur comme les collections de primitives (rectangles, disques,...), modèles déformables, modèles en squelettes ou simplement un ensemble de points comme c'est le cas dans la morphologie mathématique.

Il est clair que les deux représentations sont complètes et chacune peut être utilisée pour calculer l'autre et cela par le remplissage de la région intérieure ou par le traçage du contour. Néanmoins, les deux représentations sont interchangeables dans le sens du contenu informatif. Donc, une représentation idéale est celle qui inclut les deux représentations à la fois.

ii. Locales Vs Globales

Une autre alternative de représentation peut voir la forme d'une perspective locale ou d'une autre globale. Les anciens modèles de l'indexation par la forme utilisent des primitives comme les moments, les régions,...etc. qui sont basées sur la forme entière et par conséquent globales. Alors que les représentations locales restreignent les calculs au petit voisinage de la forme comme par exemple les points d'inflexion du contour.

Les représentations purement globales sont affectées par des variations comme l'occultation, alors que celles purement locales sont sensibles au bruit. Les deux approches sont complémentaires et l'utilisation d'un compromis peut avoir de meilleurs résultats.

iii. Composition Vs Déformation

Les formes peuvent être vues comme une composition d'éléments simples ou comme une déformation de formes plus simples. Dans les approches basées composition les formes sont composées de composants simples par exemple : une raquette de tennis est décrite par une tête ovale attachée à un bras rectangulaire.

Le partitionnement peut se baser sur la forme globale ou sur la preuve locale. Exemple de partitionnement basé forme globale est l'approche de la longueur minimale de description (MDL). Utilisant cette approche, une forme est représentée par une combinaison de primitives sélectionnées d'une collection, pour chaque combinaison deux valeurs sont calculées : l'erreur appropriée et le coût d'encodage. Le coût d'encodage exprimé en bits est nommé la longueur de description. Il mesure la complexité de description de la combinaison. Une fonction ascendante des deux valeurs - l'erreur appropriée et la longueur de description- définit l'énergie totale. La représentation de forme avec la valeur d'énergie totale la plus faible est sélectionnée.

Une forme peut aussi être décomposée en parties basées sur la preuve locale. L'approche basée sur les parties n'est pas applicable universellement. Les formes

biologiques comme le contour de corps calleux dans le cerveau, les feuilles,... sont souvent mieux décrites en utilisant la déformation de formes simples.

1.2.2. Requêtes partielles

Les moteurs de recherche d'images par le contenu existants offrent à l'utilisateur un mode de requête par l'exemple. L'inconvénient dans ce mode est que l'utilisateur lance son requête qui est une image entière. Cependant, les requêtes partielles permettent de s'affranchir de ces limitations et de s'accommoder avec les nouveaux besoins. Parfois l'utilisateur est amené à lancer une recherche sur une partie de l'image et pas sur l'intégralité de celle-ci. Le système doit dans ce cas rechercher les images de la base qui contiennent des parties similaires. La Figure I-16 présente un exemple de recherche basé sur les requêtes partielles.

Cette recherche est basée sur les requêtes partielles à la place de celles globales qui s'avère inefficaces car reposent sur des similarités calculées sur l'image entière. Une requête partielle consiste à décrire et retrouver les parties qui présentent un intérêt visuel dans l'image. Ces parties peuvent être : des régions d'intérêt (zones visuellement homogènes, par exemple la peau), il s'agit ici de la segmentation d'images, ou des points caractéristiques détectés dans les zones d'images à forte variation locale (par exemple motifs texturés, objets particuliers, etc.). Cette alternative peut être utilisée à la place de segmentation lorsque cette dernière soit difficile à réaliser.

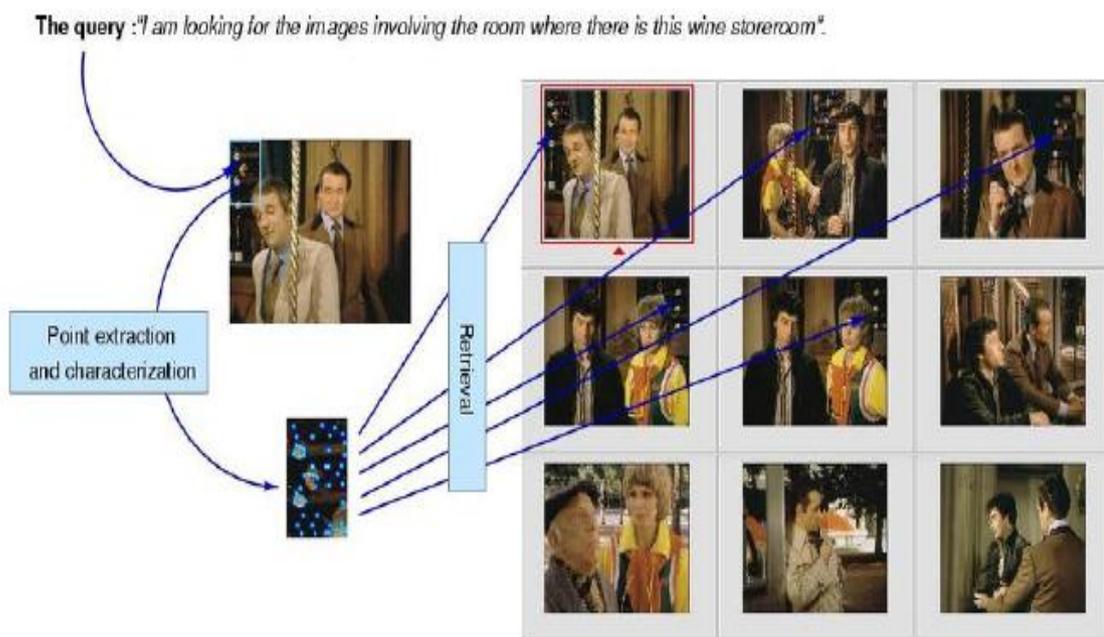


Figure I-16. Exemple de recherche d'image par le contenu à base de requêtes partielles.



Figure I-17. Processus de recherche d'image par les requêtes partielles.

Dans les approches partielles de recherche d'images, le processus de recherche passe par trois étapes principales : définition de régions, choix de paramètres et finalement l'étape de l'appariement (Figure I-17).

La définition de régions : Il existe une multitude de possibilités de traitement de bas niveau. D'un tel traitement se découle généralement des primitives de base qui sont des régions.

Le choix de paramètres : est l'étape qui permet la description des régions détectées. Plusieurs approches peuvent être utilisées pour cette fin, nous présentons quelques une dans la suite.

L'appariement : C'est la phase d'appariement proprement dite, elle consiste à trouver les descripteurs les plus semblables entre images. Cela nécessite la définition d'une mesure de similarité qui est généralement reliée à l'approche de caractérisation.

1.2.2.1. Les régions

a) Segmentation de l'image

La segmentation image est le processus de division d'image en régions ou objets. L'idée de base de la segmentation d'images c'est le groupement de pixels individuels en régions s'ils sont similaires. Similaire peut signifier qu'ils ont la même intensité, provient de la même texture, construit une forme...etc. plusieurs techniques de segmentation d'image existent et varient selon la complexité, la puissance et le domaine d'application [Philips, 2000].

La segmentation d'image en niveaux de gris est généralement basée sur l'une des deux propriétés de valeurs de niveaux de gris dans une image qui sont la discontinuité et la similarité. Donc, deux catégories d'algorithmes peuvent être distinguées :

i. Segmentation par classification ou seuillage

Cette approche trouve l'emplacement d'un objet explicitement selon la propriété de similarité. Les algorithmes appartenant à cette approche utilisent le seuillage. Le seuillage le plus simple étant de définir un intervalle d'intensité dans l'image originale et de sélectionner ensuite les pixels appartenant à cet intervalle comme appartenant à l'avant plan, et de rejeter tous les autres pixels à l'arrière plan. Plusieurs types de seuillage existent, cela peut inclure le seuillage bidimensionnel, le seuillage multi-bande ou encore le seuillage à partir de texture [Russ, 1998].

ii. Segmentation basée contours

L'un des inconvénients de la segmentation par seuillage est que les pixels sont sélectionnés principalement à base de leurs intensités et seulement secondairement par l'emplacement. Ce que signifie qu'il n'y a aucune garantie que les régions soient contiguës. Une autre alternative basée sur les contours consiste à détecter les contours des objets explicitement en utilisant la propriété de discontinuité.

Selon [Zhang, 2006] ces deux catégories peuvent être considérées comme complémentaires. D'autres méthodes existent dans la littérature, elles sont généralement associées à la compréhension d'images dans le sens de vision par ordinateur et la robotique. Les deux approches les plus largement décrites sont la décomposition/fusion et la croissance de régions. Ces approches appartiennent à la segmentation basée régions [Russ, 1998].

b) Description de régions

i. Description par couleur

Comme avec la forme et la texture, il est possible d'extraire les descripteurs couleur à partir de l'image entière, le résultat étant une caractérisation globale. Ou séparément à partir de différentes régions ce qui est une caractérisation locale. En effet, les descripteurs globaux capturent le contenu de l'intégralité de l'image tandis que les descripteurs locaux peuvent être combinés avec la position et la taille des régions correspondantes pour décrire la structure spatiale de l'image en couleur.

Un des inconvénients majeurs de l'extraction globale d'histogrammes couleur est qu'elle ne prend pas en considération la distribution spatiale de la couleur à travers les régions de l'image. Un bon nombre de méthodes ont été développées pour intégrer la couleur et l'information spatiale dans les requêtes basées contenu citant la méthode qui partitionne l'image en 5 régions non chevauchées. Le calcul de descripteurs couleur à partir de chacune des régions permet un appariement accentué sur quelques régions ou encore accommoder l'appariement d'images à celles qui ont subi une rotation. La Figure I-18(a) présente le partitionnement de l'image à 5 régions $r_0 - r_4$ dans laquelle une région centrale r_0 capture les primitives couleurs de n'importe quel objet central. La Figure I-18(b) présente un partitionnement de l'image en 16 régions distribuées uniformément $g_0 - g_{15}$. La similarité des images basée sur les descripteurs couleur spatiaux peut en suite être mesurée par le calcul de la somme pondérée de similarité de régions [Castelli, 2000].

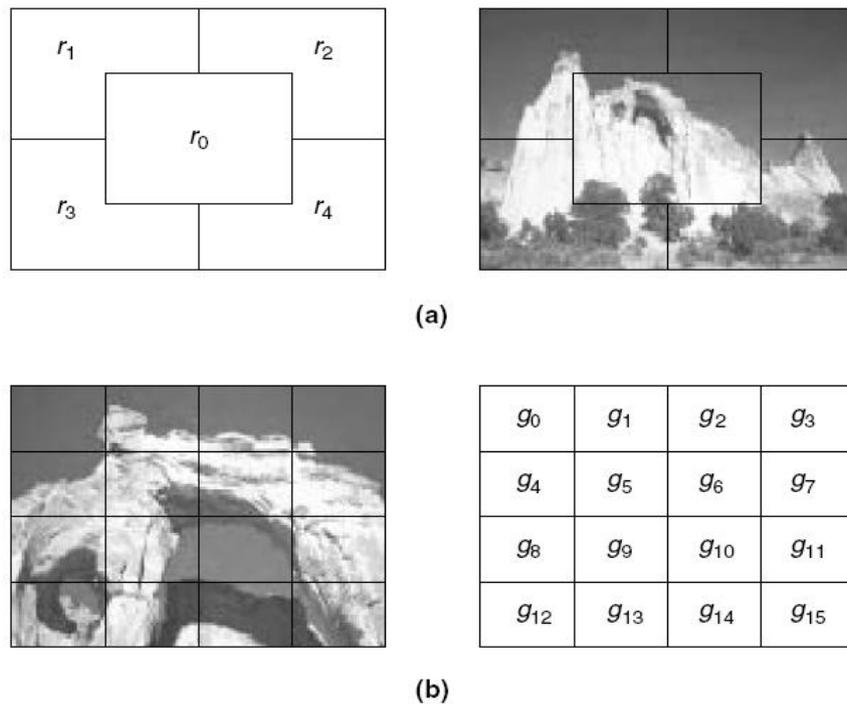


Figure I-18. Représentation en utilisant les descripteurs basés couleur de région. [Castelli, 2002]

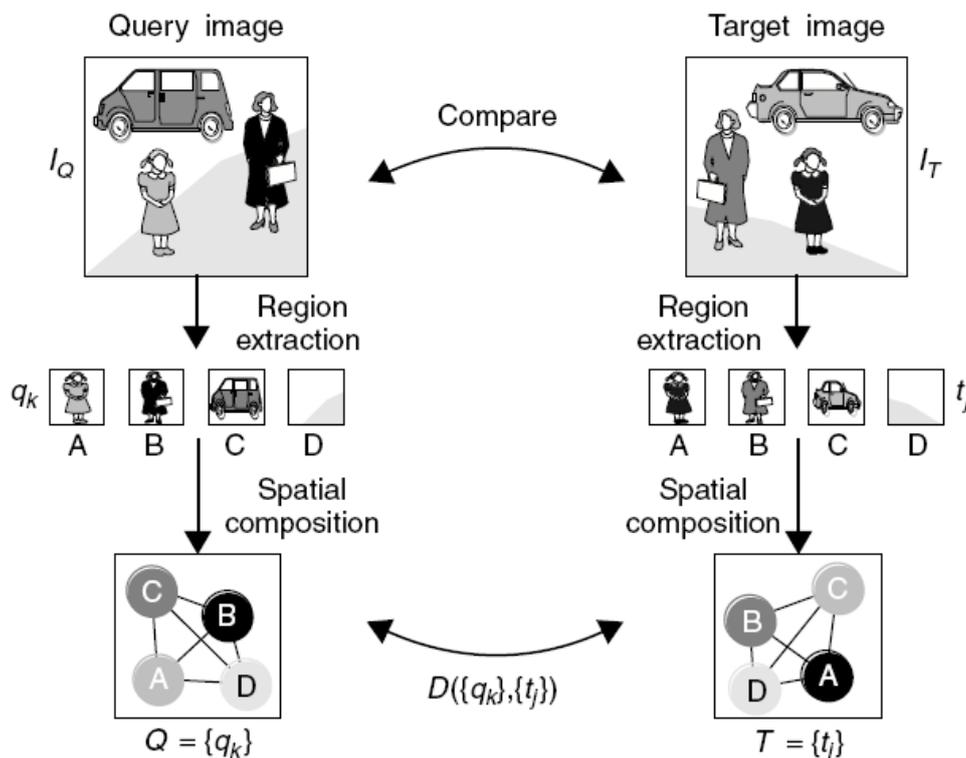


Figure I-19. Approche d'appariement d'images intégrant les primitives d'espace et de couleur par la comparaison des arrangements spatiaux des régions. [Castelli, 2002]

ii. Description par texture

Il faut noter que la texture est considérée comme un intérêt visuel important dans le domaine émergent de la recherche par le contenu dans les bases de données multimédia.

L'analyse de texture image durant les trois décades passées a concentré principalement sur la classification, la segmentation et la synthèse de texture. Dans la classification de texture l'objectif est d'assigner un label unique à chaque région homogène. Par exemple, les régions dans une image satellitaire peuvent être classifiées selon glace, eau, forêt, champ agricole...etc. Pour réaliser la classification, on associe chaque pixel dans l'image à son label de classe correspondant, celui de la région à laquelle ce pixel appartient.

La segmentation de texture se reporte au partitionnement de l'image, chacune de ces partitions soit homogène. Notant que l'homogénéité dans la couleur et la texture ne segmente pas forcément l'image à des objets significatifs. Le résultat de la segmentation déforme généralement les objets d'intérêt ce qui fait que la segmentation est souvent utilisée avec la classification. L'objectif de synthèse de texture est de générer une texture qui est perceptiblement indifférenciée de celle donnée en exemple, la texture ajoute du réalisme aux images de synthèse. Ces textures de synthèse peuvent être utilisées dans des applications comme le mapping de texture. Le mapping de texture est utilisé dans le graphisme pour générer les détails des surfaces des objets de synthèse [Castelli, 2000].

Les mêmes approches globales utilisées pour la description de texture sont utilisées dans un contexte de région.

1.2.3. Requêtes locales par points d'intérêt

Un exemple de système de recherche d'images par le contenu est IKONA de l'équipe IMEDIA (INRIA). Il est basé sur les points d'intérêt Figure I-20.



Image d'entrée



Image résultante avec les points d'intérêt

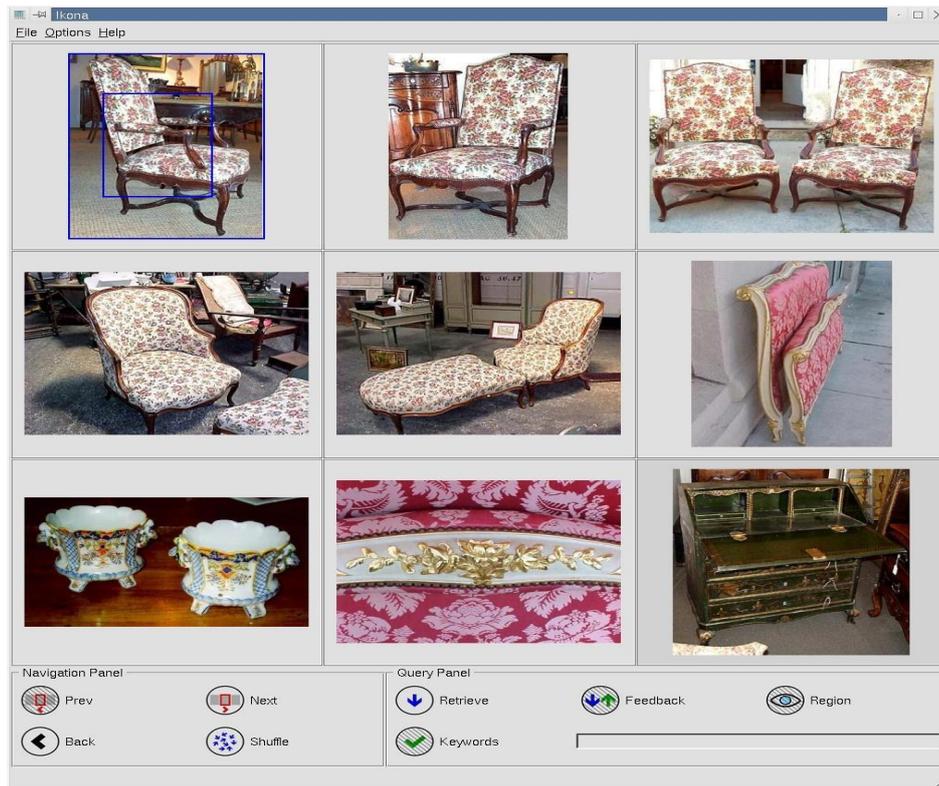


Figure I-20. Exemple de recherche d'image en utilisant les points d'intérêt.

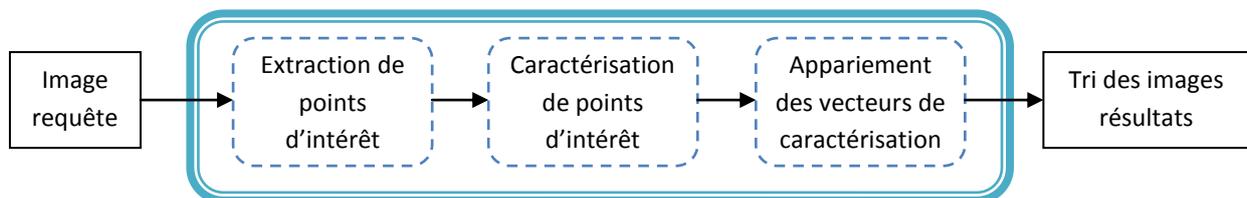


Figure I-21. Processus de recherche d'image par les points d'intérêt.

Le processus de recherche locale par points d'intérêt (Figure I-21) suit les trois étapes suivantes :

Extraction de points d'intérêt : c'est la première étape de ce processus, elle consiste à détecter les points les plus importants dans l'image requête. Ces points peuvent être des jonctions, des points d'intersection, des coins...etc. plusieurs approches pour réaliser l'extraction, ils sont détaillées dans le prochain chapitre.

Caractérisation de points d'intérêt : c'est une phase qui va permettre de valider l'hypothèse qui dit que les points d'intérêt présentent un contenu informatif important en capturant cette information. Dans cette étape, les points détectés dans la première étape doivent être caractérisés par des vecteurs caractéristiques qui capturent l'information de signal au voisinage du point détecté. Les approches de caractérisations sont nombreuses, le chapitre suivant passe en revue les plus importantes.

Appariement des vecteurs de caractérisation : c'est la dernière étape du processus, et qui consiste à comparer les vecteurs caractéristiques des points d'intérêt des deux images en correspondance. Les approches d'appariement les plus reconnues sont présentées dans le chapitre II de ce mémoire.

1.2.3.1. Les points d'intérêt

Les descriptions globales s'avère inadéquates quand il s'agit de rechercher un objet d'une image requête dans une base d'images ou de rechercher juste une partie de l'image requête. De ce fait, l'utilisation de descriptions locales constitue une étape clé de la recherche d'objet dans les bases d'images. Pour réaliser cette description il faut extraire des primitives de l'image appelées points d'intérêt découvrant les zones qui présente un contenu informatif important.

a) Définition

Un point d'intérêt comme défini dans [Schmid, 1996] correspondent à un changement bidimensionnel du signal comme par exemple les coins et les jonctions en T. Cependant, un point d'intérêt est plus général qu'un coin, pourtant ils sont utilisés dans la littérature comme équivalents. Un exemple de point d'intérêt est donné dans la Figure I-20.

b) Les avantages de points d'intérêt

Après avoir détectés, chaque point d'intérêt va être décrit par un vecteur de caractéristiques photométriques et géométriques. Ce qui permet d'avoir une caractérisation locale robuste aux occultations et éventuellement robuste à d'autres transformations comme par exemple le changement d'illumination, la rotation, le changement d'échelle...etc.

La description utilisant les points d'intérêt est dite locale car repose sur l'utilisation de points d'intérêt qui sont jugés les moins sensibles au contenu géométrique de l'image et qui sont robustes aux transformations de l'image et aux occultations.

c) Le choix des points d'intérêt comme des primitives locales

On veut pouvoir reconnaître une image même si elle était sujet de quelques transformations, une image peut être recadrée, tournée, agrandie,...etc. un des problèmes majeurs des descripteurs globaux est qu'ils sont mal adaptés pour faire face à ces transformations. On peut alors chercher à n'utiliser que des descriptions locales. Pour cela, on va rechercher des zones d'intérêt dans l'image, puis décrire chacune de ces zones. De manière générale, on peut chercher des points, des courbes ou des régions de l'image.

Les approches qui utilisent les régions présentent un problème car les petites régions sont très instables. Une solution possible est l'utilisation d'un nombre très limité de régions celles qui sont intéressantes. Mais rien ne prouve que les régions extraites aient quelques valeurs sémantiques.

D'autres travaux favorisent l'emploi de points d'intérêt car révélé plus fructueux. Un point d'intérêt est un point qui ressemble le moins possible à ses voisins ou encore un point porteur d'un maximum d'informations au sens de la théorie de l'information.

En plus, ces points résistent mieux aux transformations que l'on peut faire subir aux images, ils sont dits répétables. Les approches qui s'affilient sous cette catégorie sont discutées dans le chapitre suivant.

Conclusion

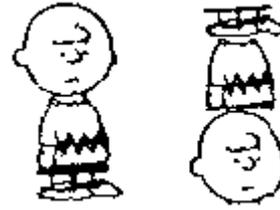
La recherche d'images par le contenu est un résultat naturel du manque d'efficacité de la recherche textuelle. Cette dernière a prouvé ses limites face aux larges bases d'images à indexer car nécessite plus de mains d'œuvre et plus de temps. Elle est également classée comme méthode subjective pour la représentation du contenu de l'image.

La communauté de la vision par ordinateur se trouve orientée vers les approches de recherches par le contenu. Ce contenu peut être une couleur, une texture ou une forme. Les approches qui utilisent ces primitives sont dites globales. Il se trouve que le but de la recherche d'images ne se porte pas toujours sur la totalité de l'image mais sur juste une partie de cette image. Ce besoin a fait émerger une nouvelle catégorie de recherche d'images par le contenu nommé requêtes partielles. Ces requêtes utilisent les mêmes primitives utilisées par les requêtes globales mais dans un contexte de région. Il s'est avéré ensuite que même ces approches ne permettent pas de représenter précisément le contenu de l'image. Le but est plutôt de représenter l'image localement (au niveau de pixel). Les approches résultantes s'appellent requêtes locales par points d'intérêt et qui se contentent de représenter l'image par un ensemble de points qui présente un intérêt par rapport aux autres points constituant l'image.

Dans ce chapitre, nous avons survolé sur ces différentes approches pour mieux situer les approches basées points d'intérêt par rapport aux autres approches. Ceci car le travail présenter dans ce mémoire propose une ébauche d'amélioration dans les détecteurs de points d'intérêt.

Le chapitre suivant présente les différentes phases de processus de recherche locale d'images par points d'intérêt cela inclut les différentes approches de détection de points d'intérêt, les approches de caractérisation et les approches d'appariement.

Chapitre II Les Points d'intérêt



Introduction

Ce chapitre met l'accent sur les points d'intérêt et les différentes phases du processus de mise en correspondance en utilisant ces points. Pour être capable de comprendre ce processus nous avons jugé intéressant de doter ce travail d'un état de l'art des différents détecteurs de points d'intérêt dédiés aux images en niveau de gris ainsi que celle en couleur. Ce chapitre contient également un état de l'art des approches de caractérisations et finalement un état de l'art des approches d'appariement. Ce chapitre est structuré comme suit : la première section contient une définition des points d'intérêt. La deuxième section présente les différents types de points d'intérêt. La dernière section présente d'une manière détaillée les différentes phases du processus d'appariement. Nous avons consacré la première partie de cette section pour répondre à la question "quelle primitive utilisée ? " La réponse ne s'avère pas étonnante car c'est la plus coutumière dans la littérature. Les points d'intérêt font la base de la plupart des applications qui exigent une description locale de l'image ; ils sont détaillés dans un état de l'art qui fait l'objet de la deuxième partie de la section. Les deux autres parties de la section présentent les états de l'art des approches de caractérisation et les approches d'appariement.

1. Définition des points d'intérêt

La notion de points d'intérêt a été introduite pour la première fois par Moravec [Moravec, 1977]. Pour lui, les points d'intérêts correspondent à un changement bidimensionnel du signal comme par exemple les coins, les jonctions en T, en Y...etc. Toutefois, un point d'intérêt est plus général qu'un coin, pourtant ils sont utilisés dans la littérature comme équivalents [Schmid, 1996].

Les points d'intérêt utilisés dans ce travail sont les points de focus. Les détecteurs de points de focus utilisent la symétrie pour calculer des points qui ont un intérêt visuel important vis-à-vis le system visuel humain.

2. Les différents types de points d'intérêt

Parmi les points d'intérêt on trouve des coins en L, des jonctions en T, des jonctions en Y ou des points de forte variation de texture. Ils correspondent à des doubles discontinuités de la fonction d'intensité provoquées par des discontinuités de la fonction réflectance ou de profondeur [Boyer, 2002].

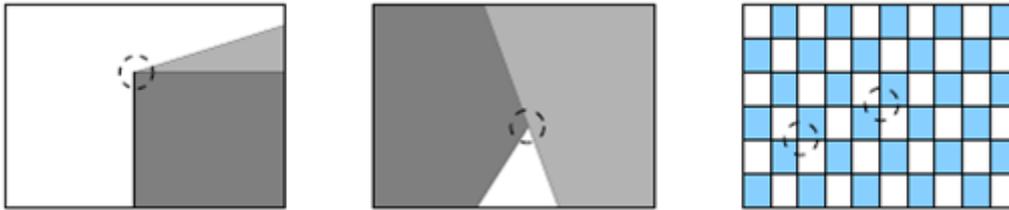


Figure II-1. Types de points d'intérêt : coins, jonction en Y et points de fortes variations de texture. [Boyer, 2002]

3. Phases du processus d'indexation par point d'intérêt

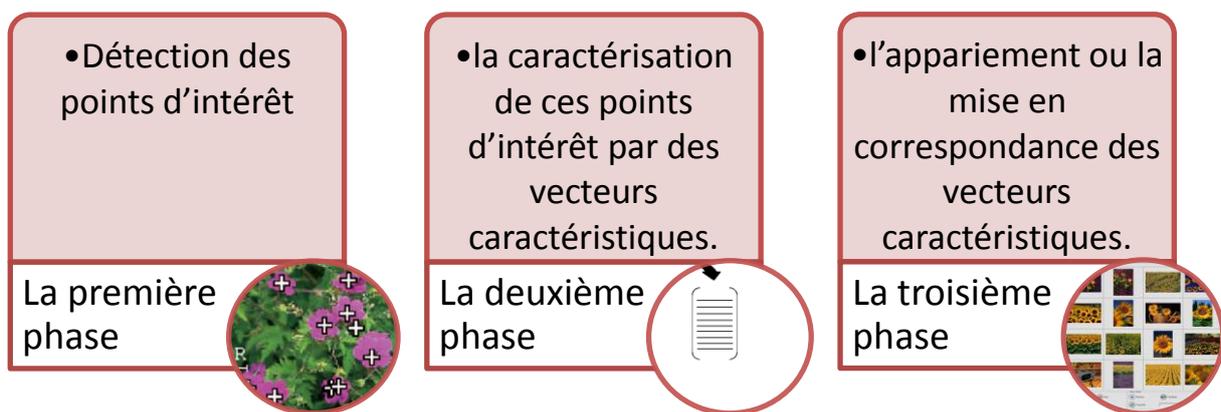


Figure II-2. Etapes du processus de mise en correspondance.

La Figure II-2 montre les différentes phases du processus de mise en correspondance par points d'intérêt. Dans cette section, nous allons détailler les trois étapes du processus de mise en correspondance en menant un état de l'art des différentes approches utilisées pour accomplir chacune des étapes.

3.1. Détection de primitives

3.1.1. Pourquoi les points d'intérêt

L'extraction d'indices dans une image est une étape indispensable en vision par ordinateur. Dans ce chapitre nous discutons les points d'intérêt.

La détection des points d'intérêt directement à partir du signal semble la technique la plus puissante. D'autre part, la détection des points d'intérêt s'avère fiable quelques soient

les images traitées, car ils donnent des résultats comparables sur des images très texturées comme sur des images contenant moins de textures.

Le choix des points d'intérêt est fait d'une part, parce que beaucoup de travaux ont montré leur utilité [Schmid, 1996], [Gouet, 2000], [Mikolajczyk, 2006]...etc. D'autre part, parce qu'ils sont locaux donc robustes en présence de transformations d'images comme l'occultation, la rotation, le changement de point de vue... etc.

3.1.2. État de l'art de points d'intérêt

Plusieurs types de détecteurs de points d'intérêt sont décrits dans la littérature. les détecteurs de coins et de jonctions : [Moravec, 1977], [Harris, 1988], [Smith, 97], les détecteurs basés sur les ondelettes [Wu, 1995], les détecteurs basés sur la différence gaussienne [Lowe, 1999], les détecteurs couleurs [Gouet, 2000], les détecteurs quasi-invariants photométriques [Weijer, 2005], les détecteurs basés sur l'intérêt visuel des points détectés [Reisfeld, 1995], [Zelinsky, 2003], [Heidemann, 2004], [Rebai, 2006a],[Rebai, 2006b], [Lakehal, 2009] ...etc. Une comparaison a été menée par [Mikolajczyk, 2006], en prenant compte de quelques détecteurs de la liste précédente, cette comparaison a montré que la performance de tous les détecteurs en question décline graduellement, avec des taux similaires, quand le changement de point de vue augmente. Un état de l'art des points d'intérêt est donné par [Tuytelaars, 2006].

La détection des points d'intérêt peut être aussi formulée comme un problème d'optimisation résolu en utilisant la programmation génétique [Trujillo, 2006a] [Trujillo, 2006b], [Ebner, 1998]. Les algorithmes génétiques peuvent aussi donner des points d'intérêt ce qui a été montré par [Kiryati, 1996].

Ces détecteurs et plusieurs autres peuvent être classés selon les catégories suivantes :

a) Méthodes basées sur les contours

Il existe plusieurs approches dans cette catégorie, dont le principe est soit de rechercher les points de courbures maximales sur les chaînes de contour, soit d'effectuer une approximation polygonale afin de déduire des points particuliers (intersections, inflexion,...).

b) Méthodes basées sur un modèle théorique du signal

Dans ces méthodes le but est d'obtenir une précision sous pixelique en approximant le signal observé par un modèle théorique. Ces méthodes ne sont utilisables que pour des types précis de points d'intérêt comme les coins.

c) Méthodes basées sur le signal

Ces méthodes n'utilisent ni les contours ni le modèle théorique du signal ; la présence ou pas d'un point d'intérêt est indiquée par une mesure calculée directement à partir du signal. Plusieurs détecteurs s'affilient sous cette catégorie comme le détecteur de Beaudet, celui de Moravec, les coins de Kitchen et Rosenfeld, le détecteur de Harris et Stephens et celui de Susan.

i. Détecteur de Beudet

Le détecteur de Beudet est l'un des premiers détecteurs de points d'intérêt. Il définit un opérateur basé sur les dérivées secondes du signal pour juger si un point est un point d'intérêt ou non. Son point faible est qu'il est basé sur les dérivées secondes ce qui le rend plus sensible au bruit [Goshtasby, 2005]. Ce détecteur utilise la formule suivante :

$$I_{xx}I_{yy} - I_{xy}^2 \quad (1)$$

ii. Détecteur de Moravec

Les travaux les plus anciens concernant la détection iconique des points d'intérêt sont ceux de Moravec. Il a défini dans [Moravec, 1977] un détecteur qui introduit pour la première fois la notion d'auto-corrélation. Le principe était de calculer les variations d'intensité lumineuse dans quatre directions parallèles aux lignes et aux colonnes de l'image et de mesurer ensuite les différences entre une fenêtre rectangulaire autour d'un pixel et quatre fenêtres voisines. Un point qui présente des variations d'intensité dans toutes les directions est un point d'intérêt.

iii. Les coins de Kitchen et Rosenfeld

Dans leur approche, Kitchen et Rosenfeld ont utilisé la courbure des isophotes, qui est la courbure de la surface d'intensité de l'image, elle est approximée localement de façon polynomiale. Ensuite, ils recherchent les maxima locaux d'un opérateur qui estime la courbure locale multipliée par la norme du gradient. Ces maxima locaux représentent les points d'intérêt. À l'inverse des deux détecteurs précédents, ce détecteur utilise les dérivées premières et secondes en même temps.

L'approche de Kitchen et Rosenfeld est l'une des approches qui travaillent à l'ordre deux. Elle consiste à maximiser la courbure des isophotes multipliée par la magnitude du gradient. La courbure des isophotes peut s'exprimer par l'opérateur K :

$$K = \frac{I_{xx}I_y^2 - 2I_{xy}I_xI_y + I_{yy}I_x^2}{\|\vec{\nabla}I\|^3} \quad (2)$$

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

$$KR = \frac{I_{xx}I_y^2 - 2I_{xy}I_xI_y + I_{yy}I_x^2}{\|\vec{\nabla}I\|^2} \quad (3)$$

iv. Détecteur de Harris et Stephens

La contribution de Harris et Stephens [Harris, 1988] intervient pour améliorer le détecteur proposé par Moravec. Ils ont corrigé plusieurs défauts parmi lesquels la prise en

compte des variations d'intensité dans toutes les directions et l'utilisation de fenêtres circulaires au lieu de fenêtres rectangulaires autour des pixels.

Ce détecteur est resté reconnu comme le détecteur de référence dans le domaine de la vision par ordinateur. Il a été sujet à plusieurs améliorations ; la plus qualitative est celle de Schmid [Schmid, 1996] qui a proposé une version précise du détecteur de Harris dans laquelle le gradient de l'image est calculé en utilisant les dérivées d'une gaussienne plutôt qu'avec un simple masque discret $[-1 \ 0 \ 1]$. Comme le bruit contenu dans une image peut perturber considérablement le calcul de ses dérivées, il faut soit faire un lissage de l'image avant de calculer ses dérivées soit effectuer un lissage de l'opérateur de dérivation. Un moyen simple pour stabiliser les calculs des dérivées est donc d'utiliser les dérivées d'une fonction de lissage. La gaussienne est la plus communément choisie pour une telle fonction.

La nouvelle version du détecteur de Harris utilise les dérivées d'une gaussienne. Les convolutions avec les dérivées des gaussiennes intervenant dans les calculs sont implémentées de manière récursive ce qui permet des calculs rapides.

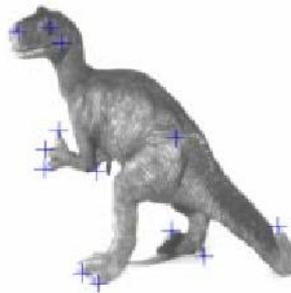


Figure II-3. Exemple de points détectés par le détecteur de HARRIS

Ce détecteur a été conçu sur le même principe du détecteur de Moravec. L'idée de base de ce détecteur est d'utiliser la fonction d'auto-corrélation pour déterminer les endroits où le signal change dans deux directions simultanément. La version précise de ce détecteur a été évaluée dans [Schmid, 1996], [Schmid, 2000]. La conclusion était que le détecteur de Harris précis donne des résultats meilleurs ou équivalents à ceux des autres détecteurs. Il a été montré qu'il est le détecteur le plus stable. Cette stabilité provient principalement du fait qu'il utilise les dérivées d'une fonction de lissage et qu'il travaille seulement à l'ordre un.

La fonction d'auto-corrélation locale mesure les changements locaux du signal. Cette mesure est obtenue par la corrélation d'un patch avec les patches voisins, c'est-à-dire avec les patches déplacés d'une petite valeur dans différentes directions

Pour un déplacement $(\Delta x, \Delta y)$ et un point (x, y) la fonction d'auto-corrélation est donnée par :

$$f(x, y) = \sum_{(x_k, y_k) \in W} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2 \quad (4)$$

Où : (x_k, y_k) sont les points dans la fenêtre centrée en (x, y) et I la fonction image et I_x, I_y représente respectivement les dérivées partielles de I par rapport à x, y .

Pour utiliser cette fonction dans la détection des points d'intérêt, il faut la calculer pour toutes les directions de changement. Toutefois, on peut éviter l'intégration à travers ces directions discrètes par l'utilisation de la matrice d'auto-corrélation. Cette matrice est dérivée en utilisant une approximation de premier ordre basée sur le développement de Taylor :

$$I(x_k + \Delta x, y_k + \Delta y) \approx I(x_k, y_k) + \begin{pmatrix} I_x(x_k, y_k) & I_y(x_k, y_k) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \quad (5)$$

Par la restitution de (5) dans (4) on aura :

$$\begin{aligned} f(x, y) &= \sum_{(x_k, y_k) \in W} \left(\begin{pmatrix} I_x(x_k, y_k) & I_y(x_k, y_k) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right)^2 \\ &= \begin{pmatrix} \Delta x & \Delta y \end{pmatrix} \begin{bmatrix} \sum_{(x_k, y_k) \in W} (I_x(x_k, y_k))^2 & \sum_{(x_k, y_k) \in W} I_x(x_k, y_k) I_y(x_k, y_k) \\ \sum_{(x_k, y_k) \in W} I_x(x_k, y_k) I_y(x_k, y_k) & \sum_{(x_k, y_k) \in W} (I_y(x_k, y_k))^2 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \\ &= \begin{pmatrix} \Delta x & \Delta y \end{pmatrix} \Lambda(x, y) \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \quad (6) \end{aligned}$$

Où :

$$\Lambda(x, y) = \begin{pmatrix} \sum_w (I_x(x_i, y_i))^2 & \sum_w (I_x(x_i, y_i) I_y(x_i, y_i)) \\ \sum_w (I_x(x_i, y_i) I_y(x_i, y_i)) & \sum_w (I_y(x_i, y_i))^2 \end{pmatrix} \quad (7)$$

La formule (6) montre que la fonction d'auto-corrélation peut être approximée en utilisant la matrice Λ qui capture la structure du voisinage d'un point d'intérêt. Cette matrice est donnée par la formule (7).

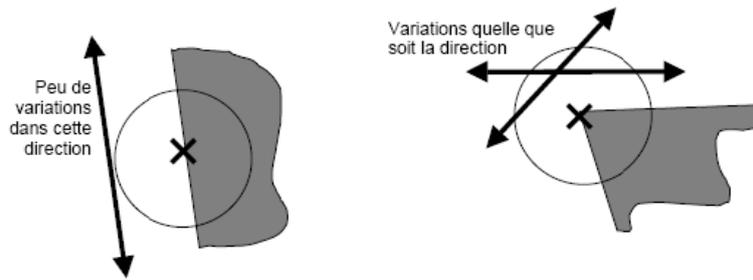


Figure II-4. Variation du signal dans le cas de contour et de coin.

Après l'obtention de cette matrice, le détecteur de Harris définit une mesure de réponse de coin R ; c'est une fonction qui utilise $Tr(M)$, $Det(M)$ dans sa formulation ce qui permet d'éviter la décomposition explicite de M et la recherche des valeurs propres de la matrice :

Où :

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \text{ matrice d'auto-corrélation,}$$

$$R = Det - k Tr^2 \quad (8)$$

R est la réponse du détecteur de Harris, tel que :

$$Tr(M) = A + B \quad (9)$$

$$Det(M) = AB - C^2 \quad (10)$$

K est une valeur empirique définie par Harris à 0.04.

v. Détecteur SUSAN

Une autre approche, introduite par Smith et Brady [Smith, 1997], a conduit au détecteur SUSAN (*Smallest Univalued Segment Assimilating Nucleus*). À la différence de toutes les méthodes déjà proposées, cette approche n'utilise pas les dérivées des intensités de l'image mais elle mesure, dans une fenêtre, le nombre de pixels ayant une intensité voisine de celle du pixel central (Figure II-5). Une forte variation des intensités signifie l'existence d'un point d'intérêt. Ce détecteur permet de détecter les contours, les coins et aussi les jonctions.

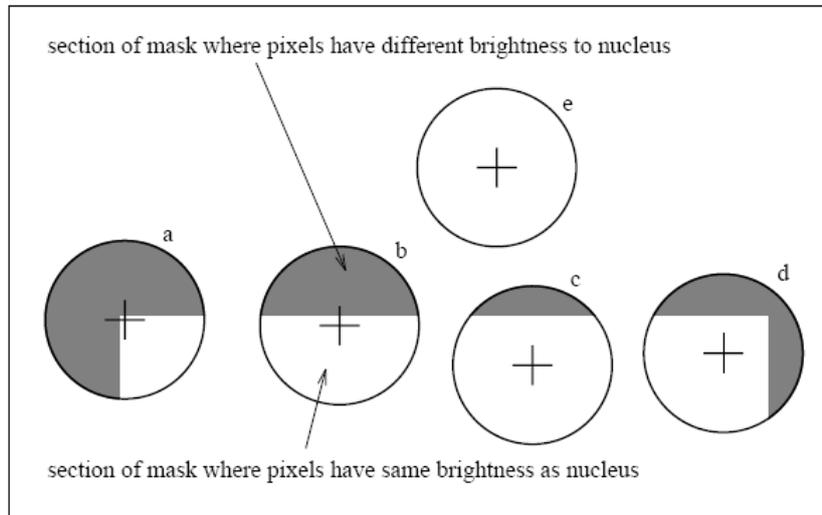


Figure II-5. Quatre masques circulaires avec coloration similaire; l'USANs sont présentés par les parties blanches des masques. [Smith, 1997]

Le principe est simple : un masque est placé sur chaque point de l'image et pour chaque point, la luminance de chaque pixel à l'intérieur du masque est comparée avec celle du centre nommé noyau. Une simple équation détermine cette comparaison :

$$c(\vec{r}, \vec{r}_0) = \begin{cases} 1 & \text{if } |I(\vec{r}) - I(\vec{r}_0)| \leq t \\ 0 & \text{if } |I(\vec{r}) - I(\vec{r}_0)| > t \end{cases} \quad (11)$$

Où \vec{r}_0 est la position du noyau dans l'image bidimensionnelle et \vec{r} est la position de n'importe quel autre pixel à l'intérieur du masque. $I(\vec{r})$ est la luminance de n'importe quel pixel, t est le seuil de variation de luminance et c est le résultat de la comparaison. Cette comparaison est faite pour chaque pixel à l'intérieur du masque et un total n des résultats de comparaison est ensuite calculé par :

$$n(\vec{r}_0) = \sum_{\vec{r}} c(\vec{r}, \vec{r}_0) \quad (12)$$

Cette valeur n donne le nombre de pixels dans l'USAN. Elle doit être minimisée en la comparant avec un seuil fixe g nommé le seuil géométrique qui est fixé à la valeur n_{max} où n_{max} est la valeur maximale prise par n . La réponse du contour est ensuite créée en utilisant la formule :

$$R(\vec{r}_0) = \begin{cases} g - n(\vec{r}_0) & \text{Si } n(\vec{r}_0) < g \\ 0 & \text{Sinon} \end{cases} \quad (13)$$

L'équation (13) est utilisée aussi pour calculer une réponse de coin mais en utilisant une valeur différente de g . Ce qui permet d'avoir un détecteur de contour et de coin en même temps.

Cette approche présente l'avantage d'être rapide mais son inconvénient majeur est la mauvaise localisation des points d'intérêt.

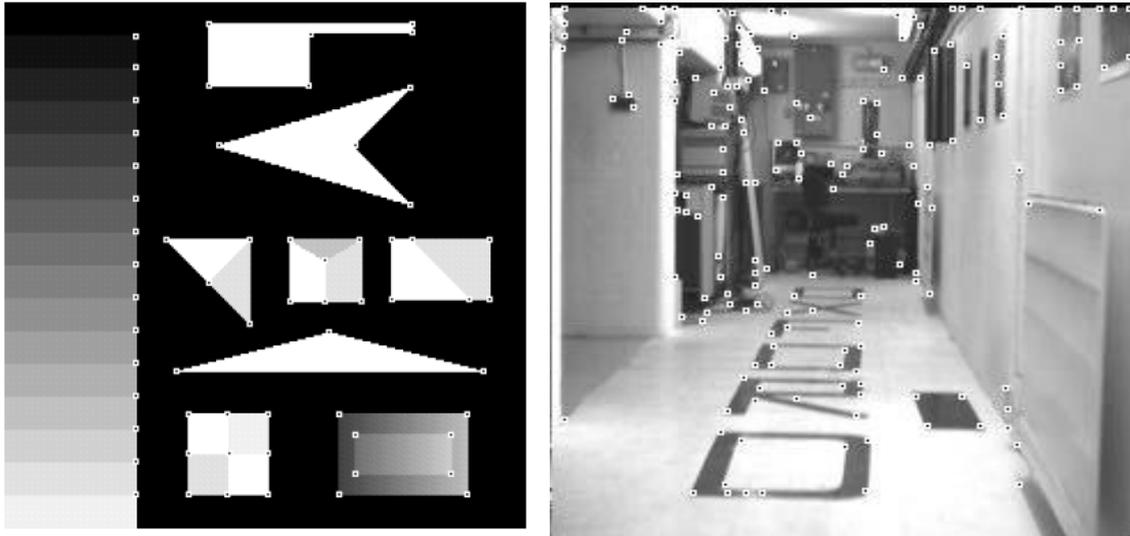


Figure II-6. Coins résultants de l'application du détecteur SUSAN sur (a) une image de test générique (synthétique) (b) sur une image réelle capturée par une caméra vidéo. [Smith, 1997]

d) Méthodes multi-échelles

La majorité des méthodes proposées souffrent du problème de changement d'échelle. Ce qui a fait émerger la notion de multi-échelles selon laquelle les points d'intérêt sont calculés sur plusieurs échelles. La totalité de ces points est ensuite utilisée pour représenter l'image. Dans l'approche proposée par [Schmid, 1997], les points sont extraits par l'opérateur Harris sur plusieurs échelles parce que ce détecteur a prouvé ses excellents résultats face à la rotation, bruit et conditions de luminance mais a échoué en présence de changement d'échelle. Le graphe ci-dessous montre l'influence du changement d'échelle sur le taux de répétabilité du détecteur de Harris. Ce taux est inversement proportionnel au facteur d'échelle.

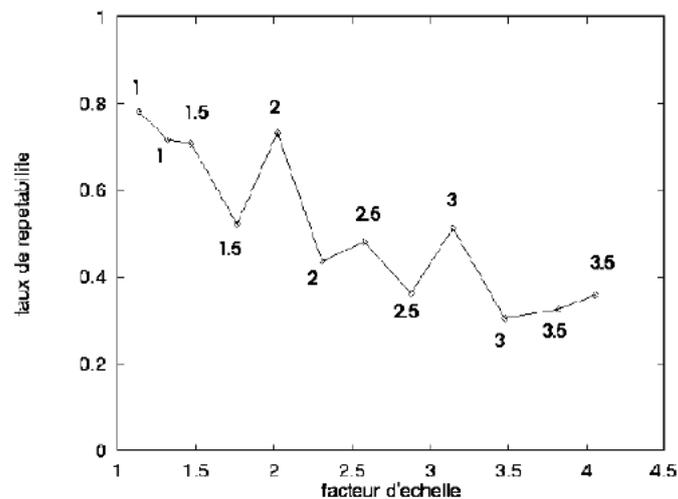


Figure II-7. Exemple d'utilisation d'un cadre multi-échelle. Il est clair que le taux de répétabilité est inversement proportionnel au facteur d'échelle. [Schmid, 1996]

e) Méthodes invariantes à l'échelle

Très peu de méthodes sont vraiment invariantes à un changement considérable de l'échelle. Etant donné une fonction f , un changement d'échelle peut être décrit par un changement de variable : $f(x) = g(u)$

Où :

$$g(u) = g(u(x)) = g(\alpha x) \quad (14)$$

De cette relation découlent les relations suivantes entre f et g :

$$f^{(n)}(x) = \alpha^n g^{(n)}(u) \quad (15)$$

Où $f^{(n)}(x)$ représente la dérivée nième de f .

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

$$\frac{[f^{(n)}(x)]^k}{f^{(k)}(x)} \quad (16)$$

Cependant des résultats expérimentaux ont montré que de tels invariants sont peu stables à un changement d'échelle supérieur à 20%.

Sachant que les méthodes invariantes à l'échelle sont incluses dans les méthodes d'invariants affines, ces dernières présentent une robustesse pour un changement très limité de l'échelle ne dépassant pas 20 %. Alors que des détecteurs standards comme celui de Harris donnent de bons résultats pour un changement d'échelle plus de 1,4. Un détecteur est considéré invariant au changement d'échelle s'il est invariant à un facteur d'échelle supérieur à 4. Les méthodes existantes utilisent une représentation tridimensionnelle de l'image (x , y et échelle) et recherche des maxima locaux dans un cube 3D qui entoure un point caractéristique ayant une valeur absolue supérieure au seuil.

Les approches existantes diffèrent essentiellement dans leurs façons de définir la représentation de l'espace d'échelle. Dans les travaux de Lindeberg [Schmid, 1996], cette représentation est construite à travers un lissage successif par des gaussiennes de différentes tailles de noyau sur l'image de haute résolution. En fait, Lindeberg propose l'utilisation de l'opérateur LoG (Laplace of Gaussien). Cet opérateur présente une symétrie circulaire ce qui le rend naturellement invariant à la rotation.

Dans le travail de [Mikolajczyk, 2004] une étude montre l'utilité de l'opérateur LoG pour la sélection automatique d'échelle. Les auteurs présentent également leur opérateur basé sur l'opérateur LoG nommé Harris-Laplace. D'autres travaux [Lowe, 1999] présentent l'opérateur DoG (Difference of Gaussien) pour calculer des maxima 3D dans la pyramide d'espace d'échelle.

f) Méthodes d'invariants affines

Les invariants affines peuvent être vus comme une généralisation d'invariants à l'échelle. Dans le cas de transformations affines, la déformation de l'échelle est différente dans chaque direction. Contrairement aux invariants à l'échelle où la variation d'échelle est uniforme pour toute l'image et dans toutes les directions. La non uniformité d'échelle a un impact sur la localisation, l'échelle et la forme de la structure locale. Beaucoup de travaux s'enregistrent dans cette catégorie citons les travaux de Mikolajczyk et Schmid, dans [Mikolajczyk, 2001].

g) Méthodes bio-inspirées

Reisfeld et al. dans [Reisfeld, 1995] propose une méthode basée sur des expérimentations psychophysiques. Il calcule un score de symétrie de signal pour chaque pixel de l'image. Les pixels avec un score maximal correspondent à des points d'intérêt qui représentent des centres de symétrie. Dans [Zelinsky, 2003], la transformée généralisée de Reisfeld a été améliorée. Au lieu de calculer le score de symétrie pour un pixel central par la contribution des pixels dans un voisinage, ce score est calculé par la participation de tous les pixels dans un Rayon r de l'image. Heidemann dans [Heidemann, 2004] cherche des points d'intérêt qu'il appelle point de focus. Il propose l'extension à la couleur de la transformée généralisée de Reisfeld. Dans l'approche proposée par [Rebai, 2006a], la méthode de Loy et Zelinsky a été généralisée; ils ont gardé le principe de vote en faisant ce vote dans plusieurs directions et non pas la direction du gradient seulement. Cette amélioration permet de détecter les centres de symétrie des objets qui ne sont pas strictement symétriques et caractériser les points détectés par l'angle de rotation du vecteur gradient nommé θ [Rebai, 2007].

Dans les travaux d'Ebner et Trujillo et Ollague [Ebner, 1998], [Trujillo, 2006a] [Trujillo, 2006b], le problème de détection de points d'intérêt est formulé autrement. Ils considèrent que les détecteurs de points d'intérêt peuvent être formulés comme un problème d'optimisation et résolu en utilisant la programmation génétique. La fonction de fitness peut être la répétabilité du détecteur ou sa localisation. Ils présentent des détecteurs très intéressants comme les détecteurs IPGP1 et IPGP2 dont les taux de répétabilité sont plus importants que ceux obtenus par le détecteur de Harris. Dans des travaux récents de Trujillo et al. [Trujillo, 2008] la fonction de fitness est devenue multi-objective car les objectifs d'un détecteur de points d'intérêt sont souvent contradictoires et la programmation génétique peut aider dans ces cas. Les algorithmes génétiques sont utilisés par Kiryati et Gofman dans [Kiryati, 1996] pour détecter la symétrie réflexionnelle locale dans les images en niveaux de gris. L'intelligence en essaim a aussi prouvé ses succès dans la détection de points d'intérêt

par l'utilisation des colonies de fourmis [Lakehal, 2009], l'approche proposée détecte les centres des objets dans une image à travers un groupe de fourmis placé initialement sur les contours de ces objets.

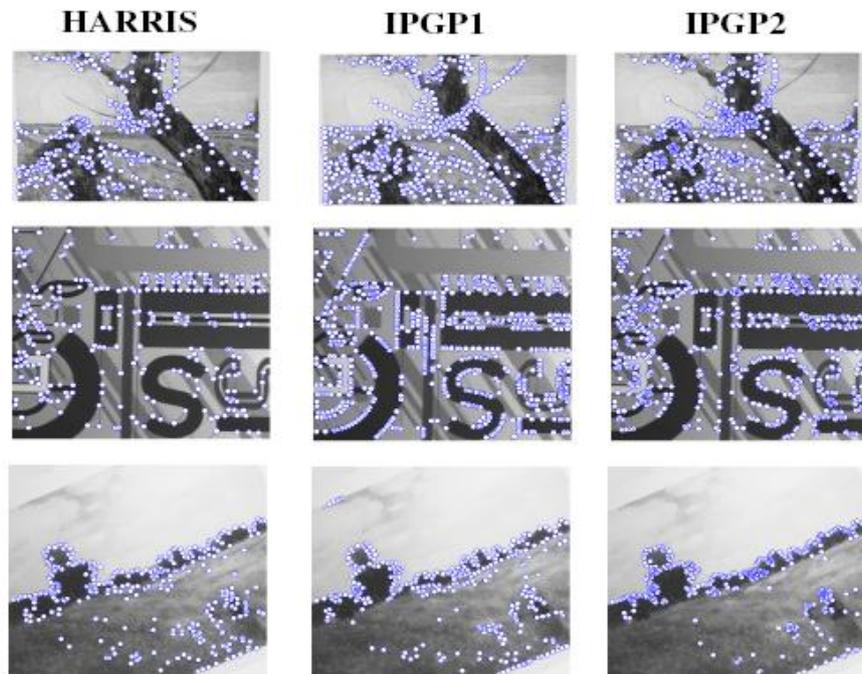


Figure II-8. Points d'intérêt extraits avec les détecteurs Harris, IPGP1, IPGP2. [Trujillo, 2006b]

h) Méthodes basées sur la couleur

Gouet, dans [Gouet, 2000], présente deux approches pour la segmentation couleur. Ce sont des généralisations au cas multi-spectral. Il est possible de définir de nombreuses méthodes pour combiner les gradients provenant des différents plans d'une image couleur. Par exemple la méthode de gradient multi-spectral de Di-Zenzo et la méthode de gradient maximal.

i. Extraction de points caractéristiques

1. Les coins

L'opérateur de Di-Zenzo généralise le gradient d'une image scalaire au cas des champs de vecteurs. Alors que dans le travail de Gouet [Gouet, 2000], on généralise la courbure des isophotes au cas multi-spectral {RVB}.

Soit une courbe tracée dans le plan image dont l'orientation de la tangente par rapport à l'axe des x effectue un angle θ_τ . La courbure multi-spectrale de cette courbe (noté CMS) s'exprime par :

$$CMS = \frac{d\theta_\tau}{d_s} \quad (17)$$

La détection des coins couleurs se fait, alors, selon le processus suivant :

- Extraction des contours couleurs à l'aide de l'opérateur de Di-Zenzo
- Calcul de la courbure couleur $K_{couleur}$
- Extraction des points ayant la plus forte courbure en maximisant $K_{couleur}$ dans la direction de $\vartheta : K_{max}$
- Recherche de points de contour parmi K_{max}
- Enfin les coins couleurs sont déterminés par seuillage parmi les maximas locaux des points trouvés.

2. Les Courbures maximales

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

$F(Kr, Kv, Kb) = \text{Max}\{Kr, Kv, Kb\}$ on peut choisir l'orientation des lignes d'iso-intensité comme l'orientation correspondante à la courbure maximale.

ii. Le détecteur de Harris Précis couleur

Il s'agit d'introduire l'information couleur {RVB} dans ce détecteur du premier ordre déjà existant et qui présente une excellente répétabilité et une forte stabilité. La matrice M de Harris utilise alors le tenseur multi-spectral et devient :

$$M = G(\tilde{\sigma}) \otimes \begin{bmatrix} R_x^2(\sigma) + V_x^2(\sigma) + B_x^2(\sigma) & R_x(\sigma)R_y(\sigma) + V_x(\sigma)V_y(\sigma) + B_x(\sigma)B_y(\sigma) \\ R_x(\sigma)R_y(\sigma) + V_x(\sigma)V_y(\sigma) + B_x(\sigma)B_y(\sigma) & R_y^2(\sigma) + V_y^2(\sigma) + B_y^2(\sigma) \end{bmatrix} \quad (18)$$

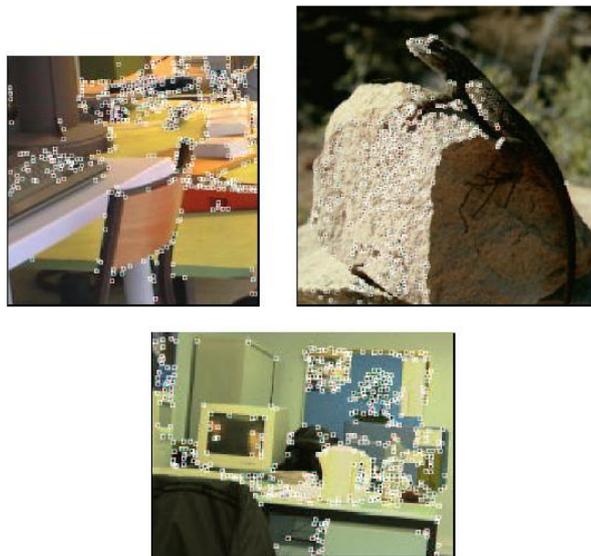


Figure II-9. Des points d'intérêt détectés par l'opérateur Harris couleur. [Gouet, 2000]

3.1.3. Méthodes d'évaluation des détecteurs des points d'intérêt

La détection des points d'intérêt est une étape primordiale dans les applications de la vision par ordinateur dont la fiabilité dépend grandement de la précision des points détectés. Cette dernière varie selon le contexte d'utilisation de ces points, toutefois la méthode d'évaluation la plus répandue est de vérifier visuellement la qualité des points extraits. Plusieurs travaux ont porté sur l'évaluation des détecteurs des points d'intérêt ; les critères d'évaluation les plus utilisés sont la répétabilité et la localisation [Schmid, 1996], [Gouet, 2000],... D'autres travaux exploitent les scores d'appariement [Mikolajczyk, 2005] et quelques travaux introduisent de nouveaux critères comme la discrimination, l'utilisation et la stabilité [Heidemann, 2004].

a) Critères d'évaluation

En général, les critères les plus utilisés sont la répétabilité et la localisation. La répétabilité signifie qu'un même point est détecté dans une suite d'images. Alors que la localisation signifie qu'un point détecté dans l'image correspond de façon précise à un point 3D de la scène.

Les critères d'évaluation diffèrent selon les applications des points détectés. Pour des applications comme le calcul de la géométrie épipolaire, l'appariement ou la reconnaissance d'objets, le seul critère important est la répétabilité. Par contre pour la reconstruction 3D ou le calibrage les deux critères sont importants.

i. Critère de répétabilité

Pour mesurer la répétabilité d'un détecteur, il faut établir une relation entre p_1 et p_2 appartenant respectivement à I_1 , I_2 . Dans le cas général, il n'existe pas de telle relation, cependant si I_1 et I_2 sont des images d'une même scène plane, cette relation est définie par une homographie. Il n'est pas possible d'avoir tous les points répétés, et un point n'est souvent pas répétable d'une façon exacte mais il est répétable dans un voisinage.

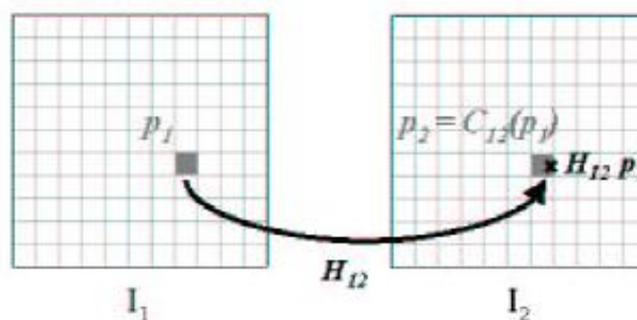


Figure II-10. Homographie entre Deux images planes.

Il faut tenir compte uniquement de la partie commune effective des images, c'est la partie de la scène vue dans les deux images. Soient $\{d\}_1$ et $\{d\}_2$ les sous ensembles des points détectés correspondant à la partie commune.

$D = \{(d_2, d_1) / d_2 = H_{12} d_1\}$ est l'ensemble des points détectés à la même position dans l'image I_2 , sachant que le support de l'image est discret, cette intersection n'existe pas dans la réalité, alors il n'y aura plus un ensemble D mais pratiquement un ensemble $D(\varepsilon)$ déterminé par un seuil de distance ε (' ε répétabilité') tel que :

$$D(\varepsilon) = \{(d_2, d_1) / \text{dist}(d_2, H_{12}d_1) < \varepsilon\} \quad (19)$$

Par conséquent, le taux de répétabilité $r_2(\varepsilon)$ pour l'image I_2 est défini par :

$$r_2(\varepsilon) = \frac{|D(\varepsilon)|}{\min(n_2, n_1)} \quad (20)$$

Tel que :

$$\begin{aligned} n_2 &= |\{d\}_2| \\ n_1 &= |\{d\}_1| \end{aligned}$$

ii. Critère de localisation

Cette mesure permet de vérifier si les points détectés par un détecteur ont une réalité physique dans la scène. Les applications liées à la reconstruction tridimensionnelle donnent une très grande importance à cette mesure. Mais tout dépend de l'application envisagée, il est possible de choisir l'un des deux critères de cette mesure de localisation qui sont : le critère de reconstruction projective qui met en jeu une distance euclidienne 2D et celui de reconstruction euclidienne qui se base sur une distance euclidienne 3D.

1. Localisation projective

Soit P l'ensemble des points de la scène 3D dont la position exacte est connue. Soit P_{proj} l'ensemble des projections de P sur le plan rétinien d'une caméra et P_{det} l'ensemble des points détectés sur l'image de la scène vue par la caméra. Il suffit de chercher pour chaque point détecté P_{det}^i le point projeté P_{proj}^j le plus proche, puis de considérer la distance euclidienne qui les sépare. Alors, il devient possible de définir un premier critère de localisation $L_{projectif}$ dont la formule est la suivante :

$$L_{projectif} = moy_i \left\{ \min_j \left\{ \text{dist}(P_{det}^i, P_{proj}^j) \right\} \right\} \quad (21)$$

Il est à remarquer que plus $L_{projectif}$ est petite plus la calibration obtenue sera précise.

2. Localisation euclidienne

Le critère de localisation projective s'avère insuffisant lorsque nous souhaitons procéder à la reconstruction tridimensionnelle de la scène. Supposons que nous avons à notre disposition au moins deux vues de la scène, la calibration des caméras est une méthode de mise en correspondance des points. Donc pour définir une mesure de localisation spécifique à la reconstruction 3D, il suffit de mettre en correspondance les

points détectés, de les reconstruire par triangulation et enfin de les comparer aux points P^j de la scène. La mesure de localisation euclidienne est alors donnée par :

$$L_{euclidien} = moy_i \left\{ \min_j \left\{ dist(p_{reconstruits}^i, P^j) \right\} \right\} \quad (22)$$

iii. Discrimination

C'est un critère défini dans [Schmid, 2000]. La discrimination détermine le contenu informatif du descripteur en mesurant la discrimination du point d'intérêt. La discrimination est basée sur la probabilité de présence d'un descripteur local en niveau de gris calculé à un point d'intérêt dans une population de tous les descripteurs locaux des points d'intérêts observés. Pour un ensemble d'images, il faut calculer les descripteurs locaux pour chacun des points d'intérêt détectés. La discrimination mesure alors la distribution de ces descripteurs. Si les descripteurs sont très proches c'est qu'ils n'ont pas importé d'information et la discrimination devient minimale. Mais, si les descripteurs sont éparpillés, le taux de discrimination est significatif et il y aura une grande chance d'avoir une mise en correspondance correcte.

La discrimination est mesurée en utilisant l'entropie, les descripteurs les plus éparpillés ont une forte entropie. L'entropie mesure la randomisation d'une variable, plus une variable est aléatoire plus l'entropie est grande. Dans ce cas il n'y a pas de variables continues, l'utilisation des partitions est indispensable.

L'entropie d'une partition $A = \{A_i\}$ est :

$$H(A) = -\sum_i p_i \log(p_i) \quad (23)$$

Où : p_i est la probabilité de A_i

L'entropie mesure le contenu informatif, dans la théorie de l'information le contenu C d'un message i est donné par :

$$C_i = \log\left(\frac{1}{p_i}\right) = -\log(p_i) \quad (24)$$

La moyenne des contenus informatifs par message pour un ensemble de messages et son entropie est donnée par:

$$-\sum_i p_i \log(p_i) \quad (25)$$

Dans le cas des points d'intérêt ; le but est de savoir quel contenu informatif apporte ce point et par conséquent quelle discrimination apporte un point d'intérêt moyennant son descripteur. Un descripteur avec un contenu informatif important est un descripteur discriminant.

3.2. Caractérisation des points d'intérêt

Nous cherchons d'une part, une caractérisation qui décrit la fonction du signal de la manière la plus précise et la plus complète possible, et d'autre part une description qui soit la plus invariante aux transformations usuelles de l'image comme le changement

d'illumination, la rotation, le changement des points de vue, le changement d'échelle et le bruit. Nous allons présenter les différentes approches de caractérisation dans la littérature aussi bien pour les images en niveaux de gris que pour les images en couleur.

3.2.1. Caractérisation en niveaux de gris

a) Dérivées

Une fonction peut être approximée localement par ses dérivées. Alors, il est possible de décrire une image (fonction) en un point en stockant dans un vecteur l'ensemble des dérivées en ce point.

Koendrink a utilisé ce vecteur, il l'a nommé jet local. Il peut être calculé de manière stable en utilisant un filtre passe bas : la gaussienne et ses dérivées.

b) Descriptions fréquentielles

Une fonction peut être caractérisée par une description fréquentielle. La transformée de Fourier, transformée de Mellin...etc. sont des exemples de descriptions fréquentielles globales.

L'intérêt des méthodes de descriptions fréquentielles est que la phase est normalisée ; c'est-à-dire indépendante de la luminosité et du contraste des images. Un autre point fort est que cette phase est stable à un changement d'échelle jusqu'à 20% [Schmid, 1996]. Alors son utilisation peut être mise en œuvre dans un contexte multi-échelle. Par ailleurs, cette phase est une variable continue ce qui permet d'obtenir des résultats sous-pixelliques.

i. La transformée de Fourier

La transformée de Fourier d'une paire de fréquences (u, v) est caractérisée par une amplitude et une phase. L'intérêt de son utilisation réside dans le fait qu'au moins l'un de ces deux attributs reste constant à travers certaines transformations de l'image. Par exemple, sa phase est indépendante de la luminosité des images et son amplitude ne varie pas avec la translation. D'autre part la phase est stable à un changement d'échelle jusqu'à 20% et peut, donc, être mise en œuvre dans un contexte multi-échelle. Le principal avantage de la caractérisation basée sur la transformée de Fourier est qu'elle peut être calculée très efficacement.

ii. La transformée de Mellin

La Transformée de Mellin, une autre transformée, s'applique sur des nombres complexes dans sa forme la plus générale. Si elle est utilisée avec des valeurs imaginaires pures, son amplitude est invariante au changement d'échelle. Il existe une variante appelée Transformée de Fourier-Mellin, qui permet d'obtenir en plus l'invariance à la rotation.

Cependant, ces transformées sont globales ; c'est-à-dire permettent une localisation en fréquence et non pas en espace. Alors elles ne permettent pas de dire quelles fréquences

appartiennent à quel point. Ce problème est connu sous le nom de principe d'incertitude. Pour le résoudre, il est préférable d'utiliser un fenêtrage.

iii. La transformée de Gabor

Gabor a proposé l'utilisation d'un fenêtrage Gaussien et a démontré qu'un tel fenêtrage est optimal pour obtenir une bonne localisation à la fois en fréquence et en espace.

Dans le cadre du calcul d'appariements, une égalité de phase entre deux points de deux images différentes signifie une grande probabilité qu'il s'agit de points à appairer. Pour qu'une égalité de phase corresponde à un appariement de façon certaine, une approche multi-résolution est nécessaire.

iv. Les ondelettes

La théorie des ondelettes consiste à remplacer le paramètre de fréquence de la transformée de Gabor par un paramètre d'échelle. L'inconvénient de la transformée en ondelettes est que sa résolution fréquentielle est d'autant plus mauvaise que l'échelle est petite. Une solution proposée donne naissance aux méthodes qui analysent les signaux suivant trois paramètres physiques : le temps, la fréquence et l'échelle.

c) Les moments

Les moments permettent aussi de caractériser un signal. D'un point de vue stochastique une image peut être vue comme une distribution aléatoire et peut donc être caractérisée de façon unique par ses moments.

Etant donné un signal bidimensionnel $I(x, y)$, le moment d'ordre $(p+q)$ est défini dans le cas discret de la manière suivante :

$$m_{pq} = \sum_x \sum_y x^p y^q I(x, y) \quad (26)$$

$$p, q = 0, 1, 2, \dots$$

Hu dans [Hu, 1961] a utilisé pour la première fois les moments pour la reconnaissance en vision par ordinateur. Il a démontré l'existence de combinaisons de moments qui sont invariantes en translation, rotation et changement d'échelle. Dans des travaux récents, Van Gool a présenté un ensemble de moments jusqu'au deuxième ordre qui sont invariants aux changements affines et aux changements d'intensité. Il est à mentionner que l'inconvénient majeur de ces approches est qu'il est nécessaire, le plus souvent, de calculer les moments à un ordre relativement grand si le souhait est d'obtenir une caractérisation suffisamment discriminante. Il faut cependant noter que les moments restent assez sensibles aux occultations.

d) Autres caractérisations

Il existe d'autres méthodes de caractérisation du signal, par exemple : Bigum propose l'utilisation d'un système non linéaire de coordonnées. Weiss propose aussi une méthode pour calculer localement des invariants affines et projectifs. D'autres caractérisations sont possibles, citons une méthode simple, mais répandue où un point est décrit par les valeurs des pixels voisins.

Les limites des méthodes de caractérisation poussent les chercheurs à exploiter un nouveau axe de recherche, celui des invariants.

e) Les invariants

Pour rendre la caractérisation locale invariante aux différentes transformations de l'image il faut utiliser les invariants.

i. Définition théorique d'un invariant

Un invariant est une propriété qui est constante pour un ensemble de fonctions. Trois types de transformations nous intéressent dans le cadre de la vision par ordinateur : les transformations de la scène tridimensionnelle vers l'image, les transformations de l'image et les transformations qui opèrent sur le signal de l'image (les changements de luminosité). Il faut donc calculer des invariants pour ces transformations.

ii. Calcul des invariants

Il existe deux types de méthodes pour calculer des invariants : les méthodes infinitésimales utilisent les groupes de Lie [Schmid, 1996] et les méthodes par généralisation et contraintes. Le principe de ces dernières méthodes est la recherche des invariants associés à un problème généralisé du problème donné, ensuite le calcul des invariants pour le problème généralisé et l'expression du fait que le problème de départ est un cas particulier. Le problème dans ces méthodes est de trouver une généralisation du problème de départ pour laquelle le calcul des invariants est connu.

f) Les Quasi-invariants

Binford démontre que les invariants aux groupes des similitudes images sont des quasi-invariants pour une transformation perspective. Notons que le dénombrement des quasi-invariants pour un problème donné est difficile, de même, la vérification qu'une fonction est un quasi-invariant est souvent complexe. Par contre, les quasi-invariants se révèlent souvent plus stables que les invariants et ils apportent une solution dans le cas où il n'existe pas d'invariants.

Théorème de Burns

Le théorème de Burns démontre qu'il n'y a pas d'invariants pour les transformations de la scène tridimensionnelle vers l'image. Mais il est également possible d'approximer les transformations perspectives par des quasi-invariants.

3.2.2. Invariance et transformations de l'image

a) Rotation image

Afin d'obtenir une invariance à la rotation, deux approches différentes sont possibles : la première est d'utiliser des mesures différentielles invariantes au groupe des déplacements. Cette caractérisation repose sur des combinaisons de dérivées invariantes à une rotation image. La deuxième propose l'utilisation du principe des filtres ajustables. Dans ce cas, des dérivées sont ajustées dans la direction du gradient et sont alors invariantes à une rotation d'image.



Figure II-11. Exemple d'images transformées par une rotation par 35° à gauche et par 325° à droite.
[Nene, 1996]

i. Les invariants différentiels

1. Le jet local

Une fonction peut être approximée localement par ses dérivées. Il est donc possible de décrire une image en un point par l'ensemble des dérivées en ce point sous forme de vecteur. Cet ensemble caractérise la géométrie locale du voisinage du point. Koendrink a nommé ce vecteur le jet local. Le principal inconvénient de cette caractérisation est que le vecteur n'est pas invariant aux différentes transformations de l'image.

Koendrink et Romeny proposent de calculer des invariants pour les groupes de déplacements $SO(2)$ à partir du jet local défini précédemment. Ils soulignent la nécessité d'implémenter le calcul des dérivées de manière stable afin de calculer ces invariants à un ordre élevé. Schmid dans [Schmid, 1996] a utilisé l'ensemble d'invariants différentiels jusqu'au troisième ordre. Ces invariants sont regroupés dans un vecteur noté \vec{V} . Ce vecteur est divisé en deux parties :

La première partie est constituée d'un ensemble complet et irréductible d'invariants différentiels jusqu'au deuxième ordre.

$$\vec{v}[0..4] = \begin{bmatrix} l \\ l_i l_i \\ l_i l_{ij} l_j \\ l_{ii} \\ l_{ij} l_{ji} \end{bmatrix} \quad (27)$$

La deuxième partie est constituée d'un ensemble complet d'invariants du troisième ordre.

$$\vec{v}[5..8] = \begin{bmatrix} \varepsilon_{ij} (l_{jkl} l_i l_k l_l - l_{jkk} l_i l_l l_l) \\ l_{ijl} l_j l_k - l_{ijk} l_i l_j l_k \\ - \varepsilon_{ij} l_{jkl} l_i l_k l_l \\ l_{ijk} l_i l_j l_k \end{bmatrix} \quad (28)$$

2. Le jet ajustable

À partir du jet local déjà défini, il est possible de calculer les dérivées dans une direction donnée, par exemple la direction du gradient afin d'avoir des invariants à la rotation. La formule (29) donne la dérivée à l'ordre 1 dans une direction θ , sachant que L_x et L_y sont les éléments du jet local :

$$L'(\theta) = L_x \cos(\theta) + L_y \sin(\theta) \quad (29)$$

3. Les filtres directionnels

Freeman et Adelson ont montré qu'il est possible de calculer ce qu'ils appellent des filtres directionnels. C'est-à-dire, des filtres définis par des dérivées calculées dans n'importe quelle direction. Par exemple la direction du gradient puisqu'elle représente l'orientation de l'image. Cette méthode présente l'inconvénient que le résultat dépend fortement du calcul de l'angle θ , qui représente une importante source d'instabilité.

ii. Les invariants de Hilbert

Hilbert a montré que n'importe quel invariant au groupe de déplacement $SO(2)$ d'ordre fini pouvait s'exprimer comme un polynôme d'invariants irréductibles reposant sur la combinaison de dérivées de jet local.

iii. Les invariants non paramétriques

C'est une méthode récente développée par Lan et Mohr dans [Lan, 1997]. Elle est basée sur des invariants à la rotation non paramétriques. Soient deux images I_1 et I_2 différentes d'une rotation d'angle C , et deux pixels (r_1, θ_1) et (r_2, θ_2) exprimés en

coordonnées polaires. L'idée part du principe que l'ordre entre deux pixels reste inchangé sous une rotation de l'image. Ainsi, si $I_1(r_1, \theta_1) > I_1(r_2, \theta_2)$ alors $I_2(r_1, \theta_1+C) > I_2(r_2, \theta_2+C)$ ce qui permet de définir pour une image I la fonction de comparaison suivante :

$$C_I(r_1, \theta, r_2, \theta + \Delta_1) = \begin{cases} 1 & \text{Si } I(r_1, \theta) > I(r_2, \theta + \Delta_1) \\ 0 & \text{Sinon} \end{cases} \quad (30)$$

Ils obtiennent l'invariant N_I défini ci- après, pour n'importe quel autre couple de points $(r_3, \theta+\Delta_2)$ et $(r_4, \theta+\Delta_3)$:

$$N_I(r_1, r_2, r_3, r_4, \Delta_1, \Delta_2, \Delta_3) = \frac{1}{2\pi} \int_0^{2\pi} |C_I(r_1, \theta, r_2, \theta + \Delta_1) - C_I(r_3, \theta + \Delta_2, r_4, \theta + \Delta_3)| d\theta \quad (31)$$

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

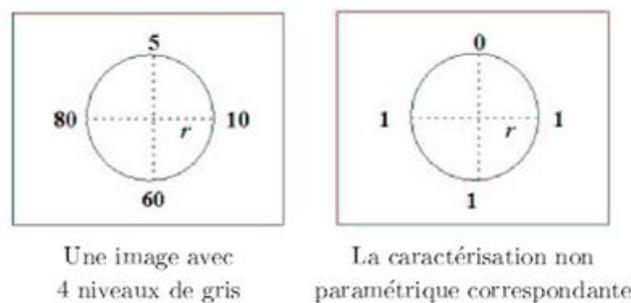


Figure II-12. Exemple de caractérisation non paramétrique

Il reste à noter qu'il faut effectuer une normalisation en taille de l'image où la forme des pixels doit être carrée pour obtenir des dérivées invariantes en rotation car la rectangularité des pixels introduit une anisotropie qui fausse le calcul des dérivées. Pour rendre les pixels carrés il faut normaliser l'image.

b) Changement d'échelle

Le changement d'échelle a deux raisons: il est dû soit à un changement de la distance entre la caméra et l'objet, soit à un changement de la distance focale de l'objectif.

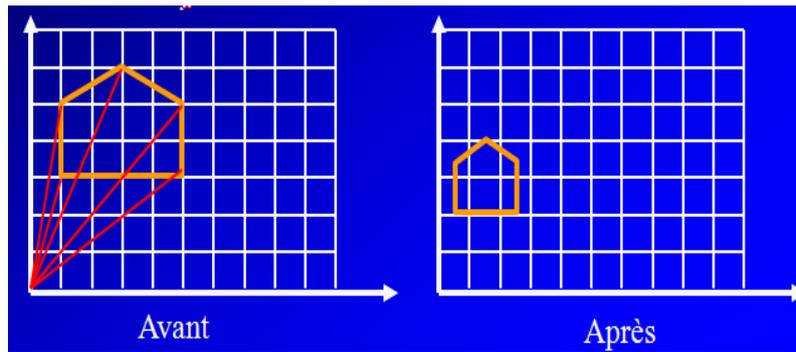


Figure II-13. Représentation d'un changement d'échelle.

i. Invariants à un changement d'échelle

Comme l'on a déjà évoqué, le changement d'échelle pour une fonction f , peut être décrit par un changement de variable : $f(x) = g(u)$ où $g(u) = g(u(x)) = g(\alpha x)$. D'où découlent les relations suivantes entre f et g :

$$f^{(n)}(x) = \alpha^n g^{(n)}(u) \quad (32)$$

Où $f^{(n)}(x)$ représente la dérivée nième de f .

L'équation (32) montre que les dérivées nièmes de f et de g présentent un facteur multiplicatif $\alpha^{(n)}$. Donc, à partir du quotient de deux dérivées, il est possible d'éliminer le facteur $\alpha^{(n)}$. Des invariants théoriques à un changement d'échelle sont donnés par l'équation suivante :

$$\frac{\left[f^{(n)}(x) \right]^{\frac{k}{n}}}{f^{(k)}(x)} \quad (33)$$

Cependant des résultats expérimentaux ont montré que de tels invariants sont peu stables à un changement d'échelle supérieur à 20%.

ii. Approche multi-échelle

Cette approche est utilisée dans [Schmid, 1997] et repose sur la notion d'espace d'échelle où un paramètre connu définit l'échelle.

L'espace d'échelle ne permet pas le calcul des invariants à toutes les échelles, ce que nécessite la discrétisation de l'espace d'échelle. Et comme la caractérisation obtenue auparavant est robuste à un changement d'échelle jusqu'à 20 %, cela exige que le choix de pas de discrétisation doit garantir que le changement entre deux échelles consécutives est inférieur à 20 %. Les calculs sont effectués pour les différentes échelles, c'est à dire pour différentes tailles σ de la gaussienne. Ce qui permet de réaliser une approche multi-échelle.

c) Changement de luminosité

La caractérisation doit également être robuste à un changement de luminosité. Il existe plusieurs possibilités pour modéliser un changement de luminosité : trois modèles de transformations de niveaux de gris sont possibles, pour chacun de ces modèles il faut définir les invariants correspondants.



Figure II-14. Exemple d'images soumises à des changements de luminosité. [Geusebroek, 2005]

i. Translation des niveaux de gris

Une translation des niveaux de gris se modélise par :

$$\tilde{I}(x, y) = I(x, y) + b \quad (34)$$

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

ii. Transformations affines de niveaux de gris

Une transformation affine des niveaux de gris est formulée par :

$$\tilde{I}(x, y) = aI(x, y) + b \quad (35)$$

Une telle transformation modifie les dérivées du signal de la manière suivante :

$$\tilde{I}^{(n)}(x, y) = a I^{(n)}(x, y) \quad (36)$$

N'importe quel quotient de deux dérivées est donc invariant à une transformation affine de la luminance. Florak a montré que jusqu'à l'ordre deux, il existe deux invariants : la courbure des isophotes k et la courbure des lignes de plus grande pente μ . Il y a différentes manières de rendre le vecteur \vec{V} invariant à une transformation affine. Dans [Schmid, 1996], il est divisé par la puissance adéquate de la magnitude du gradient. Le résultat est le vecteur d'invariant affine \vec{V}_A :

$$\vec{V}_A[0..2] = \begin{bmatrix} \frac{L_i L_{ij} L_j}{(L_i L_i)^{3/2}} \\ L_{ii} \\ \frac{L_{ij} L_{ji}}{L_i L_i} \end{bmatrix} \quad (37)$$

$$\text{Et } \vec{V}_A[3..6] = \frac{1}{(L_i L_i)^2} \begin{bmatrix} \varepsilon_{ij} (L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{iij} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ - \varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix} \quad (38)$$

iii. Transformation monotone de la luminosité

Un changement de luminosité peut également être modélisé par une fonction monotone et donc irréversible. L'irréversibilité de la fonction est nécessaire pour éviter une perte d'information par rapport à celle contenue dans l'image en niveaux de gris.

d) Changement des points de vue

Le jet local ajustable et le vecteur d'invariants sont invariants à la rotation d'images et au changement d'échelle s'ils sont calculés dans un cadre multi-échelle. Donc, ils sont invariants au groupe de similitudes de l'image, alors quasi-invariants aux transformations perspectives. Binford a démontré que de tels invariants sont des quasi-invariants à une transformation perspective, c'est à dire localement invariants à une transformation perspective.



Figure II-15. Exemple d'images prises sous différents points de vue.

e) Autres transformations image

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

3.2.3. Caractérisation en couleur

Le signal couleur peut être décomposé de diverses manières en trois composantes. Il est alors possible de considérer l'image en couleur comme la superposition d'images en niveaux de gris. Par conséquent, il est possible d'exploiter pour chacune d'elles les approches existantes en niveaux de gris.

A partir des années 90, de nouvelles méthodes dédiées à la couleur ont vu le jour. Vu le contexte dans lequel sont réalisées ces méthodes - l'indexation- une des préoccupations principales est l'obtention de l'invariance face aux changements d'illumination, nommée problème de la constance de couleurs. C'est un problème qui nécessite en premier lieu la disposition d'un modèle d'illumination.

a) Les modèles d'illumination

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

- Un changement d'intensité d'une ou de plusieurs teintes de la source émettrice, c'est le changement interne de la source lumineuse.
- Un déplacement de la source émettrice, c'est le cas d'un changement externe d'illumination.

i. Changement interne

Il existe deux types de modèles pour modéliser le changement interne : le modèle linéaire et le modèle diagonal. Le modèle linéaire est le plus complet des modèles d'illumination. Par contre, le modèle diagonal s'avère plus simple et semble suffisant dans sa version généralisée.

ii. Changement externe

Dans ce cas, Finlayson suppose que chaque pixel est multiplié par un facteur. Par exemple, un pixel (r, v, b) est transformé en (r', v', b') tel que : $(r', v', b') = \delta(r, v, b)$

δ dépend de la position et de l'orientation de la source par rapport à deux positions successives de cette source et à celle de la caméra, et varie donc de pixel en pixel. Avec une image contenant n pixel on obtient le modèle d'illumination suivant :

$$\begin{pmatrix} r'_1 & v'_1 & b'_1 \\ r'_2 & v'_2 & b'_2 \\ \vdots & \vdots & \vdots \\ r'_n & v'_n & b'_n \end{pmatrix} = \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_1 \end{pmatrix} \begin{pmatrix} r_1 & v_1 & b_1 \\ r_2 & v_2 & b_2 \\ \vdots & \vdots & \vdots \\ r_n & v_n & b_n \end{pmatrix} \quad (39)$$

iii. Normalisation de l'image

La normalisation d'une image est une solution pour rendre celle-ci indépendante d'un modèle d'illumination. Le choix de la méthode de normalisation est, bien sûr, fonction du modèle d'illumination considéré. Alors, le problème majeur consiste à réaliser une normalisation conjointe pour les deux modèles (interne et externe) à la fois.

iv. Méthode itérative de Finlayson

Finlayson a mis en place une méthode itérative pour normaliser l'image face à des changements d'illumination internes et externes. Chaque itération comporte deux étapes. Dans la première, l'image est normalisée pour les changements externes. Dans la deuxième, le résultat de la première étape est normalisé pour les changements internes. Le processus est assez rapide et converge en quelques itérations.

b) Les approches de caractérisation locale

Très peu de travaux ont été réalisés sur la caractérisation locale des points d'intérêt à partir de l'information couleur. Il existe deux approches qui utilisent des histogrammes locaux construits sur plusieurs zones de l'image. Ces approches ne sont pas satisfaisantes car elles utilisent des méthodes statistiques qui demeurent incertaines pour des petites populations.

En fait, deux autres directions semblent être intéressantes, la première exploite le paradigme du hachage géométrique dans le cadre de l'invariance aux changements d'illumination. Pour la deuxième direction, les travaux de Gouet [Gouet, 2000], [Gouet, 2001] proposent une approche qui étend à la couleur la technique basée sur les invariants différentiels de Hilbert.

i. Hachage couleur

C'est une méthode d'indexation qui utilise le paradigme de hachage géométrique de Wolfson. Elle consiste à stocker dans une table des informations géométriques indépendantes de la position, de l'orientation et de l'échelle des objets. Son inconvénient réside dans le coût important qu'exige la construction de la table.

ii. Les invariants différentiels couleurs

1. Deux invariants spécifiques à la couleur

Considérons une image $\{R, V, B\}$ ainsi que le groupe des rotations spécifié par un paramètre (l'angle de rotation). Il y a par plan couleur 5 dérivées jusqu'au deuxième ordre, sans compter l'intensité lumineuse. Donc le nombre des invariants est $(5 + 1) * 3 - 1 = 17$, et nous avons comptabilisé 5 invariants différentiels en niveaux de gris alors $5 * 3 = 15$ invariants pour la couleur il y a donc deux $(17 - 15)$ invariants supplémentaires inter-canaux pris dans l'ensemble suivant :

$$\nabla R \cdot \nabla V \quad \nabla R \cdot \nabla B \quad \nabla V \cdot \nabla B \quad (40)$$

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

2. Une caractérisation couleur robuste

Elle consiste à n'utiliser que les invariants de Hilbert à l'ordre 1. La caractérisation alors obtenue pour des images en niveaux de gris serait inutilisable puisqu'elle est réduite à deux invariants. Cependant, l'approche couleur permet d'obtenir un ensemble de $2 * 3 + 2 = 8$ invariants appelés \vec{V}_{col} et tel que :

$$\vec{V}_{col}(\bar{x}, \sigma) = \begin{pmatrix} R \\ \|\nabla R\|^2 \\ V \\ \|\nabla V\|^2 \\ B \\ \|\nabla B\|^2 \\ \nabla R \cdot \nabla V \\ \nabla R \cdot \nabla B \end{pmatrix} \quad (41)$$

Les avantages de l'utilisation de l'information couleur sont :

La caractérisation obtenue est robuste au bruit car elle n'utilise que les dérivées d'ordre 1 et c'est une méthode de faible complexité.

3. Changement d'illumination

Il faut rendre cette caractérisation invariante aux changements d'illumination (interne, externe et complexe).

➤ Changement interne

Ce changement est modélisé par le modèle diagonal enrichi d'une translation. Pour rendre la caractérisation invariante, Gouet dans [Gouet, 2000] propose la normalisation de l'image. Cette solution a l'avantage de prendre en compte les 6 paramètres de modèle sans appauvrir la caractérisation basée sur les invariants différentiels du premier ordre.



Figure II-16. Exemples de changements internes de luminosité [Gouet, 2000].

➤ Changement externe

Le modèle utilisé pour approximer les changements d'illumination externes est un modèle à n paramètres si l'image contient n pixels. Il est donc, bien sûr, impossible de rendre

le vecteur des huit composantes indépendant de ce modèle. La solution consiste alors à normaliser l'image.



Figure II-17. Exemples de changements externes de luminosité [Gouet, 2000].

➤ Changement complexe

Un changement complexe d'illumination est la combinaison des modèles internes et externes. Vu les résultats des expériences faites dans [Gouet, 2000], la décision était de ne pas normaliser l'image face aux changements d'illumination externes, afin de conserver la pertinence de chacun des invariants.



Figure II-18. Exemples de changements complexes de luminosité [Gouet, 2000].

En résumé, la normalisation d'une image n'est pas une opération simple à réaliser. En effet, l'estimation d'un modèle d'illumination réaliste est délicate.

3.2.4. Évaluation de la caractérisation

Le critère d'évaluation est le taux d'appariements corrects. En effet, un taux élevé signifie que la caractérisation choisie est discriminante. Ce taux est calculé séparément pour les différentes transformations de l'image. Le taux est défini comme suit :

$$\frac{\text{Nombre d'appariements corrects}}{\text{Nombre d'appariements obtenus}}$$

3.3. Appariement d'images

Dans la suite on présente un état de l'art des méthodes d'appariement d'images qui existe dans la littérature.

3.3.1. Appariement basé sur les données photométriques

La méthode la plus ancienne de mise en correspondance est la corrélation du signal qui peut être vue comme la caractérisation élémentaire du signal autour d'un point : tout point est caractérisé par l'ensemble des valeurs de niveaux de gris sur un voisinage. Pour que cette méthode réussisse, les images doivent avoir été prises dans des conditions voisines pour que les fenêtres de corrélation se superposent correctement par simple translation.

La méthode échoue lorsqu'une transformation (par exemple rotation) intervient. Cependant il est possible qu'elle devienne robuste en rotation en utilisant des mesures de corrélation dans plusieurs directions ce qui est difficile et coûteux. Afin de diminuer les erreurs d'appariement et le coût de corrélation, certaines méthodes utilisent des points particuliers, par exemple les points d'intérêt. Pour améliorer les résultats obtenus, pour un couple de points appariés, les points dans des voisinages respectifs de même taille doivent également se correspondre. De plus, ces points de voisinage doivent avoir la même position relative dans les deux images, ce que nécessite l'introduction d'une mesure de distance.

a) Algorithme d'appariement

L'algorithme d'appariement est basé sur la caractérisation locale du signal présenté auparavant. Cette caractérisation doit être invariante aux transformations d'images ce qui lui permet d'être, à la fois, discriminante et invariante aux transformations d'images. Pour appairer les vecteurs d'invariants il faut introduire une distance qui permet de procéder à une telle comparaison. Il est ensuite possible de choisir dans deux images les vecteurs correspondants au même point physique.

b) Distance entre images

Pour appairer deux points, il faut mettre en correspondance leurs vecteurs d'invariants \vec{V} et décider si les deux vecteurs sont similaires ou non. L'utilisation de la distance euclidienne est peu judicieuse car les différentes composantes d'un vecteur \vec{V} ne peuvent être corrélées et qu'elles n'ont pas le même ordre de grandeur. Il est proposé donc de modéliser les composantes d'un vecteur \vec{V} par des variables aléatoires gaussiennes et d'utiliser la distance de Mahalanobis pour comparer les deux vecteurs d'invariants. Cette distance statistique utilise la matrice de covariance des composantes Λ et prend ainsi en compte la différence de magnitude des composantes ainsi que leur corrélation éventuelle :

$$d_M (\vec{b} - \vec{a})^T \Lambda^{-1} (\vec{b} - \vec{a}) \quad (42)$$

En utilisant cette distance, il est possible de décider si deux vecteurs sont similaires ou non. Les appariements qui correspondent aux plus grandes valeurs de la distance ont la plus grande probabilité d'être faux et inversement.

c) Le calcul des scores d'appariement

La première étape de la mise en correspondance de deux ensembles de points des deux images est le calcul d'un score d'appariement pour chaque paire de points. À partir de ces scores nous pouvons connaître les couples qui ont de bonnes chances de constituer un appariement correct.

Ce score est calculé en comparant la caractérisation associée à chaque point.

Alors selon les méthodes de caractérisation il y aura différents types de mesures de similarité.

d) Appariement basé sur une partie du signal

Plusieurs techniques réalisent la comparaison d'images directement à partir du signal, sans passer par une caractérisation spécifique.

i. La corrélation du signal

La corrélation du signal est la méthode de mise en correspondance la plus ancienne, dont tout point est caractérisé par l'ensemble des valeurs du signal sur son voisinage. On calcule un score de corrélation entre les fenêtres de corrélation entourant les deux points à traiter. L'inconvénient majeur de la corrélation du signal c'est qu'elle est très sensible aux transformations géométriques de l'image et qu'elle gère mal les occultations.

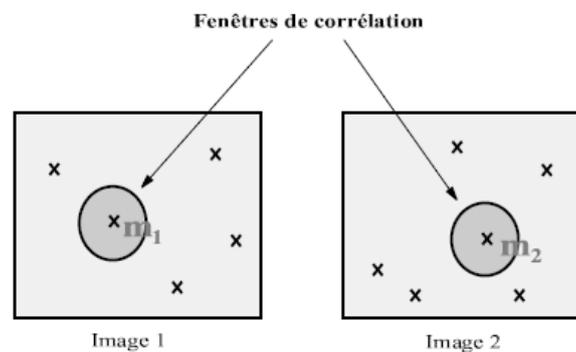


Figure II-19. La corrélation de signal entre deux points.

ii. La corrélation de phase

Comme il a été vu précédemment le signal peut être caractérisé dans le domaine spectral par la Transformée de Fourier. Alors il est possible de comparer deux images à partir de cette transformée. La corrélation de phase est robuste au bruit et aux variations d'illumination non uniformes. Ceci la rend plus appropriée que les méthodes de corrélation classiques. Elle est parfaitement applicable sur des images ayant subi un lissage. La corrélation de phase classique traite l'appariement d'images différemment par rapport à une simple translation. La méthode a été étendue récemment au cas de la rotation et du changement d'échelle. En plus, c'est une méthode qui permet de comparer deux images à travers ces transformations tout en déterminant les paramètres de translation, de rotation et les facteurs d'échelle impliqués. Cependant, cette méthode est beaucoup plus complexe que la version classique, car elle nécessite le calcul de trois Transformées de Fourier et de trois Transformées de Fourier inverses.

iii. La distance de Hausdorff

La distance de Hausdorff est une mesure de similarité permettant de comparer deux ensembles finis de points. Sa formulation est :

$$H(A,B) = \max(h(A, B), h(B, A)) \quad (43)$$

Avec :

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (44)$$

Cependant, elle ne peut pas être utilisée telle quelle est car trop fragile. Elle devient par exemple très grande lorsqu'un seul point de A est très éloigné de tous les points de B et donc n'est pas acceptable pour comparer deux images la plupart du temps bruitées. Une solution consiste à remplacer la première équation par :

$$h^f(A, B) = f_{a \in A}^{eme} \min_{b \in B} \|a - b\| \quad (45)$$

Ses principaux avantages sont d'être une robustesse au bruit et aux occultations.

e) Comparaison d'histogrammes

Il n'est pas facile de mettre en place une distance entre les histogrammes parce que l'algorithme de mise en correspondance doit être capable de gérer efficacement la masse d'information souvent imposante que contiennent les histogrammes.

i. Intersection d'histogrammes

Dans cette approche, l'histogramme du modèle est comparé aux histogrammes des images de la base selon la formule ci-après où n est le nombre de couleurs de l'espace discret des couleurs.

$$Inter(H, M) = \frac{\sum_{i=1}^n \min(h_{c_i}, m_{c_i})}{\sum_{i=1}^n m_{c_i}} \quad (46)$$

La valeur obtenue est entre 0 et 1. Le 1 signifie une bonne ressemblance des images en comparaison. Les avantages de cette méthode est qu'elle néglige explicitement les pixels du fond qui peuvent apparaître dans l'histogramme du modèle mais pas dans l'histogramme de la base. D'autre part, elle n'implique pas un lourd niveau de calcul car seulement deux opérations (minimum, addition) sont nécessaires par cellule d'histogramme.

L'inconvénient c'est que toutes les cellules sont traitées de manière égale, impliquant l'équiprobabilité des couleurs.

ii. Mesures de distance entre histogrammes

L'intersection d'histogrammes telle quelle est présentée ne peut pas être considérée comme une distance, car le calcul n'est pas symétrique à cause de la normalisation au dénominateur. Pour mesurer la distance entre deux histogrammes il existe plusieurs distances L_1 , L_2 et L_∞ .

$$d_{L_1}(H, M) = \sum_{i=1}^n |h_{c_i} - m_{c_i}| \quad (47)$$

$$d_{L_2, A}(H, M) = \sqrt{(H - M)A(H - M)^T} \quad (48)$$

$$d_{L_\infty}(H, M) = \max_{1 \leq i \leq n} |h_{c_i} - m_{c_i}| \quad (49)$$

L'utilisation de ces distances fournit de meilleurs résultats que la comparaison à partir d'histogrammes.

f) Appariement basé sur les invariants différentiels

Dans ce cas d'appariement, il faut décider si deux ensembles d'invariants sont similaires ou non. Il est possible par exemple d'utiliser la distance euclidienne L_2 mais elle s'avère inappropriée ici. Plusieurs méthodes plus sophistiquées sont mises en place comme la corrélation, la distance de Mahalanobis, ...etc.

i. La corrélation

Il est possible de calculer un score de corrélation entre les deux vecteurs à comparer, cela donne de très bons résultats surtout avec des vecteurs de grande dimension.

ii. La distance de Mahalanobis

La distance de Mahalanobis est une distance statistique qui utilise la matrice de covariance des composantes de vecteur Λ . La matrice Λ permet de tenir compte de la variation de chacune des composantes ainsi que de leur éventuelle corrélation. Toutefois, l'estimation de cette matrice est très délicate elle est donc estimée d'une façon empirique.

La distance de Mahalanobis est intéressante car c'est une variable aléatoire qui suit la distribution du χ_2 à p degré de liberté, pour des vecteurs d'invariants de dimension p. il est également possible d'utiliser une table de cette distribution pour seuiller la distance et ainsi rejeter les appariements jugés faux, c'est-à-dire les appariements qui correspondent aux plus grandes valeurs de la distance.

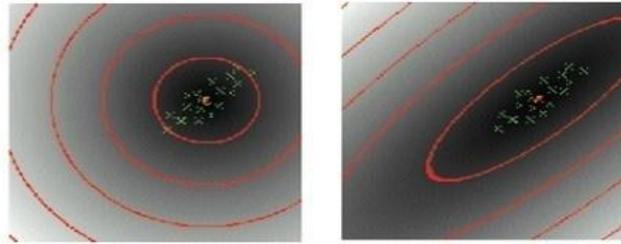


Figure II-20. Distance des points de l'espace au centre d'inertie d'un nuage de points, en distance euclidienne (à gauche) et en distance de Mahalanobis (à droite).

☛ Changement de base

Calculer la distance de Mahalanobis entre deux vecteurs revient donc à transformer ces deux vecteurs en les multipliant par la matrice \sqrt{DP} puis calculer la distance euclidienne entre les deux nouveaux vecteurs.

Selon Ivanov dans [Ivanov, 2004], la distance euclidienne est bonne pour les espaces isotropiques mais présente moins de qualité en cas de transformations linéaires. La distance de Mahalanobis est très bonne s'il y a suffisamment de données.

iii. Contexte multi-échelle

Le choix de l'approche multi-échelle doit s'accompagner de la possibilité de sélectionner pour chaque couple de points comparés les vecteurs qui correspondent au bon facteur d'échelle entre les deux images. Plusieurs travaux sont réalisés dans ce contexte, notons celui de Schmid dans [Schmid, 1997]. Ce travail propose une méthode qui ne nécessite pas la connaissance au préalable du facteur d'échelle entre les deux images appariées. Les supports σ_k de gaussiennes sont calculés par rapport à un support de référence σ_0 que l'on fait varier d'un facteur 1.2^n . Pour $\sigma_0 = 1$ et $n = 4$ on aura les supports: 0.48, 0.58, 0.69, 0.83, 1, 1.2, 1.44, 1.73, 2.07. Les invariants sont alors calculés par rapport à toutes ces échelles. Une méthode plus complète est proposée par Dufournaud qui définit un critère $C(S_i)$ pour trouver parmi tous les facteurs S_i celui qui s'approche le plus du facteur réel. Ce critère atteint son maximum lorsque l'appariement des points entre les deux images est correct pour l'échelle S_i considérée.

g) Procédure d'appariement

Après avoir calculé les scores d'appariement pour tous les couples de points, il faut déterminer ceux qui ont la plus forte probabilité d'être bien appariés. L'intérêt de cette procédure est de respecter le principe d'unicité selon lequel un point doit avoir un correspondant unique dans l'autre image. Plusieurs approches sont possibles :

i. Algorithme du "Winner Takes All"

C'est le plus simple, il consiste à fournir un seul correspondant à chaque point de la première image en ne retenant que le correspondant ayant le meilleur score d'appariement.

Les inconvénients, étant qu'il n'est pas symétrique et qu'il impose un correspondant à un point qui pourrait être occulté dans l'autre image.

ii. Mise en correspondance par appariement croisé

Ayant calculé les vecteurs des caractéristiques \vec{V} pour chaque point d'intérêt, l'algorithme recherche d'abord pour chaque point d'intérêt de la première image le point le plus semblable dans la deuxième image. Pour décider si deux points sont semblables, leurs vecteurs des caractéristiques sont comparés en utilisant une distance (exemple: distance de Mahalanobis). Le processus est ensuite inversé. Le résultat est donc deux listes de paires de points appariés. Les appariements retenus sont les paires de points qui se sont choisies mutuellement.

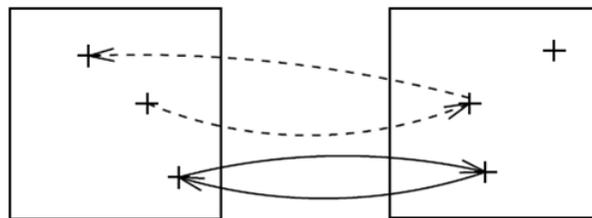


Figure II-21. La mise en correspondance par le principe d'appariement croisé.

Cette méthode a l'avantage d'être symétrique, chaque point apparié n'est impliqué que dans un seul appariement. Mais tous les points ne trouvent pas nécessairement de correspondant et les paires de point retenues sont celles qui ont été mutuellement sélectionnées.

iii. Utilisation d'une fonction de seuillage

Le seuillage des scores d'appariement obtenus consiste à ne garder que les couples d'appariement possédant un score supérieur à une certaine valeur donnée. Ce seuil est généralement choisi arbitrairement par rapport à la moyenne ou médiane de l'ensemble des scores. Pour la distance de Mahalanobis il est possible d'utiliser une table de distribution χ^2 pour calculer le seuil.

Après avoir éliminé les appariements improbables, le reste est un ensemble de couples constituant les appariements potentiels. Ces couples ne respectent pas en général la contrainte d'unicité. Il faut alors utiliser des contraintes de cohérence et des contraintes géométriques pour éliminer les ambiguïtés éventuelles.

3.3.2. Appariement à partir de données géométriques

Ces méthodes cherchent à effectuer la mise en correspondance à partir de données géométriques. Ainsi, bon nombre d'auteurs ont cherché à appairer deux images à partir de

leurs lignes de contraste. Il est possible, dans ce cas, d'avoir une structure de contours qui capture la structure globale de la scène.

Cependant, toutes les méthodes de mise en correspondance qui se basent sur les contours partagent le même défaut : l'utilisation des contours, car ce type de méthodes n'est applicable que si la segmentation des images en contours est correctement faite.

a) Contraintes géométriques

i. Contraintes semi-locales « relaxation »

En présence de bruit, un vecteur des caractéristiques donné peut être impliqué dans plusieurs appariements. Un nombre important d'appariements possibles ainsi que la similarité de certains vecteurs des caractéristiques accroissent encore la probabilité qu'un vecteur vote pour plusieurs appariements.

Califano ainsi que Rao ont suggéré l'utilisation de vecteurs caractéristiques de grande dimensionnalité pour diminuer cette probabilité d'erreur.

Dans la reconnaissance d'objets, il existe une classe d'algorithmes itératifs dits algorithmes de relaxation. Ils permettent de résoudre des problèmes qui s'expriment sous la forme de la minimisation d'une fonction d'énergie globale, appelée critère de relaxation. Le critère de relaxation généralement utilisé est basé sur des contraintes de voisinage du point.

La relaxation dans notre contexte consiste à propager des hypothèses d'appariement par rapport à leur voisinage. Au début, une probabilité initiale mesurant la qualité de l'hypothèse locale est affectée à chaque appariement et elle est révisée suivant la compatibilité du voisinage.

Un autre moyen de diminuer la probabilité de fausses correspondances est de filtrer les mises en correspondance en ajoutant une contrainte de cohérence basée sur le voisinage.

☛ Contrainte de voisinage

La contrainte de voisinage prend en compte la configuration relative des points d'intérêt. Alors, à chaque point P_j dans une image sont associés les p points d'intérêt les plus proches. Un point d'une image est alors représenté par son vecteur d'invariants et par ces p points voisins : $(\vec{V}_j, P_{j,1}, \dots, P_{j,p})$

Si le vecteur des caractéristiques \vec{V}_j issu de l'image est apparié avec le vecteur \vec{V}_k d'une seconde image, tout ou partie des p points d'intérêt voisins de P_j doivent correspondre à des voisins de P_k . Ceci est illustré par la figure ci-dessous.

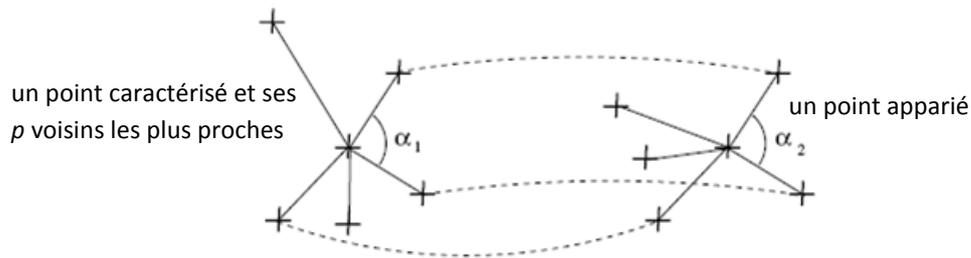


Figure II-22. Exemple de contrainte de voisinage.

Pour que tous les p voisins les plus proches de P_j correspondent aux voisins de P_k il faut qu'il n'y a aucun bruit de détection. Afin de rendre la caractérisation robuste à ces inévitables bruits, il a donc été imposé qu'au moins 50% des voisins de P_j correspondent à des voisins de P_k . L'utilisation de cette contrainte augmente le taux d'appariements corrects en réduisant significativement la probabilité des fausses mises en correspondance.

ii. Autres contraintes géométriques

La géométrie épipolaire est la contrainte géométrique bilinéaire qui existe entre deux images. S'il y en a plusieurs caméras utilisées pour la mise en correspondance, il est possible d'utiliser des contraintes multilinéaires d'ordre supérieur pour rendre les appariements moins ambigus.

L'utilisation de contraintes géométriques permet une vérification supplémentaire des appariements trouvés. Un exemple de ces contraintes est la contrainte basée sur la conservation des angles. La conservation est vérifiée dans le cas des similitudes entre images. C'est à dire dans le cas des similitudes images, l'angle défini par deux voisins d'un point doit être constant pour toutes les vues de ce point. Le point P_j est mis en correspondance avec le point P_k d'une seconde image si et seulement si :

$\vec{V}_{j,m}, \vec{V}_{k,n}$ se correspondent pour quelques (m, n) et $\|\alpha_{k,n} - \alpha_{j,m}\| < \alpha_t$ Pour les (m, n) considérés

Où $\alpha_{j,m}$ est l'angle défini entre deux voisins mis en correspondance et α_t est un seuil angulaire donné. L'ajout de cette contrainte augmente le taux d'appariements corrects.

3.3.3. Evaluation de l'appariement

a) Cadre d'évaluation

i. Critères d'évaluation

Le critère d'évaluation souvent choisi est le nombre d'appariements corrects par rapport au nombre d'appariements trouvés :

$$\frac{\text{Nombre d'appariements corrects}}{\text{Nombre d'appariements trouvés}}$$

Plusieurs autres critères peuvent être utilisés. Dans [Goshtasby, 2005] les critères utilisés sont : la conformité, la fiabilité, la robustesse et la complexité computationnelle.

La conformité se réfère à la différence entre valeurs obtenues et valeurs estimées. Plus cette différence est petite, plus la valeur estimée est conforme. Dans la mise en correspondance d'images, la conformité se réfère au minimum, médiane, maximum ou racine carré de minimum des distances entre les points de l'image de référence et les points correspondants dans l'image en correspondance. La conformité peut être mesurée en utilisant des images synthétiques ou des images de simulation où les coordonnées de vraies correspondances sont connues. Alternativement, des marqueurs fiduciels peuvent être placés dans la scène et les emplacements des fiduciels peuvent être utilisés pour évaluer la conformité de la mise en correspondance. La conformité est mesurée en pixels/voxels.

La fiabilité se réfère au nombre de fois un algorithme réussit à trouver une réponse satisfaisante comparé au nombre total des tests accomplis. Dans la mise en correspondance d'images, si n paires d'images sont testées par un algorithme et si m test des n tests effectués donnent des résultats satisfaisants, si n est suffisamment grand et les images sont représentatives des images d'entrée, donc, m/n représente la fiabilité de l'algorithme. Plus ce taux est proche de 1 plus l'algorithme est fiable.

La robustesse se réfère au degré de stabilité de la conformité ou de la fiabilité de l'algorithme sous une variation d'un ou de plusieurs de ses paramètres d'entrée. Par exemple, si un algorithme peut mettre en correspondance une image pré-opérative avec une image intra-opérative contenant des instruments chirurgicaux qui cachent quelques tissus, l'algorithme est considéré robuste à l'occultation. La robustesse peut être calculée par rapport au bruit, changement d'intensité entre images et le pourcentage des régions dissimilaires dans des images. La robustesse d'un algorithme peut être déterminée en calculant le degré de stabilité de la conformité ou de la fiabilité de l'algorithme quand un ou plusieurs de ses paramètres d'entrée sont changés. Le degré de stabilité de la conformité ou de la fiabilité peut être quantifié en utilisant la déviation standard de la conformité ou la fiabilité lorsque l'un des paramètres d'entrée change. Plus cette déviation est petite, plus l'algorithme sera robuste. S'il y a plusieurs paramètres d'entrée, chacun affectant la conformité ou la fiabilité de l'algorithme, la robustesse de l'algorithme face à plusieurs paramètres peut être déterminée. Un algorithme peut être robuste face au bruit mais pas face aux différences géométriques entre images. Le fait de dire qu'un algorithme est, en général, robuste implique que la performance de l'algorithme ne change pas considérablement lorsque tous ses paramètres d'entrées sont changés. Un algorithme qui a constamment une faible conformité ou fiabilité n'est pas considéré comme robuste. Pour être robuste il doit avoir un taux de conformité important.

La rapidité d'un algorithme dépend de la complexité computationnelle et montre la praticabilité de l'algorithme dans des situations réelles. Un algorithme de mise en correspondance qui nécessite quelques heures pour apparier deux images de cerveau n'est pas de grande valeur dans la neurochirurgie guidée par image. Pour être utile, il doit être capable de mettre en correspondance une image pré-opérative qui est utilisée pour planifier

la chirurgie avec une image intra-opérative qui montre les étapes de la chirurgie à un certain moment et cela dans quelques secondes. L'algorithme qui met en correspondance la mappe d'une région à des images aériennes de la région capturées par une plateforme en mouvement est utile pour l'intervention sur scène si les images peuvent être appariées en millisecondes. La complexité computationnelle est mesurée par le nombre d'additions/multiplications exprimé comme une fonction de dimensions d'image. Un algorithme d'appariement désiré est celui avec une complexité computationnelle qui est une fonction linéaire de dimensions d'image.

ii. Evaluation automatique

L'évaluation automatique d'un appariement diffère suivant que la scène est plane ou non. Dans le cas d'une scène plane, deux images d'une scène sont reliées par une homographie. Etant donné un appariement de points entre deux images, l'homographie existant entre ces images permet de connaître la validité de cet appariement. Dans le cas d'une scène non plane, il n'existe pas d'homographie entre deux images. Cependant, il existe une contrainte entre deux vues quelconques d'une scène, à savoir la relation épipolaire représentée par la matrice fondamentale F où à un point a d'une image correspond une droite épipolaire F_a dans la seconde image. Un appariement (a, b) peut alors être évalué comme correct si le point b appartient à la droite épipolaire correspondant au point a : $F_a * b = 0$, soit ${}^t b F a = 0$. Une telle évaluation n'est pas exempte d'erreurs puisque deux points d'un appariement faux peuvent vérifier la contrainte épipolaire. Cependant, la probabilité d'un tel événement est suffisamment faible pour que cet estimateur fournisse une bonne évaluation du nombre d'appariements corrects.

Conclusion

Dans ce chapitre, nous avons passé en revue un état de l'art des détecteurs de points d'intérêt ainsi qu'une classification de ces détecteurs selon les approches de détection utilisées. Il est clair qu'il y a un grand nombre de détecteurs de points d'intérêt chacun ayant ses qualités et défauts. Le choix du détecteur utilisé est généralement le résultat d'évaluation de ce dernier. Ceci nécessite la définition d'un cadre d'évaluation du détecteur et également des critères d'évaluations selon l'application envisagée. Pour avoir une idée sur ces critères, nous avons présenté les critères les plus fréquents dans la littérature.

Nous avons également présenté un état de l'art des méthodes de caractérisations locales de points d'intérêt en niveaux de gris comme en couleur

Après avoir présenté les deux premières étapes du processus de mise en correspondance d'images, nous avons étoffé le chapitre d'un état de l'art des différentes méthodes d'appariement. À partir de cet état de l'art, nous sommes arrivés à la conclusion que les grandeurs géométriques sont instables car reposant sur une description symbolique et peu discriminante du fait qu'une partie de l'information de niveau de gris n'est pas utilisée. Toutefois, les méthodes basées sur des grandeurs photométriques sont par nature

plus discriminantes car moins symboliques. Elles sont plus proches du signal, alors que leur défaut principal est qu'elles ne sont pas invariantes à des types de transformations qui peuvent être prises en compte par les méthodes géométriques.

Cet état de l'art a également prouvé que la qualité de l'appariement ne dépend pas uniquement de la caractérisation mais aussi des points pour lesquels cette caractérisation est calculée. Il est important que ces points soient répétables, car une telle répétabilité influence la stabilité de la caractérisation et donc la qualité des appariements obtenus. Cela revient toujours à l'idée que la discrimination et l'interprétabilité des points détectés jouent un rôle fondamental dans le processus complet de la mise en correspondance.

L'accent a également été mis sur les détecteurs des points d'intérêt inspirés de la biologie et particulièrement les détecteurs de points de focus car ils présentent un intérêt visuel et psychologique très important.

Après l'étude des différents travaux dans la littérature, nous avons remarqué que les détecteurs des points de focus ont un intérêt spécifique dans le domaine de reconnaissance d'objets, car l'interprétabilité des points détectés détermine la qualité de l'appariement d'images. En suivant cet axe de recherche, nous avons décidé de reprendre les travaux faits dans cette orientation et particulièrement le travail de Rebai [Rebai, 2006]. Il nous a semblé intéressant d'étendre à la couleur l'approche proposée par Rebai en profitant de l'information contenue dans l'image couleur. Dans l'espoir d'améliorer la détection des points de focus et d'exploiter la richesse de l'information couleur, nous avons implémenté notre nouvelle approche étendue à la couleur. La présentation de cette approche ainsi que les expérimentations effectuées fait l'objet du chapitre suivant.

Chapitre III Les points de focus couleur



Introduction

La symétrie est un phénomène dominant aussi bien dans les scènes naturelles que celles artificielles ce qui a renforcé le fait qu'elle va occuper un rôle fondamental dans les groupements perceptuels. La symétrie se présente sous différents types: elle peut être réflexive, translationnelle, radiale ou rotationnelle [Hugo, 2006]. Il faut mentionner aussi que plusieurs travaux psychologiques confirment que la symétrie attire l'attention du système visuel humain. Ce qui fait que l'utilisation des centres d'objets symétriques comme des points d'intérêt constitue une issue très importante pour remédier aux insuffisances des approches existantes qui reposent sur la simple information du signal. La détection des centres de symétrie se fait sans connaissance préalable du contenu de la scène. Les approches qui exploitent les points de focus sont alors dites indépendantes du contexte.

1. Approches de détection de points de focus

1.1. Transformée généralisée de Reifeld

Reifeld dans [Reifeld, 1995] suggère que les opérateurs attentionnels doivent être pris avec sérieux par la communauté de la vision par ordinateur. Reifeld atteste aussi que la détection des régions d'intérêt est un processus dépendant du contexte car le système adjacent est orienté tâche et en même temps indépendant du contexte car parmi les tâches les plus importantes d'un tel système est la détection des signaux imprévus. Reifeld préfère ainsi l'utilisation de primitives extraites moyennant un concept plus généralisé. Il suggère que la symétrie locale généralisée est le principal "context free attention". La transformée de Reifeld utilise la symétrie réflexive pour calculer une matrice de vote sur l'image. Cette transformée ne nécessite aucune connaissance préalable de la forme de l'objet. Elle réalise des opérations locales sur les contours de l'image en assignant un score de symétrie à chaque point dans l'image. Soit p_k un point quelconque dans l'image, soit ∇p_k le gradient du point p_k . Supposant qu'un vecteur $v_k = (r_k, \theta_k)$ est associé à chaque point p_k tel que :

$$r_k = \log(1 + \|\nabla p_k\|) \text{ et } \theta_k = \arctan\left(\frac{\frac{\partial}{\partial y} p_k}{\frac{\partial}{\partial x} p_k}\right)$$

Pour chaque couple de points p_i, p_j , l étant la ligne passant par les deux points. On note par α_{ij} l'angle que fait la ligne l avec l'Horizon. L'ensemble $\Gamma(p)$ est défini par :

$$\Gamma(p) = \left\{ (i, j) \mid \frac{p_i + p_j}{2} = p \right\} \quad (1)$$

La transformée généralisée est définie par la formule :

$$M(p) = \sum_{(i,j) \in \Gamma(p)} PWF(i, j) \cdot GWF(i, j) \cdot DWF_\sigma(i, j) \quad (2)$$

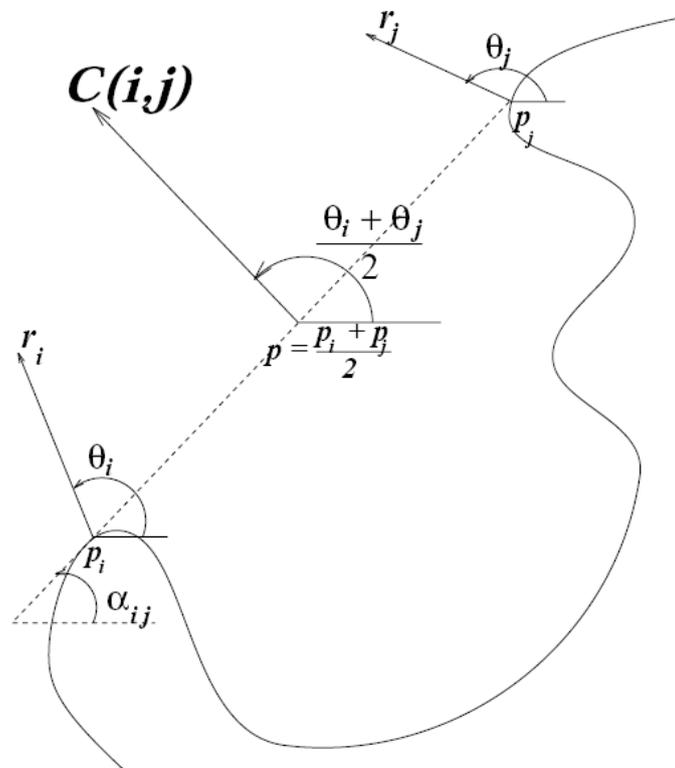


Figure III-1. Symétrie réfective de Reisfeld [Reisfeld, 1995]

Le rôle de chaque fonction dans la transformée est donné comme suit :

➤ **Original phase weight function (PWF)**

$$PWF(i, j) = \underbrace{[1 - \cos(\gamma_i + \gamma_j)]}_{PWF^+(i, j)} \cdot \underbrace{[1 - \cos(\gamma_i - \gamma_j)]}_{PWF^-(i, j)} \quad (3)$$

Où $\gamma_i = \theta_i - \alpha_{ij}$, $\gamma_j = \theta_j - \alpha_{ij}$

La fonction PWF représente la contribution la plus importante dans la transformée. Elle atteint son maximum lorsque les vecteurs gradients des points en question sont strictement opposés ou ils sont dans des directions strictement opposés.

➤ **Gradient Weight Function (GWF)**

$$GWF(i, j) = \log(1 + G_I(p_i)) \cdot \log(1 + G_I(p_j)) \quad (4)$$

L'objectif de cette fonction est d'augmenter les poids des deux points s'ils sont tous les deux sur des contours ; car les contours souvent relient les pourtours d'un objet, donc une grande importance doit être accordée à leurs votes.

➤ **Distance Weight Function (DWF)**

$$DWF_{\sigma}(i, j) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma^2}\right) \quad (5)$$

La fonction DWF permet de faire de la symétrie une mesure locale et le choix de σ définit l'échelle sur laquelle les symétries sont détectées et par conséquent les rayons des objets à détecter.

Cette formule a été sujette de plusieurs améliorations. Citons, la restriction des pixels traités aux pixels des contours, la restriction aux régions circulaires et la marginalisation du rôle de DWF_{σ} , car c'est le rayon r qui définit dans ce cas l'échelle d'intérêt et non pas le paramètre σ . Le calcul de la matrice de symétrie suit donc la nouvelle formule :

$$M(p) = \sum_{(i,j) \in \Gamma^*(p)} PWF(i, j) \cdot GWF(i, j) \quad (6)$$

Où:

$$\Gamma^* = \Gamma' \setminus \{(i, j) \mid \|p_i - p_j\| > 2R\} \quad (7)$$

Et

$$\Gamma' = \Gamma \setminus \{(i, j) \mid G_I^B(p_i) = 0 \vee G_I^B(p_j) = 0\} \quad (8)$$

1.2. Transformée généralisée de symétrie en couleur

C'est la transformée de Reisfeld étendue à la couleur en vue d'améliorer ses résultats. Heidemann a amélioré le rendement de la transformée généralisée de Reisfeld en utilisant la couleur. Il propose sa transformée couleur pour détecter les objets qui diffèrent de l'arrière-plan en couleur mais pas en valeurs de niveaux de gris. Parce que les centres de

ces objets ne peuvent pas se détecter en utilisant la transformée généralisée (Figure III-2). Heidemann dans [Heidemann, 2004] redéfinit la matrice de vote en couleur par :

$$M_{Col} = \sum_{(i,j) \in \Gamma^*(p)} \sum_{(k,l) \in \Lambda(p,i,j)} PWF_{Col}(i,j,k,l) \cdot GWF_{Col}(i,j,k,l) \quad (9)$$

Tel que:

$$\Gamma^*(p) = \left\{ (i,j) \mid \frac{p_i + p_j}{2} = p \wedge \|p_i - p_j\| \leq 2R \right\} \quad (10)$$

Et :

$$\Lambda(p,i,j) = \left\{ (k,l) \mid k,l \in \{0,1,2\} \wedge G_k(p_i) \geq \theta_k \wedge G_l(p_j) \geq \theta_l \right\} \quad (11)$$

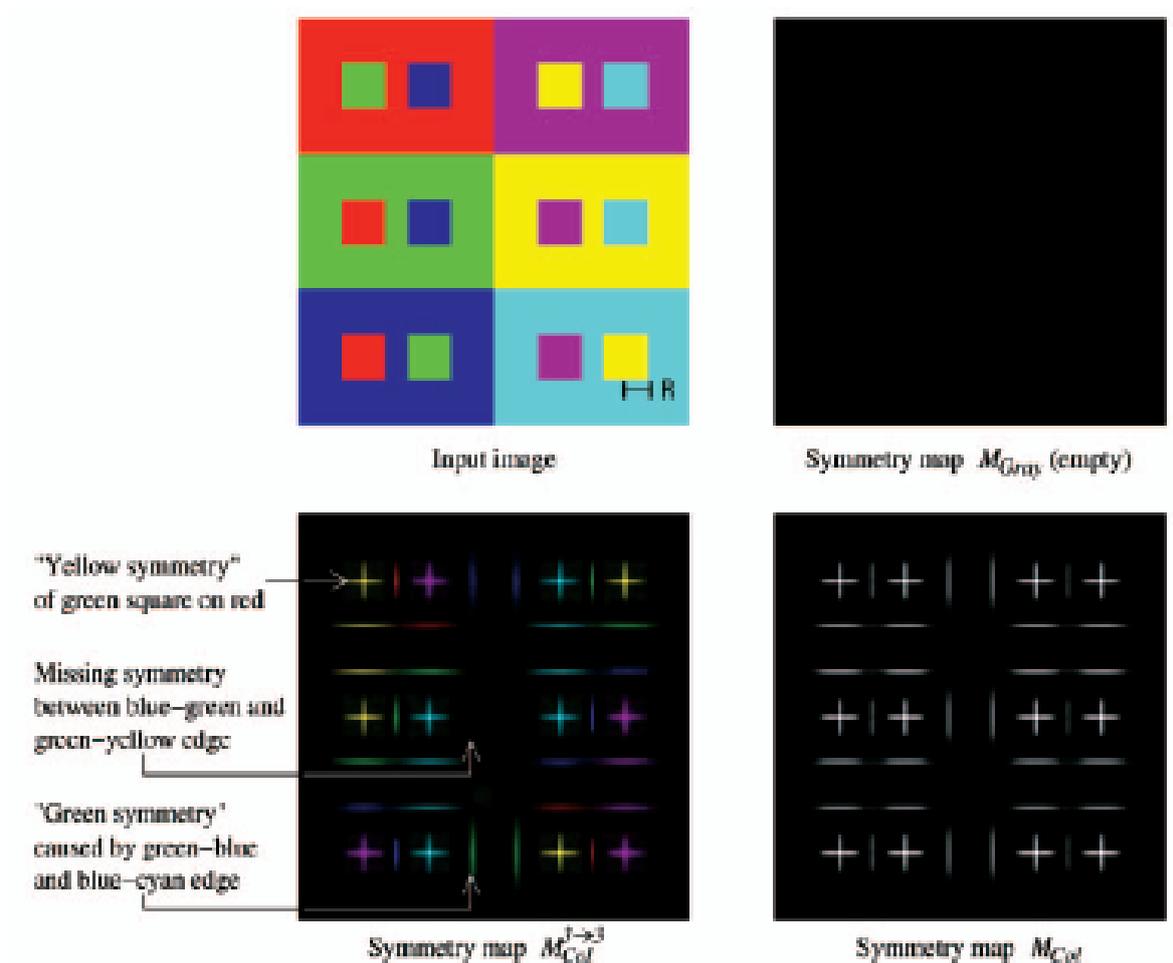


Figure III-2. Apport de la couleur dans la symétrie réflexive [Heidemann, 2004]

En utilisant le gradient en couleur, Heidemann redéfinit les fonctions *PWF* et *GWF*, la Figure III-3 représente les gradients couleur utilisés pour la fonction *PWF*, La Figure III-4

donne des exemples de valeurs que peut prendre la fonction PWF . selon les vecteurs gradient couleurs la formule de PWF est donnée par :

$$PWF_{Col}(i, j, k, l) = \underbrace{\left[\cos^2(\gamma_{ik} + \gamma_{jl}) \right]}_{PWF_{Col}^+(i, j, k, l)} \cdot \underbrace{\left[\cos^2(\gamma_{ik}) \cdot \cos^2(\gamma_{jl}) \right]}_{PWF_{Col}^-(i, j, k, l)} \quad (12)$$

La fonction GWF est reformulée ainsi:

$$GWF_{Col}(i, j, k, l) = \log(1 + G_k(p_i)) \cdot \log(1 + G_l(p_j)) \quad (13)$$

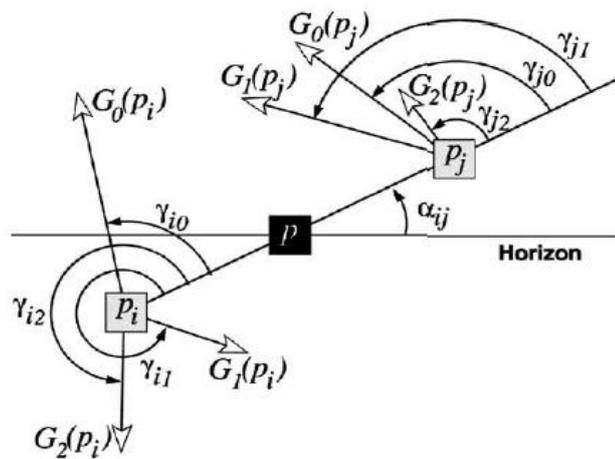


Figure III-3. Gradient multi spectral pour deux points en symetrie reflective

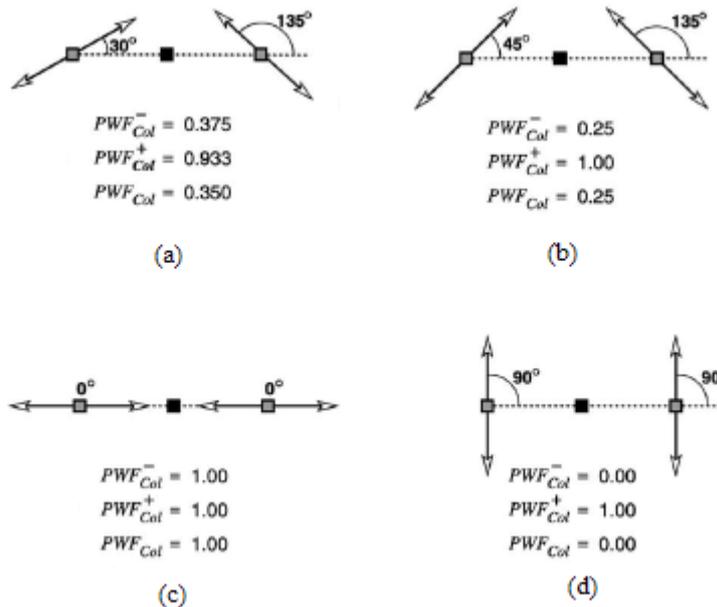


Figure III-4. Valeurs de la fonction PWF suivant l'angle d'orientation des vecteurs gradients

1.3. Transformée de Loy et Zelinsky

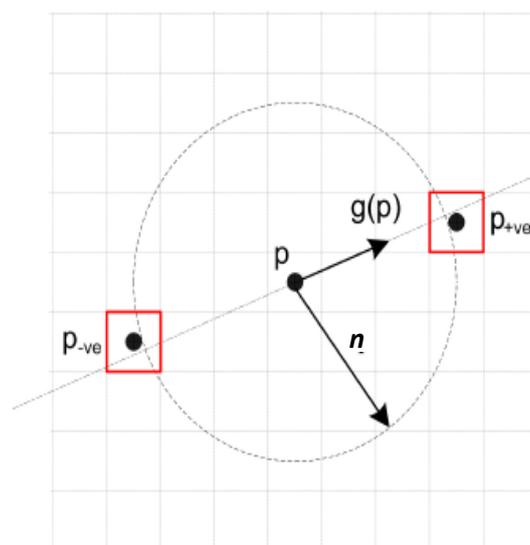
Généralement, la détection de la symétrie est faite en appliquant l'une des deux approches de vote suivantes :

La première approche fait un seul parcours de l'image durant lequel, pour chaque pixel, il faut regarder les pixels dans son voisinage pour vérifier s'il représente un point de symétrie.

Dans la deuxième approche, un score de symétrie est calculé à partir des votes faits par un pixel pour les pixels (points de symétrie) qui l'entourent. Le détecteur de Loy et Zelinsky [Zelinsky, 2003] se base sur cette deuxième approche pour déterminer les points de symétrie qui entourent chaque pixel dans un rayon n allant de 1 à n_{max} , où n_{max} représente le rayon maximal.

Avec cette transformée chaque valeur de n tel que $n < n_{max}$ entraîne le calcul de deux images de projections correspondantes : Une image d'orientation O_n et une image de magnitude M_n .

Pour chaque rayon n , le gradient g de chaque point p de l'image est examiné. Pour le point p il faut calculer deux points symétriques qui se situent à une distance n dans la direction du gradient g . l'un de ces points symétriques est dit affecté positivement et noté p_{+ve} ; parce que le vecteur gradient $g(p)$ pointe directement vers ce point. Cependant, l'autre point est dit affecté négativement et noté p_{-ve} . Les coordonnées de ces points sont données par les équations suivantes :



$$p_{+ve} = p + \text{ARRONDI} \left(n \cdot \frac{g(p)}{\|g(p)\|} \right) \quad (14)$$

$$p_{-ve} = p - \text{ARRONDI} \left(n \cdot \frac{g(p)}{\|g(p)\|} \right) \quad (15)$$

Figure III-5. Points affectés positivement et négativement

La fonction "ARRONDI" arrondit chaque élément du vecteur à l'entier le plus proche. Les images d'orientation et de magnitude sont initialement zéro. Pour chaque paire de pixels affectés, il faut incrémenter par 1 et par $\|g(p)\|$ respectivement le point correspondant à p_{+ve} dans l'image d'orientation et l'image de magnitude. Par contre, pour le point

correspondant à p_{-ve} il faut décrémenter les mêmes valeurs. Ceci est donné par les équations suivantes :

$$O_n(p_{+ve}) = O_n(p_{+ve}) + 1 \quad (16)$$

$$O_n(p_{-ve}) = O_n(p_{-ve}) - 1 \quad (17)$$

$$M_n(p_{+ve}) = M_n(p_{+ve}) + \|g(p)\| \quad (18)$$

$$M_n(p_{-ve}) = M_n(p_{-ve}) - \|g(p)\| \quad (19)$$

La contribution de la symétrie radiale au rayon n est donnée par la convolution :

$$S_n = F_n * A_n \quad (20)$$

Tel que :

$$F_n(p) = \frac{M_n(p)}{k_n} \left(\frac{|\tilde{O}_n(p)|}{k_n} \right)^\alpha \quad (21)$$

Et :

$$\tilde{O}_n(p) = \begin{cases} O_n(p) & \text{Si } O_n < kn \\ k_n & \text{Sinon} \end{cases} \quad (22)$$

A_n est une gaussienne bidimensionnelle, α est le paramètre spécifiant la précision radiale de symétrie, k_n est un facteur d'échelle qui normalise M_n et O_n à travers différents rayons n . La transformée radiale complète est donnée par l'équation :

$$S = \frac{1}{|N|} \sum_{n \in N} S_n \quad (23)$$

2. Les points d'intérêt et la couleur

En fait, la couleur a été choisie par plusieurs auteurs antérieurement pour améliorer les méthodes existantes et profiter du contenu informatif de celle-ci. Le but de notre travail est de faire un état de l'art des méthodes existantes et essayer de proposer une nouvelle approche qui exploite l'information couleur. En examinant les travaux existants, nous avons conclu que le travail de Gouet [Gouet, 2000] est une extension à la couleur du fameux détecteur de points d'intérêt Harris [Harris, 1988]. Le travail de Heidemann [Heidemann, 2004] est une extension à la couleur de l'approche proposée par Reisfeld dans [Reisfeld, 1995]. Sur ce même axe, nous proposons dans ce travail, l'extension à la couleur du détecteur proposé par Alexis et Rebai [Rebai, 2006b] qui est lui-même une généralisation du détecteur proposé par Zelinsky [Zelinsky, 2003]. Nous proposons une nouvelle approche basée sur la couleur dont le but est de renforcer la détection par l'utilisation de la combinaison Symétrie-Couleur. Notre travail porte sur deux apports principaux:

Le premier apport est l'exploitation de l'information contenue dans la couleur par l'extension de l'approche d'Alexis et Rebai à la couleur.

Le deuxième apport est la proposition d'une méthode automatique pour le calcul de l'angle θ utilisé par Alexis et Rebai. Cet angle permet de détecter des centres d'objets qui ne sont pas mathématiquement symétriques. Il peut être aussi utilisé pour la caractérisation des points détectés.

3. Extension de l'approche d'Alexis et Rebai

3.1. Approche d'Alexis et Rebai

Dans des travaux récents, ceux de Rebai et al. [Rebai, 2006a], [Rebai, 2006b], [Rebai, 2007] la méthode de Loy et Zelinsky a été généralisée. Ils ont gardé le principe de vote en le faisant dans plusieurs directions et non pas la direction du gradient seulement. Cette amélioration permet de détecter les centres de symétrie des objets qui ne sont pas forcément circulaires comme les spirales, les points d'intersection, les jonctions fictives...etc.

Le principe d'élection

Dans les travaux de Loy et Zelinsky, un principe de vote est utilisé pour détecter les centres d'objets circulaires. Ce même principe a été amélioré ultérieurement par Rebai et al. Pour détecter des centres d'objet qui ne sont pas forcément circulaires mais qui présentent à un certain degré un aspect symétrique comme les spirales, les points d'intersection ...etc. Pour chaque pixel nous voulons calculer un accumulateur dont les maximas donnent les centres de symétrie. Donc pour chaque direction de gradient, tous les points situés dans cette direction vont augmenter par un leurs accumulateurs. Contrairement au travail original où chaque pixel calcule ses deux points de symétrie dans la direction du gradient et la direction inverse.

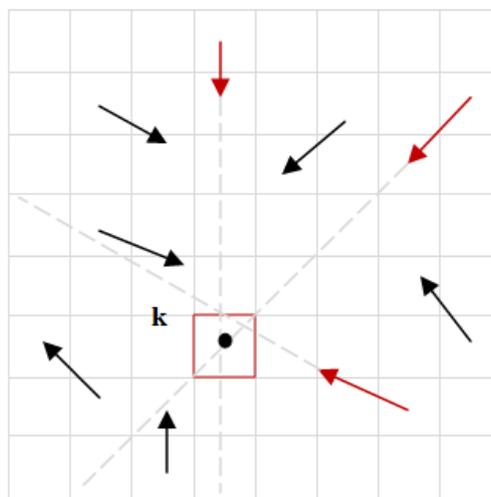


Figure III-6. Illustration du principe de vote ; dans ce cas le point k va avoir une valeur de symétrie égale à 3 parce que 3 vecteurs gradients votent pour lui comme un centre de symétrie.

Ce principe est généralisé par Rebai et al. en calculant les accumulateurs non plus dans la seule direction du gradient mais dans la direction de gradient tourné tout au long

d'un arc de cercle ayant le même centre du gradient et sa longueur L dépend du rayon r du vote.

$$L = \delta * r \quad (24)$$

δ est un angle constant utilisé pour définir l'arc.

L'idée est alors de faire varier d'un angle θ la direction du gradient, les équations (14) et (15) proposées par Loy et Zelinsky deviennent alors :

$$p_{+ve} = p + \text{ARRONDI}(n \cdot (\cos(\beta + \theta)x + \sin(\beta + \theta)y)) \quad (25)$$

$$p_{-ve} = p - \text{ARRONDI}(n \cdot (\cos(\beta + \theta)x + \sin(\beta + \theta)y)) \quad (26)$$

Où β désigne l'angle du gradient.

Quant $\theta = \pm \frac{\pi}{2}$ les vecteurs votants auront une direction perpendiculaire à celle du gradient. Les points détectés sont donc des jonctions ou encore des jonctions fictives tout dépend de n_{\max} . Si θ prend des valeurs autres que $\{ \frac{-\pi}{2}, 0, \frac{\pi}{2} \}$ les points détectés vont être des centres de spirales ou des morceaux de spirales logarithmiques et ainsi de suite.

Rebai et Joly définissent alors l'équation générale de la transformée par :

$$E_{\theta}(x, y) = \sum_i \sum_j \left| \vec{\nabla}_{\theta} I(i, j) \cdot \overrightarrow{R_{xy}}(i, j) \right| \omega \left(\left\| \overrightarrow{R_{xy}}(i, j) \right\|, N_{xy}(i, j) \right) \quad (27)$$

Où I est l'image originale ; $\vec{\nabla}_{\theta} I(i, j)$ désigne la direction du vecteur gradient tourné par θ ; $\overrightarrow{R_{xy}}(i, j)$ est le vecteur défini par les deux pixels (x, y) et (i, j) :

$$\overrightarrow{R_{xy}}(i, j) = \begin{pmatrix} x - i \\ y - j \end{pmatrix} \quad (28)$$

$$\text{Et } N_{xy}(i, j) = \frac{\left| \vec{\nabla} I(i, j) \cdot \overrightarrow{R_{xy}}(i, j) \right|}{\left\| \vec{\nabla} I(i, j) \right\| \cdot \left\| \overrightarrow{R_{xy}}(i, j) \right\|} \quad (29)$$

$w(\cdot, \cdot)$ est une fonction de pondération qui dépend de deux paramètres : la distance de vote $\left\| \overrightarrow{R_{xy}} \right\|$ et l'angle θ déduit de N_{xy} .

Pour calculer le rayon d'un objet détecté il faut utiliser un histogramme accumulateur pour chaque pixel dans l'image. La longueur de chaque histogramme est égale à la diagonale de l'image qui représente théoriquement le rayon le plus grand qui peut exister.

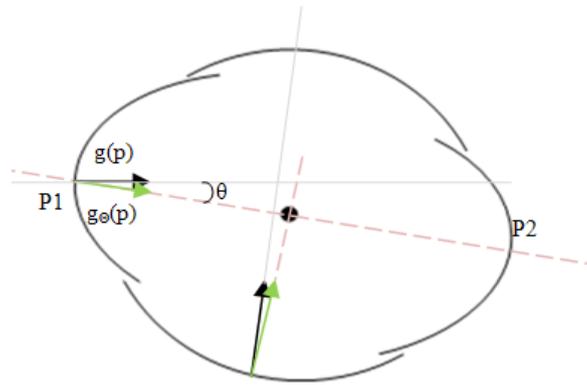


Figure III-7. Un des objets symétriques qui ne sont pas circulaires et dont les détecteurs classiques ne peuvent pas détecter le centre de symétrie.

4. Approche proposée

Notre approche est une extension à la couleur de l'approche de Rebai et al. Nous nous sommes basés sur des travaux psychologiques qui attestent que la symétrie a un très grand impact vis-à-vis du système visuel humain et sur les travaux qui attestent que la couleur est l'une des primitives qui présentent un intérêt visuel et informatif très important. Cette richesse de l'information couleur rend indispensable son utilisation dans le processus de détection des points d'intérêt. Dans ce travail, nous présentons un nouveau détecteur de point de focus qui exploite l'information couleur et l'intérêt visuel de la symétrie pour construire des détecteurs plus performants. Dans cette section nous présentons notre nouvel opérateur de symétrie enrichi par la couleur.

4.1.1. Le principe d'élection par le gradient multi-spectral

Nous avons gardé le même principe de vote par le gradient en niveau de gris. Nous avons remplacé ce gradient par le gradient multi-spectral. Dans ce cas, un vecteur gradient multi-spectral vote pour plusieurs pixels à la fois. Chaque canal (R, V, B) utilise son vecteur gradient pour faire des votes. L'utilisation du gradient multi-spectral permet de détecter les centres de symétrie des objets qui diffèrent de l'arrière plan en couleurs mais pas en niveau de gris. Plusieurs espaces couleurs peuvent être utilisés, tout dépend de la simplicité de calculs des gradients dans ces espaces. Les gradients se différencient selon l'espace couleur choisi, certains espaces génèrent des gradients forts alors que d'autres génèrent des gradients faibles, ce qui affecte considérablement la détection de symétrie. Nous avons choisi l'espace RGB.

a) Le gradient multi-spectral

De nombreuses méthodes peuvent être définies pour combiner les gradients provenant des différents plans d'une image couleur. Nous pouvons considérer que la norme du gradient multi-spectral est la norme infinie des normes quadratiques des gradients dans chacun des plans, et l'orientation est celle correspondant au maximum des trois gradients.

Nous pouvons toutefois considérer chaque gradient couleur séparément et utiliser sa magnitude et son orientation.

b) Apport d'utilisation du gradient multi-spectral

La majorité des images qu'elles soient naturelles ou artificielles présentent des objets sur des arrière-plans. Dans certains cas, ces objets ne se différencient de leurs arrière-plans qu'en valeur couleur. Donc, en niveau de gris il n'est pas possible de détecter ces objets ni de détecter leurs centres de symétrie. Un autre intérêt de l'approche proposée est le nombre de votes procurés par un pixel. Ce fait permet d'avoir des accumulateurs plus différentiables avec des taux de discrimination importants et par conséquent il devient plus facile de déterminer les centres de symétrie. En utilisant les gradients couleurs nous pouvons approfondir la différence entre les scores de symétrie des pixels constituant des centres de symétrie et les autres pixels de l'image.

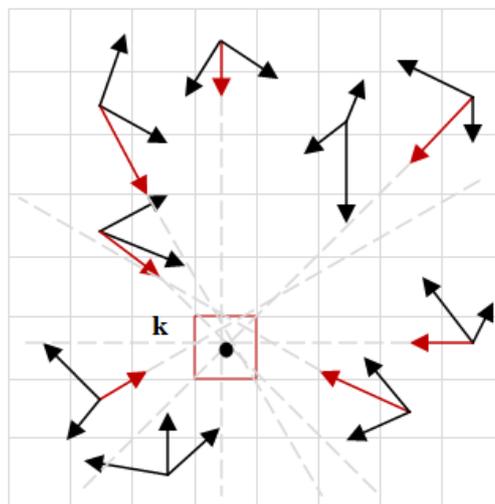


Figure III-8. Les votes en utilisant le gradient multi-spectral ; le point k aura plus de votes (7 dans ce cas) ce que renforce sa discrimination.

L'équation générale de la transformée en couleur devient donc :

$$E_{\theta}(x, y) = \sum_i \sum_j \left| \vec{\nabla}_{\theta k} I(i, j) \cdot \vec{R}_{xy}(i, j) \right| \quad (30)$$

$$k = \{0,1,2\}$$

I étant l'image originale ; $\vec{\nabla}_{\theta k} I(i, j)$ désigne la direction du vecteur gradient dans le plan couleur k tourné par un angle θ .

4.1.2. Détails d'implémentation

Le choix de l'espace couleur utilisé est en fonction de la simplicité de calcul du gradient. Cependant ce choix affecte les matrices de symétrie résultantes car des gradients

faibles calculés dans un espace couleur peuvent être forts dans d'autres espaces couleur. La Figure III-9 illustre cette différence de calcul entre l'espace RGB et l'espace CMYK.

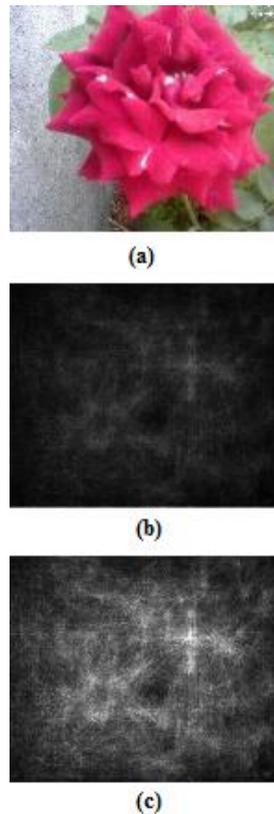


Figure III-9. (a) image original, (b) matrice de symétrie avec des gradients CMYK, (c) matrice de symétrie avec des gradients RGB ($R_{max}=40$, $\theta=0$).

Nous avons utilisé les trois canaux RGB pour effectuer les votes. Pour chaque canal nous calculons le gradient en utilisant l'opérateur de Sobel. Par le vecteur votant construit à partir du vecteur gradient dévié d'un certain angle θ , nous calculons les points affectés positivement et négativement. Après avoir déterminé les points affectés positivement et négativement nous pouvons calculer la matrice d'orientation O de l'image en utilisant les mêmes formules (25) et (26) déjà vues. À partir de la matrice d'orientation, la matrice de symétrie S peut être calculée par la sommation des différentes matrices S_n ; S_n étant la matrice d'orientation O après une convolution par une gaussienne 2D. Nous avons implémenté les trois approches (Loy & Zelinsky, Alexis & Rebai et notre approche) en utilisant le langage C++ sur la plate forme Windows XP.

4.1.3. Quelques résultats

Les premiers résultats obtenus valident les hypothèses déjà faites. La Figure III-10 montre le gain obtenu dans les images d'orientations calculées à partir d'information couleur par rapport aux images d'orientation calculées en niveau de gris. Dans la Figure III-10, (a) représente l'image d'orientation calculée à partir de votes faits par des vecteurs gradients en niveau de gris avec les points représentant les maxima locaux sur l'image originale. (b) représente l'image d'orientation calculée à partir de votes réalisés par des

vecteurs des gradients multi-spectral dans les trois canaux Rouge, Vert et Bleu avec les points représentant les maxima locaux sur la même image originale. Les figures de (11) à (19) présentent d'autres résultats sur différentes images. Ces résultats présentent le gain obtenu par l'introduction de la couleur dans les votes.

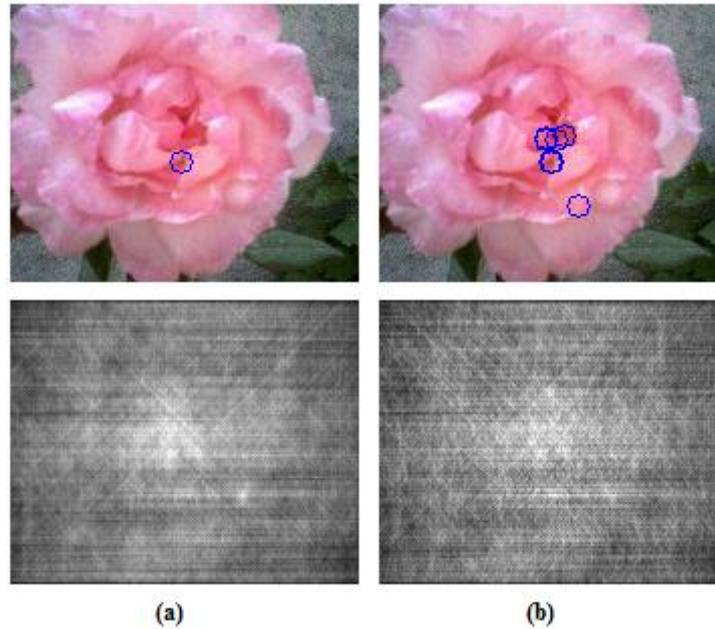


Figure III-10. (a) image d'orientation couleur, (b) image d'orientation en niveau de gris, avec les points détectés marqués par des cercles bleu sur l'image originale.

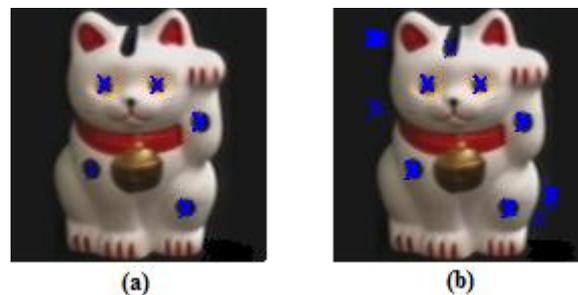


Figure III-11. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=9$, Nombre maximum de points de focus = 15, $\theta=0$)

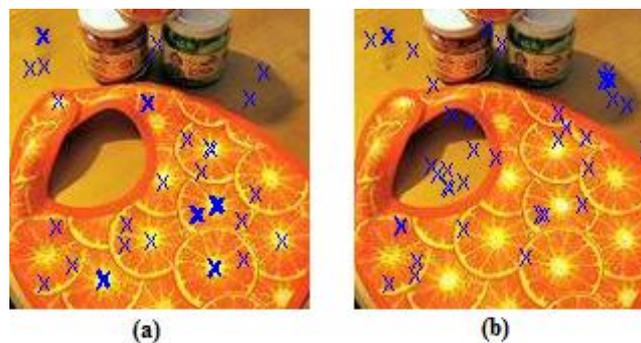


Figure III-12. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=8$, Nombre maximum de points de focus = 40, $\theta=0$)



Figure III-13. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=13$, Nombre maximum de points de focus = 13, $\theta=0$)



Figure III-14. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=70$, Nombre maximum de points de focus = 3, $\theta=0$)

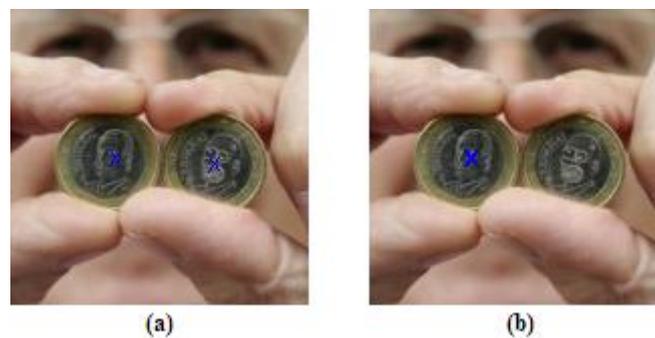


Figure III-15. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=35$, Nombre maximum de points de focus = 2, $\theta=0$)

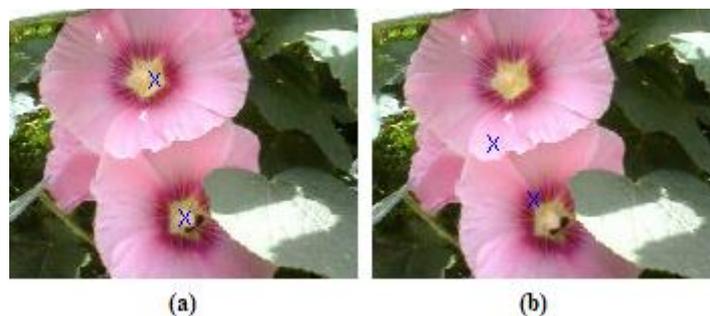


Figure III-16. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai ($R_{max}=30$, Nombre maximum de points de focus = 2, $\theta=0$)

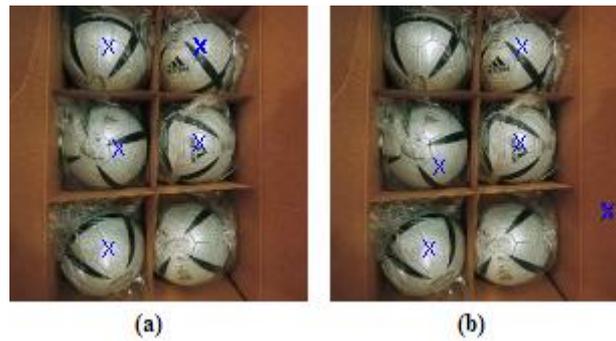


Figure III-17. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai (Rmax=20, Nombre maximum de points de focus = 6, $\theta = \pi/45$)



Figure III-18. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai (Rmax=40, Nombre maximum de points de focus 1, $\theta = \pi/2$)

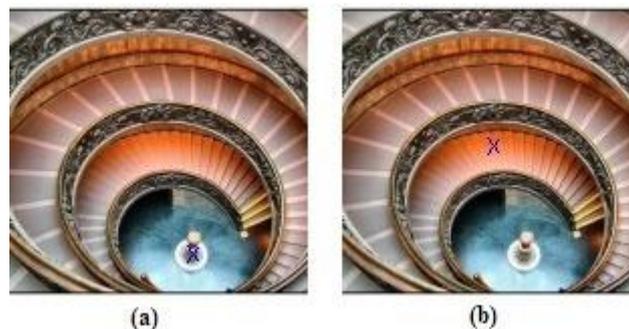


Figure III-19. (a) les points détectés par l'approche proposée. (b) les points détectés par l'approche de Rebai (Rmax=30, Nombre maximum de points de focus = 1, $\theta = \pi/16$)

5. Calcul automatique de l'angle θ

Dans ce travail, nous proposons aussi le calcul automatique de l'angle θ utilisé par Alexis & Rebai pour détecter les centres d'objets qui ne sont pas strictement symétriques ou encore pour détecter des centres de spirales, des points d'intersections voire des jonctions fictives. Alexis & Rebai définissent un espace de valeur θ et le nomment « θ space ». Toutefois, l'utilisation de toutes les valeurs de cet ensemble pour la détection est peu judicieuse. Ils proposent ainsi l'utilisation d'un sous ensemble de valeur dans cet ensemble.

Ils choisissent aléatoirement des valeurs entre $-\frac{\pi}{2}$ et $\frac{\pi}{2}$. La valeur 0 donne les centres de

cercles. Les valeurs $-\frac{\pi}{2}$ et $\frac{\pi}{2}$ donnent les points d'intersections et les jonctions fictives. Pour détecter les centres de spirales, ils accordent à l'angle θ la valeur $\Phi - \alpha$ où α est l'angle du gradient et Φ l'angle polaire de la spirale. Dans des cas d'objets qui présentent un certain degré de symétrie mais qui ne sont pas symétriques mathématiquement ; par exemple l'objet de la Figure III-7, la valeur de l'angle θ n'est pas donnée, Elle est choisie aléatoirement. Alexis et Rebai proposent également dans des travaux plus récents [Rebai, 2007], l'utilisation de l'angle θ pour la caractérisation des points détectés. Pour cela, nous avons jugé intéressant la redéfinition de cet espace sur des fondements mathématiques plutôt que de laisser le choix aléatoire. En se basant sur les travaux de Reisfeld présentés antérieurement, nous avons défini une approche qui permet le calcul automatique de cet angle pour chaque couple de pixels symétriques ce qui peut, théoriquement, aider dans le processus de caractérisation des points détectés.

Dans la Figure III-20 la valeur de θ est celle qui maximise la fonction MSF (Maximal Symmetric Function) formulée par :

$$MSF(i, j) = [1 - \cos(\alpha_i + \alpha_j - 2\theta)] \cdot [1 - \cos(\alpha_i - \alpha_j)] \quad (31)$$

Où : α_i et α_j sont les angles de gradient des pixels i, j respectivement. Le maximum théorique de cette fonction est bien 4. Car le maximum du $\cos(\beta)$ est soit 1 si $\beta = 0^\circ$ soit -1 si $\beta = 180^\circ$. La valeur 1 donne $MSF=0$ et la valeur -1 donne $MSF = 2 \cdot 2 = 4$.

Cette approche permet de calculer automatiquement cet angle pour n'importe quel couple de pixels symétriques. Ceci va permettre la sélection de l'angle θ dans un ensemble de valeurs extraites directement de l'image concernée. La simplicité du calcul et la robustesse de cet angle face à plusieurs transformations image le rend qualifié pour construire des nouveaux descripteurs locaux plus adéquats aux points de focus.

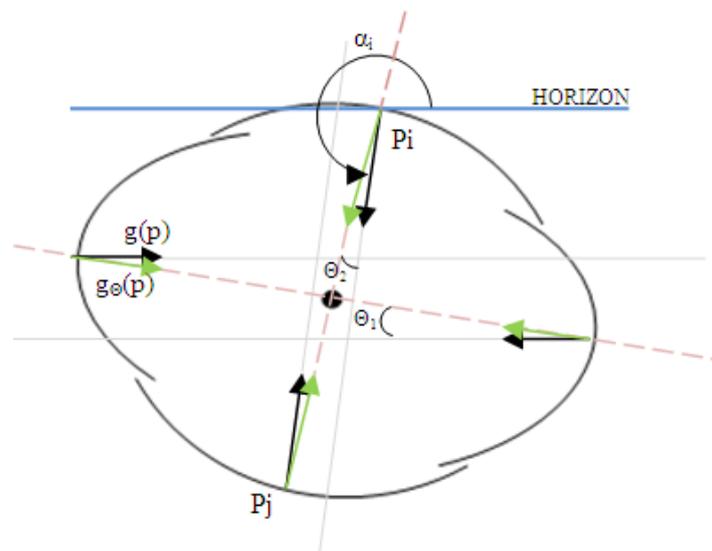


Figure III-20. Relation angulaire entre le gradient et l'angle θ .

6. Résultats expérimentaux

6.1. Intérêt de la symétrie radiale en couleur

L'utilisation des méthodes basées sur les niveaux de gris expliquées plus haut permet de détecter des centres de symétrie radiale. Cependant, si les objets ne diffèrent pas de l'arrière plan en niveau de gris, la détection de leurs centres devient impossible. Dans ce cas, l'intégration de la couleur dans les votes, devient indispensable. Les résultats obtenus valident cette hypothèse. La Figure III-21 donne les centres de symétrie radiale trouvés en niveaux de gris et ceux en couleur. Les figures (a), (d), (g) représentent les images originales, (b), (e), (h) représentent les matrices de vote en couleur, (c), (f), (i) représentent les matrices de vote en niveaux de gris. Nous remarquons que les centres des cercles (vert sur l'arrière plan bleu, jaune sur l'arrière plan magenta) et l'inverse (bleu sur l'arrière plan vert, magenta sur l'arrière plan jaune) ne peuvent pas être détectés. Mais, l'introduction de la couleur dans les votes de la symétrie radiale apporte une amélioration significative et permet de détecter les centres de tous les cercles en question. Cette approche a été introduite récemment par Heidemann et cela pour calculer les scores de symétrie réflexive. Nous l'introduisons ici pour calculer les scores de symétrie radiale. Il est clair que le gain apporté est assez bon.

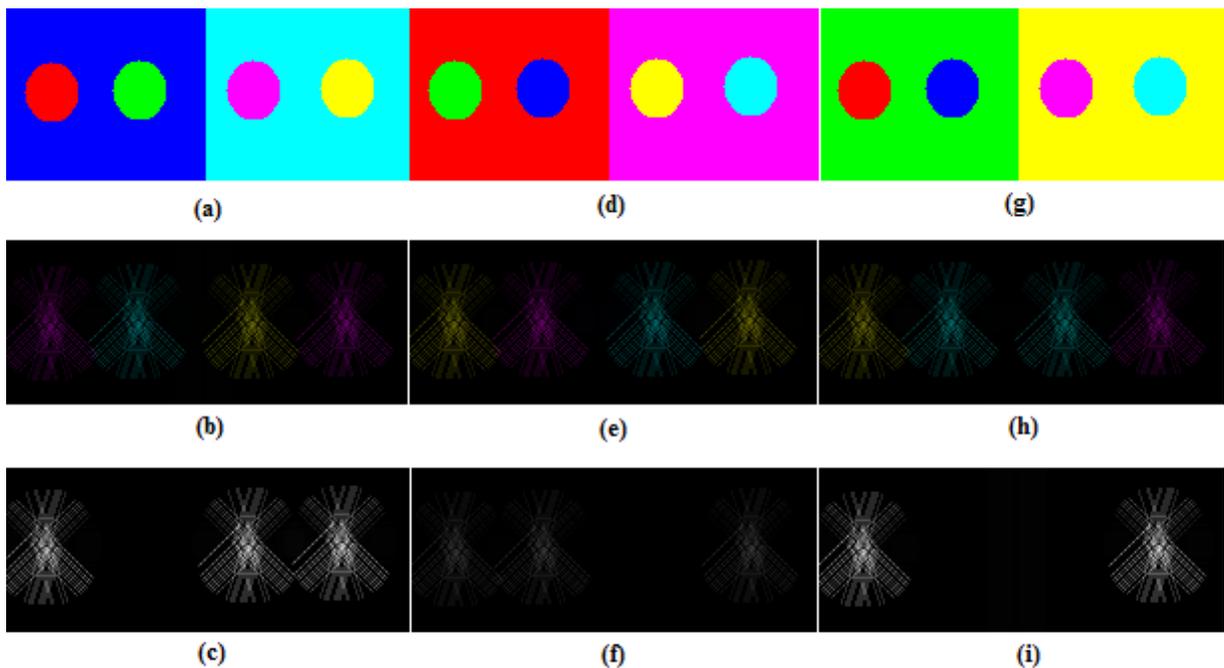


Figure III-21. Résultats des votes en niveaux de gris et en couleur pour des objets qui ne se différencient pas de leur arrière plan en niveaux de gris. (Les tests pour (a, d, g) sont réalisés avec un $R_{max}=15$, $\theta=0$)

6.2. Relation Temps de calcul /Rayon maximal

De cette expérimentation, il est clair que le temps de calcul est facteur du rayon maximal ce qui nécessite l'utilisation de rayons maximaux de petite taille (tableau II.1).

Rayon maximal	Temps de calcul (secondes)
10	1.219
20	2.156
30	3.094
40	4.047
50	4.907
60	5.875
70	6.750
80	7.609
90	8.546
100	9.453

Tableau III-1 Temps de calcul suivant Rmax, ici l'image est de taille 175x144 pixels

Cependant cette utilisation restreint le nombre d'objets détectables dans une image. Une solution consiste à utiliser les pyramides d'images ; c'est une approche qui permet d'avoir presque le même contenu de l'image mais avec une taille plus réduite.

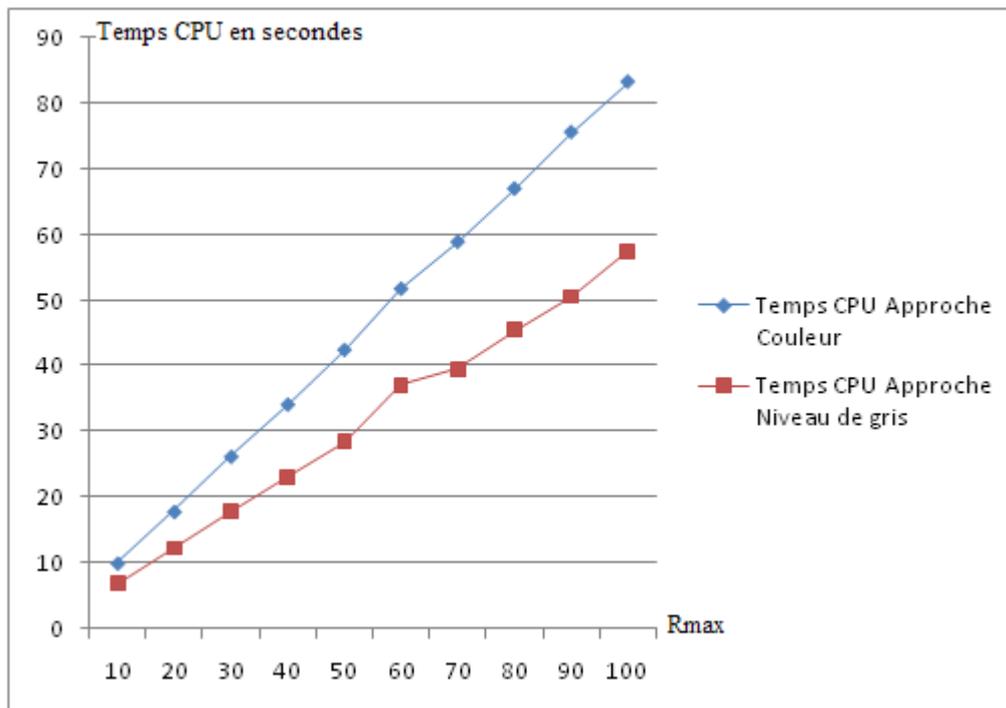


Figure III-22. Variation de temps de calcul suivant différentes valeurs du rayon maximal (Rmax)

Les temps CPU pris par le calcul des matrices de symétrie en couleur pour différentes valeurs R_{max} sont naturellement plus élevés que ceux marqués en calculant les matrices de symétrie en niveaux de gris. Cela est due au nombre d'instructions élevé car pour chaque plan couleur on refait le même ensemble d'instructions faites en niveaux de gris pour détecter les trois symétries (Rouge, Vert et Blue). Après avoir calculé les matrices d'orientation pour chaque canal couleur une sommation de ces matrices donnera la matrice d'orientation finale. A partir de cette matrice nous procédons au calcul de la matrice de symétrie. Cela explique la différence en temps de calcul entre les deux approches. Toutefois, il est possible de remarquer que les deux courbes ne sont pas trop éloignées ce qui fait que l'approche couleur est convenable pour des applications avec des contraintes temporelles pas trop fermes.

6.3. Écartement des scores de symétrie

Un autre but de l'approche proposée est d'améliorer les scores et approfondir l'intervalle entre les points représentant des centres de symétrie et les points usuels de l'image. La Figure III-23 présente une comparaison entre la matrice d'orientation calculée à partir des votes faits par les gradients en niveaux de gris et celle calculée à partir des votes faits par les gradients couleur. Dans cette expérimentation, nous avons calculé le nombre de points d'intérêt (les points ayant un grand score de symétrie) qui se trouvent entre la valeur maximale de la matrice d'orientation et cette même valeur moins une distance D , i.e. le nombre de points entre le Max et le $(Max - D)$.

Nous avons fait varier cette distance et nous avons calculé le nombre de points présents sur les deux matrices d'orientation en niveaux de gris et en couleur et nous avons tracé les courbes sous-jacentes sur le graphique ci-dessous. Il est clair que la matrice d'orientation en couleur a des valeurs très distinctes alors que les valeurs de la matrice en niveaux de gris sont très rapprochées ; c'est-à-dire il y en a beaucoup de points d'intérêt qui peuvent être choisis. Cependant, en utilisant la couleur ce nombre de points d'intérêt sera plus petit ce qui prouve que le taux de discrimination de la matrice en couleur est beaucoup plus grand que celui de la matrice en niveaux de gris.

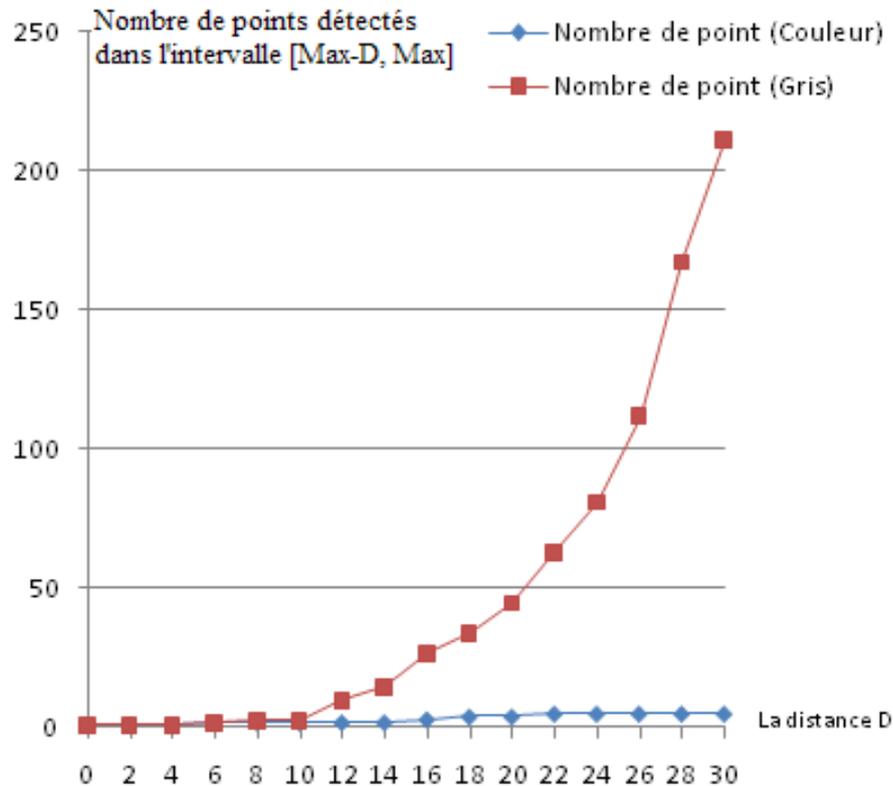


Figure III-23. Variation de nombre de points détectés entre la valeur maximale et la valeur maximale-la distance D

6.4. Evaluation face aux transformations de l'image

L'évaluation de cette approche nécessite des séquences d'images prises dans des situations très précises et cela pour toutes les transformations d'images connues. Dans le présent travail, notre but est une évaluation préliminaire qui donne plus d'information sur l'approche et ses besoins spécifiques en images de test et paramètres d'évaluation.

a) Changement d'illumination

Pour évaluer la performance de l'approche face à un changement de luminosité, nous avons utilisé un ensemble d'images prises sous différentes conditions d'illumination, et nous avons calculé les taux de répétabilité avec chaque image. L'image de référence est une image qui est ni trop sombre ni trop claire. Les autres images sont sélectionnées dans l'intervalle entre l'image trop sombre et celle trop claire (Figure III-24).

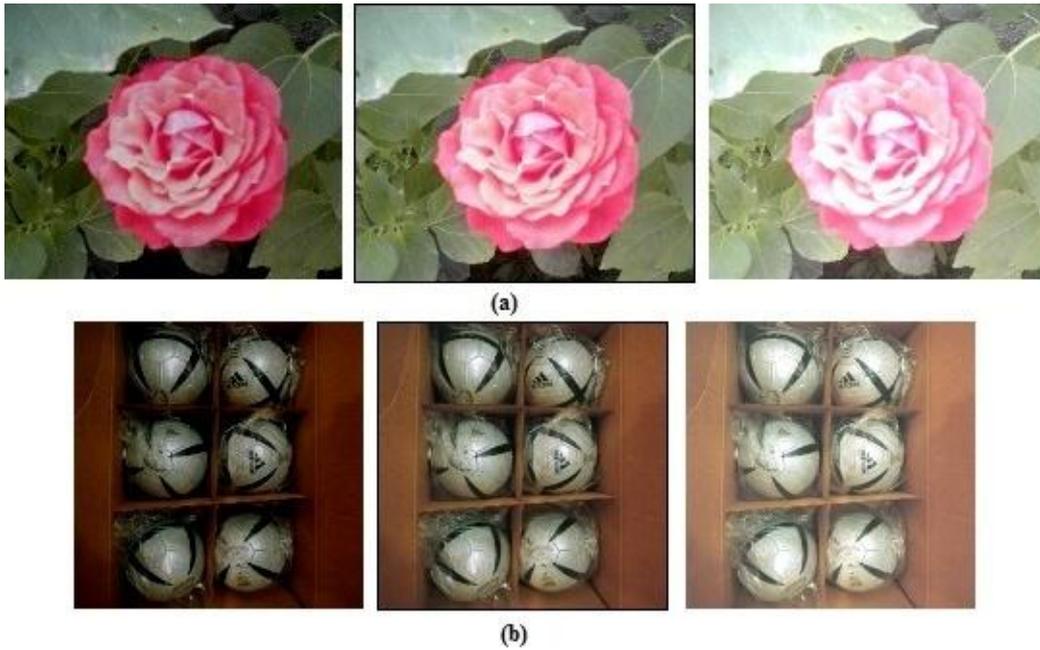


Figure III-24. Exemples d'images sous différentes conditions d'illumination (a) séquence d'images de la scène « rose », (b) séquence d'images de la scène « ballons ».

L'image encadrée est l'image de référence, celle à gauche est la plus sombre et la plus à droite est la plus claire. Nous avons utilisé une séquence de 10 images. Le graphe de la Figure III-25 présente les résultats obtenus. Ces résultats sont ceux des deux séquences. Pour la séquence 1, 100 points d'intérêt sont sélectionnés dans l'image de référence par un rayon maximal = 40 et $\theta = 0$. Pour la séquence 2, 6 points d'intérêt sont sélectionnés dans l'image de référence par un rayon maximal = 20 et $\theta = \pi/45$.

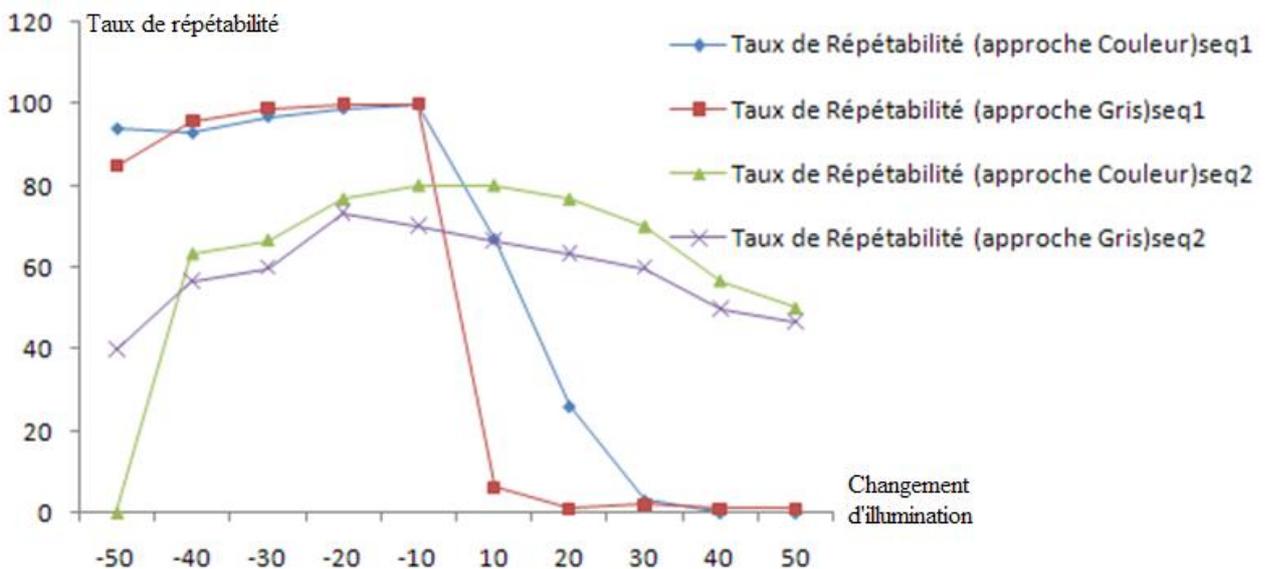


Figure III-25. Variation de taux de répétabilité suivant le changement d'illumination

Les résultats obtenus montrent que l'approche proposée donne de meilleurs résultats comparée avec l'approche en niveau de gris (sauf pour les quatre images après la première image de la séquence 1 où les taux de répétabilité sont presque équivalents).

b) Bruit

Pour évaluer notre approche face à cette transformation image, nous avons procédé de la même façon que pour le changement d'illumination. Nous avons choisi un ensemble d'images qui présente un bruit mais avec des taux différents. La Figure III-26 donne quelques images avec bruit ; le bruit introduit est un bruit gaussien. L'image au milieu est l'image de référence, celle à gauche est une image avec de faible bruit, l'image à droite est une image avec un taux important de bruit. Nous avons utilisé une séquence de 10 images.

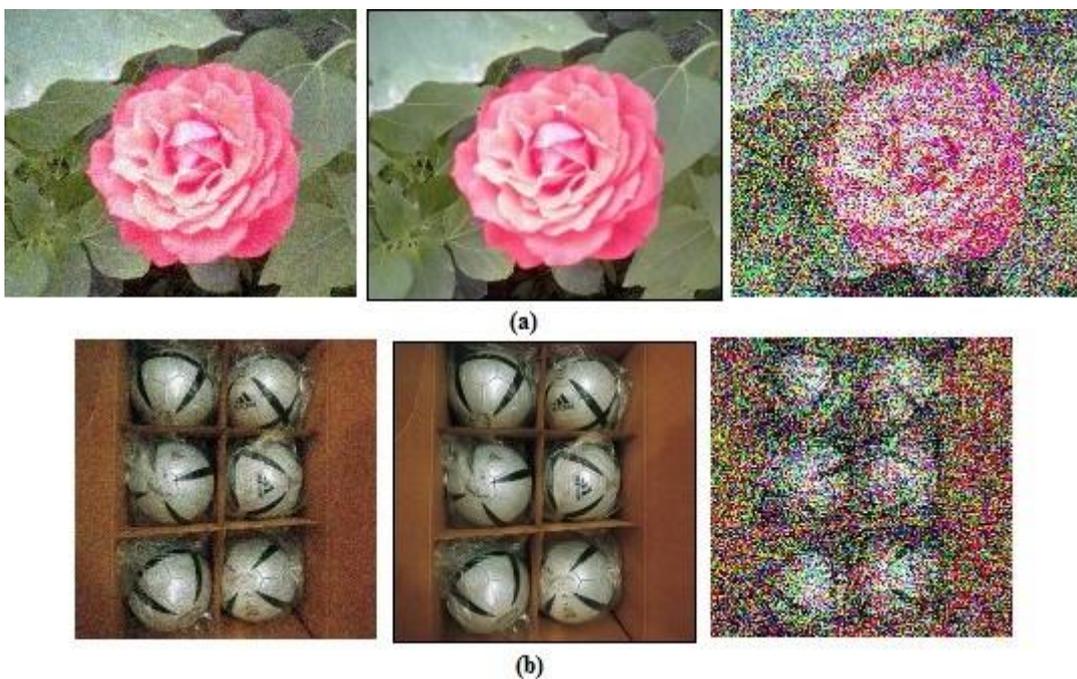


Figure III-26. Exemples d'images sous différents taux de bruit. (a) séquence d'images de la scène « rose », (b) séquence d'images de la scène « ballons ».

Les résultats de cette expérience sont représentés par le graphe de la Figure III-27. Nous avons calculé pour chaque image des deux séquences le taux de répétabilité par rapport à l'image de référence. Les expériences sont faites avec 100 points d'intérêt pour un rayon maximal = 40 et $\theta = 0$ pour la séquence 1 et avec 7 points d'intérêt pour un rayon maximal = 20 et $\theta = \pi/45$. Les résultats obtenus par cette expérience montrent que l'approche proposée donne des taux de répétabilité significativement meilleurs que ceux donnés par l'approche en niveau de gris.

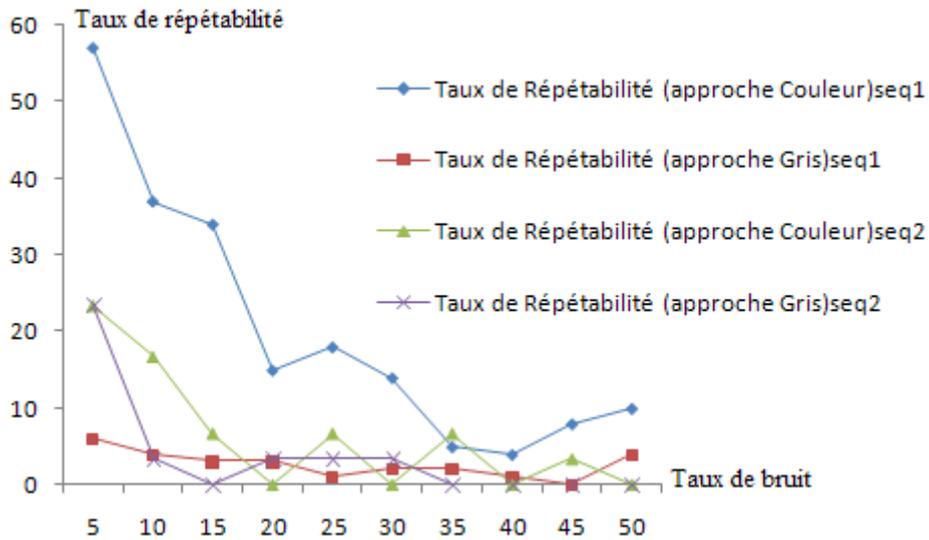


Figure III-27. Variation de taux de répétabilité face au bruit

c) Occultation

Pour ce type de transformation d'images, les images de test sont des images dont certaines parties de l'image de référence sont absentes. Nous avons calculé le taux de répétabilité de chaque image par rapport à l'image de référence. Nous avons utilisé une séquence de 6 images. Des exemples de sequences d'images utilisées sont données par la Figure III-28.

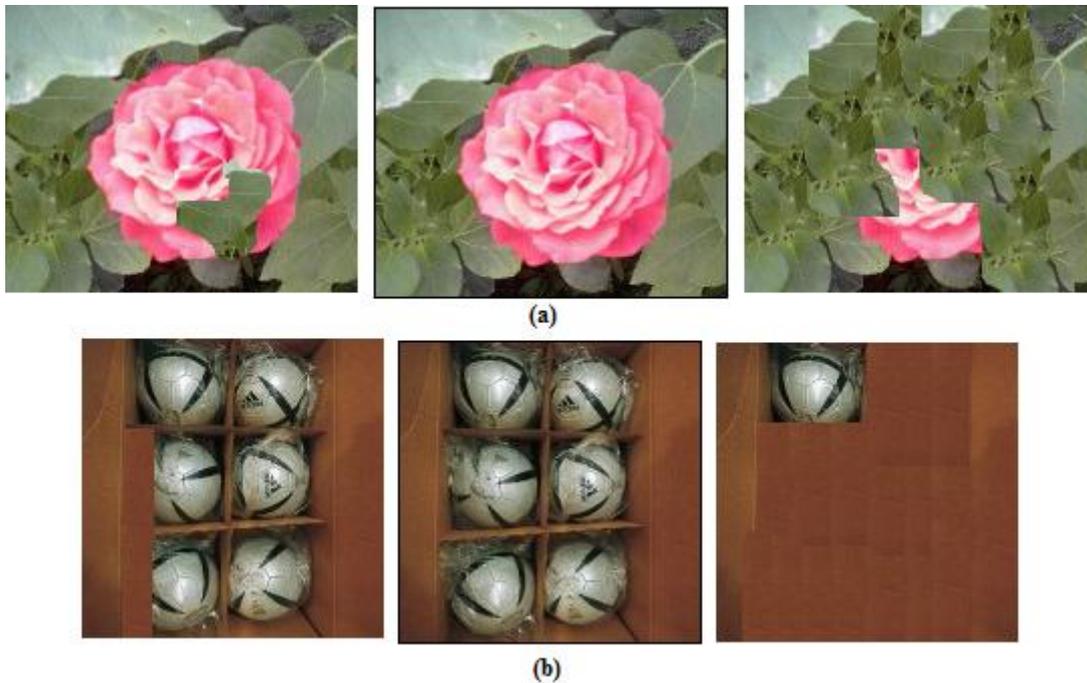


Figure III-28. Exemples d'images occultées. (a) séquence d'images de la scene « rose », (b) séquence d'images de la scene « ballons ».

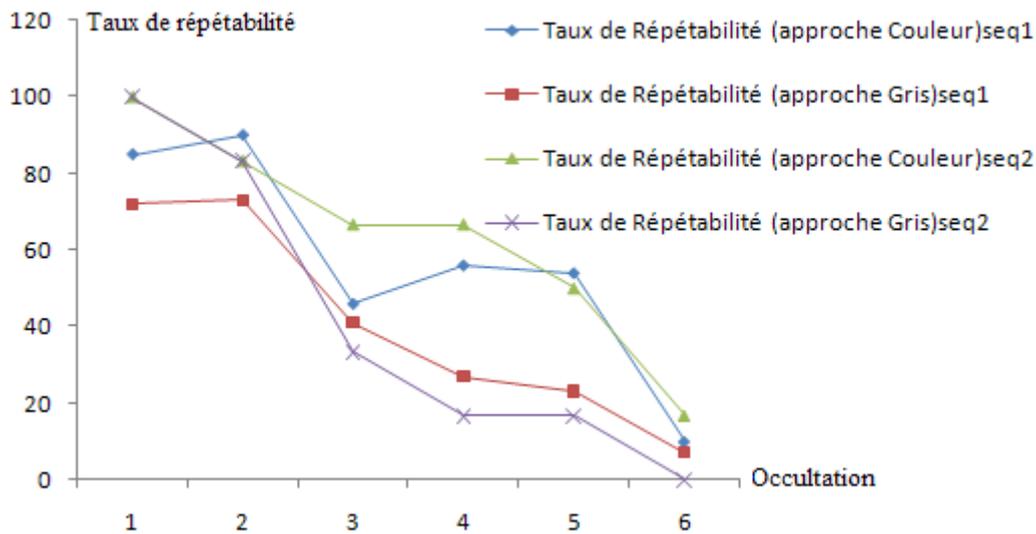


Figure III-29. Variation de taux de répétabilité face à l'occultation

Nous calculons pour chaque image des deux séquences le taux de répétabilité par rapport à l'image de référence. Les expériences sont faites avec 100 points d'intérêt pour un rayon maximal = 40 et $\theta = 0$ pour la séquence 1 et avec 7 points d'intérêt pour un rayon maximal = 20 et $\theta = \pi/45$. Les résultats de cette expérience sur le graphe de la Figure III-29 montrent que l'approche proposée donne de meilleurs résultats face à l'occultation que ceux donnés par l'approche en niveau de gris.

Conclusion

Les expériences présentées dans ce chapitre confirment les hypothèses présentées dans le chapitre précédent. Nos résultats montrent l'apport de la couleur dans les matrices de symétrie. Nous avons montré que les points de focus dans les matrices de symétrie en couleur sont très écartés. Ce qui donne plus de robustesse face aux transformations d'images. Nous avons donné également, des exemples de symétrie qui ne peut pas être détectée par la simple utilisation des matrices en niveau de gris. Nous avons comparé nos résultats avec des résultats donnés par l'approche en niveau de gris. D'autres expériences donnent une évaluation de l'approche proposée face à plusieurs transformations images pour tester sa robustesse exprimée par le taux de répétabilité des points détectés à travers les différentes images des séquences. Ces résultats et cette évaluation confirment une nouvelle fois la richesse de l'approche proposée par rapport à celle en niveau de gris.

Conclusion et Perspectives

Conclusion

A travers ce mémoire nous proposons un nouveau détecteur de points de symétrie pour la reconnaissance d'objets. Ce détecteur permet de surmonter certaines limitations des détecteurs existants. L'originalité de notre travail est l'exploitation de la richesse de l'information couleur à travers la symétrie radiale pour détecter des centres de symétrie discriminants et robustes face aux transformations usuelles de l'image. Le détecteur proposé se base sur le gradient multi-spectral pour tirer profit de la richesse de l'information couleur. Par conséquent, les points détectés sont plus discriminants. Il permet de détecter les centres de symétrie des objets qui ne se différencient de l'arrière-plan qu'en couleur, ces objets ne pouvant pas être détectés avec la simple utilisation du gradient en niveau de gris.

Le travail propose aussi une méthode pour le calcul de l'angle θ introduit précédemment par Rebai dans sa généralisation du détecteur de Zelinsky. Cet angle θ sert de paramètre caractérisant le point détecté. Il est utilisé comme caractéristique géométrique pour construire des descripteurs adaptés aux points de symétrie qui sont invariants et robustes. Auparavant cet angle était choisi arbitrairement mais en utilisant l'approche proposée il sera calculé ceci permettra plus de précision dans la caractérisation en s'affranchissant du critère stochastique du choix de l'angle θ .

Perspectives

Une première amélioration consiste à tester le détecteur proposé dans un processus d'appariement afin de mieux situer les avantages et les inconvénients de ce détecteur par rapport aux détecteurs existants. Cela peut donner des éclaircissements sur les améliorations directes qui peuvent être portées sur le détecteur couleur.

D'autres améliorations possibles peuvent être :

Une étude comparative approfondie des détecteurs en niveau de gris et en couleur

Dans ce travail, nous avons mené une étude comparative entre le nouveau détecteur et le détecteur en niveau de gris proposé par Rebai. Cependant, une étude comparative approfondie du détecteur proposé avec des détecteurs existants comme celui de Zelinsky, Heidemann, Reisfeld et Harris sera nécessaire pour valider l'hypothèse de la richesse de l'information couleur dans un processus de reconnaissance d'objets.

Utilisation d'autres espaces couleur

Comme nous avons vu dans le dernier chapitre de ce mémoire, le choix de l'espace couleur affecte significativement les matrices d'orientation résultantes et par conséquent les centres de symétrie. Le test de plusieurs espaces couleur permet de choisir l'espace couleur le plus adéquat.

Evaluation expérimentale de l'approche proposée pour le calcul de l'angle θ

Nous avons proposé une méthode pour le calcul de l'angle θ , cette méthode doit être vérifiée expérimentalement en l'introduisant dans la caractérisation des points détectés. Et en vérifiant ensuite la robustesse de cette caractérisation avec ce choix d'angle θ .

Introduction du Multi-échelle

L'extension à la multi-échelle de l'approche proposée est une perspective très intéressante. L'utilisation de plusieurs facteurs d'échelle dans le calcul des points de focus par l'approche proposée permet l'obtention d'une nouvelle approche robuste face aux changements d'échelle.

Reconnaissance d'objets

Il est intéressant de procéder à l'intégration du détecteur proposé dans un processus complet de reconnaissance d'objets. Cette intégration permettra une meilleure évaluation du détecteur face aux différentes transformations d'images entre autres les transformations d'images complexes et combinées.

Bibliographie

- [Bello, 1996] F. Bello and R.I. Kitney, "Co-occurrence based texture analysis using irregular tessellations", *In Proceedings of IEEE International Conference on Pattern Recognition*, Vienna, Austria, August 1996, pp. 780–784.
- [Besag, 1973] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *J. Royal Stat. Soc. B* 36, 192–236 (1973)
- [Boyer, 2002] E. Boyer, P. Sturm, "Synthèse d'images géographiques", Hermès Science Publications, pp. 57-89, 2002.
- [Burn, 1990] J. B. Burns, R. Weiss, E. M. Riseman, "View variation of point set and line segment features". *In Proceedings of Darpa Image Understanding Workshop*, pp. 650-659, 1990.
- [Castelli, 2002] V. Castelli L. D. Bergman "Image Databases: Search and Retrieval of Digital Imagery". Wiley, pp. 261-372, 2002.
- [Daubechies, 1990] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis", *IEEE Trans. Inf. Theory* 36, 961–1005 (1990).
- [Dunn, 1994] D. Dunn, W.E. Higgins, and J. Wakeley, "Texture segmentation using 2-D Gabor elementary functions", *IEEE Trans. Pattern Anal. Machine Intell.* 16, 130–149 (1994).
- [Eakins, 1996] J.P. Eakins, K. Shields, and J. Boardman, "ARTISAN—a shape retrieval system based on boundary family indexing", *Proc. SPIE* 2670, 17–28 (1996).
- [Ebner, 1998] M. Ebner, "On the Evolution of Interest Operators using Genetic Programming", 1998.
- [Geman, 2006] D. Geman, "*Interactive Image Retrieval by Mental Matching*", in *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, Santa Barbara, California, USA , 2006.
- [Geusebroek, 2005] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, The Amsterdam library of object images, *International Journal on Computer Vision*, 61(1), pp. 103-112, January, 2005.

- [Goshtasby, 2005] A. Goshtasby, "2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications". Wiley, pp. 46-192, 2005.
- [Gouet, 2000] V. Gouet, "Mise en correspondance d'images en couleur application à la synthèse de vues intermédiaires". Thèse de Doctorat, Université de Montpellier II, 2000.
- [Gouet, 2001] V. Gouet, N. Boujemaa, "Object-Based Queries Using Color Points of Interest". *IEEE Workshop on Content-Based Access of Image and Video Libraries (CVPR/CBAIVL 2001)*, pages 30-36, Hawaii, USA, 2001.
- [Harris, 1988] C. Harris, M. Stephens, "A Combined Corner and Edge Detector". *In Proceeding of Forth Alvey Vision Conf*, pp. 147-151, 1988.
- [Heidemann, 2004] G. Heidemann, "Focus of Attention from Local Color Symmetries". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, N° 7, pp. 817-830, 2004.
- [Hu, 1961] M. K. Hu, "Visual Pattern Recognition by Moments Invariants". *IRE Transactions on Information Theory*, pp. 179-187, 1961.
- [Hugo, 2006] C. Hugo, L. Gareth, "Detecting Rotational Symmetry under Affine Projection". *In proceedings of 18th International conference on Pattern Recognition*, 2006.
- [Ivanov, 2004] Y. Ivanov, "Pattern Recognition for Vision". Massachusetts online courses, Class VII, Part I- Techniques for Clustering, 2004.
- [Julesz, 1983] B. Julesz and J.R. Bergen, "Textons, the fundamental elements in preattentive vision and perception of textures", *Bell Syst. Tech. J.* 62(6), 1619-1645 (1983).
- [Kanjanawanishkul, 2003] K. Kanjanawanishkul, B. Uyyanonvara, "Fast Adaptive Algorithm for Time- Critical Color Quantization Application", *In Proceeding of VIIth Digital Image Computing: Techniques and Applications*, Sydney pp. 781-785, 2003.
- [Kiryati, 1996] N. Kiryati, Y. Gofman, "Detecting Symmetry in Grey Level Images: The Global Optimization Approach", *in proceeding of the 13th International Conference on Pattern Recognition, Vienna, Australia*, Vol. 1, pp. 889-894, 1996.

- [Lakehal, 2009] E. Lakehal, "A Swarm Based Approach for Image Feature Extraction", in proceeding of the IEEE International Conference on Multimedia Computing and Systems, Ouarzazate, Morocco, 2009.
- [Lan, 1997] Z. D. Lan, R. Mohr, "Non-Parametric Invariants and Application to Matching", rapport technique N° 3246, INRIA, 1997.
- [Lew, 2001] M. S. Lew, "Principles of Visual Information Retrieval". Springer-Verlag London Berlin Heidelberg, pp. 3-9, 1998.
- [Lowe, 1999] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features". *In Proceedings of the International Conference on Computer Vision (ICCV)*, Corfu, pp.1-8, 1999.
- [Mikolajczyk, 2001] K. Mikolajczyk, C. Schmid, "Indexing based on Scale Invariant Interest Points". *In Proceedings of Eighth IEEE International Conference on Computer Vision (ICCV'01)*, pp. 525-535, Vancouver, Canada, 2001.
- [Mikolajczyk, 2004] K. Mikolajczyk, C. Schmid, "Scale and Affine Invariant Interest Point Detectors". *International Journal of Computer Vision*, pp. 63-86, 2004.
- [Mikolajczyk, 2005] K. Mikolajczyk, C. Schmid, "A Performance Evaluation of Local Descriptors". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1615-1630, 2005.
- [Mikolajczyk, 2006] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky and T. Kadir, L. Van gool, "A Comparison of Affine Region Detectors". *International Journal of Computer Vision*, pp. 43-72, 2006.
- [Moravec, 1977] H. P. Moravec, "Towards Automatic Visual Obstacle Avoidance". *In Proceeding of Fifth International Joint Conference on Artificial Intelligence*, Cambridge Massachusetts USA, pp. 584-587, 1977.
- [Nene, 1996] S. A. Nene, S. K. Nayar, H. Murase, "Columbia Object Image Library (COIL-100)", Technical Report CUCS-006-96, February 1996.
- [Nieman, 1990] H. Nieman, *Pattern Analysis and Understanding*, 2nd ed., Springer Series in Information Sciences, Springer-Verlag, 1990.
- [Nixon, 2002] M. S. Nixon, A. S. Aguado, "Feature Extraction and Image Processing". Newnes, pp. 67-68, 2002.

- [Philips, 2000] D. Philips, "Image Processing in C: Second Edition". R&D Publications, pp. 103, 2000.
- [Qian, 2000] Richard. J. Qian Peter J. L. van Beek, M. Ibrahim Sezan, "Image Retrieval Using Blob Histograms", *IEEE Conference on Multimedia and Expo. ICME2000*, pp. 125-128, NY, USA, 2000.
- [Rebai, 2006a] A. Rebai, "Etude et intégration d'un nouveau détecteur de points d'intérêt pour la recherche d'images par le contenu visuel". Thèse d'ingénieur, Ecole Supérieure des communications de Tunis, 2006.
- [Rebai, 2006b] A. Rebai, A. Joly, N. Boujemaa, "Constant Tangential Angle Elected Interest Points". In *MIR'06: Proceeding of the 8th ACM International Workshop on Multimedia Information Retrieval*, ACM Press, pp. 203-211, New York, 2006.
- [Rebai, 2007] A. Rebai, A. Joly, N. Boujemaa, "Interpretability Based Interest Points Detection". In *CVIR'07: Computer Intelligence and Virtual Reality*, ACM Press, pp. 33-40. Amsterdam, 2007.
- [Reisfeld, 1995] D. Reisfeld, H. Wolfson, Y. Yeshurun, "Context Free Attentional Operators: the Generalized Symmetry Transform". *International Journal of Computer Vision*, Vol. 14, pp. 119-130, 1995.
- [Russ, 1998] J. C. Russ, "The Image Processing Handbook: Third Edition". CRC Press, pp. 47-373, 1998.
- [Schmid, 1996] C. Schmid, "Appariement D'images par Invariants Locaux de Niveaux de Gris Application à L'indexation d'une Base D'objets". Thèse de Doctorat, Institut National Polytechnique de Grenoble, 1996.
- [Schmid, 1996] C. Schmid, "Appariement D'images par Invariants Locaux de Niveaux de Gris Application à L'indexation d'une Base D'objets". Thèse de Doctorat, Institut National Polytechnique de Grenoble, 1996.
- [Schmid, 1997] C. Schmid, R. Mohr, "Local Grayvalue Invariants for Image Retrieval". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, N°5, pp. 530-535, 1997.
- [Schmid, 2000] C. Schmid, R. Mohr, C. Bauckhage, "Evaluation of Interest Point Detectors". *International Journal of Computer Vision*, 2000.

- [Smith, 1997] J.R. Smith and S.-F. Chang, SaFe: "A general framework for integrated spatial and feature image search", *IEEE 1st Multimedia Signal Processing Workshop, June 1997*.
- [Smith, 1997] S. Smith, J. Brady, "A New Approach to Low Level Image Processing". *In International Journal of Computer Vision*, pp. 45-78, 1997.
- [Swain, 1991] M.J. Swain and D.H. Ballard, Color indexing, *International Journal of Computer Visions*, 7(1) (1991).
- [Tamura, 1978] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception", *IEEE Trans. Sys. Man, and Cybernetics* 8(6), 460–473 (1978).
- [Trujillo, 2006a] L. Trujillo, G. Olague, "Synthesis of Interest Point Detectors Through Genetic Programming". *In Proceedings of GECCO'0 Seattle, Washington, USA*, pp. 887-894, 2006.
- [Trujillo, 2006b] L. Trujillo, G. Olague, "Using evolution to learn how to perform Interest point detection". *In Proceedings of 18th International Conference on Pattern Recognition (ICPR'06)*, 2006.
- [Trujillo, 2008] L. Trujillo, G. Olague, E. Lutton, F. Fernández, "Multiobjective Design of Operators that Detect Points of Interest in Images". *In Proceedings of GECCO'08, Atlanta, Georgia, USA*, pp. 1299-1306, 2006.
- [Tuytelaars, 2006] T. Tuytelaars, K. Mikolajczyk, "A Survey on Local Invariant Features", 2006.
- [Weijer, 2005] J. Weijer, T. Gevers, J. M. Geusebroek, "Edge and Corner Detection by Photometric Quasi-Invariants". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, N° 4, pp.625-630, 2005.
- [Wu, 1995] X. Wu, B. Bhanu, "Gabor Wavelets for 3-D Object Recognition". *In proceeding of the 5th International Conference on computer*, Vol. 20, N° 23, pp. 537-542, 1995.
- [Zelinsky, 2003] A. Zelinsky, G. Loy, "Fast Radial Symmetry for Detecting Point of Interest". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, N° 8, pp. 959-973, 2003.
- [Zhang, 2006] Y. J. Zhang, "Advances in Image and Video Segmentation". IRM Press, pp. 47-373, 2006.

Résumé

Ce mémoire, inscrit dans le domaine de la reconnaissance d'objets et plus particulièrement les descripteurs locaux, présente un nouveau détecteur des points d'intérêt, que nous appellerons points de focus. La conclusion à laquelle ont abouti plusieurs études psychologiques est que la symétrie attire l'attention. En conséquence, la détection des centres de symétrie donne des points ayant une grande importance par rapport au système visuel humain et sont donc plus appropriés à la reconnaissance d'objets. L'objectif de beaucoup de travaux est la recherche des points d'intérêt dans les fortes variations du signal. Malheureusement, en l'absence de ces variations les approches s'avèrent inefficaces. Par contre, l'utilisation des centres de symétrie se présente comme une solution intéressante. Le but de l'exploitation de ce nouveau détecteur est l'extraction de primitives de focus qui ont une signification visuelle. En prenant comme point de départ les travaux cités dans la littérature, nous apportons une amélioration consistant en l'utilisation de la couleur pour enrichir la détection et la robustesse des points détectés.

