

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITE MOULOUD MAMMERI, TIZI-OUZOU



FACULTE DE GENIE ELECTRIQUE ET DE L'INFORMATIQUE
DEPARTEMENT D'ELECTRONIQUE

Mémoire de Master recherche

*En vue de l'obtention du diplôme de Master recherche en
électronique,*

Option : Télécommunication et réseaux

Thème

***Détection de points d'intérêts sur les images
omnidirectionnelles***

Proposé par:

M^{me}. DJAMAA Kachi

Dirigé par :

M^{me}. Ameer Zohra

Réalisé par :

M^r. BELDI Rabah

Promotion : 2009/2010

Remerciements

Je tiens tout d'abord à remercier ma promotrice M^{me} Ameur Zohra de m'avoir fait confiance en acceptant de m'encadrer. Elle a su me soutenir, me motiver et m'encourager tout au long de ce travail.

Je tiens à exprimer ma reconnaissance à M^{me} Djemaa Kachi maitre de conférences à l'université Jules verne d'AMIENS pour son aide précieuse et pour avoir mis à notre disposition les images omnidirectionnelles utilisées dans ce mémoire.

Mes remerciements vont également vers M^r Lahdir Mourad, Mr Attaf Youcef et M^r Bitam Abdelmadjid pour le soin avec lequel ils ont lu ce manuscrit et d'avoir accepté d'être dans mon jury.

Je remercie également M^r Kannane Hocine de m'avoir motivé, soutenu et encouragé tout au long de ce travail.

Pour finir, je ne saurais exprimer, à quel point je remercie ma famille pour son soutien et ses encouragements depuis toujours. Ce sont mes parents qui m'ont appris à donner le meilleur et c'est cette exigence de soi qui m'a aidée tout au long de mes études.

Mes amis sont également à remercier, pour leur soutien inconditionnel.

Je dédie ce travail à ma famille, qu'elle soit de sang ou de cœur.

Sommaire

Table des matières

Table des figures

Introduction générale.....	1
Chapitre I Généralités sur le traitement d'image	4
Introduction	4
I.1 Définition d'une image	4
I.1.1 Image comme ensemble de points.....	4
I.1.2 Image comme signal bidimensionnel.....	4
I.1.2.1 Image binaire.....	5
I.1.2.2 Image en niveau de gris.....	5
I.1.2.3 Image couleur.....	5
I.2 Caractéristiques d'une image numérique.....	6
I.2.1 Dimension.....	6
I.2.2 Résolution.....	6
I.2.3 Luminance.....	6
I.2.4 Contraste.....	7
I.2.5 Contours et textures.....	7
I.3 Histogramme.....	7
I.4 Traitement d'images.....	8
I.4.1 Organigramme général de traitement d'image.....	8
I.4.1.1 Organigramme général de traitement d'image.....	8
I.4.1.2 Prétaitement.....	9
I.4.1.3 Traitement numérique des images.....	9
I.4.1.4 Visualisation.....	9
I.4.2 Techniques d'amélioration d'image.....	10
I.4.2.1 Filtrage.....	10
I.4.2.1.1 Filtrage linéaire.....	10

I.1.2.1.2 Filtres adaptatifs.....	10
I.4.2.1.3 Filtres morphologiques.....	11
I.4.2.2 Rehaussement.....	11
I.4.2.3 Binarisation.....	11
I.4.2.4 Notion de segmentation.....	11
I.5 Détection de caractéristiques.....	12
I.5.1 Détection de contours.....	12
I.5.2 Détection de points d'intérêts.....	12
I.5.2.1 Différentes approches.....	13
I.5.2.2 Le détecteur de Moravec.....	13
I.5.2.3 Le détecteur de harris.....	14
Conclusion.....	16
Chapitre II Vision omnidirectionnelle	17
II.1 capteurs omnidirectionnels.....	17
II.1.1 Images multiples.....	17
II.1.1.a Génération d'images panoramiques par rotation d'une camera.....	17
II.1.1.b Génération d'images panoramiques à partir de plusieurs caméras...	17
II.1.2 Lentilles spéciales.....	18
II.1.3 Utilisation de miroirs convexes.....	19
II.1.3.1 Notion de centre de projection.....	19
II.1.3.2 Aberrations.....	21
II.1.4 Comparaison des miroirs.....	23
II.1.4.1 Miroir sphérique.....	24
II.1.4.2 Miroir conique.....	25
II.1.4.3 Miroir Hyperbolique.....	26
II.1.4.4 Miroir parabolique.....	27
II.1.4.5 Combinaison des différents approches.....	27

II.2 Modélisation des caméras omnidirectionnelles.....	28
II.3 Calibrage.....	30
II.4 La recherche dans le traitement d'images omnidirectionnelles.....	31
Chapitre III Outils mathématiques sur la sphère équivalente	33
III.1 Projection stéréographique.....	33
III.2 Outils existants sur le plan.....	35
III.3 Action de groupe sur la sphère.....	35
III.4 Harmoniques sphériques.....	37
III.5 Convolution sur la sphère.....	40
Chapitre IV Traitement des images omnidirectionnelles	
Introduction	43
IV.1 Utilisation des traitements classiques.	43
IV.2 Adaptation des traitements.	45
IV.2.1 Utilisation de la géométrie du miroir.	45
IV.2.2 Projection de l'image sur la sphère.	49
IV.3 Généralités sur le traitement des images définies sur la sphère.	52
IV.3.1 Opérateur de filtrage sur la sphère.	52
IV.3.2 Gaussiennes sphériques.	53
IV.3.3 Lissage.	55
IV.3.3.1 Filtre de wiener.	55
IV.3.3.2 Régularisation de Tikhonov.	56
IV.3.4 Détection de contours.	57
IV.3.4.1 Méthode du gradient.	57
IV.3.4.2 Méthode du Laplacien.	60
Chapitre V Détection de points d'intérêts	61
Introduction.	61
V.1 état de l'art.	61

V.1.1 Moravec.	61
V.1.2 Harris et Stephens.	62
V.1.3 Lindeberg, LoG.	65
V.1.4 Harris-Laplace.	67
V.1.5 Autres travaux.	68
V.2 Méthodes d'évaluation	70
V.2.1 Critères d'évaluation.	70
V.2.2 Définition de la répétabilité.	71
V.2.3 Mesure de la répétabilité	71
V.3 Caractérisation locale	72
V.3.1 Méthodes de caractérisation locale.	73
V.3.1.1 Dérivées	73
V.3.1.2 Descriptions fréquentielles	74
V.3.1.3 Moments	76
V.3.1.4 Autres caractérisations	77
V.3.2 Vecteurs de description.	78
V.3.2.1 Voisinage de pixels	78
V.3.2.2 Distribution	80
V.3.2.3 SIFT	81
V.3.3 Introduction aux invariants.	83
V.3.3.1 Définition théorique d'un invariant	83
V.3.3.2 Calcul des invariants.	84
V.3.3.3 Dénombrement des invariants	85
V.3.3.4 Théorème de Burns	85
V.3.3.5 Quasi-invariants	86
V.3.4 Invariance et transformations de l'image	86
V.3.4.1 Rotation image	87

V.3.4.2 Changement d'échelle	89
V.3.4.3 Changement de luminosité	90
V.3.4.4 Changement de point de vue	91
Chapitre VI Tests et résultats	93
Introduction.....	93
VI.1 Généralités.....	93
VI.2 Détecteur de Harris.....	96
VI.3 Détecteur SUSAN.....	100
VI.4 Détecteur Harris Laplace.....	102
VI.5 Détecteur LoG.....	104
Conclusion générale.....	107
Bibliographie	

Tables des figures

Table des figures

I.1	Image binaire.....	5
I.2	Image en niveau de gris.....	5
I.3	Image couleur.....	6
I.4	Images en niveau de gris et son histogramme.....	7
I.5	Synoptique d'un système de traitement d'images.....	8
I.6	Différentes points d'intérêts.....	12
I.7	Différentes situations considérées par le détecteur de moravec.....	13
II.1	Image panoramique.....	17
II.2	Image omnidirectionnelle.....	17
II.3	Image omnidirectionnelle généré par Rotation de caméra.....	18
II.4	Capteurs omnidirectionnelle qui utilisent plusieurs caméra conventionnelles.....	18
II.5	Panoramic Annular Lens (PAL).....	19
II.6	Une caméra «fish-eye».....	19
II.7	Caméra catadioptrique constituée d'un miroir parabolique, d'une lentille télé centrique et d'une caméra classique.....	20
II.8	Réflexion sur un miroir parabolique concave et convexe.....	21
II.9	Aberration chromatique.....	22
II.10	Distorsions radiales.....	22
II.11	Aberration sphérique	23
II.12	Aberration coma.....	23
II.13	Les quatre types principaux de miroirs omnidirectionnels.....	23
II.14	Caméra omnidirectionnelle avec un miroir sphérique.....	24

II.15	Capteur omnidirectionnelle avec miroir sphérique	24
II.16	Caméra omnidirectionnelle avec Miroir conique.....	25
II.17	Capteur omnidirectionnel avec miroir conique.....	25
II.18	Caméra omnidirectionnelle avec miroir hyperbolique.....	26
II.19	Capteur omnidirectionnel avec miroir hyperbolique.....	26
II.20	Combinaison d'un miroir parabolique avec un miroir.....	28
II.21	Schéma générale de formation des images catadioptriques.....	29
II.22	Image omnidirectionnelle d'une mire.....	31
II.23	Schéma de traitement des images omnidirectionnelles.....	32
III.1	Angles de paramétrisation de la sphère.....	33
III.2	Projection stéréographique.....	34
IV.1	Processus de traitement d'une image omnidirectionnelle.....	43
IV.2	Définition des patchs sur le miroir.....	46
IV.3	Image anamorphosée.....	46
IV.4	Reconstruction d'une image perspective.....	47
IV.5	Projection de l'image sur le cylindre englobant.....	48
IV.6	Projection de l'image sur le cylindre englobant.....	48
IV.7	Projection de l'image sur l'hémisphère.....	50
IV.8	Exemple de plan tangent à la sphère.....	58
V.1	Détection de points d'intérêts par la méthode de Harris et Stephens.....	64
V.2	Détection de blobs par la méthode LoG sur l'image « tournesol ».....	67
V.3	26 voisinages utilisé pour la détection des maxima locaux de l'opérateur LoG.....	67
V.4	Définition de répétabilité.....	71
V.5	Fonction d'intensité lumineuse autour d'un point d'intérêt.....	73
V.6	La transformée de Fourier est directionnelle.....	75
V.7	Construction de vecteurs de description.....	79
V.8	Construction de vecteurs de description fondé sur l'estimation de la distribution de l'intensité.....	81

V.9	Illustration du problème intrinsèque à l'utilisation de distributions pour la description de points d'intérêt.....	81
V.10	Construction de la matrice de pixels utilisée pour construire le vecteur de description SIFT.....	82
V.11	Construction du vecteur de description SIFT.....	83
VI.1	Image Test1.....	94
VI.2	Image Test2.....	94
VI.3	Image Porte.....	94
VI.4	Image Tournesol.....	94
VI.5	Image Bureau.....	95
VI.6	Image Mire.....	95
VI.7	Image Radiales.....	95
VI.8	Image Droites.....	95
VI.9	Image Texture11.....	95
VI.10	Image Texture14.....	95
VI.11	Points d'intérêts détectés sur l'image Test1 par la méthode de Harris.....	96
VI.12	Points d'intérêts détectés sur l'image Test2 par la méthode de Harris.....	96
VI.13	Points d'intérêts détectés sur l'image Porte par la méthode de Harris.....	97
VI.14	Points d'intérêts détectés sur l'image Radiales par la méthode de Harris.....	97
VI.15	Points d'intérêts détectés sur l'image Droites par la méthode de Harris.....	98
VI.16	Points d'intérêts détectés sur l'image Mire par la méthode de Harris.....	98
VI.17	Points d'intérêts détectés sur l'image Bureau par la méthode de Harris.....	99
VI.18	Points d'intérêts détectés sur l'image Texture11 par la méthode de Harris.....	99
VI.19	Points d'intérêts détectés sur l'image Test2 par la méthode SUSAN	100
VI.20	Points d'intérêts détectés sur l'image Radiales par la méthode SUSAN	100
VI.21	Points d'intérêts détectés sur l'image Texture11 par la méthode SUSAN	101
VI.22	Points d'intérêts détectés sur l'image Bureau par la méthode SUSAN	101
VI.23	Points d'intérêts par le détecteur de Harris-Laplace sur l'image Porte.....	102
VI.24	Points d'intérêts par le détecteur de Harris-Laplace sur l'image Tournesol.....	102
VI.25	Points d'intérêts par le détecteur de Harris-Laplace sur l'image Bureau.....	103
VI.26	Points d'intérêts par le détecteur de Harris-Laplace sur l'image Texture14.....	103
VI.27	Détection de blobs par la méthode LoG sur l'image Tournesol.....	104
VI.28	Détection de points d'intérêts par le détecteur LoG sur l'image Porte.....	105
VI.29	Détection de points d'intérêts par le détecteur LoG sur l'image Texture11.....	105

VI.30	Détection de points d'intérêts par le détecteur LoG sur l'image Bureau.....	106
--------------	---	-----

Introduction Générale

Le traitement d'images désigne une discipline des mathématiques appliquées qui étudie les images numériques et leur transformation dans le but d'améliorer leur qualité, de réduire leur coût de stockage ou d'en extraire une information sémantique et pertinente. Une image peut tout à fait être traitée comme un signal bidimensionnel dans le cas d'images en niveaux de gris, et comme un signal tridimensionnel dans le cas d'images couleur. C'est pourquoi le traitement d'images est souvent défini comme étant un sous-ensemble du traitement du signal.

Aujourd'hui, les énormes progrès dans le domaine, aidés par l'augmentation de la puissance de calcul des ordinateurs, la miniaturisation des caméras et la diminution des prix du matériel, permettent une utilisation de plus en plus fréquente des systèmes de vision dans des domaines très variés, tels que la vidéo surveillance (détection d'intrusion, d'incendie), les applications militaires (guidage de missiles, repérage d'installations), médicales (détection de tumeurs par IRM), robotiques (perception de l'environnement du robot), spatiales (recherche d'astres), industrielles (métrologie sans contact, inspection), et même ludiques (réalité augmentée, robotique ludique).

Malgré ces différents progrès, le domaine de la vision est encore en perpétuelle évolution. Chaque année, de nouvelles méthodes sont proposées, de nouvelles applications imaginées et de nouveaux capteurs développés, faisant de ce domaine une source intarissable pour les chercheurs. L'une des particularités de la vision est sa pluridisciplinarité. En fait, elle implique un travail de programmation, mais aussi de bonnes connaissances pratiques des mathématiques, tant en algèbre qu'en géométrie et en statistiques. Le traitement du signal, l'Intelligence Artificielle et même l'électronique y sont largement utilisés.

Augmenter le champ de vision, c'est donc imposer avec comme espoir d'obtenir une vision omnidirectionnelle, c'est-à-dire un champ de vision à 360°.

Dans ce but, plusieurs systèmes ont été développés et ce sont les capteurs catadioptriques qui sont les plus utilisés. Ce capteur comporte un miroir dans lequel une caméra observe le reflet de la scène.

La première utilisation d'un tel système date de 1970. Rees [34] y déposa un brevet industriel pour un miroir hyperboloïdal couplé avec une caméra pour un opérateur assis dans une tourelle de tir. Depuis, les recherches sur la conception de tels capteurs ont continué et se sont accrues avec le développement de la robotique mobile et l'utilisation en 1990 d'un miroir conique dans le système COPIS [11] proposé par Yagi. L'avantage de l'omnidirectionnalité

pour la navigation de robots mobiles est évidente puisqu'elle offre un champ de vision à 360° avec seulement une caméra. Par la suite, beaucoup de travaux ont été réalisés sur la modélisation ou le calibrage des capteurs omnidirectionnels. Cependant, les recherches sur le traitement des images omnidirectionnelles ont commencé plus tardivement puisqu'au départ on traitait ces images comme des images perspectives sans prendre en considération les distorsions introduites par les miroirs. Nous présenterons une méthode qui consiste à utiliser un espace intermédiaire pour effectuer les traitements. Cet espace, appelé "sphère d'équivalence" permet un traitement uniforme quelque soit l'emplacement de l'opérateur de traitement sur la sphère.

La détection de points d'intérêt constitue un problème central en vision par ordinateur, de par ses diverses applications, comme la reconnaissance d'objets, l'appariement d'images, le suivi de mouvement ou le rehaussement de certains éléments particuliers. Cette étape de détection est essentielle dans la mesure où elle constitue une première étape avant d'autres plus complexes, comme la détermination de déformations géométriques ou l'extraction de caractéristiques invariantes.

Dans cette optique, des approches multi-échelles ont été introduites et plus particulièrement celles impliquant une certaine redondance, de manière à ce que les éléments les plus remarquables soient présents à différentes échelles. Parmi les différentes théories permettant de formaliser de telles approches, la théorie du Scale-Space, introduite par Witkin [114] puis approfondie par Lindeberg [115], définit des méthodes de détection automatique, qui prennent en compte certaines transformations de l'image, notamment le changement d'échelle et les déformations affines. Ceci conduit alors au concept d'invariance à une certaine transformation (d'où le terme d'invariance à l'échelle par exemple) : soit que les points extraits soient invariants à celle-ci, soit qu'elles la prennent en compte (l'invariance étant atteinte par la suite).

Dans ces méthodes classiques de détection, l'image est d'abord lissée, typiquement par un noyau Gaussien, à différentes échelles, avant d'appliquer un opérateur adapté [114]. Une fois cette opération effectuée, la sélection de certains maxima permet de mettre en évidence différentes structures plus ou moins complexes présentes dans l'image. Naturellement, la diversité et la complexité de celles-ci rendent d'autant plus difficile la formulation d'un détecteur efficace de points d'intérêt.

De nombreuses méthodes ont été proposées afin d'extraire les points les plus pertinents d'une image (algorithmes de détection) puis de calculer des grandeurs caractérisant le contenu de ces points (calcul de descripteurs) [3, 58, 92]. Lorsqu'il s'agit d'extraire de telles caractéristiques d'une image, il importe qu'elles soient invariantes (ou au moins robustes) à certaines transformations de l'image. Ces dernières pouvant être géométriques (rotations, déformations affines) ou autres (changement de luminosité ou de contraste, image bruitée). Une application majeure de la théorie des invariants a été de rendre ces détecteurs invariants à ces transformations (comme les déformations affines), et un important travail d'amélioration de ces algorithmes a permis d'obtenir des performances accrues. Notons néanmoins que la nature des objets détectés n'est pas toujours claire, et en particulier, il est difficile de caractériser les points détectés.

Dans un premier temps, nous définirons la vision par ordinateur, les différentes disciplines et méthodes de traitements.

Dans le deuxième chapitre, seront décrits les différents capteurs produisant des images omnidirectionnelles.

Le troisième chapitre présentera les différents outils mathématiques qui seront nécessaires pour la construction des opérateurs sur la sphère.

Dans le quatrième chapitre, Nous donnerons brièvement quelques éléments sur l'état de la recherche dans le traitement des images omnidirectionnelles et nous décrirons les différentes méthodes de traitements de ces images.

L'extraction des points d'intérêt, l'avantage de ces points d'intérêt par rapport à d'autres caractéristiques de bas niveau, la comparaison de ces différents détecteurs et les critères de comparaison qui sont la répétabilité et la localisation feront l'objet du chapitre cinq.

Le dernier chapitre sera consacré à nos différents tests et résultats. Dans un premier temps, une comparaison entre les différents détecteurs sera donnée, puis dans un second temps, une comparaison entre les différents résultats obtenus par ces détecteurs sur les images planes et les images omnidirectionnelles sera donnée et commentée.

Notre travail s'achèvera par une conclusion et quelques perspectives ouvertes par ce travail seront posées.

Chapitre I

Généralités sur le traitement d'image

Introduction

Le traitement d'image consiste à améliorer sa qualité en corrigeant les dégradations subies lors de son acquisition ou à faire ressortir les informations permettant une interprétation visuelle ou automatique faisant appel à des techniques destinées à mieux exploiter les informations qu'elle contient.

I.1 Définition d'une image

Une image numérique est définie comme un signal continu bidimensionnel échantillonné à valeur quantifiées dans un certain espace de couleurs. Elle est constituée de points appelés pixel (picture element en anglais). Le pixel est l'unité de base d'une image numérique. C'est le point minimal adressable par le contrôleur vidéo. La définition est tributaire de la résolution, mais également de la taille de l'image souhaitée en sortie. Autrement dit, une image est une matrice $M \times N$ de valeurs entières prises sur un intervalle borné $[0, Ng]$ ou Ng est la valeur maximale du niveau de gris.

Toute image est définie par un ensemble de propriétés indispensables à toute analyse. Les notions de voisinage et de niveau de gris constituent un support essentiel pour une étude quelconque.

I.1.1 Image comme ensemble de points

L'image est considérée comme un mélange de points dont la surface des objets diffère. La valeur de chaque point est la réalisation d'une variable aléatoire liée à l'apparition de la luminance dans un site donné de l'image. Ce pixel ainsi défini véhicule une donnée numérique informationnelle sur l'image.

I.1.2 Image comme signal bidimensionnel

Une image peut être considérée comme étant un signal bidimensionnel variant dans l'espace suivant les directions horizontale X et verticale Y .

L'analyse de ce signal sera effectuée en lui adaptant tous les éléments du traitement de signal à deux démentions. Nous pourrions citer parmi ces traitement les filtres linéaire, permettant d'éliminer le bruit, et la transformée de fourrier qui transforme l'analyse dans le domaine fréquentielle.

Il y a trois formes physiques d'existence d'une image : forme binaire, en niveau de gris et couleur.

I.1.2.1 Image binaire

Une image binaire est une image pour laquelle chaque pixel ne peut avoir pour valeur que 0 ou 1. Les pixels sont noirs (0) ou blancs (1). Le niveau de gris est codé sur un bit, dans ce cas $N_g=2$ et la relation sur les niveaux de gris devient : $p(i, j)=0$ ou $p(i, j)=1$.



Figure I.1-Image binaire

I.1.2.2 Image en niveau de gris

Une image en niveau de gris autorise un dégradé de gris entre le noir et le blanc. En général, on code le niveau de gris sur un octet soit 256 nuances de dégradé. L'expression de la valeur du niveau de gris avec $N_g=256$ devient : $p(i, j)=[0,256]$.



Figure I.2 - Image en niveau de gris

I.1.2.3 Image couleur

Une image couleur contient trois plans de couleur, le rouge, le vert et le bleu (RVB). Chaque plan est codé comme une image en niveau de gris, avec des valeurs allant de 0 à 255. Lorsque $R=V=B$, la valeur associée est un niveau de gris.



Figure I.3 - Image couleur

D'autre part, pour aller d'une image couleur à une image en niveau de gris on réalise la fonction suivante :

$$g(x,y) = \frac{R(x,y)+V(x,y)+B(x,y)}{3} \quad (I.1)$$

I.2 Caractéristiques d'une image numérique

Une image numérique est caractérisée par les paramètres suivant :

I.2.1 Dimension

C'est la taille de l'image. Cette dernière se présente sous forme de matrice dont les éléments sont des valeurs numérique représentative des intensités lumineuses (pixels). Le nombre total de pixels est donné par le produit du nombre de ligne de cette matrice avec le nombre de colonnes.

I.2.2 Résolution

C'est la clarté ou la finesse de détails atteinte lors de la production d'une image. Sur un écran d'ordinateur, la résolution représente le nombre de pixels par unité de mesure (pouce ou centimètre). Elle représente aussi le nombre total de pixels affichables sur un écran ; plus ce nombre est grand mieux est la résolution.

I.2.3 Luminance

C'est le degré de luminosité des points de l'image. Le mot luminance est souvent substitué au mot brillance.

Une bonne luminance doit se caractériser par :

- Un bon contraste : évité les images dont le contraste tend vers le noir ou le blanc ; ces images entraînent des pertes de détails dans les zones sombres ou lumineuses.
- L'absence de parasites.

I.2.4 Contraste

C'est l'opposition marquée entre les régions sombres et les régions claires d'une image. Il est défini en fonction des deux zones d'image.

I.2.5 Contours et textures

Les contours représentent la frontière entre les objets de l'image, ou la limite entre deux pixels dont les niveaux de gris représentent une différence significative. Les textures quant à elles, décrivent la structure des contours. L'extraction de ces derniers consiste à identifier dans l'image les points qui séparent deux textures différentes.

I.3 Histogramme

L'histogramme représente la distribution des fréquences d'apparition de tous les niveaux de gris dans l'image. Selon l'image à considérer, il nous renseigne sur la constitution de celle-ci.

Si l'histogramme est formé de deux modes bien séparés, on dit qu'il est bimodal (deux pics séparés par une vallée) et l'on déduit qu'il existe un objet sur un fond. Un histogramme unimodale représente soit un objet, soit un fond ; par contre un histogramme multimodale (plusieurs pics séparés par des vallées) nous renseigne sur la présence de plusieurs objets.

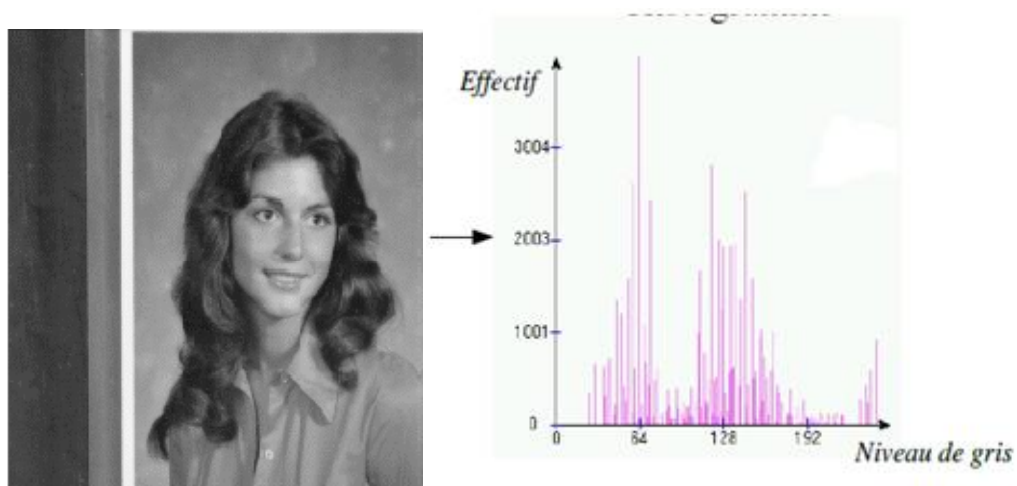


Figure I.4 -Images en niveau de gris et son histogramme .

I.4 Traitement d'images

Le traitement des images constitue actuellement l'une des grandes orientations du traitement de l'information, l'essor des potentialités des calculateurs a permis d'envisager des procédures de plus en plus complexes pour mettre en évidence ce que l'œil ne peut voir.

Le traitement d'image désigne un ensemble de méthodes dont l'objectif est, soit de transformer des images (par exemple pour en améliorer l'apparence ou pour les coder de manière plus compacte en vue d'une transmission ou de stockage), soit d'en extraire de l'information. Il s'agit donc d'un sous ensemble du traitement de signal dédié aux images et aux données dérivées comme la vidéo (par opposition aux parties du traitement du signal consacré à d'autres types de données : son et autres signaux monodimensionnels), tout en opérant dans le domaine numérique (par opposition aux techniques analogiques du traitement du signal, comme la photographie et la télévision traditionnelle).

I.4.1 Organigramme général de traitement d'image

Le traitement d'image et ses techniques sont destinés à exploiter au mieux l'information contenue dans une image.

Figure suivante montre les différentes étapes du traitement.

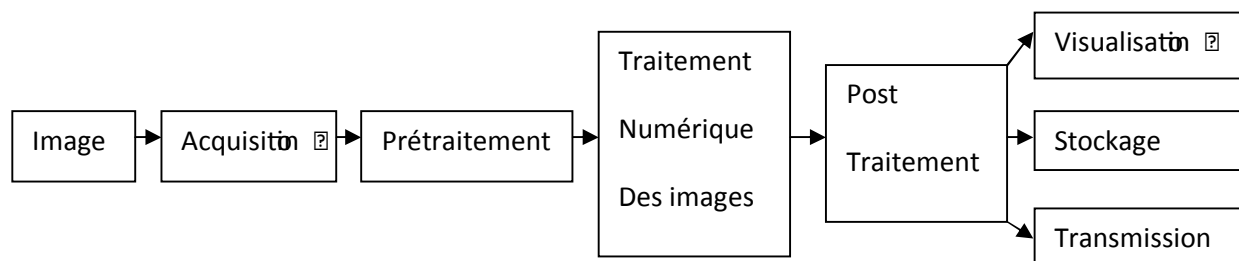


Figure I.5-Synoptique d'un système de traitement d'images

I.4.1.1 Acquisition des images

L'acquisition d'une image est l'opération qui permet le passage de l'information réel à une représentation numérique. Ce passage appelé aussi numérisation n'est autre que la transformation d'un signal analogique en un signal numérique. L'acquisition de l'image constitue l'interface entre le système et le monde extérieur. Elle est réalisée en plusieurs étapes : le transfert des images optiques qui est une opération permettant de transformer le signal en un signal analogique, l'échantillonnage de ce signal analogique qui consiste à

multiplier ce dernier par une série d'impulsions unité dans le but d'avoir des échantillons du signal, la quantification qui est une traduction des échantillons en valeur numériques selon une règle de codage choisit. Suivant l'objet ou le document à numériser et le domaine d'application dans lequel l'image va être utilisée, il existe divers dispositifs de numérisation d'image allant du simple scanner au satellite de télédétection.

I.4.1.2 Prétraitement

L'opération de prétraitement a pour but principal la correction, des imperfections du système d'acquisition. Les méthodes développées pour le prétraitement des images différentes selon la technique utilisée, les hypothèses sur le modèle du bruit, la nature de l'image à traiter et surtout le but recherché par le traitement. Toutes les approches tentent, dans leur ensemble, de retrouver à partir d'une image dégradée, celle qui est plus proche de l'originale. Ceci revient à obtenir des contours droits et francs et des zones homogènes bien lisses entre les contours. Les techniques de prétraitement les plus utilisées reposent principalement sur le filtrage et le rehaussement de contraste.

I.4.1.3 Traitement numérique des images

Ils sont principalement des opérations de convolution et de segmentation. La convolution est le remplacement de la valeur d'un pixel par une combinaison linéaire des ses voisins. Elle consiste à faire balayer une fenêtre (masque) sur l'ensemble des points de l'image. Segmenter une image revient à la partager en un certain nombre de domaines. On dit que deux points appartiennent à un même domaine s'ils possèdent une même propriété caractéristique et si on peut les joindre par une chaîne continue de points voisins ayant cette propriété. Donc une segmentation est définie par le type de propriété et la définition du voisinage.

I.4.1.4 Visualisation

Tout système de traitement d'image est doté d'un dispositif de visualisation qui permet l'affichage des images. L'utilisation de différents reconstituteurs permet de transformer le signal numérique qui est la matrice image en un signal analogique visible par l'œil de l'observateur. Pour cela, différents types de supports peuvent être employés : moniteurs vidéos, clichés photographiques, impression sur papier... Dans tous les cas et pour chaque échantillon de l'image numérique, on recrée un nouvel élément d'image ou un nouveau pixel dont on choisit

la forme de façon à reconstituer une image analogique qui soit la plus proche possible de l'image avant numérisation.

I.4.2 Techniques d'amélioration d'image

Nous développerons dans ce paragraphe les techniques de filtrage et de rehaussement de l'image.

I.4.2.1 Filtrage

L'amélioration des images est essentiellement obtenue par filtrage qui consiste soit à atténuer les bruits indésirables, soit à accentuer les contours.

On note trois catégories de filtrages : le filtrage linéaire, le filtrage adaptatif et le filtrage morphologique.

I.4.2.1.1 Filtrage linéaire

Ils ont appelés filtres linéaires car ils possèdent la propriété de linéarité du produit de convolution.

Filtrer une image $I(x, y)$ consiste à la convoluer avec une fonction $f(x, y)$ qui est la réponse impulsionnelle du filtre.

Dans le cas continu, l'image filtrée est donnée par :

$$I(x, y) = (f * I)(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, y_1) I(x - x_1, y - y_1) dx_1 dy_1 \quad (\text{I.2})$$

Dans le cas discret, l'image filtrée est donnée par :

$$I(x, y) = (f * I)(x, y) = \sum_{i'=-k/2}^{+k/2} \sum_{j'=-N/2}^{+N/2} f(x - x_1, y - y_1) I(x_1, y_1) \quad (\text{I.3})$$

I.4.2.1.2 Filtres adaptatifs

On les classe en deux catégories :

- **Filtres d'ordres** : ils classent les niveaux de gris d'un voisinage et sélectionnent parmi ces quantités une ou plusieurs valeurs. Ils sont de deux types : le filtrage médian et le filtrage par le plus proche voisin radio métrique.
- **Les filtres de position** : considèrent les sous voisinages du point. Pour ce type de filtrage, on utilise souvent des voisinages 5*5. Le sous voisinage le plus homogène est sélectionné et son niveau de gris moyen est affecté au point central.

I.4.2.1.3 Filtres morphologiques

Les filtres morphologiques constituent une alternative aux filtrages linéaires. Ils n'ont aucune propriété de linéarité. Le filtrage morphologique consiste à considérer l'image non comme une fonction bidirectionnelle mais plutôt comme un espace de représentation d'objets ayant des propriétés ensemblistes. Ce type de filtre est souvent appliqué après binarisation de l'image.

I.4.2.2 Rehaussement

Le rehaussement du contraste est la méthode la plus simple pour améliorer l'image non contrastée. Afin d'appliquer cette méthode on identifie les limites supérieures et inférieures d'intensité représentée sur l'histogramme et à l'aide d'une transformation linéaire, on étire ces valeurs sur l'ensemble des valeurs disponible.

Ce procédé rehausse le contraste dans l'image en pâlisant d'avantage les régions plus forcées ce qui facilite l'interprétation visuelle.

I.4.3 Binarisation

La binarisation est une technique classique très utilisée en traitement d'images. Elle consiste à obtenir à partir d'une image codée sur plusieurs niveaux de gris (généralement 256), une image codée sur deux niveaux de gris (noir et blanc ou bien objet et fond). Son but est de réduire la quantité d'informations en réduisant le nombre de niveaux de gris et de séparer les objets du fond. Son principe consiste à déterminer un seuil S , appelé seuil de binarisation, puis modifier les valeurs des pixels de l'image suit :

Soient I une image et $I(x, y)$ un point de I de coordonnées (x, y) .

$$I(x, y) = \begin{cases} 0 & \text{si } I(x, y) < S, (\text{noir}) \\ 255 & \text{sinon } (\text{blanc}) \end{cases} \quad (\text{I.4})$$

I.4.4 Notion de segmentation

La segmentation constitue une étape très importante dans la chaîne de la reconnaissance des formes. Elle reste un des problèmes fondamentaux dans l'analyse d'images.

On distingue la segmentation par régions et la segmentation par contours. Son principe c'est d'effectuer une partition de l'image en régions, telle que chacune d'elles possède au moins une caractéristique que ne possèdent pas les régions voisines.

I.5. Détection de caractéristiques

I.5.1 Détection de contours

Un contour peut approximativement être défini comme une zone de l'image où l'intensité des pixels change brusquement. Cette discontinuité dans l'image (variations) est le passage d'un niveau de gris à un autre, de manière plus ou moins rapide.

Le contour peut être détecté par le maximum d'une dérivée première de la fonction niveau de gris d'un point de l'image ou par le passage par zéro de la seconde dérivée.

Les principaux opérateurs utilisés pour la détection des contours sont les opérateurs gradient, Sobel et de Prewitt.

I.5.2 Détection de points d'intérêts

La détection de points d'intérêts est, au même titre que la détection de contours, une étape préliminaire à de nombreux processus de vision par ordinateur. Les points d'intérêts, dans une image, correspondent à des doubles discontinuités de la fonction d'intensités. Ce sont par exemple : les coins, les jonctions en T ou les points de fortes variations de texture.



Figure I.6- *Différentes points d'intérêts.*

Avantages des points d'intérêts :

1. Sources d'informations plus fiables que les contours car plus de contraintes sur la fonction d'intensité.
2. Robuste aux occultations (soit occulté complètement, soit visible).
3. Pas d'opérations de chaînage (-> contours !).
4. Présents dans une grande majorité d'images (\neq contours !).

I.5.2.1 Différentes approches

De nombreuses méthodes ont été proposées pour détecter des points d'intérêts. Elles peuvent être classées grossièrement suivant trois catégories :

1. Approches contours : l'idée est de détecter les contours dans une image dans un premier temps. Les points d'intérêts sont ensuite extraits le long des contours en considérant les points de courbures maximales ainsi que les intersections de contours.
2. Approches intensité : l'idée est cette fois-ci de regarder directement la fonction d'intensité dans les images pour en extraire directement les points de discontinuités.
3. Approches à base de modèles : les points d'intérêts sont identifiés dans l'image par mise en correspondance de la fonction d'intensité avec un modèle théorique de cette fonction des points d'intérêts considérés.

Les approches de la deuxième catégorie sont celles utilisées généralement. Les raisons sont : indépendance vis à vis de la détection de contours (stabilité), indépendance vis à vis du type de points d'intérêts (méthodes plus générales).

I.5.2.2 Le détecteur de Moravec (1980)

L'idée du détecteur de Moravec [1,2] est de considérer le voisinage d'un pixel (une fenêtre) et de déterminer les changements moyens de l'intensité dans le voisinage considéré lorsque la fenêtre se déplace dans diverses directions. Plus précisément. On considère la fonction :

$$E(x, y) = \sum_{u,v} \omega(u, v) |I(x + u, y + v) - I(u, v)|^2 \quad (\text{I.5})$$

Ou :

- ω spécifie la fenêtre/voisinage considérée (valeur 1 à l'intérieur de la fenêtre et 0 à l'extérieur);
- $I(u, v)$ est l'intensité au pixel (u, v) ;
- $E(x, y)$ représente la moyenne du changement d'intensité lorsque la fenêtre est déplacée de (x, y) .

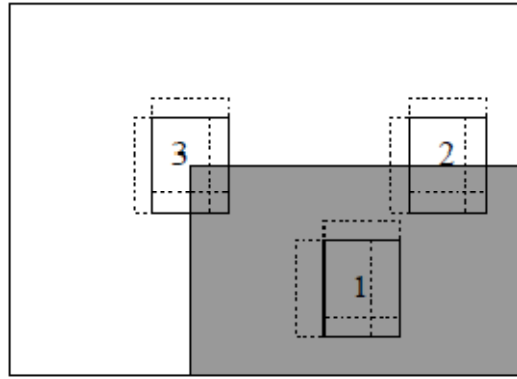


Figure I.7-Différentes situations considérées par le détecteur de moravec.

En appliquant cette fonction dans les trois situations principales suivantes (voir la figure I.7), on obtient :

1. L'intensité est approximativement constante dans la zone image considérée : la fonction E prendra alors de faibles valeurs dans toutes les directions (x, y) .
2. La zone image considérée contient un contour rectiligne : la fonction E prendra alors de faibles valeurs pour des déplacements (x, y) le long du contour et de fortes valeurs pour des déplacements perpendiculaires au contour.
3. La zone image considérée contient un coin ou un point isolé : la fonction E prendra de fortes valeurs dans toutes les directions.

En conséquence, le principe du détecteur de Moravec est donc de rechercher les maxima locaux de la valeur minimale de E en chaque pixel (au dessus d'un certain seuil).

I.5.2.3 Le détecteur de harris (1988)

Le détecteur de Moravec fonctionne dans un contexte limité. Il souffre en effet de nombreuses limitations. Harris et Stephen ont identifié certaines limitations et, en les corrigeant, en ont déduit un détecteur de coins très populaire : le *détecteur de Harris* [3]. Les limitations du détecteur de Moravec prises en compte sont :

1. La réponse du détecteur est anisotrope en raison du caractère discret des directions de changement que l'on peut effectuer (des pas de 45 degrés). Pour améliorer cet aspect, il suffit de considérer le développement de Taylor de la fonction d'intensité I au voisinage du pixel (u, v) :

$$I(x + u, y + v) = I(u, v) + x \frac{\delta I}{\delta x} + y \frac{\delta I}{\delta y} + o(x^2 + y^2) \quad (\text{I.6})$$

D'où :

$$E(x, y) = \sum_{u,v} \omega(u, v) [x \frac{\delta I}{\delta x} + y \frac{\delta I}{\delta y} + o(x^2, y^2)]^2 \quad (\text{I.7})$$

En négligeant le terme $o(x^2, y^2)$ (valable pour les petits déplacements), on obtient l'expression analytique suivante :

$$E(x, y) = Ax^2 + 2Cxy + By^2 \quad (\text{I.8})$$

Avec :

- $A = \frac{\delta I^2}{\delta x} * \omega$
- $B = \frac{\delta I^2}{\delta y} * \omega$
- $C = \left(\frac{\delta I}{\delta x} \frac{\delta I}{\delta y} \right) * \omega$

2. La réponse du détecteur de Moravec est bruitée en raison du voisinage considéré. Le filtre ω utilisé est en effet binaire (valeur 0 ou 1) et est appliqué sur un voisinage rectangulaire. Pour améliorer cela, Harris et Stephen propose d'utiliser un filtre Gaussien :

$$\omega(u, v) = \exp -(u^2 + v^2) / 2\sigma^2 \quad (\text{I.9})$$

3. Enfin, le détecteur de Moravec répond de manière trop forte aux contours en raison du fait que seul le minimum de E est pris en compte en chaque pixel. Pour prendre en compte le comportement général de la fonction E localement, on écrit :

$$E(x, y) = (w, y) \cdot M \cdot (x, y)^t \quad (\text{I.10})$$

Avec :

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \quad (\text{I.11})$$

La matrice M caractérise le comportement local de la fonction E , les valeurs propres de cette matrice correspondent en effet aux courbures principales associées à E :

- Si les deux courbures sont de faibles valeurs, alors la région considérée a une intensité approximativement constante.
- Si une des courbures est de forte valeur alors que l'autre est de faible valeur alors la région contient un contour.
- Si les deux courbures sont de fortes valeurs alors l'intensité vari fortement dans toutes les directions, ce qui caractérise un coin.

Par voie de conséquence, Harris et Stephen propose l'opérateur suivant pour détecter les coins dans une image :

$$R = \text{Det}(M) - k \text{Trace}(M)^2 \quad (\text{I.12})$$

Avec :

$$\text{Det}(M) = AB - C^2 \quad \text{et} \quad \text{Trace}(M) = A + B.$$

Les valeurs de R sont positives au voisinage d'un coin, négatives au voisinage d'un contour et faibles dans une région d'intensité constante.

Conclusion

Nous avons présenté dans ce chapitre quelques notions essentielles pour l'étape de prétraitement et l'étape de détection de points d'intérêts qui sera détaillée dans le cinquième chapitre.

Chapitre II

Vision omnidirectionnelle

II.1. Capteurs omnidirectionnels

La vision omnidirectionnelle (Figure II.2) correspond à un champ de vision couvrant toute une sphère dont le centre correspondrait à la place de l'observateur. La vision panoramique (Figure II.1) quant à elle fournit un champ de vue réduit à une bande.



Figure II.1 –Image panoramique



Figure II.2-Image omnidirectionnelle

Pour obtenir des images omnidirectionnelles, on distingue trois classe de systèmes possibles :

- L'utilisation d'images multiples
- L'utilisation de lentilles spéciales
- L'utilisation de miroirs convexes.

II.1.1 Images multiples

Le principe ici est de recomposer une image omnidirectionnelle à partir d'images multiples, en exploitant, soit les images produites par la même caméra à différentes positions, soit les images produites par différentes caméras.

II.1.1. a Génération d'images panoramiques par rotation d'une camera

Dans la littérature, on peut trouver des systèmes qui fournissent des images omnidirectionnelles en utilisant une seule caméra. Cette caméra étant montée autour d'un axe vertical à vitesse angulaire constante. Il faut alors construire l'image omnidirectionnelle à partir des différentes images prises successivement par la caméra. Cette technique de reconstruction est appelée mosaïque d'image (voir [4] pour plus de détails).

L'inconvénient de ces systèmes réside dans le temps d'acquisition qui est assez long. Ces derniers sont plutôt préconisés pour des scènes statiques et non pour des scènes en mouvement.

II.1.1.b Génération d'images panoramiques à partir de plusieurs caméras

Pour ce type de capteurs, nous pouvons citer le système développé dans [5]. Ce capteur utilise quatre caméras conventionnelles et quatre miroirs triangulaires (Figure II.4 (a)). Le capteur acquiert quatre images qui sont utilisées pour calculer une image panoramique. Le résultat est une image à haute résolution avec un centre unique de projection, qui permet donc de calculer n'importe quelle projection sur n'importe quelle surface.

Un autre exemple de ce type de système est présenté dans [6] et [7] : Le capteur génère une image omnidirectionnelle grâce à un système composé de plusieurs caméras (Figure II.4(b)).

L'inconvénient de ces techniques réside d'une part dans les angles morts, c'est-à-dire les parties de l'environnement qui ne sont pas acquises par le système, et d'autre part dans les zones de l'environnement qui sont dupliquées à cause de l'intersection de l'angle de vue entre les caméras adjacentes. .

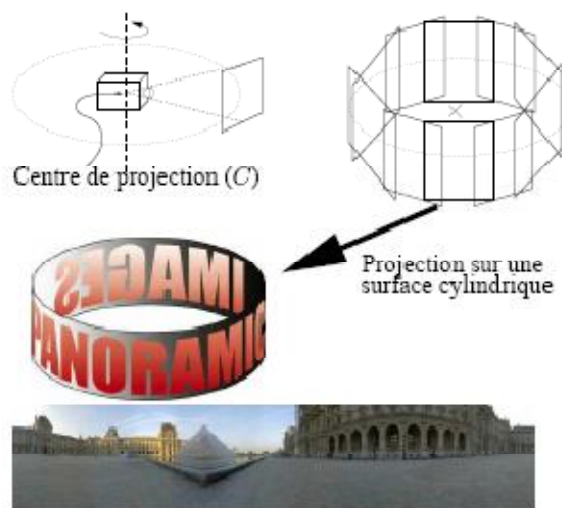


Figure II.3 Image omnidirectionnelle générée par Rotation de caméra

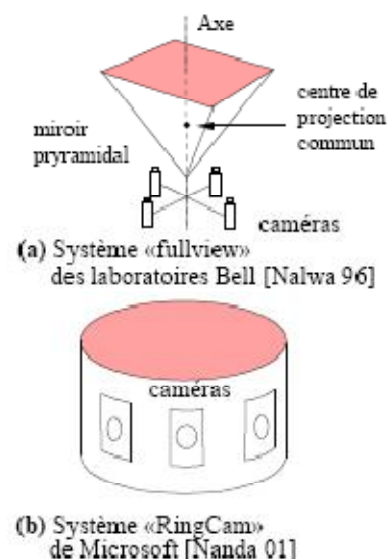


Figure II.4 Capteurs omnidirectionnelle qui utilisent plusieurs caméra conventionnelles

II.1.2 Lentilles spéciales

Cette méthode consiste à ajouter un ensemble optique sur la caméra CCD, afin de dévier les rayons lumineux. Une caméra œil de poisson (Fish-eyes) utilise un objectif fish-eye au

lieu d'un objectif conventionnel. Les objectifs fish-eyes sont des objectifs de très petite focale dont l'angle de champ est très large.

Les auteurs dans [8], présentent une méthode stéréoscopique de l'analyse du mouvement utilisant un tel système. Les images acquises par ce genre d'objectifs sont déformées et l'analyse de ces images est difficile principalement sur la périphérie où la résolution est faible. Elles ne peuvent pas être transformées en images perspectives, car les objectifs fish-eyes n'ont pas un centre de projection unique, et ils sont difficiles à modéliser.

Greguss a proposé une lentille appelée «Panoramic Annular Lens» (PAL) qui acquiert une image panoramique sans distorsion [9]. Le système est composé d'un bloc optique d'une seule pièce de verre sur laquelle deux plans de réflexion sont créés : le trajet des rayons lumineux comprend deux réflexions et deux réfractions avant d'être projeté sur la lentille de la caméra (figure 1.3). Le système peut être facilement miniaturisé mais sa modélisation est complexe. Par ailleurs, le champ de vue vertical est assez faible et difficilement augmentable.

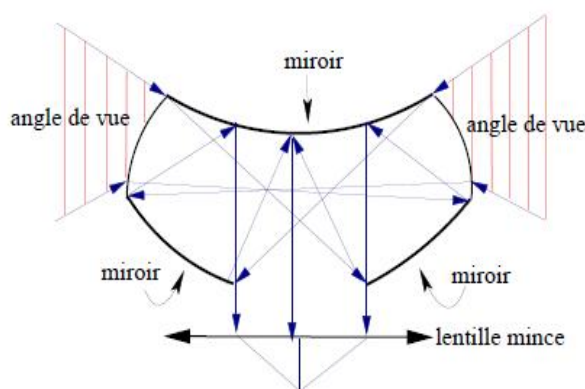


Figure II.5- «Panoramic Annular Lens» (PAL). Le parcours des rayons optiques sont indiquées par les flèches en bleu



Figure II.6 Une caméra «fish-eye» composée d'une lentille fish-eyes et d'une caméra sténopé.

II.1.3 Utilisation de miroirs convexes

Le dernier procédé que nous décrivons est le capteur catadioptrique : dioptrique pour la réfraction (lentilles) et catoptrique pour la réflexion (miroirs). Ce capteur se compose d'un miroir de révolution (dans lequel se reflète une scène) qui est filmé par une caméra. C'est ce

dernier système, à priori plus simple à mettre en œuvre et avec des propriétés intéressantes, que nous utiliserons dans la suite.

Avant de comparer les différents types de miroirs, nous présentons la notion de centre de projection, et recensons les différentes aberrations qui peuvent survenir dans tout système d'imagerie.

II.1.3.1 Notion de centre de projection

La réflexion est le changement de direction de la propagation d'une onde ou d'une particule se produisant lorsque cette onde arrive à la frontière du milieu dans lequel elle se déplace sans franchir cette frontière. La plupart des corps, transparents ou non, réfléchissent dans toutes les directions une fraction de la lumière qu'ils reçoivent. Certaines surfaces polies provoquent dans l'espace une réflexion régulière et concentrée : c'est le cas des miroirs. Par réflexion, ils renvoient la totalité de la lumière incidente dans une direction précise, que l'on peut déterminer par des constructions géométriques simples.

On caractérise cette réflexion par :

- les angles d'incidence et de réflexion α_i et α_r . La réflexion est dite spéculaire lorsque $\alpha_i = \alpha_r$, comme c'est le cas d'une onde lumineuse sur un miroir (figure II.6). Si $\alpha_i \neq \alpha_r$, c'est une réflexion diffuse ;
- La proportion R de l'onde ou des particules qui sont effectivement réfléchies.

Prenons l'exemple des miroirs paraboliques : selon leur profil, on distingue les miroirs concaves (la face creuse est réfléchissante) et les miroirs convexes (c'est la face bombée qui est réfléchissante) (voir figure II.7). Notons S le sommet du miroir (le sommet de la calotte parabolique), F le foyer du paraboloïde à partir de laquelle la calotte a été obtenue. Un simple raisonnement géométrique montre que dans le cas du miroir concave, tout rayon lumineux passant par le centre revient sur lui-même. On peut aussi montrer que tout rayon parallèle à l'axe FS passe après réflexion par le point F appelé *foyer*. Dans le cas du miroir convexe, tous les rayons parallèles à l'axe FS intersectent "virtuellement" le point F . Ce foyer F est appelé *centre de projection* : il est défini comme le point où tous les rayons entrants ou sortants traversent le système.

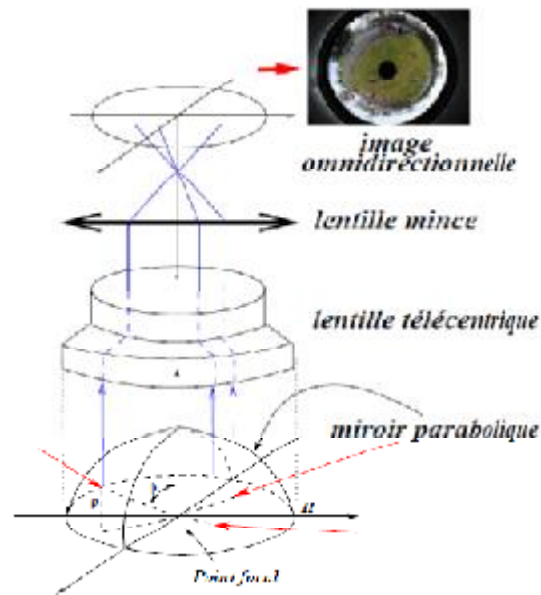


Figure II.7 -Caméra catadioptrique constituée d'un miroir parabolique, d'une lentille télécentrique et d'une caméra classique. Le parcours des rayons optiques est indiqué en bleu.

La présence d'un centre de projection unique (ou foyer unique) des faisceaux lumineux a l'avantage de permettre la reconstruction d'images selon différents types de projection, et ce, sur un type quelconque de surface. La connaissance des paramètres géométriques de formation des images panoramiques permet ainsi de générer des images en projection perspective, ou panoramique selon un cylindre par exemple. En outre, les systèmes avec centre de projection unique sont aisés à modéliser géométriquement (ce qui rend aussi leur calibration aisée), et permettent de définir simplement la géométrie épipolaire pour deux images panoramiques.

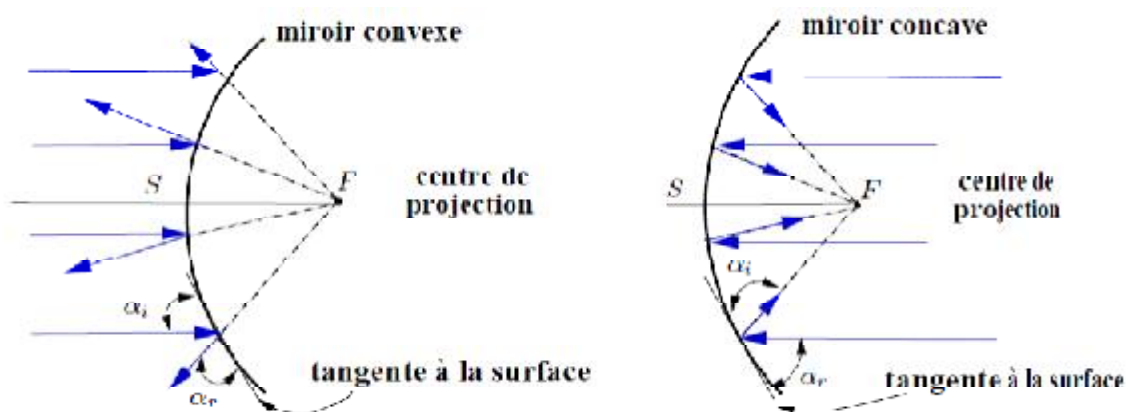


Figure II.8-Réflexion sur un miroir parabolique concave et convexe.

II.1.3.2 Aberrations

Les lentilles sont les éléments les plus largement répandus dans la fabrication des instruments optiques. On en trouve dans les oculaires, les lunettes, les microscopes, les lecteurs CD et bien sûr dans l'œil. Une lentille est un composant qui a la propriété de faire converger (lentille convergente) ou diverger (lentille divergente) les faisceaux lumineux qui la traversent. Elles possèdent en général, des surfaces sphériques pour des questions de procédés de fabrication. Les lentilles souffrent de défauts inhérents à leur construction que l'on appelle aberrations. On distingue notamment les aberrations suivantes :

- Aberration chromatique : l'indice de réfraction dépend de la longueur d'onde de la lumière, ce qui fait que toutes les couleurs ne sont pas déviées de la même façon (voir figure 1.9), le bleu l'étant plus que le rouge. La lumière blanche étant composée de la superposition de toutes les couleurs de l'arc-en-ciel, l'image d'une étoile blanche sera étalée le long de l'axe optique et on ne pourra voir nettement qu'une seule couleur à la fois. Cela provoque un effet arc-en-ciel autour de l'étoile. Ce défaut peut être réduit par l'utilisation de jeux de lentilles dits achromats.
- Distorsion radiale : ce défaut résulte de la différence de grossissement entre le centre et la périphérie de la lentille. Suivant que le grossissement augmente ou diminue, on l'appelle distorsion en coussinet ou en barillet (voir figure II.9).

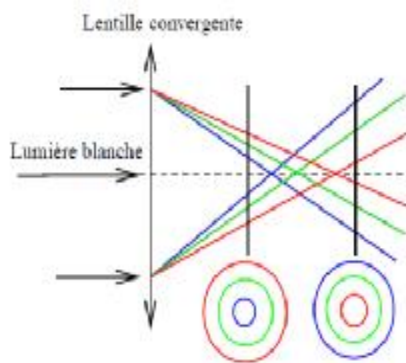


Figure II.9-Aberration chromatique

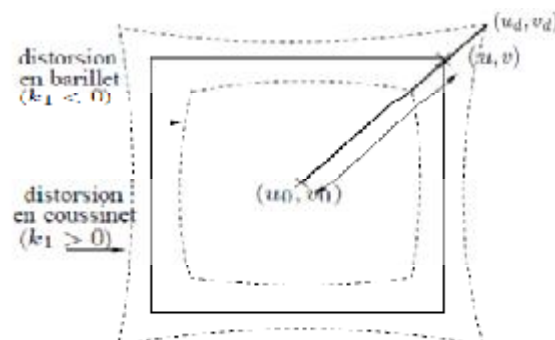


Figure II.10-Distorsions radiales

- Aberration sphérique : cette aberration résulte car les rayons qui frappent le bord des lentilles convergentes sphériques sont plus fortement réfractés que ceux passant à proximité de l'axe optique. Ces rayons périphériques passent donc à côté du point focal (voir figure II.10). Cette aberration devient de plus en plus importante à mesure que l'on s'éloigne de l'axe optique. La solution pour éviter ces aberrations est d'utiliser des lentilles améliorées avec des profils plus adéquats (dites "asphériques").

- Courbure de champ : pour une lentille idéale, la surface de focalisation est un plan, ce qui n'est pas le cas pour une lentille réelle. La courbure de champ est l'écart entre la surface de focalisation réelle et le plan focal idéal.
- Coma : cette aberration produit sur les images une sorte de traînée conique dirigée vers le centre du champ (voir figure II.11). Ceci s'explique par le fait que les rayons qui passent par la périphérie de la lentille ne sont pas focalisés au même endroit que ceux qui passent près du centre. La forme de la traînée résulte de la superposition des images partielles qui passent à des distances différentes de l'axe optique. On peut limiter cette aberration en diaphragmant la lentille (en réduisant le diamètre de la zone où la lumière peut pénétrer).
- Astigmatisme : on dit qu'une lentille est astigmatique lorsque la distance focale n'est pas la même pour toutes les directions. Ainsi, les lignes horizontales et verticales d'une grille ne seront pas focalisées au même endroit.

Dans un système omnidirectionnel, les aberrations peuvent être provoquées par les lentilles utilisées, le miroir, ou même la combinaison lentille-miroir.

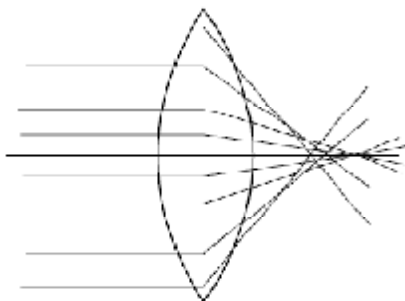


Figure II.11 -Aberration sphérique

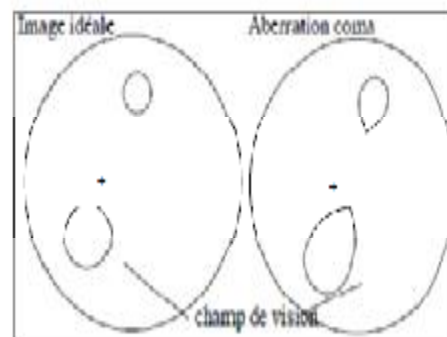


Figure II.12-Aberration coma

Les miroirs permettent de faire converger ou diverger la lumière, de manière similaire aux lentilles. Ils sont incomparablement plus légers et plus faciles à produire que les lentilles, notamment pour les grands diamètres et ils ne présentent pas d'aberration chromatique.

II.1.4 Comparaison des miroirs

Parmi les différents systèmes catadioptriques, nous pouvons déterminer quatre principales catégories, selon le type de miroir convexe utilisé : conique, sphérique, hyperbolique et parabolique (voir figure 1.13). Dans cette section nous décrivons les caractéristiques de ces quatre types de miroirs pour constituer une caméra panoramique catadioptrique, en présentant leurs avantages et inconvénients respectifs. Les avantages et inconvénients de chaque miroir

peuvent être discutés à partir des points de vue suivant : existence d'un point de projection unique, importance des aberrations et facilité de réalisation (prix).

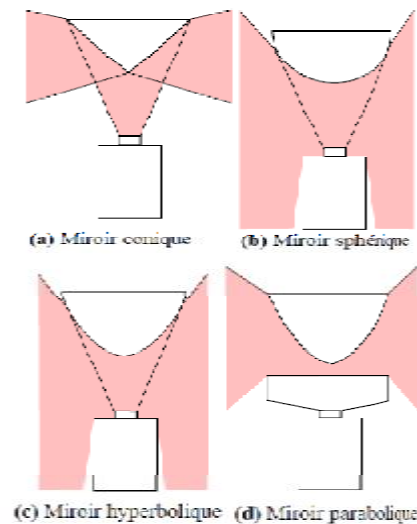


Figure II.13- Les quatre types principaux de miroirs omnidirectionnels

II.1.4.1 Miroir sphérique

Une image acquise à partir d'un miroir sphérique a une résolution qui est bonne dans la région centrale mais faible dans la périphérie de l'image. La figure II.13 montre le parcours des rayons dans une caméra omnidirectionnelle avec miroir sphérique. Parmi les capteurs omnidirectionnels qui utilisent des miroirs convexes, ceux qui utilisent des miroirs sphériques ont le plus large angle de vue.

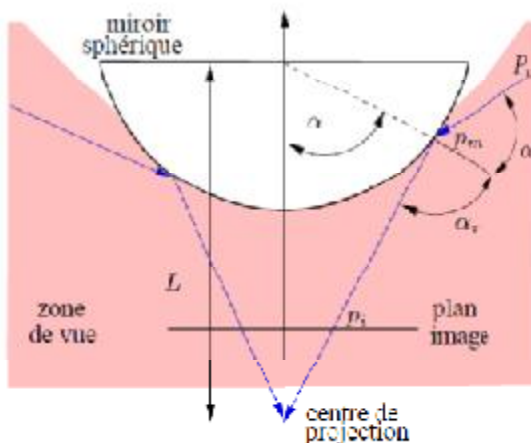


Figure II.14- Caméra omnidirectionnelle avec un miroir sphérique. le champ de vue correspond à la zone colorée, la projection des rayons est représenté par les flèches bleues.



Figure II.15- Capteur omnidirectionnel avec miroir sphérique de <<Institut of System and Robotics>> de Lisbonne, Portugal.

Avantage

- Un point important dans la fabrication du miroir est la facilité de fabrication : de façon générale, les miroirs sont faits en déposant de l'aluminium sur une surface de verre. Une lentille normale est une pièce de verre sphérique, d'où la facilité de la construction du miroir sphérique en utilisant des processus conventionnel de lentilles.
- Un autre point important du miroir sphérique est l'astigmatisme. L'astigmatisme est faible près de l'axe optique de la caméra puisqu'il peut être considéré comme une surface plane. Bien sûr, l'astigmatisme est considérable sur la périphérie.
- Le miroir sphérique n'a pas besoin d'une grande distance focale pour faire converger des images.

Inconvénients

- Les miroirs sphériques n'ont pas un centre de projection unique
- Plus l'angle par rapport à la verticale est grand (α dans la figure II.13), plus la distorsion est grande sur la périphérie.

II.1.4.2 Miroir conique

Une image acquise à partir d'un miroir conique a une résolution qui est bonne dans la périphérie de l'image. Yagi dans [10,11] développe un système de perception basé sur un miroir conique pour la navigation d'un robot mobile, appelé COPIS (*CO*nic *P*rojection *I*mage *S*ensor). Dans [12, 13,14] le système de perception est aussi basé sur un miroir conique SYCLOP «ConicalSYstem for LOcalization and Perception». Les images acquises dans des environnements d'intérieur par le capteur SYCLOP sont très intéressantes. En effet, si l'axe optique correspond avec l'axe de révolution du cône, les droites parallèles à ceux-ci se projettent de façon radiale dans l'image. Autrement dit, si on place ce capteur dans un environnement de façon à ce que la base du cône soit parallèle au sol, les droites verticales de l'environnement se projetteront en droites radiales dans l'image.

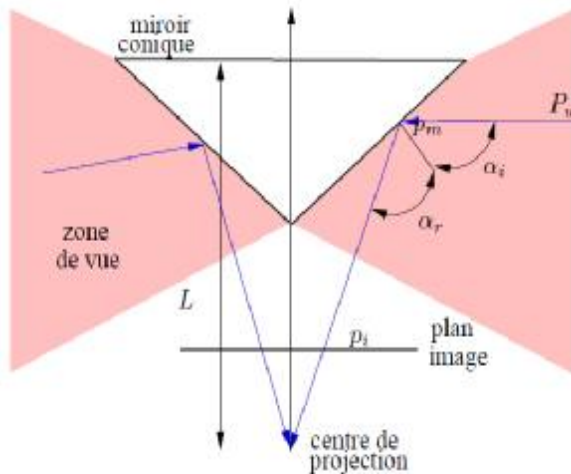


Figure II.16 -Caméra omnidirectionnelle avec Miroir conique.

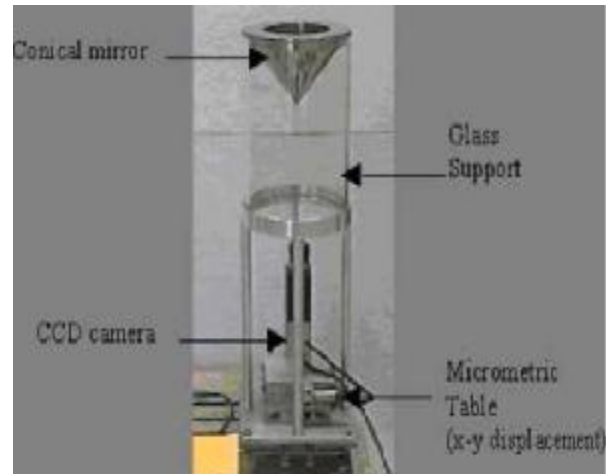


Figure II.17-Capteur omnidirectionnel avec miroir conique<< système conique pour la localisation et la perception (SYCLOP)>>

Avantages

- Ils sont relativement faciles à construire.
- La principale caractéristique est que la normale de réflexion est toujours suivant la direction verticale. Par conséquent, il est facile à combiner avec différents miroirs. Par exemple, dans [15], les auteurs combinent un miroir conique et un miroir plan.
- Les miroirs à courbure non nulle produisent l'effet *fish-eyes*: les objets sont de taille plus grande près du centre du miroir, typiquement la caméra et le robot qui la supporte. D'un autre côté, les objets sur la périphérie sont réduits. Cet effet n'existe pas sur les miroirs coniques du fait de l'absence de courbure.

Inconvénients

- L'astigmatisme est très grand.
- À cause de l'absence d'un point de projection unique, les images qu'ils produisent ne peuvent pas être transformées en images perspectives. Notons cependant que dans [16] est introduit le concept de *dimensionally separable view points*, grâce auquel il est possible de reconstruire une pseudo image perspective.

II.1.4.3 Miroir Hyperbolique

Yamzawa et al. ont proposé un capteur d'images omnidirectionnelles en utilisant un miroir hyperbolique, appelée *HyperOmni Vision* [17]. Le miroir hyperbolique a deux points focaux : si le centre optique de la caméra est placé sur un de ces deux points focaux, l'unicité du centre optique est satisfaite.

Avantages

- Ils ont un centre de projection unique.
- Si la courbure est faible, l'astigmatisme est faible.

Inconvénients

- Ils sont difficiles à construire.
- La conception n'est pas flexible car le point focal de l'hyperboloïde nécessite d'être placé au centre de la caméra.

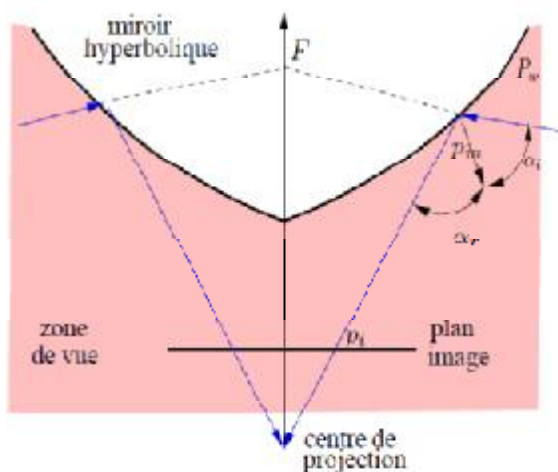


Figure II.18-Caméra omnidirectionnelle avec miroir hyperbolique



Figure II.19-Capteur omnidirectionnel avec miroir hyperbolique.

II.1.4.4 Miroir parabolique

Le capteur omnidirectionnel avec miroir parabolique a été un des derniers à être utilisé. Une des premières études intégrant ce type de miroir a été réalisée par Péri et Nayar, qui présentent dans [18,19] une caméra omnidirectionnelle composée d'une caméra conventionnelle, d'un miroir parabolique et d'une lentille télécentrique. Les lentilles télécentriques sont des lentilles spéciales conçues avec des éléments multiples qui permettent de sélectionner un faisceau lumineux de rayons parallèles à l'axe optique de la lentille. Avec une lentille télécentrique alignée sur l'axe du miroir, seuls les rayons incidents lumineux qui arrivent sur le miroir et qui traversent le foyer du paraboloïde sont transmis (voir figure 1.20). Le capteur nommé OmniVideo a été proposé pour des applications de vidéo-conférence.

Avantages

- Ils sont compacts.
- Ils ont un centre de projection unique.

Inconvénients

- Difficulté d'assemblage du système, la caméra devant être précisément localisée sur le foyer du miroir sphérique.
- Si le miroir parabolique et sphérique ne sont pas bien alignés, les aberrations de coma et l'astigmatisme sont grands.

II.1.4.5 Combinaison de différentes approches

Il est possible de concevoir un montage qui présente les mêmes caractéristiques que le miroir parabolique et la lentille télécentrique mais de moindre coût et de petite taille en combinant un miroir parabolique avec un miroir sphérique. Un tel système est introduit dans [20] par Orghidan, Mouaddib et Salvi. Les auteurs présentent les caractéristiques et la modélisation géométrique de ce capteur visuel panoramique capable d'estimer la profondeur de l'environnement et sa modélisation. Les deux composants omnidirectionnels qui forment le dispositif ont été calibrés et le modèle résultant est utilisé pour estimer la profondeur de la scène. Ce dispositif peut être utilisé dans la navigation des robots mobiles comme le propose Gonzalez-Barbosa dans [21]. Il s'intéresse au problème de la localisation des robots mobiles par indexation des images dans des environnements naturels à l'aide d'une caméra panoramique. L'objectif est de donner au robot la capacité de modéliser et de reconnaître des éléments de son environnement afin de se localiser qualitativement.



Figure II.20-*Combinaison d'un miroir parabolique avec un miroir*

Dans ce capteur, le centre optique de la caméra doit être placé sur le foyer de la sphère (Figure II.19). Seuls les rayons parallèles à l'axe optique du miroir sphérique traversent le foyer. Le miroir sphérique joue donc le rôle de la lentille télécentrique. Cependant, ce système est difficile à assembler : la focale de la caméra doit coïncider avec le foyer du miroir sphérique et l'alignement des miroirs parabolique et sphérique doivent être parfaits pour éviter de fortes aberrations de coma 2 et d'astigmatisme.

Avantages

- Ils sont compacts.
- Ils ont un centre de projection unique.

Inconvénients

- Difficulté d'assemblage du système, la caméra devant être précisément localisée sur le foyer du miroir sphérique.
- Si le miroir parabolique et sphérique ne sont pas bien alignés, les aberrations de coma et l'astigmatisme sont grands.

II.2 Modélisation des caméras omnidirectionnelles

La modélisation d'une caméra catadioptrique panoramique relève de la même démarche que celle des caméras perspectives, à ceci près que les rayons lumineux sont réfléchis par le miroir de révolution. Cette réflexion introduit une transformation supplémentaire dont il faut tenir compte. La transformation dépend de la surface du miroir.

Dans tout ce qui va suivre, nous considérons des systèmes catadioptriques comprenant des miroirs paraboliques.

Dans un premier temps, il s'agit de modéliser le processus de formation de l'image par un capteur catadioptrique. On obtient alors les relations qui existent entre les points 3D réels et leurs projections sur l'image.

Nous montrons (figure II.20) le schéma des transformations qui considère que l'axe du miroir et l'axe optique de la caméra sont confondus. La forme de la quadratique dépend du miroir utilisé.

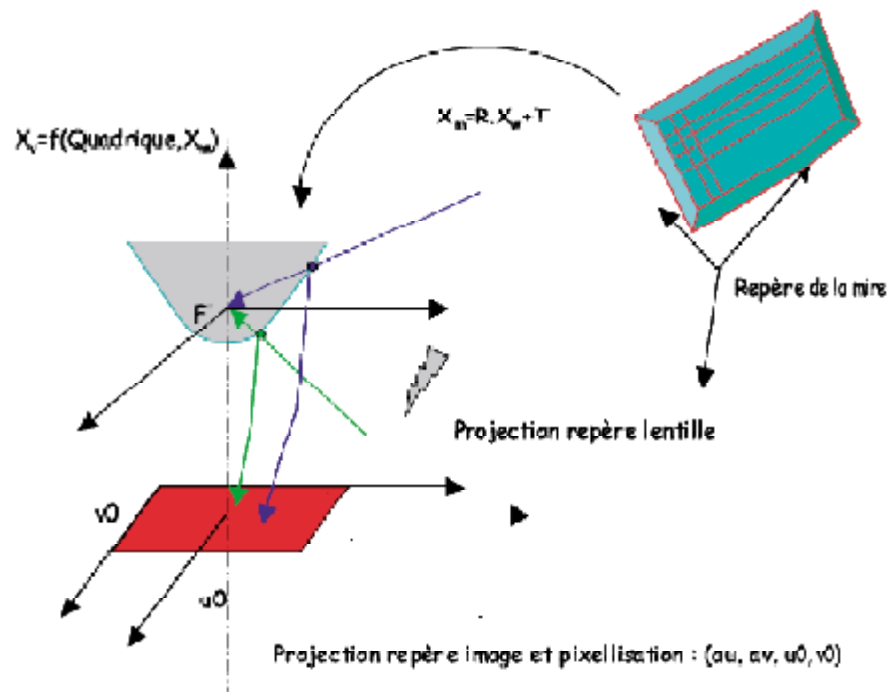


Figure II.21 -Schéma générale de formation des images catadioptriques.

Pour plus de détails les lecteurs peut ce référé au [31].

Le modèle Ad hoc utilise directement les équations du miroir ainsi que celles des caméras associées.

Pour les modèles génétiques, les auteurs dans [22] ont introduit une théorie d'unification pour les caméras catadioptriques ayant un point de vue unique (*i.e.* single viewpoint, central catadioptric image). En substance, ils ont montré qu'il est possible de modéliser ces caméras en utilisant d'abord une projection sur une sphère (de rayon égal à un par exemple) suivie d'une projection sur un plan à partir d'un point. Cette modélisation introduit deux paramètres qui sont reliés aux paramètres des miroirs en fonction de chaque forme. Une version modifiée a été proposée dans [23]. Ces modèles sont intéressants car ils permettent d'exhiber des propriétés intéressantes pour le calibrage et pour la détection de droites par exemple. Mais il faut que la condition du point de vue unique soit vérifiée, ce qui est parfois délicat à cause de l'assemblage miroir+caméra.

II.3 Calibrage

Le calibrage consiste à estimer les paramètres intrinsèques du capteur. Dans le cas de caméras catadioptriques, ces paramètres englobes les paramètres de la caméra, ceux du miroir, ceux du dispositif optique (objectif), et la carte de numérisation du signal vidéo.

Afin de pouvoir réaliser le calibrage, il est indispensable de modéliser tout le processus de formation de l'image. Le modèle doit faire l'objet d'un compromis entre simplicité et précision. La simplicité est indispensable à la résolution analytique ou numérique des équations et la précision est garante de la fidélité du comportement du modèle. Le calibrage en vision omnidirectionnelle est l'aspect le plus étudié (abouti) à l'heure actuelle.

Nous proposons aux lecteurs de ce référencier au [24], [23] pour plus de détails sur les différentes méthodes de calibrage.

Le calibrage intrinsèque consiste à utiliser l'image du miroir et les données du fabricant du miroir pour estimer les paramètres intrinsèques de la caméra catadioptrique. S.B. KANG [25] a proposé une technique simple et rapide pour calibrer les miroirs paraboloides. Les paramètres estimés sont les coordonnées du centre optique et le paramètre du miroir. La précision obtenue n'est pas bonne. Mais la technique est intéressante pour l'initialisation des paramètres avant de les affiner avec une méthode non linéaire. Les auteurs de [26] estiment les paramètres intrinsèques de la caméra et la position du miroir par rapport à la caméra (défauts d'alignements). Mais les paramètres propres du miroir sont supposés connus.

D'autres méthodes utilisent pour calibrer le capteur des mires externes, qui sont totalement connues où seulement certains éléments (droites, points...) sont connus. Les mires représentent des motifs connus que l'on observe pour déterminer les paramètres. La figure (II.21) présente un exemple d'image omnidirectionnelle obtenue quand le capteur catadioptrique observe une mire externe.

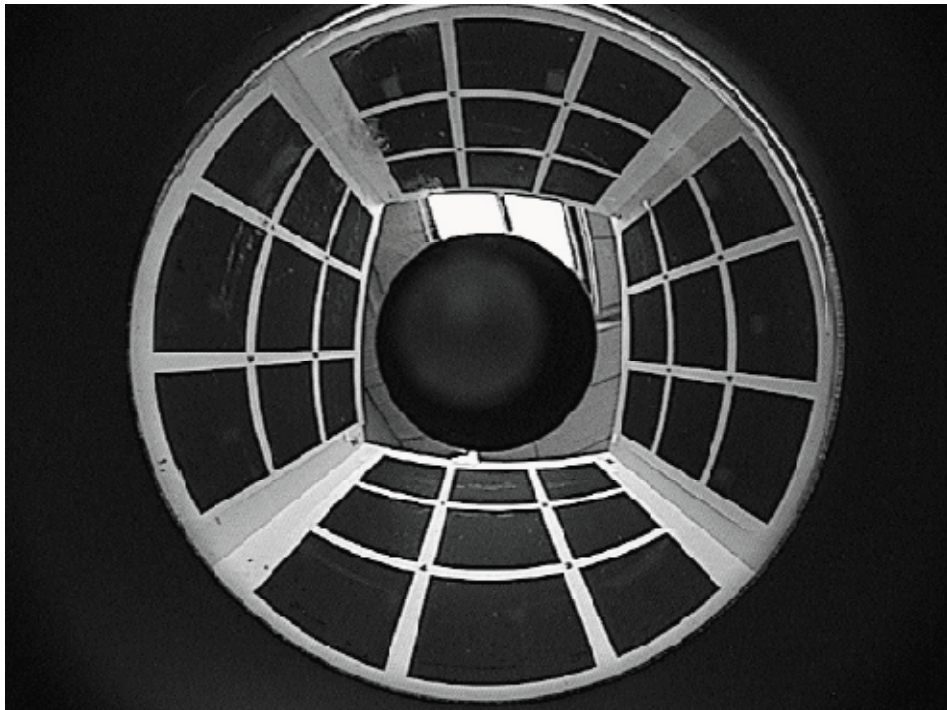


Figure II.22-*Image omnidirectionnelle d'une mire.*

La dernière méthode est l'auto calibrage qui n'utilise pas de mires externes. Il nécessite deux images et leur mise en correspondance.

II.4 La recherche dans le traitement d'images omnidirectionnelles

Appliquer de manière abrupte, les outils classiques sur des images omnidirectionnelles introduisent des erreurs. Diverses méthodes ont été proposées afin de remédier à ce problème. Dans [27], les auteurs proposent de calculer le voisinage en projetant un voisinage conçu sur un cylindre, puis utilisent des opérateurs morphologiques flous à noyaux variables pour images omnidirectionnelles. Cette idée a été utilisée dans [28] pour calculer le gradient.

Bogdanova et al. dans [29], ont proposé de corriger le gradient classique en fonction du miroir utilisé dans le capteur catadioptrique. D'autres auteurs ont également choisi d'utiliser la sphère d'équivalence, par exemple, dans [25], pour la détection de droites et pour la détection de rectangles dans le travail de Bazin et al. [30]. Cependant, ces traitements haut-niveau nécessitent une étape de détection de contours qui est effectuée sur l'image omnidirectionnelle avant d'être projetée sur la sphère pour que le traitement haut-niveau soit réalisé. Les détails concernant cette section seront repris dans le chapitre IV.

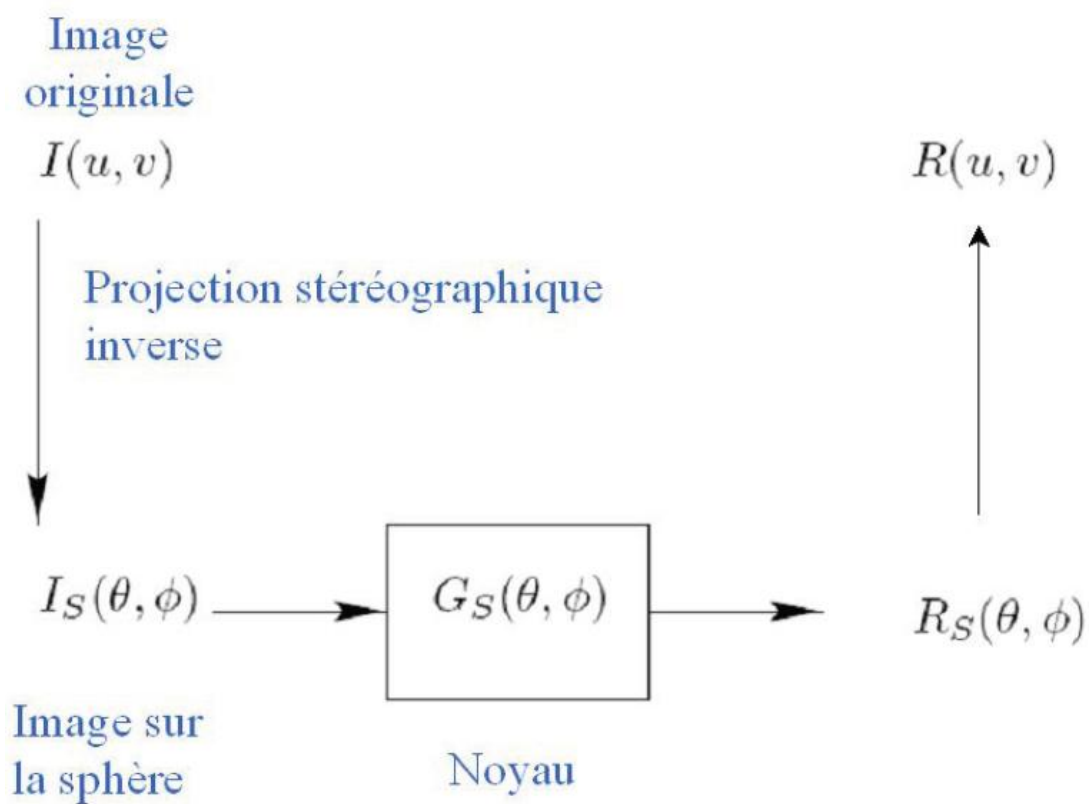


Figure II.23 -Schéma de traitement des images omnidirectionnelles.

Chapitre III

Outils mathématiques sur la sphère

Le traitement d'images planes implique dans la plupart des cas une étape de filtrage avec un opérateur adapté. Dans le plan, on dispose de multiples outils pour exprimer mathématiquement ces opérations de filtrage.

Dans ce chapitre on va essayer de présenter quelques outils mathématiques adaptés sur la sphère équivalente.

III.1 Projection stéréographique

Avant de parler plus précisément des outils mathématiques définis sur la sphère, rappelons brièvement la définition de la projection stéréographique.

On paramètre la sphère unité, plongée dans R , en utilisant les coordonnées sphériques (figure(III.1)) : soit $P \in S^2$

$$P(\theta, \varphi) = (\cos(\varphi) \sin(\theta), \sin(\varphi) \sin(\theta), \cos(\theta)) \quad (\text{III.1})$$

Avec $\varphi \in [0, 2\pi[$, angle de longitude et $\theta \in [0, \pi]$, angle de colatitude (latitude + $\pi/2$).

La définition de la projection stéréographique à partir du pôle Nord vers le plan équatorial est la suivante (figure (III.2)) :

A tout point P de la sphère, autre que le pôle Nord P_N , on associe le point P' intersection du plan équatorial et de la droite $(P_N P)$. Dans notre cas, les points de la demi-sphère supérieure se projettent à l'extérieur du cercle équatorial et les points de la demi-sphère inférieure se projettent à l'intérieur.

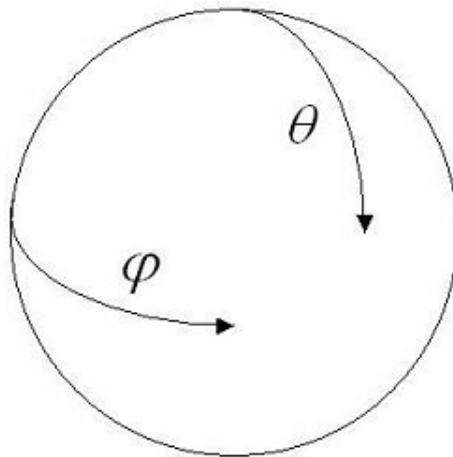


Figure III.1 –Angles de paramétrisation de la sphère.

Soient (x, y, z) les coordonnées cartésiennes du point P sur la sphère et (θ, φ) les coordonnées équivalentes sphériques.

Soient (X, Y) les coordonnées de P_0 dans le plan équatorial de la sphère unité.

En utilisant le théorème de Thalès, on montre que :

$$X = \frac{x}{1-z}, \quad Y = \frac{y}{1-z} \quad (\text{III.2})$$

D'où avec les coordonnées sphériques

$$X = \frac{\sin(\theta)\cos(\varphi)}{1-\cos(\varphi)}, \quad Y = \frac{\sin(\theta)\sin(\varphi)}{1-\cos(\varphi)} \quad (\text{III.3})$$

On peut aussi chercher P' en fonction des ses coordonnées polaires $P'(R, \gamma)$.

$$R = \frac{\sin(\theta)}{1-\cos(\varphi)}, \quad \gamma = \theta \quad (\text{III.4})$$

Cette dernière relation nous rappelle que la projection stéréographique est une application conforme, c'est-à-dire qu'elle conserve les angles.

Pour la projection stéréographique inverse, qui à un point $P'(X, Y)$ du plan équatorial associe le point $P(x, y, z)$ (ou $P(\theta, \varphi)$ en coordonnées sphériques) de la sphère unité, les formules sont les suivantes :

$$(x, y, z) = \left(\frac{2X}{1+X^2+Y^2}, \frac{2Y}{1+X^2+Y^2}, \frac{-1+X^2+Y^2}{1+X^2+Y^2} \right) \quad (\text{III.5})$$

$$\theta = \arccos\left(\frac{-1+X^2+Y^2}{1+X^2+Y^2}\right), \quad \varphi = \arctan\left(\frac{X}{Y}\right) \quad (\text{III.6})$$

Notons que $x^2+y^2+z^2=1$ puisque P appartient à la sphère unité.

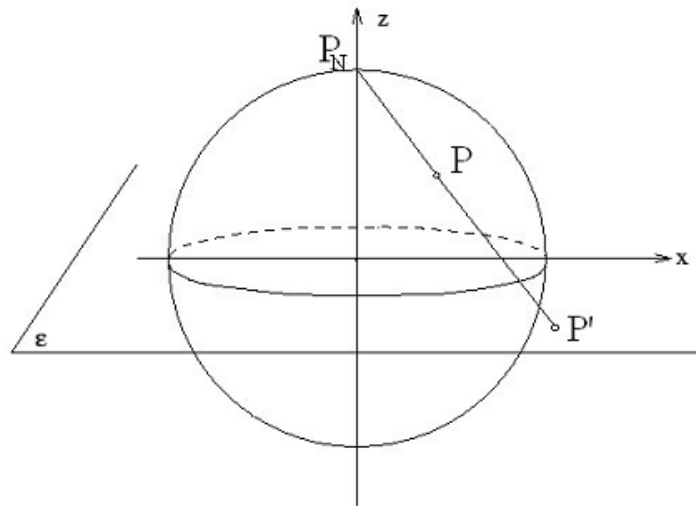


Figure III.2- Projection stéréographique.

La projection stéréographique est une application bijective de la sphère privée du pôle Nord vers le plan équatorial, donc en considérant le schéma de traitement que l'on a proposé précédemment, si aucun traitement n'est effectué sur l'image sphérique, on retrouvera l'image omnidirectionnelle originale.

Après avoir décrit la projection stéréographique et son application inverse, nous présenterons quelques opérateurs sur la sphère. Nous nous proposons de travailler par analogie avec ce qui existe pour le plan. Commençons par redonner les outils nécessaires au traitement d'images planes.

III.2 Outils existants sur le plan

Le traitement d'images classiques nécessite généralement d'effectuer des filtrages. Ces opérations s'effectuent mathématiquement par un produit de convolution entre l'image et un opérateur, et se calculent plus facilement avec la transformée de Fourier. Etant donné que nous implémentons la plupart des filtres dans le domaine de Fourier, nous allons tout d'abord rappeler la transformée de Fourier pour des images planes.

Soit $f \in (R^2)$. On définit la transformée de Fourier de f par :

$$\hat{f}(\varepsilon) = \int_{R^2} f(x) e^{-ix\varepsilon} dx. \quad (\text{III.7})$$

Notons que la transformée de Fourier de f est encore une fonction définie sur R^2 .

Etant donnée la formule de convolution sur le plan :

$$f * g(x) = \int_{R^2} f(x - y) g(y) dy. \quad (\text{III.8})$$

(avec f et g deux fonctions définies sur R^2), il est facile de montrer que $\widehat{f * g} = \hat{f} \hat{g}$.

Cette propriété est essentielle pour une implémentation rapide des opérateurs de filtrage.

De plus, la convolution sur le plan est commutative et associative.

III.3 Action de groupe sur la sphère

Un objet géométrique peut être défini comme un objet invariant par l'action d'un groupe. Nous allons donc commencer par rappeler brièvement quelques notions générales sur la théorie des groupes, avant de nous concentrer sur le cas qui nous intéresse : la sphère unité.

Pour plus d'information le lecteur pourra se référer à [31].

On dit qu'un groupe G agit sur l'ensemble X si à tout $g \in G$ correspond une action de $G \times X$ vers X qui au couple (g, x) associe l'élément $g(x)$ dans X et qui vérifie les propriétés suivantes :

- 1) pour tout x de X , on doit avoir $Id(x) = x$ ou Id est l'élément neutre de G
- 2) pour tout couple (g, g') d'éléments de G et pour tout x de X , il faut que $(gg')x = g(g'x)$

Par exemple, le groupe des isométries affines (rotations, translations) agit sur le plan R^2 .

Si pour tout couple (x, y) d'éléments de X , il existe g dans G tel que $gx = y$, alors on dit que l'action de G est transitive. Par exemple, il y a toujours une rotation qui envoie un point de la sphère sur un autre point donné.

Un espace X sur lequel agit un groupe de Lie transitif est appelé espace homogène. Rappelons qu'un groupe de Lie est une variété différentielle réelle ou complexe munie d'une structure de groupe, où les applications de multiplication et d'inversion sont différentiables.

Considérons tous les éléments h de G qui laissent a (un point fixé de X) invariant : $ha = a$. Ces éléments h constituent un sous-groupe H de G que l'on appelle stabilisateur ou sous-groupe stationnaire du point a . Quand G agit transitivement sur X , tous les sous-groupes stationnaires des points de X sont conjugués les uns aux autres. Le sous-groupe H devient le sous-groupe stationnaire de la classe $eH = H$. Dans ce cas, on note X par G/H .

Dans le cas de la sphère qui nous intéresse ici, G est $SO(3)$ le groupe des rotations autour du centre de la sphère. Ce groupe est isomorphe au groupe des matrices orthogonales d'ordre 3 et de déterminant 1.

$SO(3)$ agit transitivement sur la sphère S^2 , donc S^2 est un espace homogène pour $SO(3)$.

Soit $n_0 = (0, 0, 1)$ le pôle Nord de la sphère, le stabilisateur associé à ce point est le groupe $SO(2)$.

Par conséquent, on peut écrire $S^2 = SO(3)/SO(2)$.

De plus, nous avons les propositions suivantes :

Proposition 1 : $SO(n)$ est compact.

Proposition 2 : Tout groupe topologique localement compact G possède une mesure invariante à gauche ($\mu(gB) = \mu(B)$ pour toute partie borélienne B de G et pour tout $g \in G$), qui prend des valeurs finies et que l'on appelle mesure de Haar.

Nous ne donnerons pas ici la preuve de la proposition précédente sur l'existence de la mesure de Haar, mais signalons que c'est Haar [32] qui le premier a montré l'existence sous des hypothèses plus restrictives et von Neumann [33] a montré l'unicité à un scalaire près de cette mesure.

Dans le cas qui nous intéresse, nous en déduisons donc qu'il existe une mesure de Haar sur $SO(3)$.

De plus, comme le rappelle Vilenkin [34], on peut également définir une mesure de Haar sur S^2 .

En effet, si H est un sous-groupe localement compact d'un groupe compact G , alors par l'invariance à gauche de la mesure dg sur G , on peut définir une mesure dx sur l'espace homogène $X = G/H$ en posant $\int_X f(x)dx = \int_G f(\phi(g))dg$ où $\phi : g \rightarrow gH$ est une application définie de G vers X . La mesure dx ainsi définie est également invariante [31].

Définition 1 : Une mesure μ sur X est dite invariante à gauche par rapport au groupe G , qui agit sur X si $\mu(x) = \mu(gx)$ pour tout $x \in X$ et pour tout $g \in G$. Cela implique que $\int_X f(x) d\mu(x) = \int_X f(g(x)) d\mu(x)$ pour tous $g \in G$ et pour toute fonction f continue à support compact sur X .

Pour ce qui nous concerne, $H = SO(2)$ et $G = SO(3)$ qui sont tous les deux des groupes compacts d'après la proposition vue précédemment, donc l'existence d'une mesure de Haar sur S^2 est bien justifiée.

D'un point de vue pratique, pour le groupe $G = SO(3)$ que nous considérons, nous allons paramétrer les rotations par les angles d'Euler.

Soient $R \in SO(3)$, et α, β, γ les angles d'Euler avec α et $\gamma \in [0, 2\pi]$ et $\beta \in [0, \pi]$.

$$R(\alpha, \beta, \gamma) = R_z(\alpha)R_y(\beta)R_z(\gamma) \quad (\text{III.9})$$

Où R_z et R_y sont les rotations par rapport aux axes z et y respectivement.

$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ et } R_y = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \quad (\text{III.10})$$

$dR = \sin\beta d\gamma d\beta d\alpha$ représente alors la mesure de Haar sur $SO(3)$ que nous utiliserons par la suite. On vérifie aisément que c'est une mesure invariante par rotation.

Etant donné les résultats que nous venons de rappeler, nous pouvons maintenant étudier plus précisément ce qui se passe sur la sphère unité.

III.4 Harmoniques sphériques

Rappelons que l'on paramètre la sphère unité, plongée dans \mathbb{R}^3 , en utilisant les coordonnées sphériques : soit $\eta \in S^2$

$$\eta(\theta, \varphi) = (\cos(\varphi) \sin(\theta), \sin(\varphi) \sin(\theta), \cos(\theta)), \quad (\text{III.11})$$

avec $\varphi \in [0, 2\pi]$, angle de longitude et $\theta \in [0, \pi]$, angle de colatitude (latitude + $\pi/2$).

Par conséquent, l'intégration d'une fonction $f \in L^2(S^2)$ (l'espace de Lebesgue sur la sphère) est définie par :

$$\int_{\eta \in S^2} f(\eta) d\eta = \int_0^{2\pi} \int_0^\pi f(\theta, \varphi) \sin(\theta) d\theta d\varphi. \quad (\text{III.12})$$

A partir du paramétrage des rotations avec les angles d'Euler, on obtient facilement l'égalité suivante :

$$\int_{SO(3)} f(Rn_0) dR = 2\pi \int_{S^2} f(\eta) d\eta. \quad (\text{III.13})$$

Où $n_0 = (0, 0, 1)$ représente le pôle Nord de la sphère unité.

Dans ce qui suit, nous procédons comme pour le plan, pour définir une transformée de Fourier sphérique.

Les harmoniques sphériques $Y_{lm} : S^2 \rightarrow \mathbb{C}$ sont des fonctions propres de l'opérateur de Laplace sphérique (ou opérateur de Laplace-Beltrami) Δ_{S^2} :

$$\text{Avec } \Delta_{S^2} = \frac{1}{\sin(\theta)} \frac{\partial}{\partial \theta} \left(\sin(\theta) \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2(\theta)} \frac{\partial^2}{\partial \varphi^2}, \quad (\text{III.14})$$

$$\Delta_{S^2} Y_{lm} = -l(l+1) Y_{lm}. \quad (\text{III.16})$$

Soit Y_{lm} l'harmonique sphérique de degré l et d'ordre m :

$$Y_{lm}(\theta, \varphi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos(\theta)) e^{im\varphi} \text{ pour } m \geq 0, \quad (\text{III.17})$$

$$Y_{lm}(\theta, \varphi) = (-1)^m \overline{Y_{l,-m}(\theta, \varphi)} \text{ pour } m \leq 0. \quad (\text{III.18})$$

Où $P_l^m(x)$ sont les polynômes de Legendre associés de degré l et d'ordre m .

Les fonctions $P_l^m(x)$ se construisent à partir des polynômes de Legendre P_l qui vérifient la formule de Rodrigues suivante :

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l. \quad (\text{III.19})$$

Par conséquent, on définit les polynômes de Legendre associés par :

$$P_l^m(x) = (-1)^m (1-x^2)^{\frac{m}{2}} \frac{d^m}{dx^m} P_l(x). \quad (\text{III.20})$$

En pratique, il suffit de calculer $P_l^m(x)$ pour $m \geq 0$ étant donné que l'on a :

$$P_l^{-m}(x) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(x). \quad (\text{III.21})$$

Les harmoniques sphériques de degré l forment un sous-espace de $L^2(S^2)$ de dimension $2l+1$ qui est invariant sous l'action des rotations (autour de l'axe z) de la sphère [31].

Remarquons que :

$$\int_0^{2\pi} \int_0^\pi Y_{lm}(\theta, \varphi) \overline{Y_{l'm'}(\theta, \varphi)} \sin(\theta) d\theta d\varphi = \delta_{l,l'} \delta_{m,m'}. \quad (\text{III.22})$$

Où $\delta_{l,l'}$ vaut 1 si $l=l'$, et 0 sinon.

Etant donné que les harmoniques sphériques forment une base ortho normale pour $L^2(S^2)$, chaque fonction $f \in L^2(S^2)$ s'écrit à l'aide des harmoniques sphériques sous la forme suivante :

$$f = \sum_{i \in \mathbb{N}} \sum_{|m| \leq l} \widehat{f}_{lm} \widehat{Y}_{lm}, \quad (\text{III.23})$$

Avec :

$$\widehat{f}_{lm} = \hat{f}(l, m) = \langle f, Y_{lm} \rangle,$$

Où le produit scalaire sur la sphère est défini par :

$$\langle f, h \rangle = \int_0^{2\pi} \int_0^\pi f(\theta, \varphi) \overline{h(\theta, \varphi)} \sin(\theta) d\theta d\varphi. \quad (\text{III.24})$$

L'ensemble des coefficients \widehat{f}_{lm} est appelé transformée de Fourier sphérique ou spectre de f .

Le théorème de Parseval se généralise de la façon suivante :

$$\int_0^{2\pi} \int_0^\pi f^2(\theta, \varphi) \sin(\theta) d\theta d\varphi = \sum_{l=0}^{+\infty} \sum_{m=-l}^l \widehat{f}_{lm}^2, \quad (\text{III.25})$$

$$\int_0^{2\pi} \int_0^\pi f(\theta, \varphi) g(\theta, \varphi) \sin(\theta) d\theta d\varphi = \sum_{l=0}^{+\infty} \sum_{m=-l}^l \widehat{f}_{lm} \widehat{g}_{lm}. \quad (\text{III.26})$$

Définissons maintenant les fonctions à bande-limitée (ou à spectre borné).

On dit qu'une fonction $f \in L^2(S^2)$ est à bande-limitée de largeur de bande $B \geq 0$ si $\widehat{f}_{lm} = 0 \forall l \geq B$:

$$f(\theta, \varphi) = \sum_{l=0}^{B-1} \sum_{m=-l}^l \widehat{f}_{lm} Y_{lm}(\theta, \varphi). \quad (\text{III.27})$$

Notons une propriété intéressante de ces fonctions : la rotation d'une fonction bande-limitée est aussi une fonction à bande-limitée avec la même largeur de bande.

Nous venons d'introduire la notion de transformée de Fourier sphérique, définie comme dans le cas du plan à l'aide des solutions de l'équation de Laplace. Dans la partie qui suivra, nous présenterons un autre outil mathématique sphérique indispensable au traitement d'images : le produit de convolution.

III.5 Convolution sur la sphère

Nous présentons ici deux définitions qui ont été implémentées pour effectuer un produit de convolution sur la sphère : celle proposée par Driscoll et Healy et celle utilisée, entre autres, par Daniilidis et Wandelt.

Commençons par quelques notations.

On représente la sphère S^2 comme le quotient $SO(3)/SO(2)$ Où $SO(3)$ est le groupe des rotations qui agit sur la sphère.

La rotation d'une fonction f définie sur la sphère par un élément $g \in SO(3)$ est alors définie avec l'opérateur Λ_g tel que :

$$\Lambda_g f(\eta) = f(g^{-1}\eta) \forall \eta \in S^2. \quad (\text{III.28})$$

Soit $L^2(SO(3)) \equiv L^2(SO(3), dR)$ l'espace de Lebesgue sur $SO(3)$ où dR est la mesure de Haar sur $SO(3)$ vue précédemment.

Définition 2 Soit $f \in L^2(S^2)$ et $g \in L^1(S^2)$, la convolution entre f et g est la fonction sur $SO(3)$ définie par :

$$(f \tilde{*} g)(R) = \int_{S^2} f(R^{-1}\eta)g(\eta)d\eta. \quad (\text{III.29})$$

On a :

$$(f \tilde{*} g) \in L^2(SO(3)),$$

Et

$$\|f \tilde{*} g\|_2 \leq \|f\|_2 \|g\|_1.$$

Un problème important se pose : alors que les fonctions f et g sont définies sur la sphère, le produit de convolution est défini sur le groupe des rotations, par conséquent, il est évident que le produit de convolution n'est pas associatif. De plus $f \tilde{*} \delta_0(R) = f(R^{-1}n_0)$, où δ_0 est défini de la façon suivante :

$$f(n_0) = \int_{S^2} f(\eta) \delta_0(\eta) d\eta \quad (\text{III.30})$$

D'où un problème de symétrie dans cette définition du produit de convolution.

Voyons maintenant une autre définition de la convolution, introduite par Driscoll et Healy.

Définition 3 Soit $f, h \in L^2(S^2)$, on a :

$$(f * g)(\eta) = \int_{SO(3)} f(Rn_0)h(R^{-1}\eta)dR, \quad (\text{III.31})$$

Où $\eta \in S^2, n_0 = (0, 0, 1)$ représente le pôle Nord de la sphère unité.

Pour arriver à cette définition, on peut suivre le raisonnement suivant :

Considérons la convolution de deux fonctions f et h définies sur R comme

$$(f * g)(x) = \int_R f(x') h(x - x') dx'. \quad (\text{III.32})$$

Introduisons l'opérateur de translation $T_{x'}$ tel que $T_{x'}(x) = x - x'$, alors

$$(f * g)(x) = \int_R f(T_{x'}(0))h(T_{x'}^{-1}(x))dx. \quad (\text{III.33})$$

Sur la sphère, ce sont les rotations qui agissent alors que ce sont les translations qui opèrent sur la droite réelle.

On identifie aussi le pôle Nord dans le cas sphérique avec l'origine dans le cas linéaire. On retrouve ainsi une analogie entre la convolution sur la droite réelle et la convolution définie sur la sphère par Driscoll et Healy.

Théorème 3 Soient $f, h \in L^2(S^2)$, en utilisant la définition 2, on a

$$(\widehat{f * g})_{lm} = 2\pi \sqrt{\frac{4\pi}{2l+1}} \hat{f}_{lm} \hat{h}_{l0}. \quad (\text{III.34})$$

On peut remarquer que seuls les coefficients \hat{h}_{lm} avec $m = 0$ interviennent dans le théorème.

Ces coefficients correspondent aux harmoniques zonales Y_{lm} et représentent donc la partie invariante par rotation autour de l'axe (Oz) du filtre h .

$$Y_{l0}(\eta) = \sqrt{\frac{2l+1}{4\pi}} P_l(\cos(\theta)); \quad P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l. \quad (\text{III.35})$$

Notons également que ce théorème prouve que le produit de convolution défini par Driscoll et Healy n'est pas commutatif, ce qui est logique étant donné que $SO(3)$ n'est pas un groupe commutatif [31].

Cependant, en utilisant la définition même de la convolution et par un simple changement de variable, on peut montrer que cette convolution est associative. Cette propriété interviendra par exemple dans la construction de filtre, puisque convoluer une image bruitée revient à additionner la convolée de l'image originale et la convolée du bruit.

Une autre différence avec la convolution sur la droite réelle est que l'on a ici

$$f * \delta_0(\eta) = 2\pi f(\eta) \quad (\text{III.36})$$

Pour des applications, la deuxième définition du produit de convolution permet une construction théorique des filtres plus simple et une implémentation rapide grâce au théorème de convolution. Dorénavant et sauf mention explicite du contraire, le produit de convolution sphérique désignera par défaut le produit de convolution de Driscoll-Healy.

Discussion : les différents outils ainsi définis seront nécessaires pour la construction des différents opérateurs sur la sphère qui seront définis dans la deuxième partie du quatrième chapitre.

Chapitre IV

Traitement des images omnidirectionnelles

Intorduction

Les systèmes de vision omnidirectionnelle sont de plus en plus utilisés pour la surveillance, la navigation de robots ou encore la vidéoconférence. Le problème principal avec les images obtenues par ces systèmes est leur différence de résolution: élevée à la périphérie et faible au centre de l'image. Appliquer directement les opérateurs classiques dédiés aux images perspectives introduira donc par essence des erreurs dans le traitement des images omnidirectionnelles. Le traitement des images catadioptriques en passant par une sphère virtuelle, objet mathématique sous-jacent à ces images, permet de prendre en compte la spécificité géométrique de ces images. Il existe donc de nouveaux opérateurs de traitement (lissage, détection de contours...) d'images sphériques qui agiront de la même façon quelque soit l'emplacement du noyau de l'opérateur sur la sphère. Cela permet donc d'obtenir des outils bas-niveaux adaptés aux images omnidirectionnelles.

Le schéma de traitement des images omnidirectionnelles est représenté sur la figure suivante

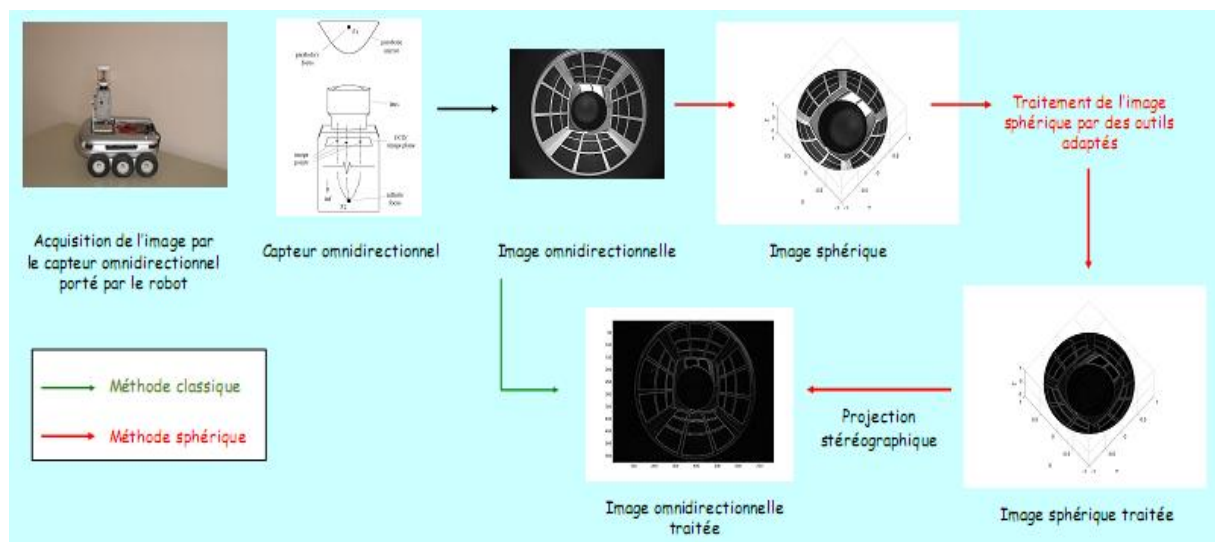


Figure IV.1 -Processus de traitement d'une image omnidirectionnelle.

IV.1 Utilisation des traitements classiques

Si la caméra catadioptrique accroît le champ de vision, en revanche, la géométrie des images formées est relativement complexe. Comme nous l'avons vu dans l'introduction, la résolution des images panoramiques varie en fonction de la position dans l'image. Certains auteurs pensent que les outils classiques donnent de bons résultats et qu'il est inutile d'augmenter le temps de calcul pour gagner en précision. Les images omnidirectionnelles sont alors traitées de la même manière que les images perspectives.

Dans l'article [35] par exemple, Vlasilis, Motomura, Hara et Asoh travaillent sur la détection de contours. Leur objectif est d'extraire l'information de bas-niveau d'une séquence d'images omnidirectionnelles pour la localisation et la navigation d'un robot mobile. La méthode proposée suit trois étapes : la détection de contours dans l'image, l'estimation de Parzen de la densité spatiale de ces contours et l'analyse des composantes principales. Pour la détection de contours, deux filtres de Sobel décalés de 90° sont utilisés. La taille de ces filtres est toujours la même quelle que soit leur position dans l'image. L'image omnidirectionnelle est donc traitée de la même manière qu'une image perspective. Cette méthode fournit de bons résultats; cependant, elle n'est pas robuste aux phénomènes d'occultation, ce qui engendre de sérieux problèmes en situation réelle.

Gonzalez-Barbosa et Lacroix [36] présentent un moyen de rectifier un banc stéréoscopique panoramique de sorte que les lignes épipolaires soient parallèles. Cette méthode permet d'exécuter un algorithme optimisé de stéréovision par corrélation de pixels. La méthode proposée est indépendante de la configuration du banc stéréo. Les équations qui permettent de faire de la reconstruction 3D en utilisant des caméras panoramiques ont été déterminées. Dans la plupart des modèles proposés dans la littérature, l'alignement entre la caméra et le miroir est supposé parfait. Dans cet article, une caractérisation complète d'une caméra catadioptrique avec miroir parabolique et lentille télécentrique a été réalisée. L'originalité de ce travail réside dans la prise en compte des paramètres intrinsèques du système (surface du miroir et paramètres intrinsèques de la caméra), des erreurs d'alignement entre le miroir et le repère caméra et de la distorsion optique radiale. Pour l'extraction des indices visuels, un extracteur de croix adapté aux images perspectives a été utilisé. Les auteurs soulignent l'importance d'améliorer la précision de ce détecteur en développant des extracteurs adaptés aux images panoramiques.

Dans [37], Demonceaux et Kachi-Akkouche s'intéressent au problème d'estimation de mouvement dans une séquence d'images acquise avec un capteur catadioptrique. Généralement, le lien entre les gradients d'intensités et les composantes du champ de vitesse est donné par contrainte du flot optique. Cette contrainte repose sur l'hypothèse que le flot est localement constant sur un voisinage. Cependant, les distorsions introduites par le capteur invalident cette hypothèse. Les auteurs ont donc développé une nouvelle méthode de décomposition de l'équation de contrainte du mouvement apparent dans une base d'ondelettes. La base d'ondelettes discrète permet de résoudre le problème d'ouverture généralisée et de diminuer le temps de calcul. Pour prendre en compte les déformations créées par le capteur, l'hypothèse de constance du flot optique pour un voisinage de points, utilisée

pour les images perspectives, est remplacée par une hypothèse d'affinité du mouvement dans ce voisinage. Cette hypothèse permet d'obtenir de meilleurs résultats ; cependant, elle ne caractérise pas correctement les déformations non-linéaires des images omnidirectionnelles, tenant pas compte de la géométrie du miroir.

IV.2 Adaptation des traitements

Les images omnidirectionnelles, présentent des déformations. Nombreux auteurs ont souligné les limites des approches classiques et l'importance d'adapter les traitements à leur géométrie. Dans le cadre de l'imagerie omnidirectionnelle, plusieurs solutions ont été proposées afin d'obtenir des résultats significatifs. Cependant, les nouveaux algorithmes créés sont souvent partiellement adaptés aux images. On peut distinguer deux principales catégories d'adaptations :

- l'utilisation du miroir afin d'obtenir des informations sur les distorsions de la scène observée.
- la projection et le traitement de l'image dans l'espace de la sphère,

IV.2.1 Utilisation de la géométrie du miroir

Ieng, Benosman et Devars ont étudié la mise en correspondance de points dans des images acquises avec deux caméras catadioptriques ayant des points de vue différents. La mise en correspondance d'images consiste dans un premier temps à extraire des primitives (points, lignes, courbes) caractéristiques des images et à faire une association entre les deux images, le plus souvent à l'aide de fonctions de corrélation ou de similarité. La non-uniformité de l'espace des images omnidirectionnelles nécessite une adaptation de ce traitement. Dans leur article [38], l'algorithme proposé s'effectue suivant quatre étapes. Tout d'abord, les points d'intérêt sont extraits avec un détecteur de Harris. Les auteurs précisent que ce détecteur ne tient pas compte de la géométrie des images catadioptriques acquises avec un miroir hyperbolique et qu'une approche mieux adaptée serait souhaitable. Les points détectés sont ensuite projetés sur le miroir hyperbolique. Des quadrilatères de différentes tailles en forme de diamant, centrés autour de chacun des points projetés, sont définis. Chaque diamant représente un angle solide correspondant à une portion de l'espace projeté sur le miroir hyperbolique (Figure IV.2(c)). Les sommets des diamants sont projetés sur l'image omnidirectionnelle (Figure IV.2(a)). La projection des diamants permet d'éliminer les faux appariements entre les points des deux images. Toutes les combinaisons possibles entre les diamants des différentes tailles des deux images sont analysées (Figure IV.2(b)). Le vainqueur est celui auquel on accorde le plus grand coefficient de confiance. Cette méthode permet d'estimer la géométrie épipolaire et de faire de la reconstruction 3D. Les expérimentations ont

mis en évidence la fiabilité et la robustesse de cet algorithme qui fournit des résultats similaires à ceux obtenus avec des images perspectives.

Avec un capteur catadioptrique central, la contrainte du point de vue unique est vérifiée. A partir de l'image anamorphosée (Figure IV.3), il est donc possible de reconstituer n'importe quelle image perspective de la scène observée (Figure IV.4). Sur un plan choisi dans champ de vision du capteur, on définit un maillage (échantillonnage de l'image), puis on calcule la valeur de niveau de gris de chaque point du maillage par interpolation.

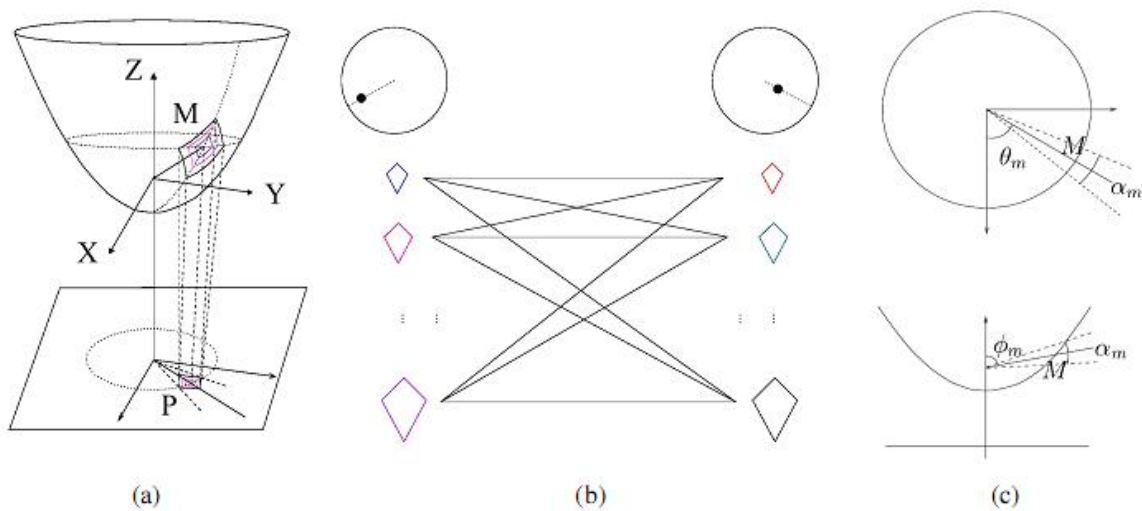


Figure IV.2– (a) Définition des patches sur le miroir puis projection sur l'image omnidirectionnelle.(b) Les pixels sont appareillés lorsque leurs patches de différentes tailles ont le plus grand coefficient de confiance. (c) Définition de l'angle solide.



Figure IV.3 -Image anamorphosée.

Pour traiter une image omnidirectionnelle, il est donc possible de recréer toutes les images perspectives issues de cette image et de leur appliquer le traitement désiré. Dans certains domaines d'application où les utilisateurs des images ne sont pas habitués aux déformations des images omnidirectionnelles, par exemple l'inspection, la vidéo-surveillance ou la visio-conférence, la restitution d'images perspectives est souhaitable. Cependant, il existe une infinité d'images perspectives ; autant que de lignes radiales dans l'image omnidirectionnelle. Un très grand nombre d'images perspectives devra être reconstruit, ce qui rend le processus très lourd. De plus, le choix arbitraire du maillage de l'image perspective a évidemment une forte influence sur les valeurs interpolées des niveaux de gris des pixels. La partie extérieure de l'image aura donc une bonne résolution, mais la perte d'information dans la partie centrale de l'image produira une forte pixellisation dans l'image perspective obtenue.

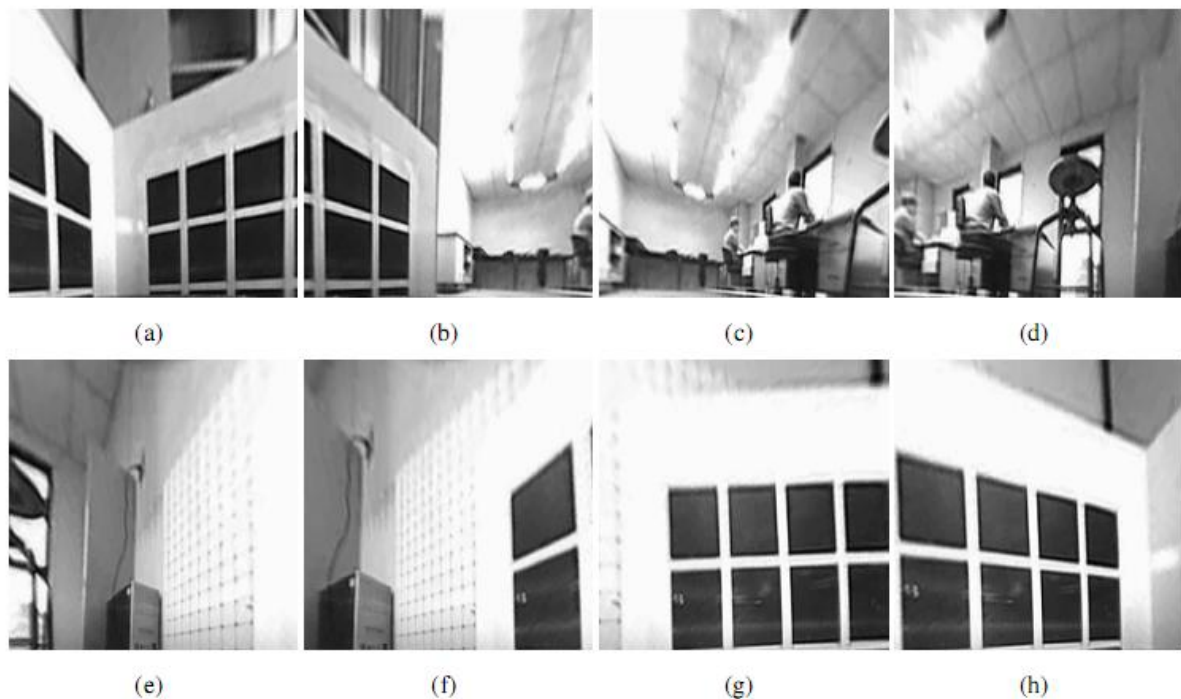


Figure IV.4– *Reconstruction d'une image perspective.*

Cependant, la reconstruction des images perspectives ne permet qu'une vision partielle de la scène dans chaque image. Dans la littérature, la projection sur le cylindre englobant le miroir de révolution est fréquemment utilisée pour visualiser l'ensemble de l'image (Figure IV.5). La Figure IV.6 représente l'image obtenue par projection sur le cylindre.

Bunschoten et Krose [39] proposent une méthode de reconstruction 3D de l'environnement à partir d'une séquence d'images panoramiques. Cette méthode est basée sur des lignes de niveaux. Les images sont acquises avec un capteur catadioptrique monté sur un robot mobile puis projetée sur un cylindre virtuel. Les lignes de références stéréo utilisent la redondance

contenue dans ces images. Dans leur article [40], Pajdla et Hlavàc s'intéressent à la localisation dans une image panoramique. Pour se faire, ils proposent de déplier l'image omnidirectionnelle sur le cylindre. Lorsque la caméra effectue une rotation, la scène observée sur l'image cylindrique est la même, à un décalage près. Lors d'une translation, la caméra change de position. Il est alors impossible de trouver une transformation (rotation et/ou translation) permettant de mettre en correspondance ces deux images. En effet, on constate une variation de la profondeur, ainsi que des phénomènes d'occultation de la scène. Ils proposent d'éviter cette étape de recherche de l'orientation en trouvant une représentation invariante par translation pour chaque classe d'image acquises à la même position.

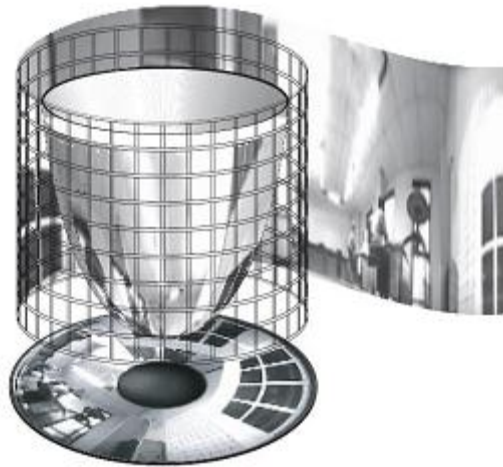


Figure IV.5- *Projection de l'image sur le cylindre englobant.*



Figure IV.6- *Projection de l'image sur le cylindre englobant.*

L'espace du cylindre permet une vision panoramique de la scène, très proche de la vue perspective. C'est un support avantageux car il est possible de l'échantillonner avec une grille régulière dans un repère cylindrique. Cette projection est très lisible par l'œil humain. Cependant, l'image reste anamorphosée au sens où des lignes droites perpendiculaires à l'axe de rotation du miroir ne se projettent pas comme des droites sur l'image projetée. Les lignes

horizontales dans la scène 3D sont donc déformées dans l'image cylindrique. De plus, comme pour l'image perspective, la reconstruction génère une interpolation des valeurs des niveaux de gris des pixels, liée au lissage et à l'échantillonnage. L'information contenue dans l'image est donc modifiée et l'application d'un traitement ajoute des imprécisions supplémentaires.

Le fait que la géométrie du capteur joue un rôle central pour les traitements des images et a servi de fil directeur à Bogdanova, Bresson, Thiran et Vanderghelynst [41] dans leur méthode de segmentation. La contrainte géométrique pour segmenter des images catadioptriques demeure importante. Une nouvelle fonction d'énergie est introduite de manière à ce que les contours actifs obtenus prennent en compte les distorsions de l'image. Pour la classe des surfaces de Riemann étudiées, les calculs comportent seulement des différences finies euclidiennes standard sur le plan de l'image. Une autre approche proposée par Demonceaux et Vasseur dans [42], utilisent les champs aléatoires de Markov, sur des images catadioptriques. La topologie d'un voisinage de pixels étant modifié par la projection sur le miroir, un nouveau voisinage est défini pour les champs aléatoires de Markov, à l'aide du modèle de projection unifié. Avec ce système de voisinage, les interactions entre deux pixels diffèrent s'ils sont situés à la périphérie ou au centre de l'image. Cependant, cette technique peut poser des problèmes liés à la minimisation markovienne d'énergie, en termes de temps de calcul. Des algorithmes déterministes, plus rapides, sont mis en place, mais ne permettent pas de trouver de minimum global.

Quant à Strauss et Comby [27], ils proposent de projeter, sur chaque pixel de l'image omnidirectionnelle, un noyau dont la géométrie est définie dans l'espace du cylindre en utilisant les paramètres d'étalonnage du système. Des opérateurs morphologiques flous à noyaux variables pour images omnidirectionnelles permettent de compenser les imprécisions liées à l'échantillonnage de l'image et la localisation du voisinage. Cette méthode est utilisée pour le traitement morphologique d'image omnidirectionnelle.

IV.2.2 Projection de l'image sur la sphère

Dans [43], Daniilidis, Makadia et Bulow travaillent sur l'estimation de mouvement pour des images omnidirectionnelles. Ils proposent de projeter l'image omnidirectionnelle sur une sphère virtuelle (Figure IV.7) pour définir l'opérateur gradient. La fonction Gaussienne est introduite sur la sphère et permet de calculer les dérivées par convolution avec ses dérivées spatio-temporelles. Les dérivées sur la sphère sont utilisées pour estimer le flot optique en conservant l'échantillonnage de l'image omnidirectionnelle. En effet, la projection transforme l'échantillonnage régulier de l'image omnidirectionnelle en un échantillonnage non-régulier

sur la sphère. Les convolutions sont effectuées sur la sphère mais leur réalisation préserve les échantillons discrets de l'image catadioptrique originale. Cependant, l'imprécision liée à cet échantillonnage irrégulier n'est pas prise en compte. Les résultats ont montré une amélioration sur le calcul de la direction et la qualité du champ de vecteur pour une même densité.

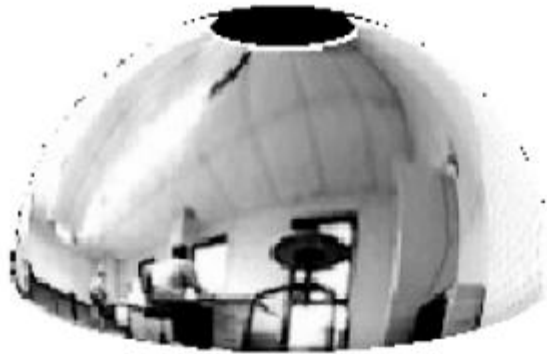


Figure IV.7– *Projection de l'image sur l'hémisphère.*

Dans le même esprit, Shakernia, Vidal et Sastry [44] généralisent la notion de rétroprojection pour l'estimation du flot optique avec des caméras catadioptriques centrales. La projection sur un capteur catadioptrique central peut être considérée comme une projection sur une rétine incurvée virtuelle dépendante de la géométrie de la caméra. Le flot optique est obtenu en retro-projetant les points de l'image sur une rétine incurvée virtuelle. Cette méthode donne de meilleurs résultats que ceux obtenus avec l'hémisphère présenté dans [33] et les auteurs suggèrent qu'il existe un espace correct qui dépend du mouvement de la caméra. La rétine virtuelle proposée dans cet article est le support le plus approprié lorsque la caméra évolue dans le plan parallèle à la rétine. Pour appuyer cette conclusion, des tests ont été proposés pour différents mouvements de la caméra, différents paramètres de la caméra catadioptrique centrale et différents niveaux de bruit.

Bigot, Kachi, Durand et Mouaddib [31] proposent de créer les opérateurs sur la sphère virtuelle de [43]. L'image omnidirectionnelle est présentée comme une fonction de la sphère. Une nouvelle méthode de lissage est décrite. Il s'agit d'une adaptation du filtre de Wiener suivi d'une régularisation de Tikhonov. Cette approche est comparée avec les méthodes de convolution des noyaux usuels sur la sphère pour des images de synthèses et des images réelles. Le filtre est invariant par rotation et les résultats sont indépendants de la position sur la sphère. De plus, cette méthode donne un bon rapport signal sur bruit et peut être adaptée en temps réel.

Dans la continuité de leurs travaux [45], Daniilidis et Makadia considèrent que les images catadioptriques centrales peuvent être projetées sur la sphère. Ils proposent une méthode permettant de déterminer les paramètres invariants pour des mouvements de rotation 3D par généralisation de la transformée de Fourier (connue pour la sphère sous le nom de transformée en harmonique sphérique). Les angles d'Euler de ces rotations sont calculés par une généralisation du théorème de décalage. Le mouvement apparent est correctement détecté pour les images omnidirectionnelles et cet algorithme est capable de détecter correctement de faibles mouvements de rotation (il n'y a pas de confusion possible avec de faibles mouvements de translation). Cependant, comme nous venons de le dire, l'échantillonnage des images omnidirectionnelles sur la sphère n'est pas régulier. Une interpolation des valeurs des niveaux de gris de ces images est donc nécessaire pour obtenir un échantillonnage uniforme en coordonnées sphériques traditionnelles.

Geyer, Sastry et Bajcsy [46] présentent une méthode d'estimation de mouvement pour les caméras omnidirectionnelles. Cette méthode est analogue à une transformation de Radon ou une transformation de Hough. Elle combine les notions de géométrie, traitement du signal et analyse harmonique. Dans cet article, les auteurs démontrent que la fonction de vraisemblance peut être obtenue dans l'espace des matrices essentielles par convolution de deux signaux. Le premier signal exprime la géométrie épipolaire de deux points de vue, et le second encode les similitudes des intensités entre deux pixels appartenant à deux images différentes. Le calcul de la convolution en utilisant les rotations harmoniques sphériques et les représentations des signaux permet de déterminer les ambiguïtés lors de la détection et de prendre en compte des bruits d'acquisition non-gaussien des capteurs. Etant donné les déformations importantes contenues dans les images omnidirectionnelles, Tosic, Bogdanova, Frossard et Vandergheynst [47] ont également choisi de travailler dans le domaine sphérique pour réaliser une estimation locale du mouvement. Les images acquises ont des points de vue différents. Une approche de multi-résolution permettant d'associer les angles solides des deux images sphériques a été proposée afin d'améliorer l'exactitude du mouvement détecté, et de limiter la complexité informatique. La détection locale du mouvement est assez efficace. L'erreur résiduelle est relativement faible et souvent à proximité des contours ou des composantes hautes fréquences dans l'image prédite.

IV.3 Généralités sur le traitement des images définies sur la sphère

Nous présenterons dans cette partie quelques opérateurs de filtrage sur la sphère, dans un premier temps pour le débruitage, puis à la détection de contours sur les images sphériques.

IV.3.1 Opérateur de filtrage sur la sphère

Dans ce qui suit, nous noterons, L l'opérateur de filtrage et h la réponse impulsionnelle du filtre sphérique.

Comme il a été vu précédemment, le groupe des rotations $SO(3)$ agit sur la sphère S^2 . De plus, avec le produit de convolution choisi, $\forall R \in SO(3)$, on a

$$(f \circ R) * g = (f * g) \circ R. \quad (\text{IV.1})$$

En effet par un changement de variable, on montre que

$$\begin{aligned} (f \circ R) * g(\eta) &= \int_{SO(3)} f \circ R(\rho n_0) g(\rho^{-1} \eta) d\rho = \int_{SO(3)} f(\gamma n_0) g(\gamma^{-1} R \eta) d\gamma \\ &= ((f * g) \circ R)(\eta) \end{aligned} \quad (\text{IV.2})$$

Par conséquent, si on définit notre opérateur de filtrage sphérique par

$$L(f) = \frac{1}{2\pi} f * h, \quad (\text{IV.3})$$

Et si on note $\gamma_R(f)(\eta) = f(R^{-1}\eta)$, l'opérateur de rotation, on a

$$\begin{aligned} L\gamma_R(f)(\eta) &= \frac{1}{2\pi} \gamma_R(f) * h(\eta) = \frac{1}{2\pi} (f \circ R^{-1}) * h(\eta) = \frac{1}{2\pi} (f * h) \circ R^{-1}(\eta) \\ &= \gamma_R(L(f))(\eta). \end{aligned} \quad (\text{IV.4})$$

Donc l'opérateur de filtrage est invariant sous les actions des rotations. Une différence est à noter cependant. Etant donné qu'ici le produit de convolution n'est plus commutatif.

On a donc

$$\widehat{L(f)}(l, m) = \sqrt{\frac{4\pi}{2l+1}} \hat{f}(l, m) \hat{h}(l, 0). \quad (\text{IV.5})$$

IV.3.2 Gaussiennes sphériques

On se propose de trouver l'équivalent de la gaussienne sur la sphère. Plusieurs constructions sont disponibles dans la littérature. Commençons par celle que nous allons privilégier dans la suite.

Fonction de Green

Thomas Bülow [48] a proposé de déterminer la fonction de Green, en tant que solution de l'équation de diffusion sphérique, puis de considérer cette fonction comme l'extension de la gaussienne sur la sphère. Cette construction est intuitive puisque sur le plan, la gaussienne est solution de l'équation de diffusion.

L'équation de diffusion sphérique est

$$\Delta_{S^2} u(\theta, \varphi, t) = \frac{1}{k} \frac{\partial}{\partial t} u(\theta, \varphi, t), \quad (\text{IV.6})$$

avec Δ_{S^2} l'opérateur de Laplace-Beltrami vu précédemment, (θ, φ) les coordonnées sphériques, k le coefficient de diffusion et t représentera un paramètre d'échelle.

Rappelons que les harmoniques sphériques sont les fonctions propres de l'opérateur de Laplace sphérique :

$$\Delta_{S^2} Y_{l,m} = -l(l+1)Y_{l,m}. \quad (\text{IV.7})$$

Par conséquent, les fonctions

$$u_{l,m}(\theta, \varphi, t) = Y_{l,m}(\theta, \varphi) \exp(-l(l+1)kt) \quad (\text{IV.8})$$

Sont solutions de l'équation de diffusion sphérique.

Pour obtenir la fonction de Green G , on impose comme condition initiale :

$$u(\theta, \varphi, 0) = G(\theta, \varphi, 0) := \delta_0(\theta, \varphi). \quad (\text{IV.9})$$

En utilisant la décomposition du Dirac sphérique dans la base des harmoniques sphériques, on obtient :

$$G(., 0) = \delta_0 = \sum_{l \in \mathbb{N}} \sqrt{\frac{2l+1}{4\pi}} Y_{l,0} = \sum_{l \in \mathbb{N}} \sqrt{\frac{2l+1}{4\pi}} u_{l,0}(., 0). \quad (\text{IV.10})$$

D'où finalement pour la fonction de Green

$$\begin{aligned}
G(\theta, \varphi, t) &= \sum_{l \in \mathbb{N}} \sqrt{\frac{2l+1}{4\pi}} u_{l,0}(\theta, \varphi, t) = \sum_{l \in \mathbb{N}} \sqrt{\frac{2l+1}{4\pi}} Y_{l,0}(\theta, \varphi) e^{-l(l+1)kt} \\
&= \sum_{l \in \mathbb{N}} \frac{2l+1}{4\pi} P_1(\cos(\theta)) e^{-l(l+1)kt}.
\end{aligned} \tag{IV.11}$$

On effectue ensuite la transformée de Fourier sphérique pour trouver :

$$G(., t)(l, m) = \begin{cases} \sqrt{\frac{2l+1}{4\pi}} e^{-l(l+1)kt} & \text{si } m = 0, \\ 0 & \text{sinon.} \end{cases} \tag{IV.12}$$

Gaussienne sphérique

En partant de la définition de la gaussienne sur le plan, on peut obtenir la gaussienne sphérique en opérant une projection stéréographique inverse :

$$G_S(\theta, \varphi, t) = e^{-(\tan^2(\frac{\theta}{2}))/t}. \tag{IV.13}$$

Noyau de Poisson

Le noyau de Poisson est donné par la fonction

$$P(\theta, \varphi, h) = \sum_{n \in \mathbb{N}} (2n+1) h^n P_n(\cos(\theta)) \tag{IV.14}$$

Cette fonction est solution de l'équation

$$\Delta_{S^2} P(\theta, \varphi, h) = -h \frac{\partial^2}{\partial h^2} (h P(\theta, \varphi, h)). \tag{IV.15}$$

La transformée de Fourier de cette fonction est donnée par

$$\widehat{P(., h)}(l, m) = \begin{cases} \sqrt{(2l+1)4\pi} h^l & \text{si } m = 0 \\ 0 & \text{sinon.} \end{cases} \tag{IV.16}$$

IV.3.3 Lissage

On utilise l'opération de lissage dans plusieurs cas. Dans le cas de débruitage de l'image, ce traitement a pour but d'améliorer la qualité de l'image et de retrouver l'image originale à partir de l'image bruitée. La deuxième utilisation importante du lissage d'images sert à les simplifier en enlevant par exemple des détails insignifiants. Le filtre de Wiener est un outil

classique de lissage pour les images planes, nous donnons dans ce qui suit un filtre de Wiener adapté aux images sphériques.

IV.3.3.1 Filtre de Wiener

Soit $f \in L^2(S^2)$, notre image originale, bruitée par un bruit additif n .

On supposera dans la suite de cette partie que n est un bruit blanc gaussien.

On notera $d=f+n$, l'image bruitée.

On cherche à obtenir la meilleure estimation possible g de f à partir de notre donnée d .

Pour cela, on essayera d'obtenir le maximum du SNR (Signal to Noise Ratio, c'est-à-dire rapport signal à bruit) :

$$SNR = 10 \log_{10} \frac{E \|f\|^2}{E (\|f-g\|^2)}, \quad (IV.17)$$

ce qui revient à minimiser l'erreur quadratique moyenne

$$e = E (\|f - g\|^2) \quad (IV.18)$$

Où $\|\cdot\|$ représentera la norme de $L^2(S^2)$.

L'image f et l'estimateur g sont des fonctions de $L^2(S^2)$, dont les harmoniques sphériques forment une base orthonormée. Par conséquent, on peut utiliser le théorème de Riesz-Fisher :

$$e = E (\sum_{l \in N} \sum_{|m| < l} |\hat{f}(l, m) - \hat{g}(l, m)|^2). \quad (IV.19)$$

On se propose de chercher g comme le résultat d'un filtre (de réponse impulsionnelle h) appliqué à la donnée d , ce qui équivaut à écrire g sous la forme : $g = \frac{1}{2\pi} d * h$.

Théorème *La meilleure estimation linéaire, au sens des moindres carrés sous l'hypothèse d'un bruit blanc gaussien additif, est donnée par le filtre de Wiener :*

$$h(\eta) = \sum_{k \in N} \sqrt{\frac{2k+1}{4\pi}} \frac{1}{1 + \frac{\delta^2 k^2}{c}} Y_{k,0}(\eta), \quad (IV.20)$$

où les $Y_{k,0}$ représentent les harmoniques zonales.

Pour les démonstrations du théorème précédent le lecteur peut se référer au [31].

Nous allons maintenant présenter l'analogue d'une autre méthode "classique" de débruitage : la méthode de régularisation de Tikhonov.

IV.3.3.2 Régularisation de Tikhonov

Le débruitage est un exemple type de problème inverse, une méthode très utilisée pour les problèmes inverse est la régularisation de Tikhonov.

Arsenin et Tikhonov [49] ont proposé une formulation qui a été adaptée par S.Bigot, D.Kashi [31] dans le cas de fonctions définies sur la sphère.

Il s'agit de déterminer f minimisant la fonctionnelle J :

$$\min_{f \in H^1(S^2)} J(f) = \min_{f \in H^1(S^2)} \int_{S^2} |\nabla f|^2 + \lambda \|f - d\|^2. \quad (\text{IV.21})$$

λ est un paramètre devant le terme d'attache aux données et l'autre terme est un terme de régularisation.

Comme pour le filtre de Wiener, l'estimateur est le résultat d'un filtrage de la donnée, c'est-à-dire, $g = \frac{1}{2\pi} d * h$, avec h la réponse impulsionnelle du filtre de restauration.

On a donc

$$h(k, 0) = \sqrt{\frac{2k+1}{4\pi}} \frac{1}{1 + \frac{k(k+1)}{\lambda}}, \quad (\text{IV.22})$$

où λ est un paramètre que l'on fera varier pour modifier l'intensité du lissage.

Remarquons que l'on retrouve un filtre de restauration de forme similaire au filtre de Wiener, ce qui est un résultat classique dans le cadre du débruitage d'images planes.

IV.3.4 Détection de contours

La détection de contours est une technique de réduction d'information dans les images, qui consiste à transformer l'image en un ensemble de courbes, pas forcément fermées, formant les frontières significatives de l'image.

Tout comme nous l'avons fait pour les outils de débruitage, nous présentons dans ce paragraphe des détecteurs de contours pour les images sphériques.

Dans une image en niveau de gris, un contour est caractérisé par une discontinuité de la fonction d'intensité de l'image.

Soit I une image sphérique (en niveaux de gris), c'est-à-dire une fonction de S^2 dans R , avec $\forall \eta \in S^2 I(\eta)$: niveaux de gris par exemple. On se propose tout d'abord de présenter une méthode basée sur le gradient sphérique, pour détecter les contours de I .

IV.3.4.1 Méthode du gradient

Soit a et b , deux points de la sphère telle que la différence $I(b)-I(a)$ soit grande, et $\gamma(t)$ une courbe paramétrée à valeurs dans la sphère telle que $\gamma(0) = a$ et $\gamma(\varepsilon) = b$. quand $\varepsilon \rightarrow 0$, on a, $I(b) - I(a) \approx DI(a)(\dot{\gamma}(0))$, avec DI la différentielle et $\dot{\gamma}$ la dérivée de γ .

Par hypothèse, $I(b)-I(a)$ est grand, donc il existe $\dot{\gamma}(0)$ tel que $DI(a)(\dot{\gamma}(0))$ soit grand. Par conséquent, par définition de l'opérateur gradient, $|\nabla I(a)|$ est grand.

Dans le but de détecter les contours de I , il suffit de trouver les points de I , où le module de gradient est maximum.

Pour détecter les contours, on doit au préalable filtrer l'image pour en retirer le bruit et les textures, avant de calculer le gradient de l'image lissée. Par le lien entre le gradient et le produit de convolution dans le plan, on peut cependant directement convoluer l'image avec le gradient de la réponse impulsionnelle du filtre.

$$\frac{\partial}{\partial x} (I *_{R^2} f) = I *_{R^2} \frac{\partial f}{\partial x} = \frac{\partial I}{\partial x} *_{R^2} f \quad (\text{IV.23})$$

Où $(*_R^2)$ est le produit de convolution sur R^2 .

Pour des images sphériques, le gradient se trouve sur le plan tangent de la sphère (figure IV.8)), qui est différent de la sphère. On ne peut donc pas convoluer une image sphérique avec les composantes du gradient sphérique, car ils ne sont pas dans le même espace.

Néanmoins, on peut lisser notre image I avec un filtre sphérique f , puis calculer la norme du gradient sur cette image lissée, c'est-à-dire $|\nabla_{S^2}(I * f)|$.

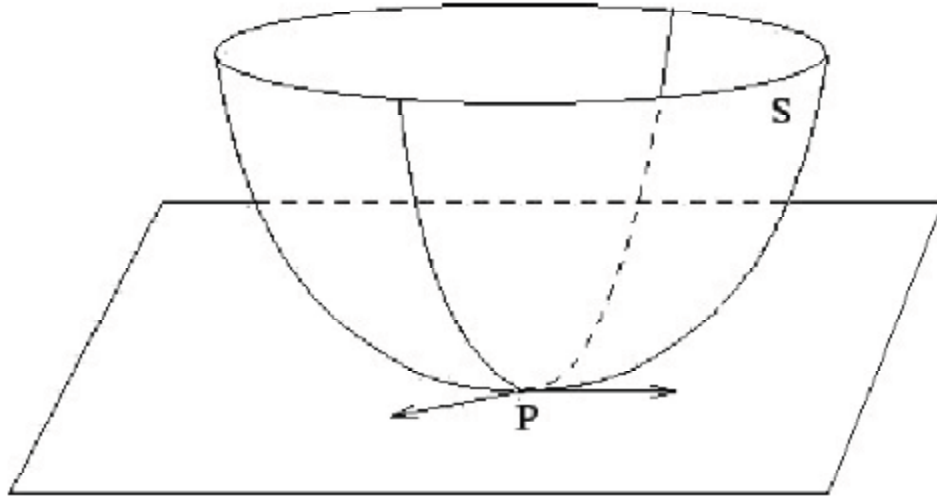


Figure IV.8 -Exemple de plan tangent à la sphère.

La définition du gradient en coordonnées sphériques, dans le cadre de la sphère équivalente est donnée par

$$\nabla_{S^2} I(\theta, \varphi) = \frac{\partial I}{\partial \theta} e_\theta + \frac{1}{\sin(\theta)} \frac{\partial I}{\partial \varphi} e_\varphi \quad (\text{IV.24})$$

Avec e_θ et e_φ les vecteurs des coordonnées sphériques.

Filtre de Robert

On suppose que la fonction d'intensité est différentiable afin de pouvoir calculer ses dérivées partielles. On notera (θ_i, φ_j) les points de la grille latitude-longitude.

Le filtre de Roberts, est une approche par différences finies d'ordre 1.

$$\frac{\partial I}{\partial \theta} = I(\theta_{i+1}, \varphi_j) - I(\theta_i, \varphi_j) \quad (\text{IV.25})$$

$$\frac{\partial I}{\partial \varphi} = I(\theta_i, \varphi_{j+1}) - I(\theta_i, \varphi_j) \quad (\text{IV.26})$$

Nous travaillerons avec la norme euclidienne.

$$\|\nabla_{S^2} I(\theta_i, \varphi_j)\|^2 = |I(\theta_{i+1}, \varphi_j) - I(\theta_i, \varphi_j)|^2 + \frac{1}{\sin^2(\theta_i)} |I(\theta_i, \varphi_{j+1}) - I(\theta_i, \varphi_j)|^2. \quad (\text{IV.27})$$

On peut remarquer qu'avec ce schéma, les contours type "marche" seront bien détectés. Cependant, en présence de bruit, qui correspond aussi à une brusque déviation des niveaux de gris, le filtre sera sûrement moins efficace, comme c'est le cas pour le filtre de Roberts pour les images planes. Pour contrer cette sensibilité au bruit, d'autres filtres ont été développés.

Filtre de Sobel et Prewitt

Pour améliorer la robustesse du détecteur de contours au bruit, Sobel, dans une conférence en 1968 [50], puis Prewitt en 1970 ont proposé des filtres dérivatifs qui reposent toujours sur le principe de calcul de gradient avec des approximations plus fines.

L'approximation $\frac{\partial I}{\partial \theta}$ au point (θ_i, φ_j) est donnée par le filtre :

$$\frac{1}{8} \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

Appliqué à la matrice $I(\theta_i, \varphi_j)$.

L'approximation de $\frac{1}{\sin(\theta)} \frac{\partial I}{\partial \varphi}$ au point (θ_i, φ_j) est donnée par le filtre :

$$\frac{1}{\sin(\theta_i)} \frac{1}{8} \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

Appliqué à la matrice $I(\theta_i, \varphi_j)$.

En pratique l'approximation $\frac{\partial I}{\partial \theta}$ au point (θ_i, φ_j) est donnée par :

$$A = \frac{1}{8} \left(-I(\theta_{i-1}, \varphi_{j-1}) - 2I(\theta_{i-1}, \varphi_j) - I(\theta_{i-1}, \varphi_{j+1}) + I(\theta_{i+1}, \varphi_{j-1}) + 2I(\theta_{i+1}, \varphi_j) + I(\theta_{i+1}, \varphi_{j+1}) \right). \quad (\text{IV.28})$$

Pour le filtre de Prewitt, les matrices d'approximations de $\frac{\partial I}{\partial \theta}$ et de $\frac{1}{\sin(\theta)} \frac{\partial I}{\partial \varphi}$ au point (θ_i, φ_j) , sont respectivement :

$$\frac{1}{6} \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\frac{1}{\sin(\theta_i)} \frac{1}{6} \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$

appliquées à la matrice $I(\theta_i, \varphi_j)$.

On peut remarquer que pour le filtre de Sobel, aussi bien que pour celui de Prewitt, seul un voisinage de taille 3*3 autour du point intervient dans le calcul. Les implémentations de ces filtres sont donc faciles.

IV.3.4.2 Méthode du Laplacien

Une autre méthode utilisée pour la détection de contours pour les images planes est la méthode de Laplacien. On se propose de présenter une méthode analogue sur la sphère.

L'opérateur de Laplace-Beltrami sphérique, est par définition le Laplacien sphérique,

$$\Delta_{S^2} = \frac{1}{\sin(\theta)} \frac{\partial}{\partial \theta} \left(\sin(\theta) \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2(\theta)} \frac{\partial^2}{\partial^2 \varphi} \quad (\text{IV.29})$$

Les harmoniques sphériques sont les vecteurs propres de cet opérateur. Ainsi, pour une fonction f définie sur la sphère et avec $(*)$ le produit de convolution sur la sphère, nous avons les relations,

$$\widehat{\Delta_{S^2} f}(l, m) = -l(l+1) \Delta_{S^2} \hat{f}(l, m), \quad (\text{IV.30})$$

$$\Delta_{S^2}(I * f) = I * \Delta_{S^2}(f). \quad (\text{IV.31})$$

Un détecteur souvent utilisé dans le cas d'images planes est le filtre de Canny, qui revient à utiliser la dérivée de la gaussienne plane. D.Kashi et S.Bigot Dans [31] ont développé l'analogue sphérique du filtre de Canny.

Chapitre V

Détection de points d'intérêts

Introduction

En vision par ordinateur, le premier et difficile problème est celui de l'extraction d'informations caractéristiques contenues dans une ou plusieurs images. En effet, la représentation que notre œil se fait d'une droite ou d'un coin n'est pas nécessairement facile à interpréter du point de vue du signal image. Depuis de nombreuses années maintenant, les chercheurs se penchent sur cet aspect de traitement « bas-niveau ». Si les progrès sont sans conteste indéniables, il n'en reste pas moins que certains cas posent encore des problèmes tant théoriques que pratiques. Nous évaluerons différentes méthodes de détection de points d'intérêts sur les images omnidirectionnelles que nous a fournis LAMFA.

Dans ce chapitre, nous présentons un état de l'art sur la détection des points d'intérêts en mettant l'accent sur les méthodes récentes et robustes. Les méthodes d'évaluation sont présentées dans la section V.2, les vecteurs de descriptions dans la section V.3. Nous renvoyons le lecteur aux références proposées dans ce chapitre pour des compléments de bibliographie.

V.1 Etat de l'art

Historiquement, la détection de points d'intérêt a débuté avec la détection de coins. Une définition simple possible serait de dire que les points d'intérêts correspondent à des doubles discontinuités de la fonction d'intensité dans une image.

Dans la suite, nous présentons les principaux travaux sur la détection de points d'intérêt. Nous utiliserons l'ordre chronologique de manière à donner un aperçu des améliorations qui sont apparues au cours des années.

V.1.1 Moravec

Le détecteur développé par Hans P. Moravec [1,2] en 1977 est l'un des plus anciens algorithmes de détection de points d'intérêt. Moravec a défini le concept de points d'intérêt comme étant une région distinctive des images et a conclu que ces points pouvaient être utilisés pour apparier des régions dans des images consécutives.

Moravec définit un point d'intérêt comme étant un point où il y a une large variation de l'intensité dans certaines directions. Son détecteur est connu comme étant un détecteur de coins. L'auteur propose d'utiliser la fonction d'auto-corrélation (corrélation du signal avec lui-même décalé spatialement) pour détecter des variations de la fonction d'intensité dans un voisinage carré W (typiquement 3×3 , 5×5 , ou 7×7 pixels). Même si le terme auto-corrélation est clairement indiqué, la fonction utilisée pour détecter les coins est la somme du carré de la différence du signal avec lui-même décalé spatialement. Ainsi, étant donné un pixel (x, y) de l'image et un décalage spatial (δ_x, δ_y) , la fonction f calculée en (x, y) est définie par :

$$f(x, y; \delta_x, \delta_y) = \sum_W (I(x_k, y_k) - I(x_k + \delta_x, y_k + \delta_y))^2 \quad (\text{V.1})$$

Où (x_k, y_k) sont les pixels du voisinage carré W centré en (x, y) et I est la fonction intensité. Le décalage spatial (δ_x, δ_y) est à valeurs dans $[-1, 1]^2 \setminus \{0, 0\}$ (8-voisinage).

L'auteur définit une fonction C à partir de laquelle les points d'intérêt vont être détectés. La valeur de la fonction en chaque pixel correspond à la valeur minimale de la fonction $f(x, y; \delta_x, \delta_y)$ pour les 8 décalages possibles :

$$C(x, y) = \min_{(\delta_x, \delta_y) \in [-1, 1]^2 \setminus \{0, 0\}} f(x, y; \delta_x, \delta_y) \quad (\text{V.2})$$

La détection des points d'intérêt est simplement réalisée en exhibant les maxima locaux de C supérieurs à un seuil fixé. Le détecteur de Moravec a une faible complexité ce qui le rend intéressant pour des applications nécessitant une implémentation temps-réel. Cependant, du fait des problèmes exposés dans la section V.1.2, ce détecteur est généralement considéré comme obsolète.

V.1.2 Harris et Stephens

Le détecteur de Moravec souffre de nombreuses limitations ce qui lui vaut de fonctionner dans un contexte limité. Harris et Stephen ont identifié certaines de ces limitations et, en les corrigeant, ont proposé en 1988 un détecteur de coins connu sous le nom de détecteur de Harris [3].

La réponse du détecteur de Moravec est anisotropique du fait du caractère discret de l'ensemble des décalages considérés (tous les 45 degrés). Cet aspect peut être résolu en considérant le développement de Taylor de la fonction intensité I au voisinage du pixel (x, y) . Étant donné un décalage (δ_x, δ_y) et un pixel (x, y) , la fonction $f(x, y; \delta_x, \delta_y)$ est définie par :

$$f(x, y; \delta_x, \delta_y) = \sum_W (I(x_k, y_k) - I(x_k + \delta_x, y_k + \delta_y))^2 \quad (\text{V.3})$$

où (x_k, y_k) sont les points dans la fenêtre W centrée en (x, y) et I la fonction intensité. En considérant la fonction w définie par

$$w(x, y) = \begin{cases} 1 & \text{si } (x, y) \in W \\ 0 & \text{sinon} \end{cases} \quad (\text{V.4})$$

la fonction $f(x, y; \delta_x, \delta_y)$ se réécrit comme une somme d'un produit de convolution sur l'image entière :

$$f(x, y; \delta_x, \delta_y) = \sum_{x, y} w(x_k, y_k) * \left(I(x_k, y_k) - I(x_k + \delta_x, y_k + \delta_y) \right)^2 \quad (\text{V.5})$$

Considérons maintenant le développement de Taylor d'ordre 1 de l'image :

$$I(x_k + \delta_x, y_k + \delta_y) \approx I(x, y) + \left[\frac{\partial I}{\partial x}(x, y) \frac{\partial I}{\partial y}(x, y) \right] \cdot \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix} \quad (\text{V.6})$$

Où $\frac{\partial I}{\partial x}$ et $\frac{\partial I}{\partial y}$ représentent les dérivées partielles de la fonction intensité relativement aux axes des abscisses et des ordonnées. La dérivation est approximée en utilisant une méthode par différence finie :

$$\frac{\partial I}{\partial x}(x, y) \approx \frac{I(x+1, y) - I(x-1, y)}{2} \quad (V.7)$$

$$\frac{\partial I}{\partial y}(x, y) \approx \frac{I(x, y+1) - I(x, y-1)}{2} \quad (V.8)$$

En substituant les deux équations précédentes, nous obtenons l'expression suivante :

$$f(x, y; \delta_x, \delta_y) \approx \sum_{x, y} w(x_k, y_k) * \left(\left[\frac{\partial I}{\partial x}(x_k, y_k) \frac{\partial I}{\partial y}(x_k, y_k) \right] \cdot \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix} \right)^2 = [\delta_x \ \delta_y] A(x, y) \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix} \quad (V.9)$$

Où

$$A(x, y) = \begin{bmatrix} \left(w * \left(\frac{\partial I}{\partial x} \right)^2 \right)(x, y) & \left(w * \left(\frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \right) \right)(x, y) \\ \left(w * \left(\frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \right) \right)(x, y) & \left(w * \left(\frac{\partial I}{\partial y} \right)^2 \right)(x, y) \end{bmatrix}$$

La matrice A capture la structure du voisinage local des pixels de l'image d'où son nom de « matrice de structure ».

La seconde limitation provient de la fonction w . En effet, w étant une fonction binaire, la réponse de la fonction f est bruitée. Harris et Stephens ont alors proposé de remplacer w par une fonction Gaussienne

$$w(x, y) = G_\sigma(x, y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (V.10)$$

Où σ (déviation standard = racine carrée de la variance) est un paramètre généralement fixé à $\sigma = 2$. La matrice A s'exprime alors comme suit :

$$A(x, y) = \begin{bmatrix} \left(G_\sigma * \left(\frac{\partial I}{\partial x} \right)^2 \right)(x, y) & \left(G_\sigma * \left(\frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \right) \right)(x, y) \\ \left(G_\sigma * \left(\frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \right) \right)(x, y) & \left(G_\sigma * \left(\frac{\partial I}{\partial y} \right)^2 \right)(x, y) \end{bmatrix}$$

Enfin, le détecteur de Moravec répond de manière trop forte aux contours car seul le minimum de f est pris en compte en chaque pixel. Les auteurs ont proposé d'étudier les valeurs propres (λ_1 et λ_2) de la matrice A . Ces deux valeurs propres sont proportionnelles aux courbures principales de la fonction f et forment une description de A invariante aux rotations [55]. Il y a trois cas à considérer :

- Si les deux courbures (λ_1 et λ_2) sont faibles, la fonction f est alors approximativement constante ce qui indique que la région de l'image considérée soit homogène.
- Si une des courbures est forte alors que l'autre est faible, cela indique la présence d'un contour dans l'image.
- Enfin, si les deux courbures sont fortes, un coin est détecté.

Harris et Stephens ont proposé d'éviter le calcul fastidieux des valeurs propres en étudiant le déterminant et la trace de la matrice A . Les auteurs définissent alors une fonction C définie en chaque pixel (x, y) de l'image et à partir de laquelle les coins et les contours de l'image sont détectés :

$$\begin{aligned} C(x, y) &= \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \\ &= \det(A(x, y)) - k \cdot \text{trace}^2(A(x, y)) \end{aligned} \quad (\text{V.11})$$

Où $k \in [0.04, 0.15]$. Les coins sont détectés en exhibant les maxima locaux de C supérieurs à un seuil donné.



Figure V.1-Détection de points d'intérêts par la méthode de Harris et Stephens

Le détecteur de Harris-Stephens, est sans doute le détecteur de points d'intérêt le plus connu et le plus utilisé en traitement d'images. De nombreuses implémentations de ce détecteur existent car cinq paramètres doivent être choisis lors de son utilisation : le filtre dérivatif, le filtre Gaussien, le paramètre k , le voisinage de l'extraction des maxima locaux, et le seuillage final. Son succès provient premièrement de sa facilité d'implémentation, et deuxièmement de la qualité des résultats obtenus. En effet, de récents travaux [51,52] prouvent que le détecteur de Harris donne les meilleurs résultats. De nombreux travaux sont venus améliorer l'approche initiale. Ainsi, Schmid et al proposent dans [52] de calculer les dérivées partielles en convoluant l'image I par la dérivée d'une Gaussienne 1D. L'implémentation récursive du

filtre Gaussien [53] permet une détection rapide. Montesinos et al. [54] proposent de généraliser l'approche de Harris et Stephens en l'adaptant aux images couleur. La matrice de structure est alors constituée de la somme des dérivées partielles des composantes rouge, verte et bleue.

V.1.3 Lindeberg, LoG

Schmid et al. [51,52] ont montré que le détecteur de Harris a d'excellentes performances en comparaison d'autres détecteurs (au sens des critères de rappel et de précision [55]). Cependant, comme tous les autres détecteurs coins, les performances du détecteur de Harris chutent lors de l'apparition d'un changement d'échelle entre deux images.

Lindeberg proposa dans [56] en 1998 son détecteur de points d'intérêt multi-échelles nommé *LoG* (pour Laplacian of the Gaussian) ; les points détectés sont ainsi invariants aux changements d'échelle. On parle de détection de blobs plutôt que de détection de points car la détection exhibe des régions d'intérêt et non des points d'intérêt comme nous pouvons le voir sur la figure V.2. Pour ce faire, Lindeberg décrit l'image dans l'espace échelle Gaussien et utilise l'opérateur Laplacien. Le Laplacien est un opérateur différentiel scalaire d'ordre 2 égal à la somme de toutes les deuxièmes dérivées partielles non mixtes d'une variable dépendante. En coordonnées cartésiennes dans R^n , le Laplacien est défini par :

$$\Delta f(x_1, \dots, x_n) = \sum_{k=1}^n \frac{\partial^2 f}{\partial x_k^2}(x_1, \dots, x_n) \quad (\text{V.12})$$

Le Laplacien d'une image permet de mettre en évidence ses changements rapides d'intensité, ce qui en fait un opérateur particulièrement adapté à la détection de contours en étudiant les zéros de la fonction. La représentation de l'image I à l'échelle σ , notée L_σ , correspond à la convolution de I par une fonction Gaussienne bidimensionnelle G_σ d'écart type σ :

$$L_\sigma(x, y) = (G_\sigma * I)(x, y) \quad (\text{V.13})$$

Avec

$$G_\sigma(x, y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (\text{V.14})$$

Par définition, le Laplacien de $\Delta L_\sigma(x, y)$ est égal à la somme de ses deux dérivées partielles secondes non mixtes. Cependant, il peut être directement calculé à partir de la fonction intensité I par convolution avec l'opérateur LoG_σ :

$$\begin{aligned}
\Delta L_\sigma(x, y) &= \frac{\partial^2}{\partial x^2} L_\sigma(x, y) + \frac{\partial^2}{\partial y^2} L_\sigma(x, y) \\
&= \left(\underbrace{\left(\frac{\partial^2}{\partial x^2} G_\sigma + \frac{\partial^2}{\partial y^2} G_\sigma \right) * I}_{LoG_\sigma} \right) (x, y)
\end{aligned} \tag{V.15}$$

Pour un facteur d'échelle σ donné, la détection de blobs est réalisée en extrayant les extrema locaux (minima et maxima) en espace de la fonction $\Delta L_\sigma(x, y)$. Cependant, l'efficacité de cette détection est intimement liée à la taille des blobs qui n'est pas supposée être connue. La détection des blobs de différentes tailles nécessite l'utilisation d'une approche multi-échelles. Une façon simple d'obtenir un détecteur de blobs multi-échelles est de considérer l'opérateur Laplacien normalisé :

$$\Delta_{norm} L_\sigma(x, y) = \sigma^2 \Delta L_\sigma(x, y) \tag{V.16}$$

Le Laplacien $\Delta_{norm} L_\sigma$ est calculé à plusieurs échelles, c'est-à-dire pour plusieurs valeurs de σ (e. $g \sigma \in [0.100]$). Les blobs sont les extrema locaux en espace et en échelle de $\Delta_{norm} L_\sigma$ sur un 26-voisinage (9+9+8) (voir figure V.2) :

$$(\hat{x}, \hat{y}; \hat{\sigma}) = \arg \max_{(x, y; \sigma)} \min(x, y; \sigma) local \Delta_{norm} L_\sigma(x, y) \tag{V.17}$$

L'intérêt de la méthode est la détection simultanée de la position (x_i, y_i) des blobs et de leur échelle caractéristique σ_i . La probabilité de dévier de plus de $3\sigma_i$ est inférieure à 1% [55]. Chaque blob correspond à un disque centré en (x_i, y_i) et de rayon $3\sigma_i$ (voir figure V.2).



Figure V.2-Détection de blobs par la méthode LoG sur l'image « tournesol ». Les cercles représentent les 120 maxima en échelle et en espace de la fonction $\Delta_{norm}L$. La taille des cercles indique l'échelle caractéristique de chaque blob .

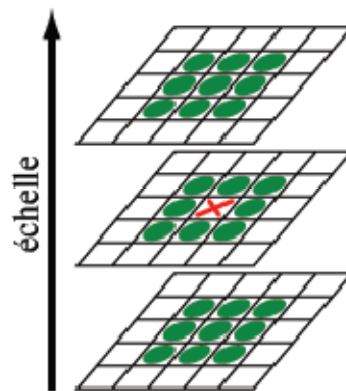


Figure V.3 -26-voisinage utilisé pour la détection des maxima locaux de l'opérateur LoG.

V.1.4 Harris-Laplace

Suite aux travaux de Harris et Stephens [3] et de Lindeberg [56], Mikolajczyk et Schmid proposent dans [57] un détecteur de points d'intérêts multi-échelles nommé Harris-Laplace utilisant les avantages de ces deux méthodes, à savoir la performance du détecteur de Harris couplée à l'aspect multi-échelles utilisé par Lindeberg avec le détecteur LoG.

Soit L_{σ_D} , la représentation de la fonction d'intensité I à l'échelle σ_D . La matrice de structure A se définit par :

$$A(x, y; \sigma_D, \sigma_I) = \sigma_D^2 \cdot \begin{bmatrix} \left(G_{\sigma_I} * \left(\frac{\partial L_{\sigma_D}}{\partial x} \right)^2 \right) (x, y) & \left(G_{\sigma_I} * \left(\frac{\partial L_{\sigma_D}}{\partial x} \cdot \frac{\partial L_{\sigma_D}}{\partial y} \right) \right) (x, y) \\ \left(G_{\sigma_I} * \left(\frac{\partial L_{\sigma_D}}{\partial x} \cdot \frac{\partial L_{\sigma_D}}{\partial y} \right) \right) (x, y) & \left(G_{\sigma_I} * \left(\frac{\partial L_{\sigma_D}}{\partial y} \right)^2 \right) (x, y) \end{bmatrix}$$

Où σ_I est appelé échelle d'intégration et σ_D échelle de dérivation. L'approche proposée ne détecte pas les points 3D directement comme le fait l'approche de Lindeberg. Dans un premier temps, les auteurs proposent de détecter, pour chaque échelle, les points d'intérêt 2D au sens de Harris en utilisant la matrice de structure A et en exhibant les maxima locaux en espace de la fonction définie dans l'équation V.11. Parmi les points d'intérêt détectés, ceux qui correspondent à un extremum local en échelle au sens de Lindeberg sont retenus comme étant les points d'intérêt de Harris-Laplace.

Le détecteur de Harris-Laplace est considéré à l'heure actuelle comme étant le meilleur détecteur multi-échelles au sens des critères de répétabilité et de précision [57]. Les régions détectées sont, par construction, circulaires. Mikolajczyk et Schmid améliorent dans [58] le détecteur Harris-Laplace en utilisant des voisinages ellipsoïdaux. L'appariement d'images est alors plus robuste aux changements de point de vue.

V.1.5 Autres travaux

Il existe dans la littérature d'autres travaux sur les détecteurs de points d'intérêts parmi eux on peut citer pour les méthodes basées sur les contours, Asada et Brady [59] qui extraient des points d'intérêt pour des objets 2D à partir de courbes planes. Ils constatent que les courbes planes ont des caractéristiques significatives : les changements de courbure. Ces changements sont classés en plusieurs catégories : coin, terminaison, etc. Pour pouvoir les détecter d'une manière robuste, l'algorithme est intégré dans un cadre multi-échelle. Une approche similaire a été proposée par Mokhtarian et Mackworth [60]. Au lieu d'utiliser les changements de courbure d'une courbe plane, ils utilisent les points d'inflexion de celle-ci.

Medioni et Yasumoto [61] approximent les contours avec des B-splines. Les points d'intérêt sont des maxima de courbure calculés à partir des coefficients de ces B-splines. Horaud et al. [62] recherchent des groupements dans une image de contours pour établir une représentation intermédiaire. Cette représentation repose sur la structuration de segments extraits dans l'image. L'intersection de ces segments donne les points d'intérêt.

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

Kitchen et Rosenfeld [63] ont proposé un détecteur de points d'intérêt qui repose sur la courbure de courbes planes. Ils recherchent les maxima de courbure des isophotes du signal. Cependant, un isophote peut présenter une courbure importante du fait du bruit sans que cela corresponde à un point d'intérêt. Cela peut par exemple survenir sur une zone quasi-uniforme, d'autant plus que le calcul fait de la courbure est très approximatif. Kitchen et Rosenfeld proposent donc de multiplier la courbure par la magnitude de gradient de l'image.

Forstner [64,65] propose une approche basée sur la statistique locale d'une image. Ceci lui permet d'estimer les paramètres de son algorithme de manière automatique. La première étape de son algorithme est d'estimer la variance du bruit. Il utilise ensuite cette estimation pour restaurer le signal. Puis, les pixels sont classés dans les catégories régions, contours et points d'intérêt. Pour ce faire, il utilise la fonction d'auto corrélation de la même manière que Harris. Enfin, il classe les points d'intérêt en jonctions ou points isolés. De plus, il effectue une estimation sous-pixelique.

Reisfeld et al. [66] ont proposé un opérateur motivé par des évidences psychophysiques. Cet opérateur est basé sur la notion de symétrie. Ils calculent une carte de symétrie qui contient pour chaque pixel une " magnitude de symétrie " et une orientation. Cette symétrie est calculée localement en regardant la magnitude et la direction des dérivées des points voisins. Cette carte de symétrie peut être appliquée à des tâches diverses, notamment à l'extraction de points d'intérêt. Les endroits avec une symétrie importante sont des points d'intérêt, des lignes de symétrie représentent les axes de symétrie.

En ce qui concerne les méthodes basées sur un modèle théorique du signal, le but est d'obtenir une précision sous-pixelique en approximant le signal par un modèle théorique. De telles méthodes ne sont utilisables que pour des types bien précis de points d'intérêt, par exemple des coins.

Rohr [67,68] modélise les jonctions de plusieurs lignes. Pour ce faire, il convolue un modèle binaire de jonction avec une gaussienne afin de modéliser le flou. Dans le cas d'un coin les paramètres du modèle sont l'angle définissant l'orientation de l'axe de symétrie, l'angle définissant l'ouverture du coin, les niveaux de gris, la position du point et le flou. Ces paramètres sont ajustés pour que le signal théorique soit le plus proche possible du signal observé. Cette recherche repose sur une minimisation au sens des moindres carrés. Les coins obtenus par cette méthode sont très précis. Toutefois, la qualité de l'approximation repose sur une bonne estimation initiale de la position. Rohr utilise les segments extraits pour déterminer les paramètres initiaux du modèle : le type de coin, les angles initiaux ainsi que la position du coin.

Deriche et Blaszkia [69] ont proposé une amélioration de la méthode de Rohr au niveau du temps de calcul en remplaçant la fonction gaussienne de lissage par une fonction exponentielle. Deriche propose, en outre, une solution pour l'initialisation des paramètres. Une fois que la position du coin a été donnée de façon interactive par l'utilisateur, l'ouverture du coin est estimée de façon automatique à partir du gradient sur les bords de la fenêtre. Le point donné par l'utilisateur est ensuite amélioré avec une méthode de descente de la variance des niveaux de gris. Deriche et al. ont montré la bonne précision en position des résultats en présence de bruit synthétique. Cette étude a aussi permis de voir que pour assurer la convergence de la minimisation, le support utilisé doit être assez grand. Ceci constitue un défaut de ces méthodes, car lorsque le signal réel comporte plusieurs signaux sur ce support, la convergence est fortement affectée.

V.2 Méthodes d'évaluation

Précédemment, nous avons vu qu'il existe beaucoup de travaux sur les détecteurs de points d'intérêts. Toutefois la méthode d'évaluation la plus répandue a été de vérifier visuellement la qualité des résultats. Ceci n'est pas systématique et risque en plus de donner des résultats subjectifs : on compare le résultat avec ce qu'on évalue comme étant un point d'intérêt et non pas avec ce qui est important pour l'étape qui suit l'extraction de ces points. Un exemple pour une évaluation systématique est le travail de Brand et Mohr [70].

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

V.2.1 Critères d'évaluation

D'une manière générale un détecteur est caractérisé par sa répétabilité et sa localisation. La répétabilité signifie qu'un même point est détecté dans une suite d'images. La localisation, par contre, signifie qu'un point détecté dans l'image correspond de façon précise à un point 3D de la scène. Plusieurs travaux ont constaté [71], [72] qu'il existe une contradiction entre répétabilité et localisation. En effet, un lissage améliore la répétabilité, mais rend plus mauvaise la localisation, comme l'a constaté Canny [73].

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

V.2.3 Définition de la répétabilité

Définition V.1 Répétabilité

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

La figure V.4 illustre cette définition. Pour mesurer la répétabilité d'un détecteur, il faut établir une relation entre p_1 et p_i à partir des images. Dans le cas général, il n'existe pas de telle relation. Toutefois, si I_1 et I_i sont des images d'une scène plane, cette relation est définie par une homographie :

$$p_i = H_{i1} p_1 \quad \text{où } H_{i1} = M_i M_1^{-1} \quad (\text{V.18})$$

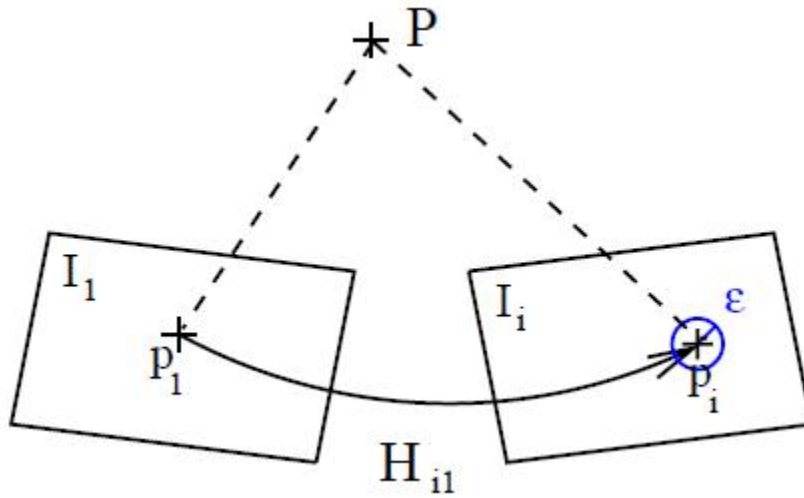


Figure V.4-Définition de répétabilité

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

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

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

V.2.4 Mesure de la répétabilité

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

vue dans les deux images. Soient $\{d\}_i$ et $\{d\}_l$ les sous-ensembles de points détectés correspondant à la partie commune :

$$\{d\}_1 = \{p_1/H_{11}p_1 \in I_i\} \text{ et } \{d\}_i = \{p_i/H_{1i}p_i \in I_1\} \quad (\text{V.19})$$

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

$$D = \{(d_i, d_1)/d_i = H_{1i}d_1\} \quad (\text{V.20})$$

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

$$D(\varepsilon) = \{(d_i, d_1)/\text{dist}(d_i, H_{1i}d_1) < \varepsilon\} \quad (\text{V.21})$$

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

$$r_i(\varepsilon) = \frac{|D(\varepsilon)|}{\min(n_i, n_1)} \quad (\text{V.22})$$

V.3 Caractérisation locale

Dans la section précédente nous avons vu comment détecter des points d'intérêt. Ces points ont été retenus parce qu'à priori ils sont à des endroits où le signal présente un contenu informatif important. Il s'agit maintenant de valider cette hypothèse en capturant cette information. Pour ce faire, il faut décrire localement la fonction signal au voisinage d'un point d'intérêt. La figure V.6 représente cette fonction $I(x, y)$, encore appelée fonction d'intensité lumineuse.

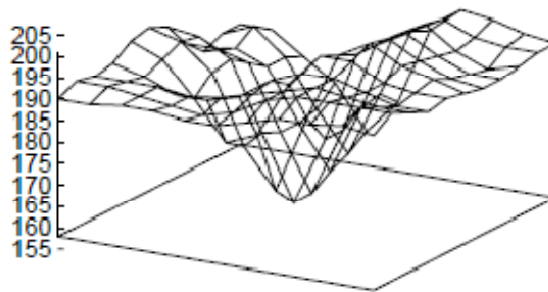


Figure V.5 -Fonction d'intensité lumineuse autour d'un point d'intérêt.

Le but de la caractérisation présentée dans cette section est de décrire cette fonction de la manière la plus précise et la plus complète possible. D'autre part, nous recherchons une description qui soit invariante aux transformations usuelles de l'image. Après avoir décrit à la section V.3.1 les méthodes existantes permettant d'obtenir une caractérisation du signal, la section V.3.2 présente les vecteurs de description, la section V.3.3 présente une brève introduction aux invariants. Il est alors montré à la section V.3.4 comment des mesures différentielles - la méthode de caractérisation retenue - peuvent être invariantes aux différentes transformations image considérées.

V.3.1 Méthodes de caractérisation locale

De nombreuses méthodes de caractérisation locale sont possibles. Dans le cas d'images de niveaux de gris des exemples en sont les dérivées, les filtres de Gabor et plus généralement les ondelettes ainsi que les moments. Des caractéristiques basées sur la couleur sont également utilisables.

V.3.1.1 Dérivées

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

$$\begin{aligned} f(x_0 + x, y_0 + y) &= f(x_0, y_0) + x \frac{\partial}{\partial x} f(x_0, y_0) + y \frac{\partial}{\partial y} f(x_0, y_0) \dots \\ &+ \sum_{p=1}^N x^p y^{N-1} \frac{\partial^N}{\partial x^p \partial y^{N-p}} f(x_0, y_0) + O(x^N, y^N) \end{aligned} \quad (V.23)$$

De ce fait il est possible de décrire une image en un point en stockant dans un vecteur l'ensemble des dérivées en ce point. Un tel vecteur a été utilisé par Koenderink [104] qui l'a nommé jet local. Koenderink calcule en outre le jet local de manière stable en utilisant un filtre passe-bas : la gaussienne et ses dérivées. La définition du jet local est la suivante :

Définition V.2 Jet local

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

$$J^N[I](\vec{x}, \partial) = \{L_{i_1 \dots i_n}(\vec{x}, \partial) | (\vec{x}, \partial) \in I \times R^+, n = 0, \dots, N\} \quad (V.24)$$

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

$$L_{i_1 \dots i_n}(\vec{x}, \partial) = G_{i_1 \dots i_n}(\vec{x}, \partial) * I(\vec{x}) \quad (V.25)$$

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

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

V.3.1.2 Descriptions fréquentielles

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

$$f(u, v) = \int \int f(x, y) e^{i(ux+vy)} dx dy \quad (\text{V.26})$$

La transformée de Fourier est un cas spécial de décomposition d'une fonction dans une base infinie de fonctions complètes et orthogonales. Cette base de fonctions est constituée des fonctions sinusoïdales donc périodiques et infinies. D'autre part, cette transformée est directionnelle : la transformée est calculée dans la direction définie par u et v (figure V.7).

Il existe pour chaque paire de fréquences $(u; v)$ une direction, une magnitude et une phase. L'intérêt de la transformée de Fourier, ainsi que des autres méthodes de description fréquentielle, réside dans le fait que la phase est "normalisée", c'est à dire indépendante de la luminosité des images ainsi que de leur contraste. D'autre part, la phase est stable à un changement d'échelle jusqu'à 20% [55]. Du fait de cette stabilité, l'utilisation de cette phase peut être mise en œuvre dans un contexte multi-échelle. Enfin, la phase est une variable continue qui permet d'obtenir des résultats sous-pixelles, c'est-à-dire de précision supérieure à celle du signal.

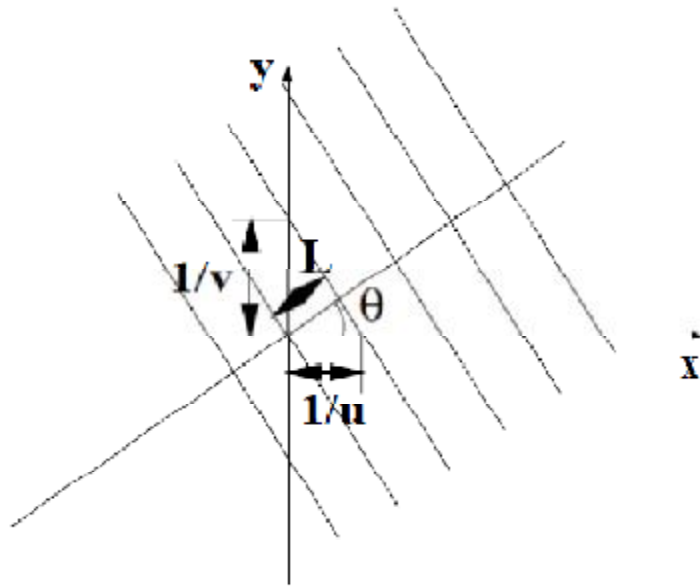


Figure V.6 -La transformée de Fourier est directionnelle.

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

$$\Delta_x \Delta_\omega \geq \text{constante}$$

Où x est la variable spatiale et ω la variable fréquentielle. Les dispersions Δ_x et Δ_ω représentent respectivement l'incertitude spatiale et fréquentielle. Ce principe montre donc que l'on ne peut pas être à la fois précis en espace et en fréquence. En fait, si un filtre est très précis en espace il l'est très peu en fréquence, et réciproquement. Pour remédier à ce problème et minimiser à la fois l'incertitude en espace et en fréquence, il est préférable d'utiliser un fenêtrage. Gabor [74] a proposé d'utiliser un fenêtrage gaussien et a démontré qu'un tel fenêtrage est optimal pour obtenir une bonne précision à la fois en fréquence et en espace. La transformation de Gabor est donc la convolution du signal par un filtre dont l'expression est la suivante :

$$g_{\sigma\omega_x\omega_y}(x, y) = e^{i(\omega_x x + \omega_y y)} \frac{1}{\sigma 2\pi} e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (\text{V.27})$$

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

est donc préférable d'effectuer une décomposition multi-échelle afin d'obtenir une description riche du signal.

Le filtre de Gabor est souvent utilisé pour les problèmes de stéréo-correspondance ou pour réaliser la vergence d'une tête stéréoscopique [75, 76, 77]. Ce filtre permet d'obtenir une information locale de la phase du signal et sert ainsi à estimer la disparité entre deux images. Dans le cadre du calcul d'appariements, une égalité de phase entre deux points de deux images différentes signifie une grande probabilité qu'il s'agisse de points à appairer, modulo le fait qu'une même valeur de phase peut apparaître plusieurs fois. Ainsi, pour qu'une égalité de phase corresponde à un appariement de façon certaine, une approche multi-résolution est nécessaire. Toutefois cette mesure est locale en fréquence et directionnelle en espace. Aussi, une telle approche résiste uniquement à des rotations et des changements d'échelle faibles. Pour résoudre ce problème, Wu [78] a proposé une implémentation qui utilise des filtres de Gabor dans plusieurs directions et à plusieurs échelles.

Morlet, Grossmann et Meyer [79] ont construit une théorie reposant sur l'idée de caractériser un signal par différentes échelles et différentes résolutions : cette théorie est à l'origine des ondelettes. L'idée principale des ondelettes est que l'apparition de hautes fréquences est de faible durée en espace. Cette idée se justifie par l'hypothèse que les hautes fréquences correspondent à des discontinuités du signal et doivent donc être considérées uniquement de manière locale. Mallat [80] a étendu ces travaux dans le domaine de la vision par ordinateur au cas des signaux discrets. La transformée en ondelettes de la fonction f est de la forme suivante :

$$CWT_h(\tau, a) = \frac{1}{\sqrt{c_h}} \frac{1}{a} \int f(x) h^* \left(\frac{x-\tau}{a} \right) dx \quad (V.28)$$

Où

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

V.3.1.3 Moments

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

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

Hu [81] a utilisé pour la première fois les moments pour des problèmes de reconnaissance en vision par ordinateur. Il a par ailleurs démontré l'existence de combinaisons de moments qui sont invariantes en translation, rotation et changement d'échelle.

Teague [82] a introduit la notion de moments orthogonaux pour caractériser une image. Pour ce faire, il utilise la théorie des polynômes orthogonaux et introduit la notion de moments de Zernike. Ces moments sont indépendants et peuvent être facilement construits pour un ordre quelconque. D'autres moments orthogonaux sont les moments de Legendre basés sur les polynômes du même nom. Teh [83] a montré que les moments de Zernike sont les moins sensibles au bruit et les moins redondants en information. Une comparaison similaire a été menée par Kim [84] entre les moments de Zernike et ceux de Hu. Les résultats de cette comparaison confirment la supériorité des moments de Zernike pour différencier deux modèles.

Parmi les travaux plus récents, Van Gool [85] a présenté un ensemble de moments jusqu'au deuxième ordre qui sont à la fois invariants aux changements affins et aux changements d'intensité

V.3.1.4 Autres caractérisations

Il existe d'autres méthodes de caractérisation d'un signal que celles présentées précédemment. Par exemple, Bigun [86,87] propose l'utilisation d'un système non linéaire de coordonnées. Son système de base doit satisfaire l'équation de Laplace et il doit être conjugué. Puis il cherche à tourner son système de coordonnées de telle manière que les isophotes dans ce nouveau repère soient les plus parallèles possible. Il utilise ces coordonnées pour caractériser des textures et pour reconnaître des objets.

Weiss [88] quant à lui propose une méthode pour calculer localement des invariants affines et projectifs. Pour ce faire, il utilise une représentation implicite de la courbe des contours. Pour un point donné il définit un voisinage puis il approxime localement la courbe passant par ce point. Toutefois il est difficile de calculer une représentation implicite à des ordres élevés. Il utilise un système de coordonnées canonique qui est localement défini par les propriétés de la forme. Un exemple simple est l'utilisation de la tangente et de la normale dans le cas des

transformations rigides. Dans le cas général, il utilise des courbes osculatrices pour obtenir l'invariance aux transformations.

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

V.3.2 Vecteurs de description

Soit $\{(x_i, s_i)\}_{1 \leq i \leq n}$ un ensemble de n points d'intérêt dont la position est notée x_i et l'échelle caractéristique s_i . Dans le cas d'une détection mono-échelle par exemple Harris), s_i est commune à tous les points détectés.

Le vecteur de description associé au point d'intérêt (x_i, s_i) est un ensemble de valeurs extraites à partir de l'image I dans le voisinage local de la position x_i . La taille du voisinage est idéalement définie par s_i . L'espace des vecteurs de description est un espace vectoriel normé où une distance permet de mesurer la similarité entre les vecteurs.

Plusieurs types de vecteurs de description existent. Dans cette partie, nous présentons trois types classiques.

V.3.2.1 Voisinage de pixels

Le vecteur de description le plus simple est la matrice contenant les valeurs de l'image dans un voisinage (par exemple carré) de chaque point d'intérêt. La taille du voisinage utilisé est généralement de $6s_i \times 6s_i$ pixels. La figure V.8 illustre un exemple de la construction de vecteurs de description. Pour chaque point, le voisinage considéré (encadré en rouge) est de taille 3×3 pixels. Ce voisinage est généralement exprimé sous forme de vecteur.

L'utilisation de plusieurs composantes permet d'augmenter le pouvoir discriminant des vecteurs de description. Ainsi, l'intensité de l'image est sensible aux changements d'illumination tandis que la norme et la direction du gradient ne le sont pas. Le choix des composantes est crucial et dépend de l'application.

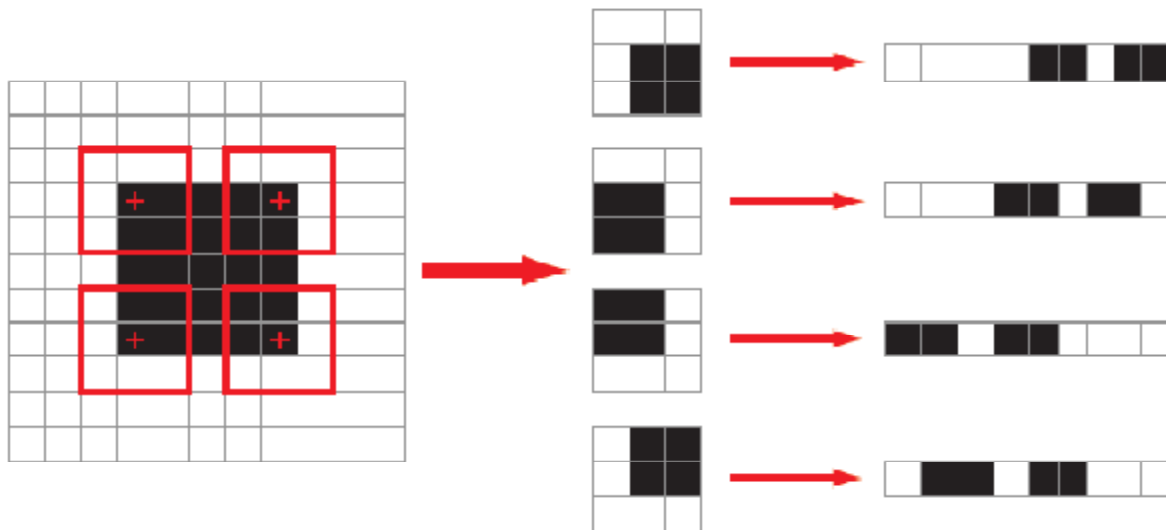


Figure V.7 -Construction de vecteurs de description pour 4 points d'intérêt indiqués par des croix rouges. Le voisinage de chaque point, délimité en rouge, est de taille 3×3 . La matrice de pixel correspondant à chaque voisinage est extraite, puis est exprimée sous forme de vecteur.

L'utilisation d'un voisinage de pixels pour la mise en correspondance de points d'intérêt utilise une information géométrique stricte, c'est-à-dire que la position et la valeur de chaque pixel sont importantes. Ce type de vecteur de description est très discriminant spatialement mais il est sensible aux changements géométriques (zoom, rotation). Deux points seront identiques si et seulement si les voisinages de pixels sont identiques. Lors de la mise en correspondance de deux nuages de points d'intérêt détectés dans deux images consécutives d'une vidéo, les propriétés géométriques des objets changent peu, surtout si la fréquence d'acquisition est élevée. La mise en correspondance des points d'intérêt à partir de la matrice de voisinage est par conséquent très fiable car très discriminante. En revanche, dans un cas plus général comme rencontré en recherche d'images dans une base de données, les images d'une même scène sont acquises avec des angles de vue très différents. L'orientation, l'échelle, la couleur et l'intensité des objets peuvent changer ce qui implique que l'utilisation d'une matrice de voisinage donne de mauvais résultats. Le but des méthodes présentées par la suite est de pallier à ces problèmes.

Enfin, il faut noter un problème intrinsèque à l'utilisation de voisinages de pixels pour décrire un point d'intérêt. D'une part, la détection de points d'intérêt par une méthode multi-échelles implique que les voisinages extraits pour l'ensemble des points soient de taille différente. Souvent, une mesure de similarité, va comparer les vecteurs de description élément par élément. Une telle comparaison est impossible si les voisinages utilisés sont de taille

différente. D'autre part, la taille du vecteur de description (taille du voisinage) est généralement très grande. Par exemple, pour un point d'intérêt dont le voisinage est de taille 15×15 , le vecteur de description est de taille 225. Cette taille importante a un impact sur le temps nécessaire pour comparer deux vecteurs de description. L'utilisation de tels vecteurs pose de gros problèmes en temps de calcul pour la recherche d'images par le contenu dans une grande base d'images.

V.3.2.2 Distribution

L'utilisation de voisinage de pixels n'est pas un bon choix de descripteur dans un cadre général d'appariement d'images. Toute transformation autre qu'une simple translation modifie la structure géométrique globale et locale de l'image. En revanche, les transformations usuelles ont un impact limité sur les propriétés statistiques de l'image. La figure V.9 illustre la construction d'un vecteur de description fondé sur la distribution de l'intensité. La première étape consiste à extraire la matrice de pixels dans le voisinage local du point d'intérêt. La seconde étape consiste à calculer la distribution des éléments de cette matrice. Plusieurs approches sont possibles. Cette distribution est généralement approximée par un histogramme normalisé. Dans l'exemple présenté, l'histogramme a 8 classes de largeur 32, l'intensité d'un pixel s'exprimant par un nombre entier naturel compris dans l'intervalle $[0, 255]$. Swain et Ballard [89] ont utilisé des histogrammes couleur pour leur méthode d'indexation. D'autres attributs que la couleur, comme l'intensité des réponses obtenues par des dérivées de Gaussienne ou par des filtres de Gabor sont utilisés dans [90].

La distribution est une information purement statistique. Ces histogrammes ne prennent pas en compte la répartition spatiale des attributs dans le voisinage du point, ce qui leur confère l'invariance aux déformations géométriques usuelles. L'utilisation de distribution pour la description locale pose néanmoins un problème : deux matrices de pixels visuellement très différentes peuvent avoir la même distribution (voir figure V.10).

L'utilisation d'une matrice de pixels comme vecteur de description pose problème si la détection des points est multi-échelles ; il est impossible de comparer point à point des vecteurs de description de taille différente. Ce problème ne se rencontre pas dans le cas de distributions. En effet, le nombre de pixels contenus dans le voisinage local d'un point d'intérêt ne modifie en rien le nombre de classes de l'histogramme. Les caractéristiques des histogrammes sont fixées et sont identiques pour tous les vecteurs de description construits. De plus, à la différence des vecteurs de description utilisant des matrices de pixels, l'utilisation de distribution permet de définir des vecteurs de taille réduite. Cette propriété est

particulièrement intéressante pour la recherche d'images par le contenu dans une grande base de données.

Il est naturellement possible d'utiliser plusieurs composantes. Dans un tel cas, nous pouvons soit construire une distribution en dimension supérieure, ou, plus simplement, nous pouvons coller bout à bout les distributions calculées pour chacune des composantes. Parmi les composantes images existantes, l'utilisation de la norme et de la direction du gradient permet de rendre l'appariement robuste aux changements d'intensité.

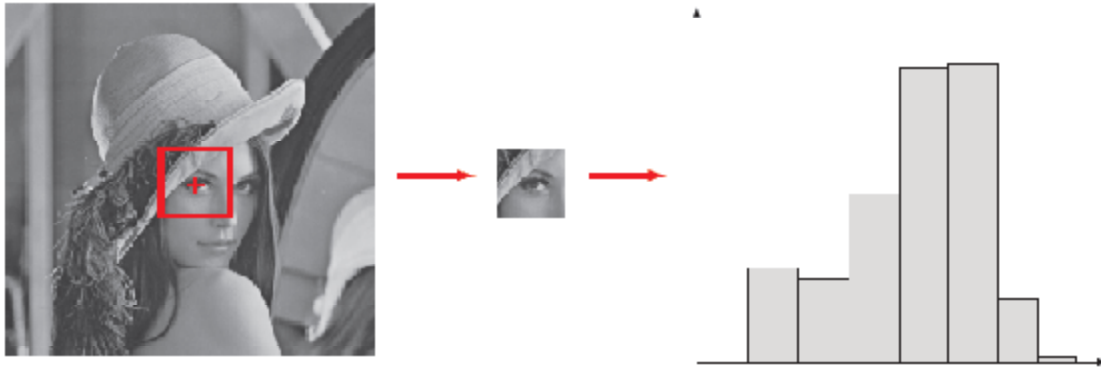


Figure V.8 Construction de vecteurs de description fondée sur l'estimation de la distribution de l'intensité. Le point d'intérêt est localisé par une croix rouge et son voisinage, dépendant de son échelle caractéristique et délimité en rouge, est de taille 97×97 . La distribution est approximée par un histogramme normalisé à 8 classes.

V.3.2.3 SIFT

La méthode SIFT [91,92] (pour Scale-Invariant Feature Transform), est une méthode permettant de décrire un point à partir des orientations locales du gradient. Les vecteurs de description construits par cette méthode sont spatialement discriminants (contrairement aux distributions) et robustes aux déformations géométriques usuelles et aux variations d'intensité.

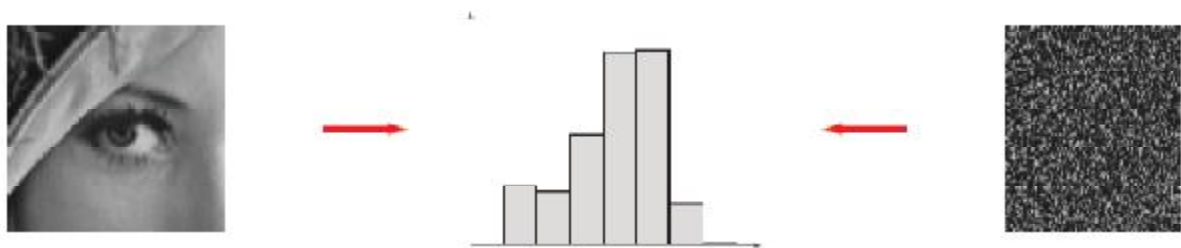


Figure V.9 Illustration du problème intrinsèque à l'utilisation de distributions pour la description de points d'intérêt. Deux images visuellement très différentes ont exactement la même distribution.

Chaque point d'intérêt est décrit à partir de son voisinage de taille 16×16 à son échelle caractéristique. Dans ce voisinage est calculé l'histogramme des orientations locales. Leur contribution à l'histogramme est pondérée par la norme du gradient et par un facteur Gaussien d'éloignement au centre du voisinage (point d'intérêt). L'invariance à la rotation est assurée en tournant le voisinage et les orientations qu'il contient d'un angle égal à l'orientation dominante (orientation maximisant l'histogramme des orientations). Les figures V.11 et V.12 illustrent la construction du vecteur de description SIFT autour d'un point d'intérêt. Le voisinage de taille 16×16 tourné est divisé en 16 (4×4) sous-ensembles de dimension 4×4 pixels. Par commodité, le voisinage représenté sur la figure est de taille 8×8 et il est découpé en 2×2 carrés. Un histogramme d'orientations est construit pour les 16 sous ensemble, chaque histogramme comporte 8 classes. Le vecteur de description est la concaténation des 16 histogrammes. Un vecteur de description SIFT comporte $16 \times 8 = 128$ éléments.

L'utilisation de l'orientation du gradient permet à la méthode d'être robuste aux variations d'intensité. L'intérêt de cette méthode réside dans le découpage du voisinage. Utiliser un histogramme pour chacun des 16 sous voisinages permet d'ajouter une information spatiale qui est perdue si l'on utilise un histogramme sur tout le voisinage. L'information spatiale n'est cependant pas aussi contraignante que dans le cas d'une simple matrice de pixels. De plus, la rotation du voisinage (et des orientations qu'il contient) et l'extraction des orientations à l'échelle caractéristique rend le vecteur de description robuste aux rotations et aux changements d'échelle. Des vecteurs de description ainsi construits donnent de très bons résultats dans tous les cas. Le descripteur SIFT est l'un des descripteurs les plus robustes et les plus discriminants [93].

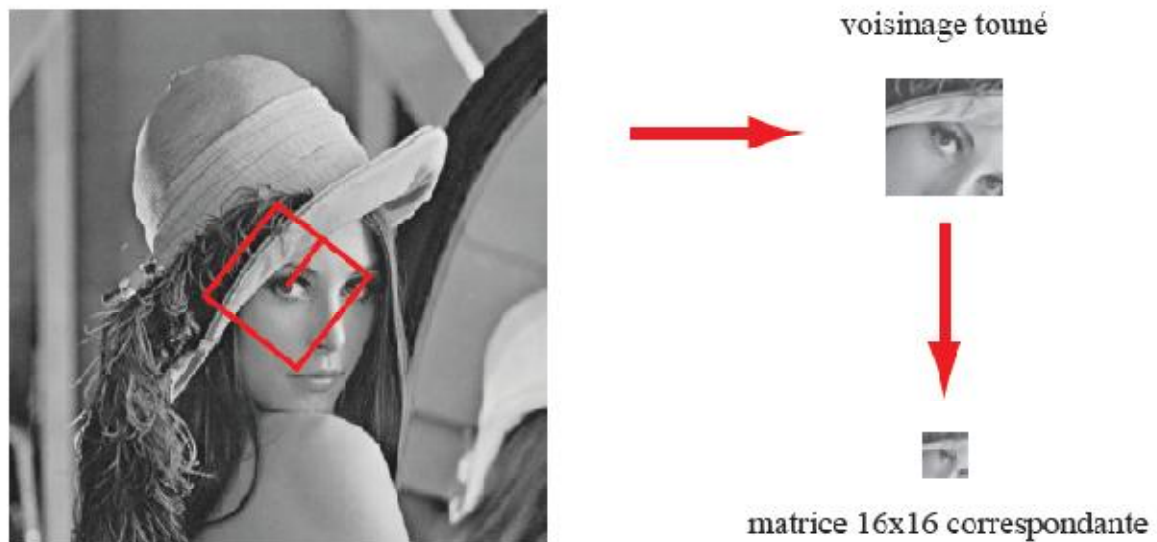


Figure V.10- Construction de la matrice de pixels utilisée pour construire le vecteur de description SIFT. Le voisinage (encadré en rouge) est tourné relativement à la direction dominante du voisinage (segment rouge). La matrice utilisée pour construire le vecteur de description est de taille 16×16 . Un simple interpolation de l'image d'origine permet d'extraire un voisinage tourné et de taille 16×16 . Pour des raisons de compréhension, l'image utilisée ici représente la composante d'intensité. La method SIFT utilise l'orientation du gradient.

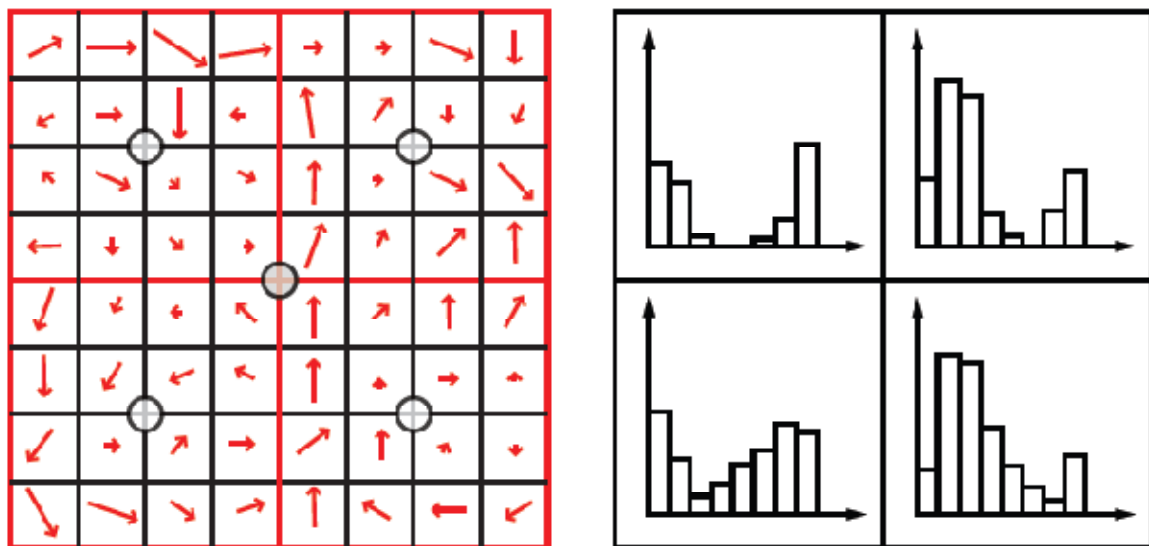


Figure V.11- Construction du vecteur de description SIFT à partir de la matrice d'orientations de taille 16×16 tournée. Nous représentons ici une matrice de taille 8×8 pour clarifier l'illustration. De même, nous découpons la matrice en 2×2 parties au lieu des 4×4 habituelles.

V.3.3 Introduction aux invariants

Afin de rendre la méthode de caractérisation locale invariante aux différentes transformations de l'image considérées, cette section présente une introduction à la théorie des invariants. Pour plus d'information le lecteur pourra se référer à [94] et à [95].

V.3.3.1 Définition théorique d'un invariant

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

Définition V.3 *Etant donnés deux ensembles E et F , un ensemble T de transformations de E dans F et I une fonction dont l'ensemble de départ est F , I est invariante par T si et seulement si :*

$$\forall e \in E \forall t, t' \in T I(t(e)) = I(t'(e)) \quad (\text{V.29})$$

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

V.3.3.2 Calcul des invariants

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

Méthodes infinitésimales

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

Définition V.4 *Un groupe de Lie est un ensemble qui est à la fois une sous-variété de R^n ou de C^n et un groupe tel que la multiplication et l'inversion sont continues.*

Les groupes de Lie sont des ensembles de fonctions paramétrées dont les paramètres définissent une structure de groupe. Dans le cadre de la définition V.4 précédente, ces fonctions vont de E dans E et l'ensemble des transformations est un groupe noté G .

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

Définition V.5

$$O(e) = \{e' \in E | \exists g \in G e' = g(e)\} \quad (\text{V.30})$$

Est l'orbite de e selon G

L'utilisation de groupes de Lie permet de calculer théoriquement des invariants par résolution d'équations différentielles. En effet, les invariants sont des fonctions analytiques f_i constantes sur les orbites, mais qui peuvent distinguer ces différentes orbites. Comme ces fonctions f_i

sont constantes sur chaque orbite, leur gradient est donc orthogonal aux espaces tangents aux orbites. Soit f_j une fonction invariant dont le gradient est noté $\vec{\nabla} f_j$, et soit $V_i(e)$ une base de l'espace tangente à l'orbite de e , on a alors :

$$\forall i \forall j \forall e \vec{\nabla} f_j v_i(e) = 0 \quad (\text{V.31})$$

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

Méthode par généralisation et contrainte

La résolution de l'équation différentielle présentée à la section précédente s'avère parfois délicate. Gros [95] a donc proposé d'utiliser une méthode classique pour calculer des invariants : pour chercher des invariants associés à un problème donné, on généralise le problème à résoudre. Ensuite on calcule les invariants pour le problème généralisé et on exprime le fait que le problème de départ en est un cas particulier. Le problème de cette méthode est de trouver une généralisation du problème de départ pour laquelle on sache calculer des invariants.

Mirbach [96] a proposé une méthode de généralisation qui exprime les solutions de l'équation différentielle sous la forme d'une intégrale calculée sur toute l'image. Pour une transformation donnée, il propose en fait de calculer la moyenne des valeurs sur une orbite associée à cette fonction.

V.3.3.3 Dénombrement des invariants

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

Définition V.6 *Des invariants sont dits indépendants si leurs dérivées partielles sont linéairement indépendantes.*

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

$$n = \dim E - (\dim G - \min_{e \in E} (\dim G_e)) \quad (\text{V.32})$$

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

$$G_e = \{g \in G \mid g(e) = e\} \quad (\text{V.33})$$

V.3.3.4 Théorème de Burns

Dans le cas de la vision par ordinateur, les ensembles E et F peuvent être différents. L'ensemble des transformations ne forme alors plus un groupe. Dans ce cas, les orbites ne forment plus une partition : elles peuvent se croiser et la relation $R(e, e')$ n'est plus une relation d'équivalence. Dans ce cas, il n'existe pas d'invariant. Ceci a été énoncé par Burns [118], puis également par Moses [97] et Clemens [98] :

Théorème V.1 *Dans le cas des configurations de n points et des projections perspectives, affines ou orthogonales, quel que soit l'entier n , les orbites se croisent de telle manière que toute fonction constante sur les orbites est constante sur tout l'ensemble d'arrivée. Il n'y a donc pas d'invariants pour ce problème exception faite des fonctions constantes.*

Dans le cas général, la recherche d'invariants pour les transformations de la scène tridimensionnelle vers l'image est donc vaine. Toutefois, il existe des invariants pour des classes géométriques d'objet 3D, comme l'ont montré Zisserman et al. [99]. Ces classes incluent les surfaces de révolution, les tubes, les objets symétriques ainsi que les polyèdres. Il est également possible d'approximer les transformations perspectives par des quasi-invariants comme nous le verrons à la section suivante.

V.3.3.5 Quasi-invariants

L'apparition des quasi-invariants dans la communauté en vision remonte à la fin des années 1960 et est due à Binford. Dans [100] il reprend et précise la définition des quasi-invariants. Il démontré également que les invariants au groupe des similitudes image sont des quasi-invariants pour une transformation perspective. En effet, la théorie des quasi-invariants permet de définir des invariants au premier ordre dans le cas des transformations perspectives.

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

Définition V.7 *Soit G un ensemble de transformation, la fonction f est quasi-invariante pour une transformation $g \in G$ si son développement de Taylor est constant au second ordre pour l'ensemble de transformations considérées et si son développement au premier ordre est égal au développement de g .*

Le dénombrement des quasi-invariants pour un problème donné est difficile. De même, vérifier qu'une fonction est un quasi-invariant est souvent complexe. Par contre, les quasi-invariants se révèlent souvent plus stables que les invariants et ils apportent une solution dans le cas où il n'existe pas d'invariants [55].

V.3.4 Invariance et transformations de l'image

Dans cette section les invariants obtenus pour une caractérisation basée sur les dérivées sont présentés. Les différentes transformations considérées sont une rotation image, un changement d'échelle, un changement de luminosité et un changement de point de vue. Obtenir des descriptions invariantes est également possible pour les autres caractérisations vues en section V.3.1. Par exemple, dans le cas d'une description fréquentielle, la transformée de Fourier-Mellin [101,102] permet de rendre une transformation de Fourier invariante aux rotations image et aux changements d'échelle. De même, la méthode proposée par Hu [103] rend invariants les moments aux rotations et changements d'échelle.

V.3.4.1 Rotation image

Dans cette section, nous présentons deux approches différentes pour obtenir une invariance à la rotation. La première est d'utiliser des mesures différentielles invariantes au groupe des déplacements. Cette caractérisation repose sur des combinaisons de dérivées invariantes à une rotation image. D'une manière théorique, de tels invariants ont été proposé par Koenderink [104] et Romeny et al. [105, 106, 107,119]. La deuxième approche pour obtenir une caractérisation invariante au groupe des déplacements consiste à utiliser le principe des filtres ajustables. Dans ce deuxième cas, les dérivées sont ajustées dans la direction du gradient et ainsi invariantes à une rotation image.

Invariants différentiels

A partir du jet local défini à la section V.1, Koenderink ainsi que Romeny proposent de calculer des invariants pour le groupe des déplacements $SO(2)$. Pour ce faire ils ont repris des résultats mathématiques, formulés entre autres par Hilbert [108]. Ils soulignent également la nécessité d'implémenter le calcul des dérivées de manière stable pour pouvoir calculer ces invariants à un ordre élevé.

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

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

Les L_i sont les éléments du jet local défini par la définition V.1. L représente par exemple la fonction de luminosité convoluée avec une gaussienne. La formulation de cette première partie du vecteur est donnée en notation d'Einstein et en coordonnées cartésiennes. En

notation cartésienne, les indices x et y représentent respectivement la dérivation par rapport aux variables x et y , par exemple $L_{xy} = \frac{\partial^2}{\partial x \partial y} L$. En notation d'Einstein, un indice i signifie la sommation des dérivations par rapport à l'ensemble des variables :

$$L_i = \sum_i L_i = L_x + L_y \text{ et } L_{ij} = \sum_i \sum_j L_{ij} = L_{xx} + L_{xy} + L_{yx} + L_{yy} \quad (\text{V.34})$$

On peut constater que la deuxième composante de ce vecteur est la magnitude du gradient et la quatrième le Laplacien. Il est possible de calculer les invariants pour différentes tailles σ de la gaussienne, car ils sont définis à partir des L_i . La deuxième partie du vecteur est constitué d'un ensemble complet d'invariants du troisième ordre. Ces invariants en notation d'Einstein sont :

$$\vec{\gamma}[5..8] = \begin{bmatrix} \varepsilon_{ij}(L_{jkl}L_iL_kL_l - L_{jkk}L_iL_lL_l) \\ L_{iij}L_jL_kL_k - L_{ijk}L_iL_jL_k \\ \varepsilon_{ij}L_{jkl}L_iL_kL_l \\ L_{ijk}L_iL_jL_k \end{bmatrix}$$

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

En notation cartésienne nous obtenons :

$\vec{\gamma}[5..8]$

$$= \begin{bmatrix} L_{xxx}L_yL_yL_y + 3L_{xyy}L_xL_xL_y - 3L_{xxy}L_xL_yL_y - L_{yyy}L_xL_xL_x \\ L_{xxx}L_xL_yL_y + L_{xxy}(-2L_xL_xL_y + L_yL_yL_y) + L_{xyy}(-2L_xL_yL_y + L_xL_xL_x) + L_{yyy}L_xL_xL_y \\ L_{xxy}(-L_xL_xL_x + 2L_xL_yL_y) + L_{xyy}(-2L_xL_xL_y + L_yL_yL_y) - L_{yyy}L_xL_yL_y + L_{xxx}L_xL_xL_y \\ L_{xxx}L_xL_xL_x + 3L_{xxy}L_xL_xL_y + 3L_{xyy}L_xL_yL_y + L_{yyy}L_yL_yL_y \end{bmatrix}$$

Jet ajustable

À partir du jet local défini à la section V.1, il est possible de calculer les dérivées dans une direction donnée. Pour être invariant à la rotation, cette direction est par exemple la direction du gradient. Nous donnons par la suite, les formules pour calculer jusqu'à l'ordre 3, les dérivées dans une direction θ donnée [57]. Les L_x, L_y, \dots sont les éléments du jet local.

$$L'(\theta) = L_x \cos(\theta) + L_y \sin(\theta) \quad (\text{V.36})$$

$$L''(\theta) = L_{xx} \cos^2(\theta) + 2L_{xy} \sin(\theta) \cos(\theta) + L_{yy} \sin^2(\theta) \quad (\text{V.37})$$

$$L'''(\theta) = L_{xxx} \cos^3(\theta) + 3L_{xxy} \cos^2(\theta) \sin(\theta) + 3L_{xyy} \sin^2(\theta) \cos(\theta) + L_{yyy} \sin^3(\theta) \quad (\text{V.38})$$

Dans ces formules, la dérivée d'ordre n dans une direction donnée dépend des $n + 1$ dérivées d'ordre n . Pour représenter de façon complète l'ensemble des dérivées à un ordre donné n , il faut utiliser $n + 1$ dérivées directionnelles correspondant à $n + 1$ directions $\theta_{n,i} i=0, \dots, n$. Pour des raisons de stabilité, les $n + 1$ directions $\theta_{n,i}$ utilisées doivent être espacées régulièrement. Afin d'obtenir des dérivées indépendantes de la rotation existante entre deux images, la

direction $\theta_{n,0}$ doit de plus être rapportée à l'image. Si cette direction correspond à la direction du gradient, les orientations sont alors : $\theta_{n,i} = i\pi/(n+1) + \theta_g$ où $\theta_g = \arctan(L_y/L_x)$. Le calcul de cette direction θ_g est une source d'instabilité des méthodes utilisant les jets ajustables.

Normalisation en taille de l'image

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

V.3.4.2 Changement d'échelle

Un changement d'échelle peut être dû soit à un changement de la distance entre la caméra et l'objet soit à un changement de la longueur focale de l'objectif (dans le cas d'un zoom). Nous noterons dans la suite un changement d'échelle par α .

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

Invariants à un changement d'échelle

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

$$f^{(n)}(x) = \alpha^n g^{(n)}(u) \quad (\text{V.39})$$

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

L'équation v.39 précédente montre que les dérivées n -ièmes de f et de g sont égales à un facteur multiplicatif α^n près. A partir du quotient de deux dérivées il est donc possible d'éliminer ce facteur α^n . Des invariants théoriques à un changement d'échelle sont donnés par l'équation suivante :

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

Schmid a montré dans [55] que de tels invariants sont peu stables à un changement d'échelle supérieur à 20%, ainsi que les invariants à la rotation étaient eux aussi robustes à un

changement d'échelle de 20%. Cette robustesse est cohérente avec les observations faites par Fleet dans [110] dans le contexte des filtres de Gabor.

En fait, les invariants à l'échelle n'apportent pas de stabilité supplémentaire. Ceci est dû au fait que le calcul numérique est effectué sur un support. En effet, dans le cas où les dérivées sont calculées par convolution avec les dérivées de la gaussienne, l'équation V.39 précédente se réécrit de la manière suivante :

$$\int_{-\infty}^{+\infty} I_1(\vec{x}) G_{i_1 \dots i_n}(\vec{x}, \sigma) d\vec{x} = \alpha^n \int_{-\infty}^{+\infty} I_2(\vec{u}) G_{i_1 \dots i_2}(\vec{x}, \sigma \alpha) d\vec{u} \quad (\text{V.40})$$

Où $G_{i_1 \dots i_2}$ représentent les dérivées de la fonction gaussienne définie par l'équation v.10.

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

Approche multi-échelle

Dans la littérature, il existe de nombreuses approches multi-échelle. Parmi les premières approches, il faut citer les pyramides qui ont été proposées par Burt [111] et Crowley [112,113]. Ces approches effectuent un sous-échantillonnage de l'image. En effet, Schmid a proposé dans [58] une approche basée sur la notion d'espace d'échelle où un paramètre continu définit l'échelle. Cette notion a été introduite par Witkin [114] et Koenderink [104]. Plus tard, Lindeberg [115] a étendu et résumé leur approche.

L'espace d'échelle permet de calculer les invariants à une échelle donnée. Il est cependant impossible de calculer les invariants à toutes les échelles. La discrétisation de l'espace d'échelle est donc nécessaire. De nombreux auteurs ont proposé une discrétisation par octave ou par demi-octave. Avec un tel pas de discrétisation la caractérisation obtenue s'est révélée imprécise et instable.

V.3.4.3 Changement de luminosité

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

Translation des niveaux de gris

Une translation des niveaux de gris se modélise par :

$$\tilde{I}(x, y) = I(x, y) + b \quad (\text{V.41})$$

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 $\vec{\gamma}$ sans la composante $\vec{\gamma}[0]$ est un invariant. Il est dans la suite référencé par $\vec{\gamma}_t$.

Transformation affine des niveaux de gris

Une transformation affine des niveaux de gris se modélise par :

$$\tilde{I}(x, y) = a(x, y) + b \quad (\text{V.42})$$

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

$\tilde{I}^{(n)}(x, y) = aI^{(n)}(x, y)$. N'importe quel quotient de deux dérivées est donc invariant à une transformation affine de la luminance. Il y a différentes manières de rendre le vecteur $\vec{\gamma}$ invariant à une transformation affine. Schmid [58] a choisi de diviser par la puissance adéquate de la magnitude du gradient :

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

Transformation monotone de la luminosité

Un changement de luminosité peut également être modélisé par une fonction monotone et donc inversible. L'invisibilité de la fonction (ou sa stricte monotonie) est nécessaire pour éviter une perte d'information par rapport à celle contenue dans l'image de niveaux de gris.

Florack [116] montre que le jet local permet également de calculer des invariants par rapport à n'importe quelle transformation inversible de la luminosité. Il fait remarquer que les isophotes ne sont pas modifiés sous l'action d'une transformation inversible de luminosité. Jusqu'au deuxième ordre, il existe deux invariants indépendants, notamment la courbure des isophotes k et la courbure des lignes de plus grande pente μ :

$$k = \frac{\varepsilon_{ij} \varepsilon_{jk} L_i L_{jk} L_l}{(L_m L_m)^{3/2}} = \frac{2L_x L_y L_{xy} - L_x^2 L_{yy} - L_y^2 L_{xx} - L_y^2 L_{xx}}{L_x^2 + L_y^2} \quad (\text{V.43})$$

$$\mu = \frac{\varepsilon_{ij} L_j L_k L_{ik}}{(L_m L_m)^{3/2}} = \frac{L_{xy} (L_y L_y - L_x L_x) + L_x L_y (L_{xx} - L_{yy})}{L_x^2 + L_y^2} \quad (\text{V.44})$$

Berthod et al. [117] présente une famille d'invariants à une transformation monotone de la luminosité. Ces invariants sont basés sur les orientations du gradient et ses dérivées partielles.

V.3.4.4 Changement de point de vue

Les invariants à une rotation image sont calculés dans un cadre multi-échelle. On obtient donc une caractérisation invariante au groupe des similitudes dans l'image. Binford[118] a montré que de tels invariants sont des quasi-invariants à une transformation perspective, c'est-à-dire qu'ils sont localement invariants à une telle transformation. Ceci est confirmé par Schmid [55].

Conclusion

Dans ce chapitre, nous avons mené une étude comparative de différents détecteurs de points d'intérêt. Le critère d'évaluation de ces détecteurs est la répétabilité de détection. Ce chapitre a également présenté un état de l'art des méthodes de caractérisation d'une image. La théorie des invariants nous permet d'introduire une caractérisation invariante aux différentes transformations. Celle-ci est basée sur les dérivées gaussiennes du signal. L'utilisation de combinaisons de ces dérivées permet d'être invariant à une rotation image. Puisqu'il n'existe pas d'invariant à l'échelle qui soit numériquement applicable, l'utilisation d'un cadre multi-échelle permet d'obtenir une robustesse à un changement d'échelle [55]. Une robustesse à un changement de luminosité peut être obtenue de deux manières différentes. La première manière consiste à ne pas utiliser la moyenne d'intensité des niveaux de gris comme composante du vecteur de caractéristiques. La deuxième repose sur des invariants spécifiques à un changement de luminosité. Le nombre de ces invariants est cependant réduit.

Chapitre VI

Tests et résultats

Introduction

Il existe des logiciels de traitements d'images dans lesquels on réalise un traitement en assemblant graphiquement des boîtes correspondants à des traitements élémentaires ou en utilisant des menus. Ces logiciels conviennent parfaitement lorsqu'on doit traiter une application simple et/ou standard. Par contre, dès que le problème à traiter s'éloigne un peu d'un problème standard, on ne peut se dispenser de comprendre le traitement d'image, et de développer ses propres traitements. Dans le travail que nous avons effectué, les algorithmes de détection de points d'intérêts ont été implantés grâce au langage de programmation Matlab, version 7.0.

Nous avons utilisé des images de test omnidirectionnelles capturées par des caméras catadioptrique fournies par LAMFA (Laboratoire Amiénois de Mathématiques Fondamentale et Appliquée AMIENS).

Dans ce qui suit, nous présentons les différents tests et résultats obtenus afin de comparer les méthodes de détection de points d'intérêts proposées auparavant.

VI.1 Généralités

L'objet de ce travail étant la détection de points d'intérêts sur les d'images omnidirectionnelles, plusieurs images ont été choisies pour vérifier l'efficacité des différents détecteurs. Nous avons utilisé le traitement classique pour prouver ces limites et en ouvrant ainsi la perspective de créer des détecteurs analogues sur la sphère équivalente.

Deux types d'images sont étudiées : des images de tests classiques (plane) et des images omnidirectionnelles, dans le but de comparer les résultats obtenus par les différents détecteurs sur les images classiques et ceux obtenus pour les images omnidirectionnelles.

Nous présentons dans ce qui suit les images retenues pour nos tests.

Images de tests classiques

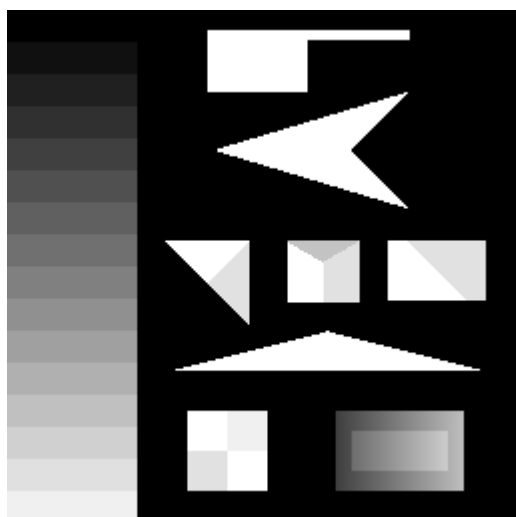


Figure VI.1 –*Image Test1*

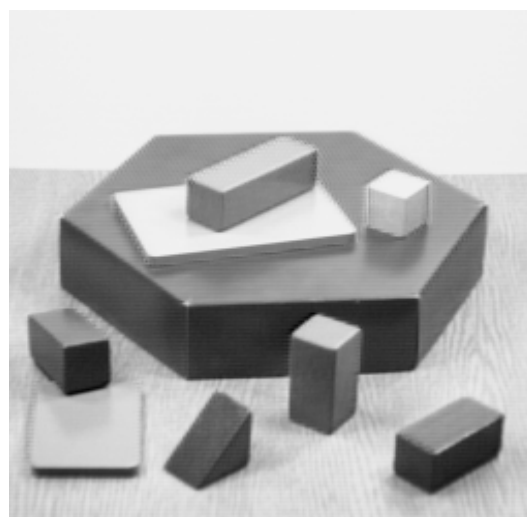


Figure VI.2 –*Image Test2*



Figure VI.3 –*Image Porte*



Figure VI.4 –*Image Tournesol*

Images de tests omnidirectionnelles

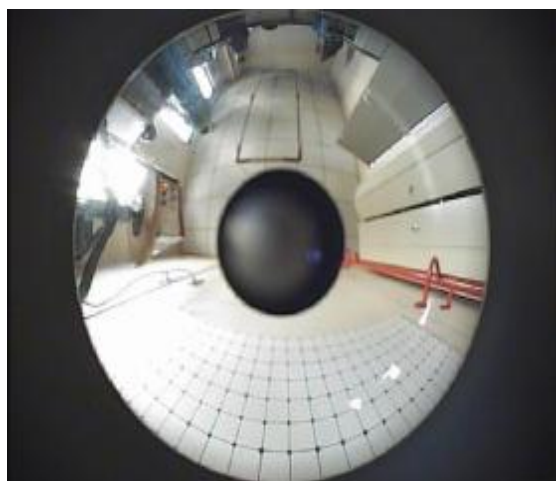


Figure VI.5-Image Bureau

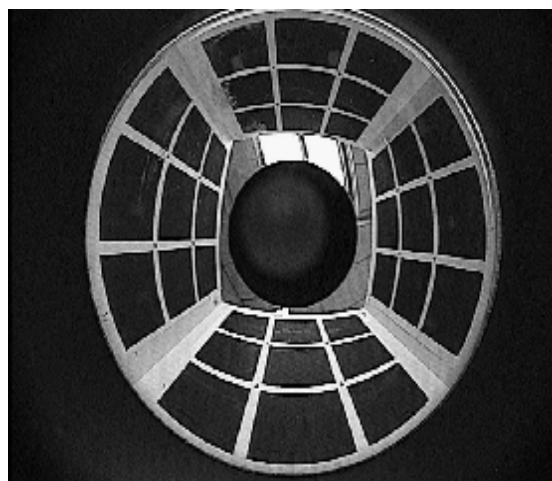


Figure VI.6 –Image Mire

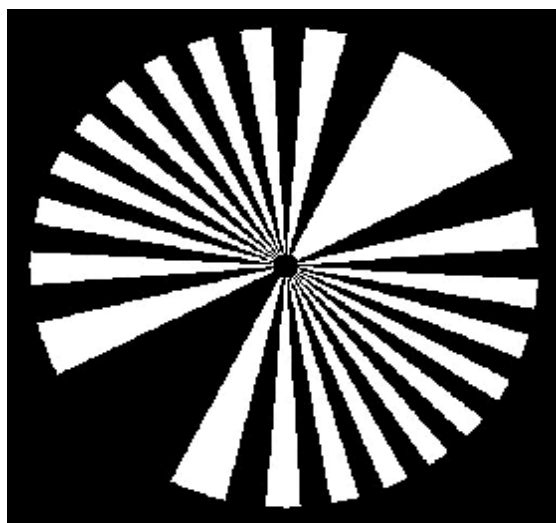


Figure VI.7 –Image Radiales



Figure VI.8 –Image Droites

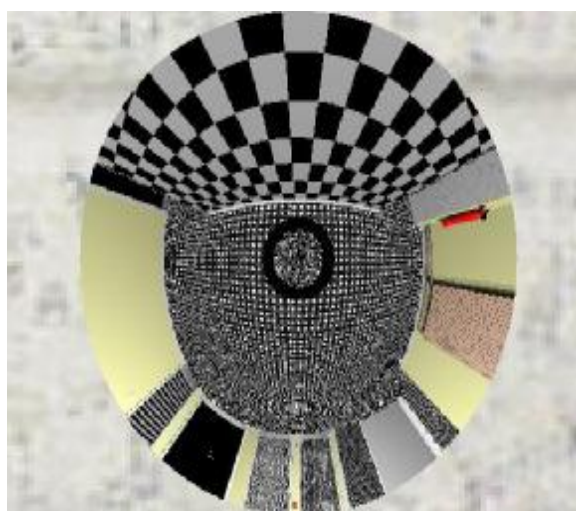


Figure VI.9 –Image Texture11



Image VI.10 –Image Texture14

Pour nos tests, nous avons retenu les détecteurs suivants : Harris, SUSAN, Harris Laplace et LoG (Lindeberg).

VI.2 Détecteur de Harris

Résultats obtenus sur les images classiques,

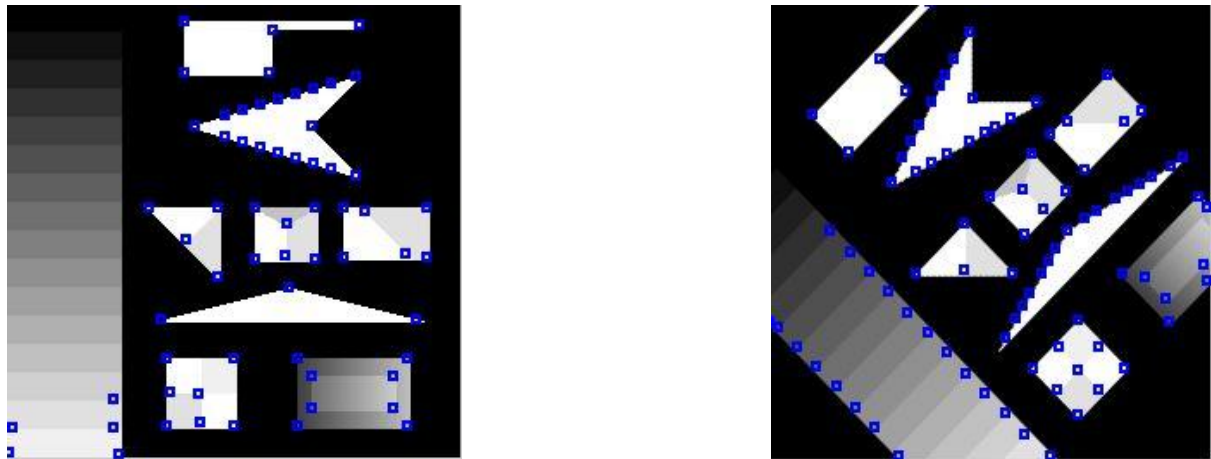


Figure VI.11 -Points d'intérêts détectés sur l'image Test1 par la méthode de Harris pour $\sigma=1.5$, la rotation entre l'image de gauche et l'image de droite est de 45 degrés.

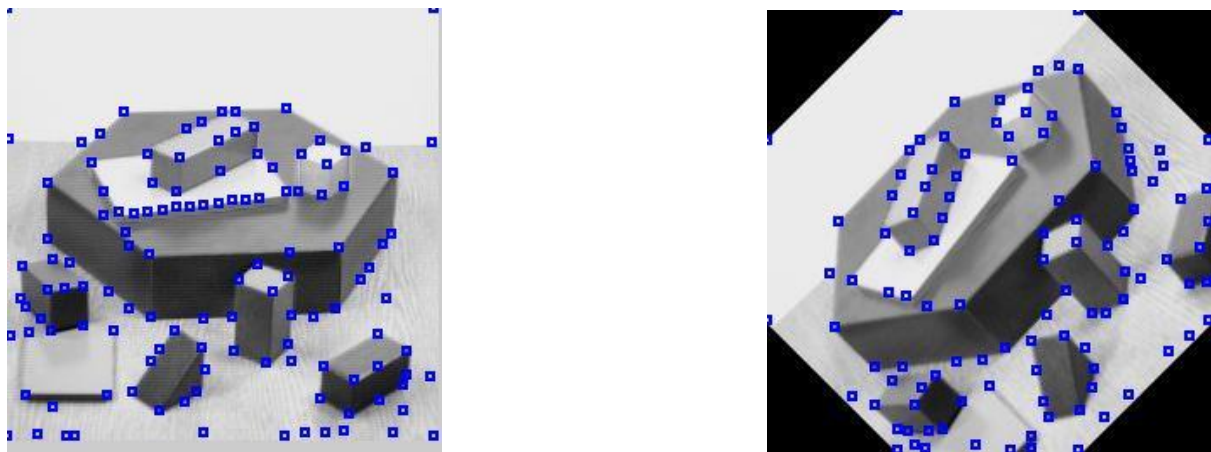


Figure VI.12 -Points d'intérêts détectés sur l'image Test2 par la méthode de Harris pour $\sigma=1.5$, la rotation entre l'image de gauche et l'image de droite est de 45 degrés.

(Figure VI.11, VI.12) Montre le détecteur de Harris appliqué à l'image Test1 avec $\sigma=1.5$, L'opérateur Harris détecte presque tous les coins, mais on constate qu'il y a une mauvaise localisation de certains coins et détecte quelques faux coins. Il a un taux de détection plus élevé que l'opérateur Moravec qui détecte beaucoup de faux coins. On peut constater que la répétabilité obtenue est bonne. Même avec une rotation de l'image, les points obtenus sont répétables sauf qu'il y a quelques points de plus qui sont apparus après la rotation de 45°, qui

sont dus au fait que la réponse de Harris est anisotrope, donc les bords diagonaux peuvent être détectés comme étant des faux coins. On peut bien remarquer ça sur la figure VI.14 .

Figure VI.13 montre l'opérateur Harris appliqué à l'image Porte. Les paramètres ont été choisis en vue de détecter la plupart des coins tout en essayant de minimiser le nombre de faux coins détectés. L'opérateur Harris détecte la majorité des coins vrais. On peut constater qu'en augmentant le paramètre σ , on minimise le nombre de faux coins détectés. Donc le nombre de coins détectés dépend de ce paramètre.

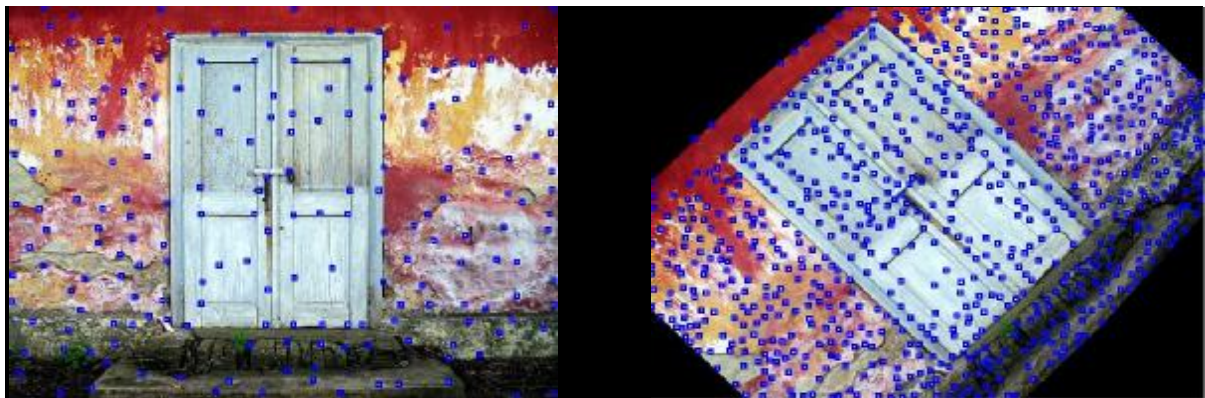


Figure VI.13 -Points d'intérêts détectés sur l'image Porte par la méthode de Harris $\sigma=3.5$ pour l'image de gauche et $\sigma=1.5$ pour l'image de droite, la rotation entre l'image de gauche et l'image de droite est de 45 degrés.

Résultats obtenus pour les images omnidirectionnelles :

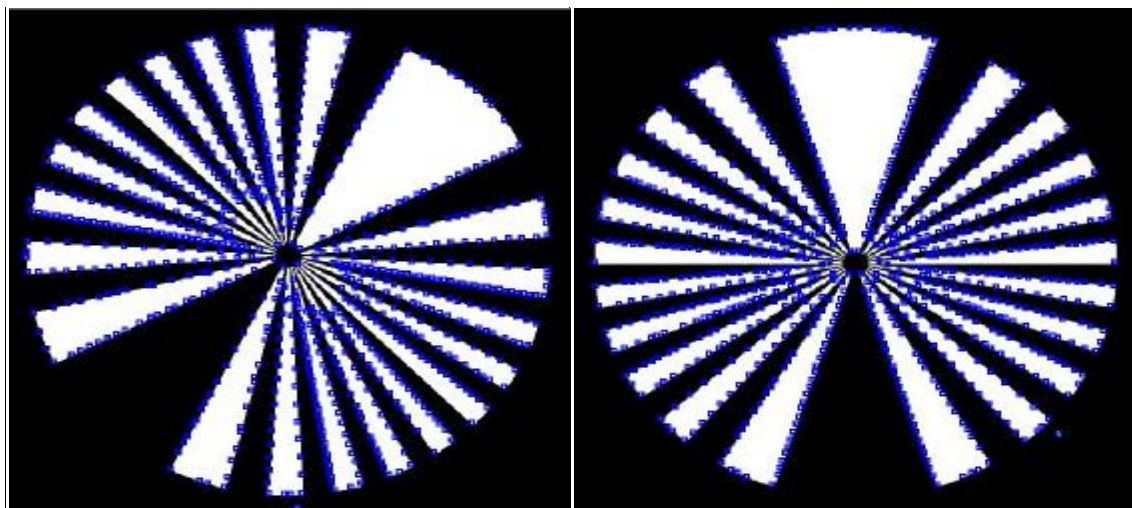


Figure VI.14 -Points d'intérêts détectés sur l'image Radiales par la méthode de Harris pour $\sigma=1.5$, la rotation entre l'image de gauche et l'image de droite est de 45 degrés.

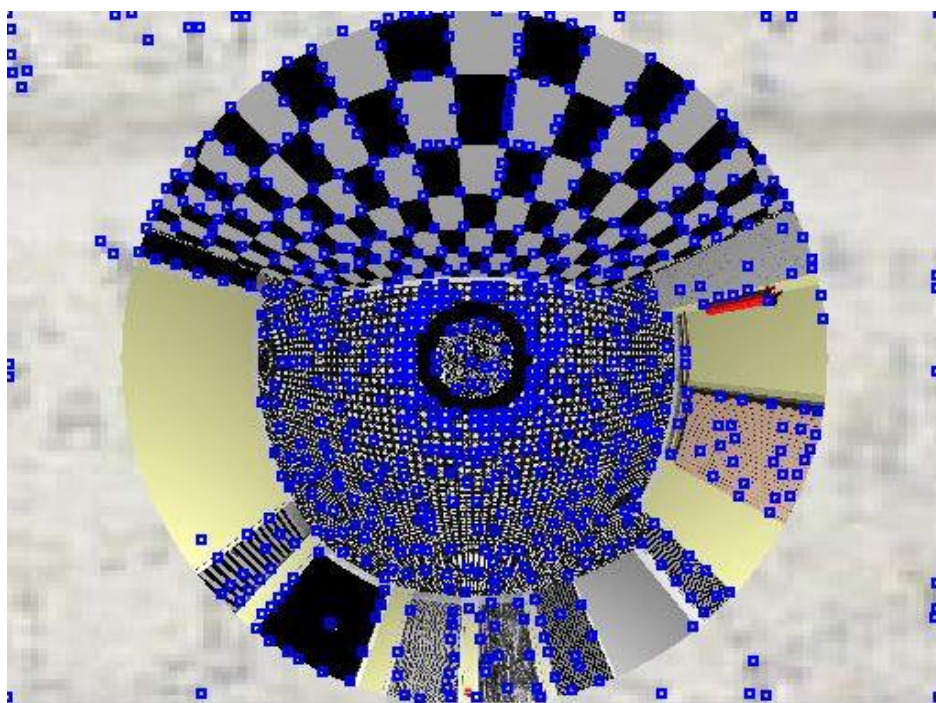


Figure VI.18 -Points d'intérêts détectés sur l'image *Texture11* par la méthode de Harris pour $\sigma=1.5$,

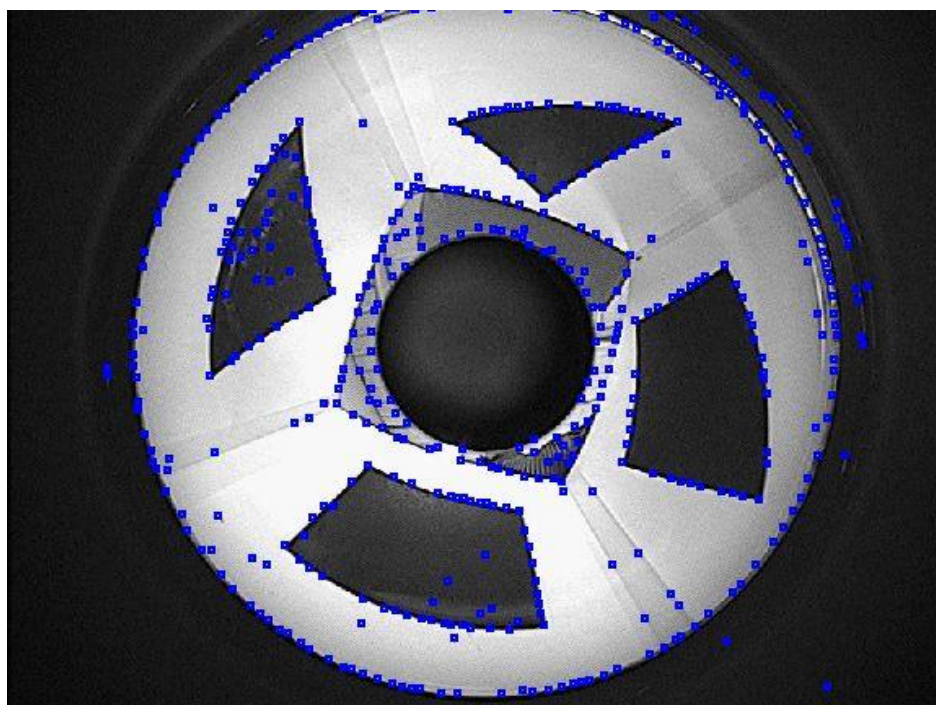


Figure VI.15 -Points d'intérêts détectés sur l'image *Droites* par la méthode de Harris pour $\sigma=1.5$,

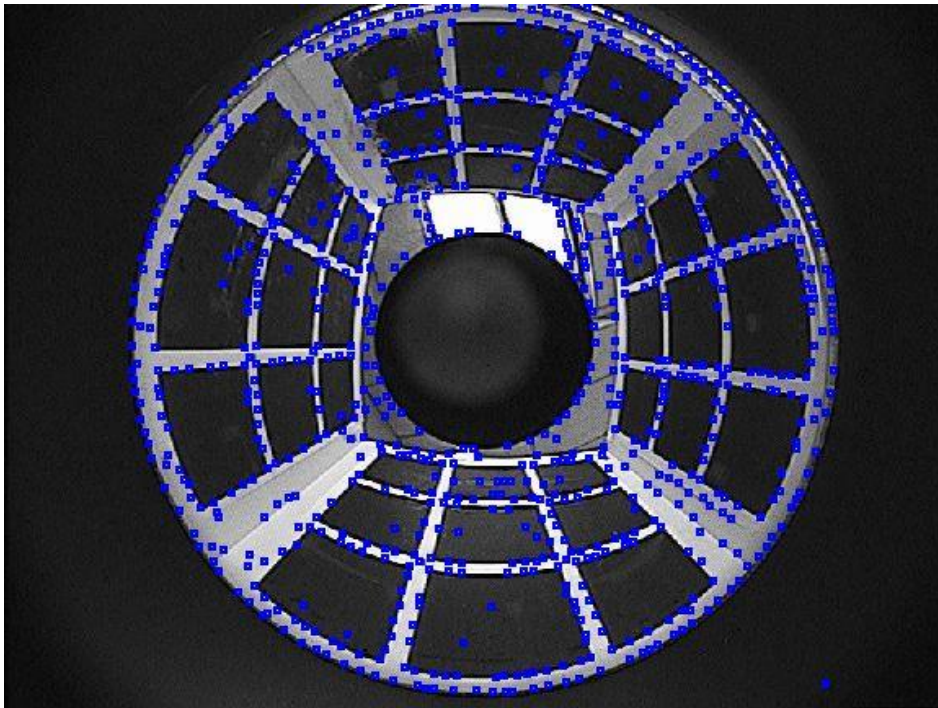


Figure VI.16-Points d'intérêts détectés sur l'image Mire par la méthode de Harris pour $\sigma=1.5$,

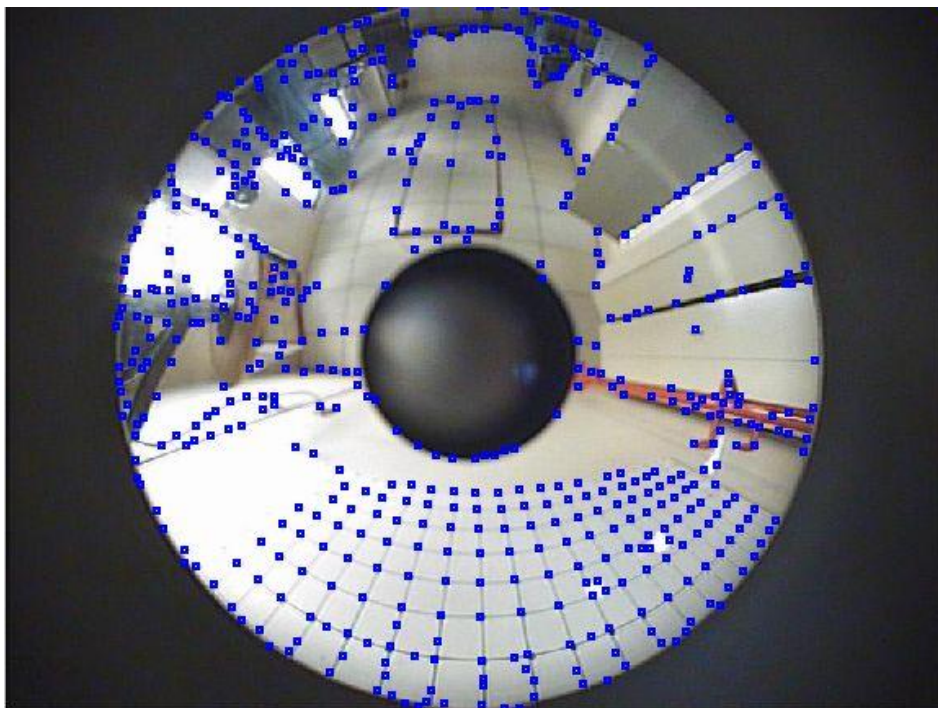


Figure VI.17 -Points d'intérêts détectés sur l'image Bureau par la méthode de Harris pour $\sigma=1.5$,

Pour les images omnidirectionnelles, on constate que le détecteur de Harris détecte beaucoup de faux coins et une mauvaise localisation de certains coins. En fait les distorsions induites par la géométrie du miroir, introduit des erreurs, surtout sur les diagonales comme on peut le voir sur la figure VI.14 ou beaucoup de faux coins sont détecté. Cependant, la grande variation angulaire près du centre des images rend les résultats délicats, donc la majorité des coins détectés sont sur la périphérie de l'image. Deux problèmes alors se posent : si on veut détecter les coins du centre de l'image, il faut diminuer σ , mais on aura des chevauchements entre les points comme on peut le voir sur la figure VI.15.

Conclusion

Le détecteur de Harris est largement utilisé, du fait qu'il possède un taux de détection élevé et une bonne répétabilité. Toutefois, il est sensible au bruit, mauvaise détection des jonctions en T et il n'est pas invariant par rotation. La sensibilité au bruit peut être réduite en utilisant une taille de fenêtre plus grande, mais cela augmente le temps de calcul de l'algorithme, nuit à la localisation et augmente le temps de calcul. Pour les images omnidirectionnelles, les distorsions induites par la géométrie du miroir rend les résultats délicats.

VI.3 Détecteur SUSAN

Résultats obtenu pour les images classiques :

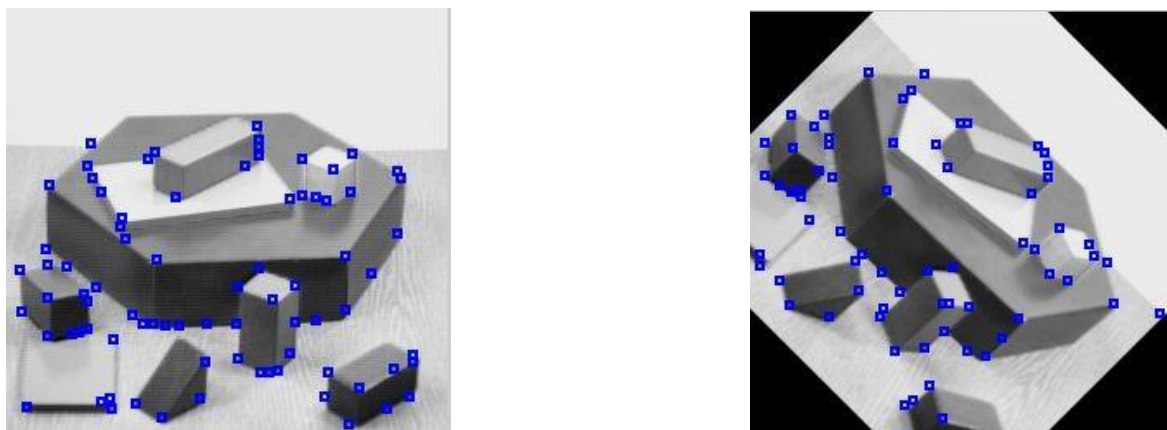


Figure VI.19 -Points d'intérêts détectés sur l'image Test2 par la méthode SUSAN pour $\sigma=3$, la rotation entre l'image de gauche et l'image de droite est de 45 degrés.

Résultats pour les images omnidirectionnelles :

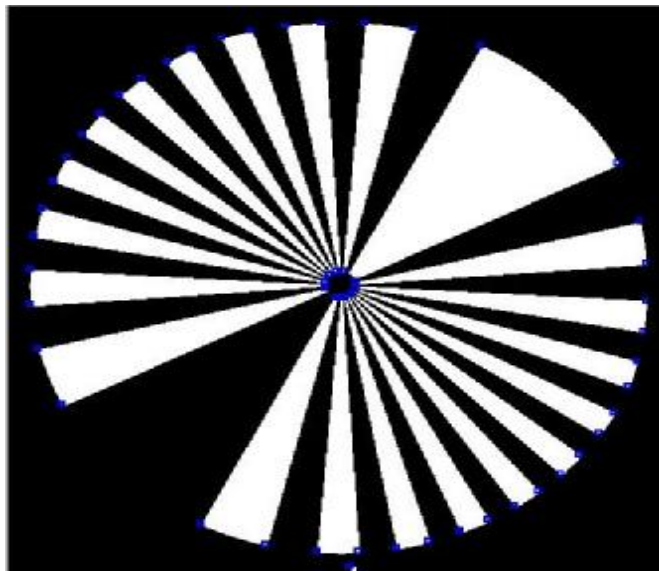


Figure VI.20 -Points d'intérêts détectés sur l'image *Radiales* par la méthode *SUSAN* pour $\sigma=3$,

Les figures VI.19 et VI.20 nous montrent la fiabilité et la précision de l'algorithme *SUSAN*. On constate qu'il n'y a pas beaucoup de faux coins détectés, et même avec une rotation de 45° les points détectés sont répétables, et il n'y a pas d'effet de bords qu'on avait trouvé avec l'algorithme *Harris*.

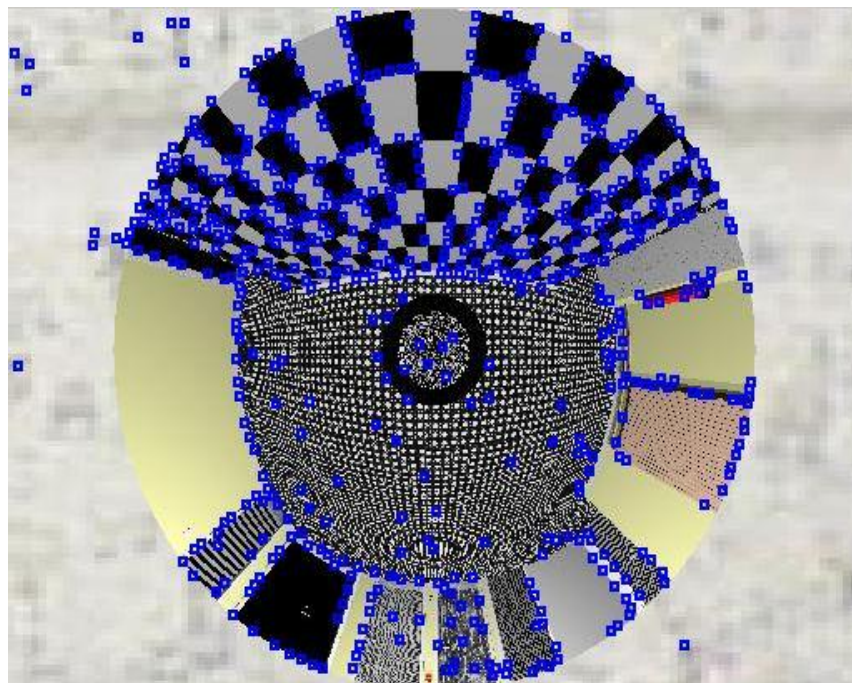


Figure VI.21 -Points d'intérêts détectés sur l'image *Texture11* par la méthode *SUSAN* pour $\sigma=3$,

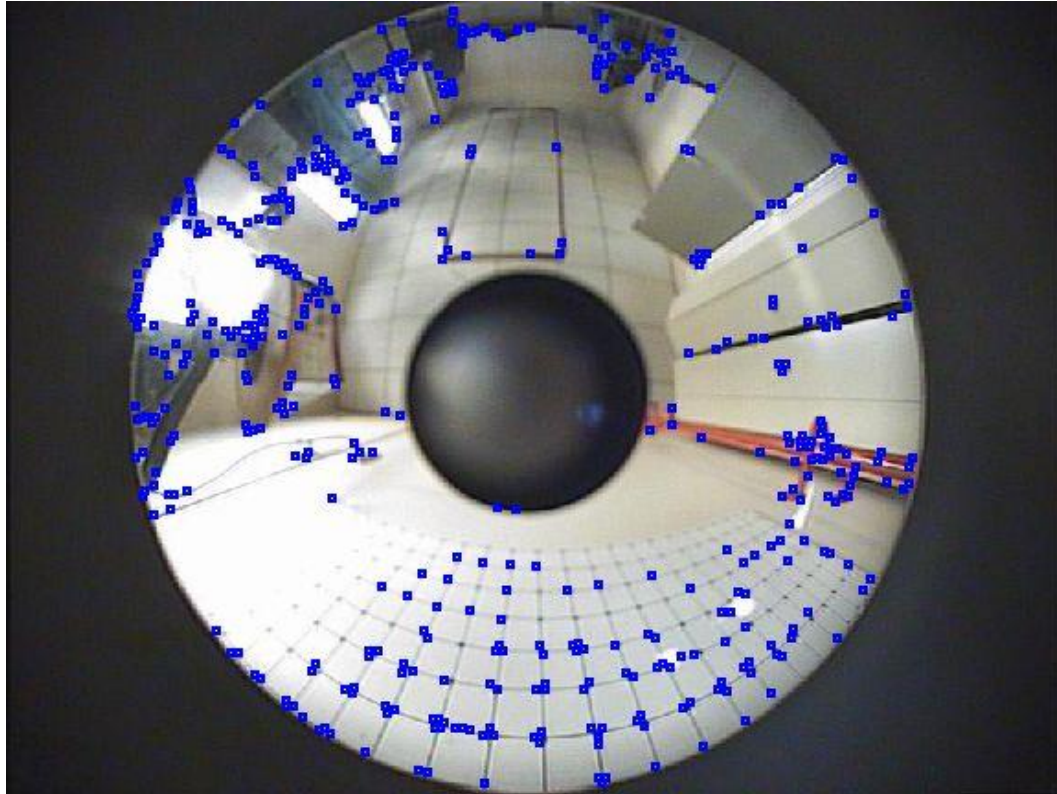


Figure VI.22 -Points d'intérêts détectés sur l'image Bureau par la méthode SUSAN pour $\sigma=3$

Pour les images omnidirectionnelles, on constate toujours que la majorité des points d'intérêts détectés sont sur la périphérie de l'image. Le détecteur SUSAN détecte moins de faux coins que le détecteur de Harris. Il élimine surtout les effets de bord comme on peut bien le voir sur la figure VI.14 et de plus, il n'est pas sensible aux bruits car le détecteur SUSAN n'utilise pas des dérivées.

Conclusion

Le détecteur SUSAN a une très bonne détection et précision, stable et efficace. La localisation des éléments est indépendante de la taille du masque utilisé. La suppression du bruit s'avère bonne. Les jonctions complexes sont détectées correctement. Dans ces tests, SUSAN a réduit beaucoup plus le temps de calcul par rapport à l'algorithme de Harris (en moyenne 10 fois plus rapide). Ce facteur est très important pour les applications temps réel.

VI.4 Détecteur Harris-Laplace

Résultats obtenus pour les images classiques :



Figure VI.23 -Détection de points d'intérêts par le détecteur de Harris-Laplace sur l'image *Porte*.

Les figures VI.23 et VI.24 montrent un résultat de détection de points d'intérêt par la méthode de Harris-Laplace sur les images classiques (porte et tournesol). Les points extraits correspondent effectivement pour la plupart à des coins dans l'image. La taille des cercles indique l'échelle caractéristique de chaque point d'intérêt. Cette approche est différente de celle de Lindeberg pour laquelle, seuls les blobs sont détectés.



Figure VI.24 -Détection de points d'intérêts par le détecteur de Harris-Laplace sur l'image *Tournesol*.

Résultats obtenus pour les images omnidirectionnelles :

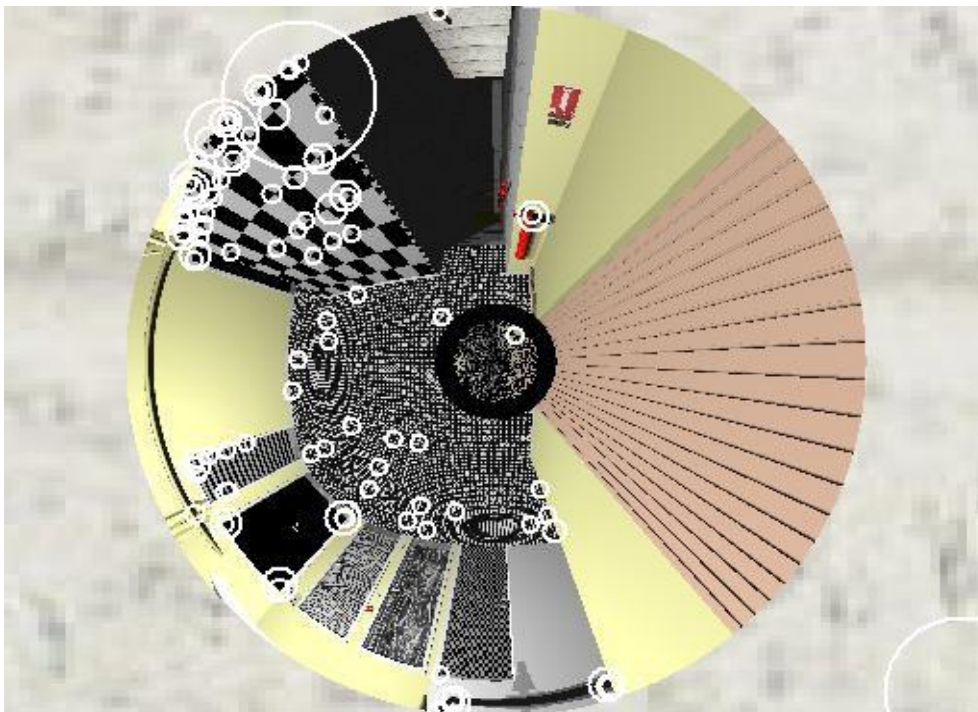


Figure VI.25 -Détection de points d'intérêts par le détecteur de Harris-Laplace sur l'image *Texture14*.

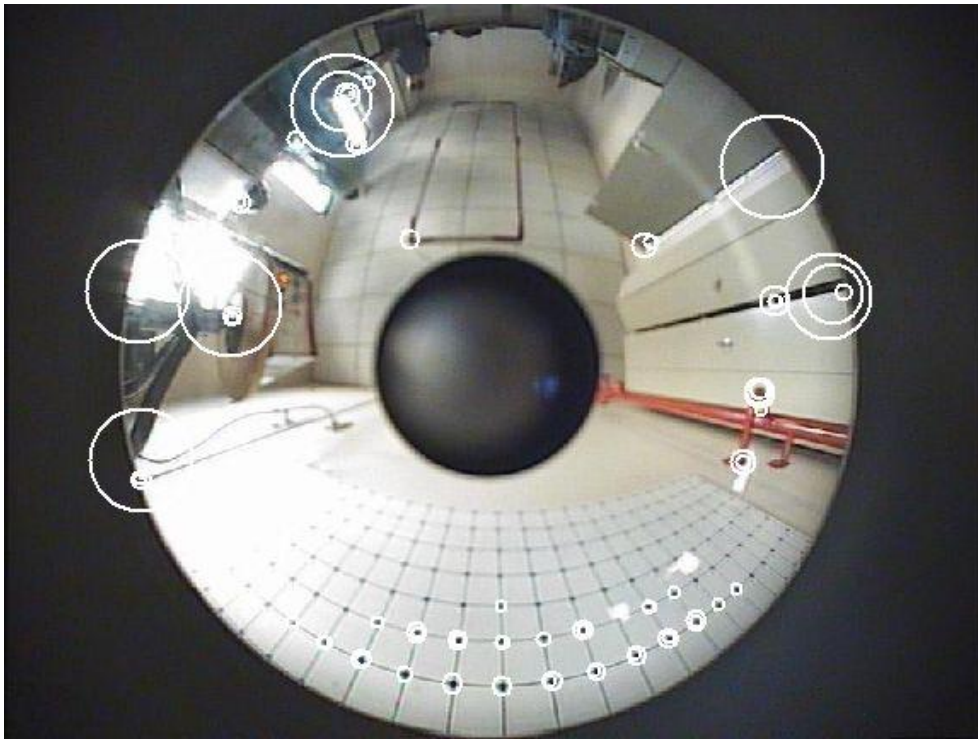


Figure VI.26 Détection de points d'intérêts par le détecteur de Harris-Laplace sur l'image Bureau.

Les figures VI.25 et VI.26, montrent les points d'intérêts par la méthode de Harris Laplace sur les images omnidirectionnelle. On constate que la majorité des points détectés sont sur la périphérie de l'image.

Conclusion

Nous pouvons remarquer que le détecteur de Harris Laplace est beaucoup moins sensible au bruit de l'image que le détecteur de Harris. Cette robustesse est induite par la détection multi-échelles. Le détecteur de Harris-Laplace est considéré à l'heure actuelle comme étant le meilleur détecteur multi-échelles au sens des critères de répétabilité et de précision. Pour les images omnidirectionnelles, le détecteur de Harris Laplace détecte la majorité des points sur la périphérie, ce qui est dû aux déformations induites par la géométrie du miroir.

VI.5 Détecteur LoG (Leindeberg)

Résultats obtenus pour les images classiques :



Figure VI.27 -Détection de blobs par la méthode LoG sur l'image Tournesol.



Figure VI.28 -Détection de points d'intérêts par le détecteur LoG sur l'image Porte.

Résultats obtenus pour les images omnidirectionnelles

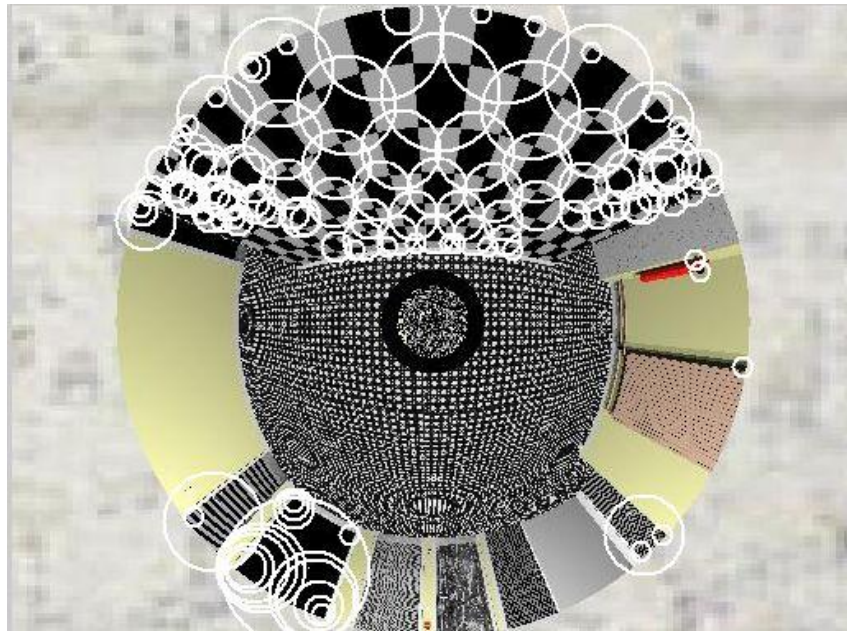


Figure VI.29 Détection de points d'intérêts par le détecteur LoG sur l'image Texture11.

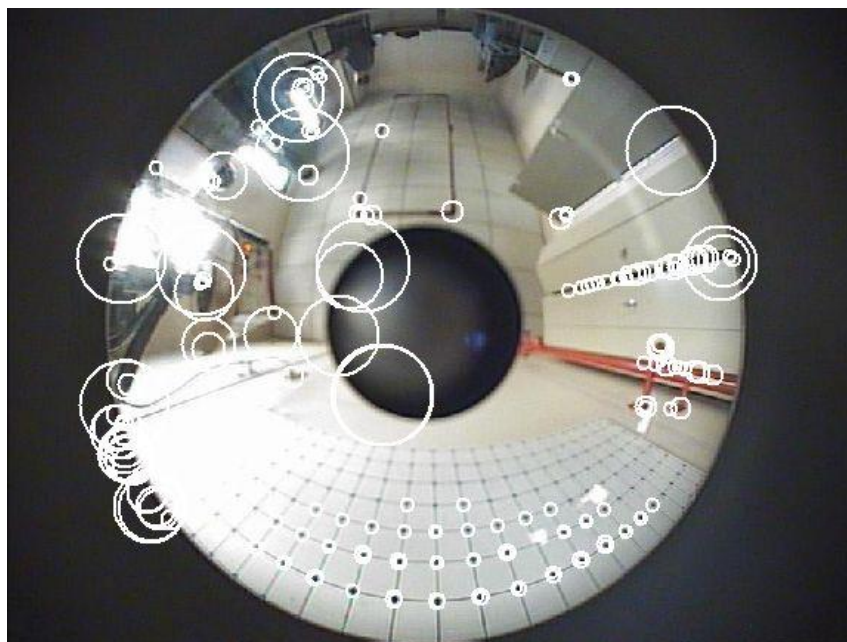


Figure VI.30 Détection de points d'intérêts par le détecteur LoG sur l'image Bureau.

Les figures VI.27 et VI.28, montrent un résultat de détection de points d'intérêts par la méthode LoG sur les images classiques. Les points détectés sont ainsi invariants aux

changements d'échelle. On parle de détection de blobs plutôt que de détection de points car la détection exhibe des régions d'intérêt et non des points d'intérêt.

Les figures VI.29 et VI.30, montrent les résultats obtenus sur les images omnidirectionnelles. La majorité des blobs détectés sont sur la périphérie de l'image. La taille des cercles indique l'échelle caractéristique de chaque blob.

Conclusion

Le détecteur LoG est un détecteur multi-échelle. Les points détectés sont ainsi invariants aux changements d'échelle. On parle de détection de blobs plutôt que de détection de points car la détection exhibe des régions d'intérêt et non des points d'intérêt comme nous pouvons le voir sur la figure VI.27. Cependant, pour les images omnidirectionnelles, la majorité des blobs détectés sont sur la périphérie de l'image.

Conclusion Générale

Conclusion générale

Dans ce travail, nous avons étudié la détection de points d'intérêts sur les images omnidirectionnelle. Nous avons présenté plusieurs approches et méthodes existantes. Nous avons jugé utile de faire un état de l'art sur les différentes méthodes et différents détecteurs. Nous avons donné le premier détecteur proposé par Moravec et en suite, les différentes méthodes qui sont apparues au cours des années pour améliorer ce dernier.

D'après les tests que nous avons effectués, nous avons constaté que la majorité des points d'intérêts détectés par les différents détecteurs sur les images omnidirectionnelles sont sur la périphérie de l'image. En effet, une image omnidirectionnelle présente une résolution plus faible au centre de l'image qu'à la périphérie.

Traiter donc les images omnidirectionnelles comme des images planes classiques aboutit à des erreurs puisque la géométrie du miroir n'est pas prise en compte.

Une perspective intéressante consisterait alors à développer des détecteurs analogues dans la sphère équivalente qui vont agir de la même manière quelque soit l'emplacement de l'opérateur sur la sphère.

Bibliographie

- [1] H. Moravec. *Towards automatic visual obstacle avoidance*. In Proceedings of the 5th International Joint Conference on Artificial Intelligence, page 584, August 1977.
- [2] H. Moravec. *Obstacle avoidance and navigation in the real world by a seeing robot rover*. Technical report, Robotics Institute, Carnegie Mellon University & doctoral dissertation, Stanford University, USA, September 1980.
- [3] C. Harris and M. Stephens. *A combined corner and edge detector*. In Proceedings of The Fourth Alvey Vision Conference, pages 147–151, Manchester, UK, 1988.
- [4] Vandeportaele B., *Contributions à la vision omnidirectionnelle : Etude, Conception et Etalonnage de capteurs pour l'acquisition d'images et la modélisation 3D*, Thèse, 2006.
- [5] V Nalwa. *A true Omnidirectional Viewer*. Rapport technique, Bell Laboratoires, Holmdel NJ 07733, USA, February 1996.
- [6] Ross Cutler, Yong Rui, Anoop Gupta, JJ Cadiz, Ivan Tashev, Li wei He, Alex Colburn, Zhengyou Zhang, Zicheng Liu & Steve Silverberg. *Distributed Meetings : A Meeting Capture and Broadcasting System*,. In ACM Multimedia, 2002.
- [7] Harsh Nanda & Ross Cutler. *Practical Calibrations for a real-time digital omnidirectional camera*. In Computer Vision and Pattern Recognition, Decembre 2001.
- [8] T. Morita, Y. Yasukawa, Y. Inamoto, T. Uchiyama & S. Kawakami. *Measurement in three dimensions by motion stereo and spherical mapping*. In IEEE International Conference on Computer Vision and Pattern Recognition, pages 422–434, 1989.
- [9] P. Greguss. *PAL-optic based instruments for space research and robotics*. Laser and Optoelektronik, pages 43–49, 1996.
- [10] Y. Yagi & S. Kawato. *Panorama scene analysis with conic projection*. In Int. Workshop on Intelligent Robotic and Systems, pages 181–187, 1990.
- [11] Y. Yagi, S. Kawato & Tsuji. *Real-time omnidirectional image sensor (COPIS) for vision-guided navigation*. IEEE Trans. Robotics and Automation, vol. 10, no. 1, pages 11–22, 1994.
- [12] Cyril Drocourt, Laurent Delahoche, Eric Brassart & Cyril Cauchois. *Experimental Result Got with Omnidirectional Vision Sensor :SYCLOP*. In IEEE Workshop on Omnidirectional Vision, Hilton Head Island, S C, 2000.
- [13] Cyril Drocourt, Laurent Delahoche, Eric Brassart & Cyril Cauchois. *Simultaneous Localization and Map building paradigm based on omnidirectional stereoscopic vision*. In 10th International conference on advanced robotics (ICAR2001), Budapest, Hungary, August 2001.

- [14] B. Marhic, E. Mouaddib, D. Fofi & E. Brassart. *Localisation absolue par le capteur omnidirectionnel SYCLOP*. Traitement du Signal, vol. 17, pages 195–206, 2000.
- [15] Zhigang Zhu, Shiquiang Yang, Guanyou Xu, Xueyin Lin & Dingji Shi. *Fast Road Clasification and Orientation Estimation Using Omni-View Image an Neural Networks*. IEEE Transactions on Image Processing, vol. 7, pages 1182–1197, 1998.
- [16] Libor Spacek. *Omnidirectional Catadioptric Vision Sensor with Conical Mirrors*. In Towards Intelligent Mobile Robotics TIMR03, Bristol, UK., August 2003.
- [17] Kazumasa Yamazawa, Yasushi Yagi & Masahiko Yachida. *Omnidirectional Imaging with Hyperboloidal Projection*. Prof. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, vol. 2, pages 1029–1034, 1993.
- [18] Shree K. Nayar. *Omnidirectional Video Camera*. Proc. of DARPA Image Understanding Workshop, May 1997.
- [19] Shree K. Nayar. *Catadioptric Omnidirectional Camera*. In Computer Vision and Pattern Recognition, 1997.
- [20] J. Salvi R. Orghidan, E. Mouaddib. *Système de vision 3d omnidirectionnelle*. In 9èmes journées ORASIS, Puy-de-Dôme, France, May 2005.
- [21] J.J. Gonzalez-Barbosa. *Vision panoramique pour la robotique mobile : stéréovision et localisation par indexation d’images*. PhD thesis, Universite de Toulouse III, 2004.
- [22] C. GEYER and K. DANIILIDIS, *A unifying theory for central panoramic systems*. In In Proc. Europ. Conf. on Computer Vision, Dublin, Ireland, 2000, 2000.
- [23] J.P. BARRETO and H. ARAUJO, *Geometric properties of central catadioptric line images*. In ECCV02, page IV: 237 ff., 2002.
- [24] Mouaddib E.M., *La Vision Omnidirectionnelle*, Journées Nationale de la Recherche en Robotique, 2005.
- [25] S.B. KANG, *Catadioptric self-calibration*. In CVPR00, pages I: 201-207, 2000.
- [26] J. FABRIZIO, J.P. TAREL and R. BENOSMAN, *Calibration of panoramic catadioptric sensors made easier*. In OMNIVIS02, pages 45-52, 2002
- [27] Strauss O., Comby F., *Variable structuring element based fuzzy morphological operations for single viewpoint omnidirectional images*. Pattern Recognition, volume 40, number 12, 3578-3596, 2007.
- [28] Jacquey F., Loquin K., Comby F., Strauss O., *Non-Additive Approach for Gradient-Based Edge Detection*. IEEE International Conference on Image Processing, volume 3, III49-III52, 2007.

-
- [29] Bogdanova I., Bresson X., Thiran J.P., Vandergheynst P., *Scale Space Analysis and Active Contours for Omnidirectional Images*. IEEE Transactions on Image Processing volume 16, number 7, 1888-1901, 2007.
 - [30] Bazin J.C., Kweon I.S., Demonceaux C., Vasseur P., *Rectangle Extraction in Catadioptric Images*. 7th Workshop on Omnidirectional Vision, Camera Networks, and Non-classical Cameras (OMNIVIS 07), volume 1, number 1, 1957-1967, 2007.
 - [31] S. Durand E. Mouaddib S. Bigot, D. Kachi. *Spherical image denoising and its application to omnidirectional imaging*. pages 8–11, March 2007.
 - [32] Haar A., Der Maassbegriff in *der Theorie der kontinuierlichen Gruppen*, Annals of Mathematics (II), t.34, p 147, 1933.
 - [33] Neumann J. von, *The uniqueness of Haar's measure*, Math. Sbornik, t 1/43, p 721, 1936.
 - [34] Vilenkin N.Ja., *Fonctions spéciales et théorie de la représentation des groupes*, 1969.
 - [35] I. Hara H. Asoh T. Matsui N. Vlassis, Y. Motomura. *Edge-based features from omnidirectional images for robot localization*. volume 2, pages 1579– 1584, Seoul, Korea, 2001.
 - [36] S. Lacroix J-J. Gonzalez-Barbosa. Un algorithme rapide de stéréovision panoramique dense. 2003.
 - [37] D. Kachi-Akkouche C. Demonceaux. *Optical flow estimation in omnidirectional images using wavelet approach*. volume 7, 2003.
 - [38] J. Devars S. Ieng, R. Benosman. *An efficient dynamic multi-angular feature points matcher for catadioptric views*. CVPRW'03 : Conference on Computer Vision and Pattern Recognition Workshop OMNIVIS'03 . *Omnidirectional Vision and Camera Networks*, 7, June 2003.
 - [39] B. Krose R. Bunschoten. *Robust scene reconstruction from an omnidirectional vision system*, 2003.
 - [40] S. Thompson T. Matsui, H. Asoh. *Mobile robot localization using circular correlations of panoramic images*. Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems IROS2000, November 2000.
 - [41] J. Thiran P. Vandergheynst I. Bogdanova, X. Bresson. *Scale-space analysis and active contours for omnidirectional images*. IEEE Transactions on Image Processing, pages 1888–1901, July 2007.
 - [42] P. Vasseur C. Demonceaux. *Markov random fields for catadioptric image processing*. Pattern Pattern Recognition Letters, 27(16) :1957–1967, 2006.

- [43] T. Bulow K. Daniilidis, A. Makadia. *Image processing in catadioptric planes : Spatiotemporal derivatives and optical flow computation*. OMNIVIS'02 : Workshop on Omnidirectional Vision, pages 3–10, June 2002.
- [44] S. Sastry O. Shakernia, R. Vidal. *Omnidirectional egomotion estimation from back-projection flow*. Madison Wisconsin, USA, June 2003.
- [45] K. Daniilidis A. Makadia. *Direct 3d-rotation estimation from spherical images via a generalized shift theorem*. CVPR'03 : Proceedings of the IEEE conference on computer vision and pattern recognition, 2 :217, 2003.
- [46] S. Sastry C. Geyer, R. Bajcsy. *Euclid meets fourier : Applying harmonic analysis to essential matrix estimation in omnidirectional cameras*. OMNIVIS'04 : Proceedings of Omnidirectional Vision, Camera Networks and Non-classical Cameras in conjunction with ECCV'04 : European Conference on Computer Vision, June 2004.
- [47] P. Frossard P. Vanderghenst I. Tosic, I. Bogdanova. *Multiresolution motion estimation for omnidirectional images*. 2005.
- [48] Bulow T, *Spherical Diffusion for 3D Surface Smoothing*, 3D Data Proc. Vis. And Transm., Juin 2002.
- [49] Tikhonov AN. and Arsenin VA., *Solution of Ill-posed Problems*, Winston and Sons, Washington, 1977.
- [50] Sobel I., Feldman G., *A 3*3 isotropic gradient operator for image processing*, Talk at the Stanford Artificial Project, 1968.
- [51] C. Schmid, R. Mohr, and C. Bauckhage. *Comparing and evaluating interest points*. In IEEE International Conference on Image Processing, Bombay, India, January 1998.
- [52] C. Schmid, R. Mohr, and C. Bauckhage. *Evaluation of interest point detectors*. IEEE Transactions Pattern Analysis Machine Intelligence, 37(2) :151–172, 2000.
- [53] R. Deriche. *Recursively implementing the Gaussian and its derivatives*. Technical Report 1893, INRIA, may 1993.
- [54] P. Montesinos, V. Gouet, and R. Deriche. *Differential invariants for color images*. In International Conference on Pattern Recognition, 1998.
- [55] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris*. Doctoral thesis, Institut National Polytechnique de Grenoble, GRAVIR, July 1996.
- [56] T. Lindeberg. *Feature detection with automatic scale selection*. IEEE Transactions Pattern Analysis Machine Intelligence, 30(2) :77–116, 1998.
- [57] K. Mikolajczyk and C. Schmid. *Indexing based on scale invariant interest points*. In IEEE International Conference on Computer Vision, volume 1, pages 525–531, 2001.

- [58] K. Mikolajczyk and C. Schmid. *Scale & affine invariant interest point detectors*. IEEE Transactions Pattern Analysis Machine Intelligence, 60(1) :63–86, 2004.
- [59] F. Ackermann. *Digital image correlation : performance and potential application in photogrammetry*. Photogrammetric Record, 64(11): 429-439, 1984.
- [60] F. Mokhtarian et A. Mackworth. *Scale-based description and recognition of planar curves and two-dimensional shapes*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(1): 34-43, 1986.
- [61] G. Medioni et Y. Yasumoto. *Corner detection and curve representation using cubic B-splines*. Computer Vision, Graphics and Image Processing, 39(1): 267-278, 1987.
- [62] R. Horaud, T. Skordas et F. Veillon. *Finding geometric and relational structures in an image*. Dans Proceedings of the 1st European Conference on Computer Vision, pages 374-384, 1990.
- [63] L. Kitchen et A. Rosenfeld. *Gray-level corner detection*. Pattern Recognition Letters, 1: 95-102, 1982.
- [64] W. Förstner et Gülch. *A fast operator for detection and precise location of distinct points, corners and circular features*. Dans Intercommission Conference on Fast Processing of Photogrammetric Data, pages 281-305, 1987.
- [65] W. Förstner. *A framework for low level feature extraction*. Dans Proceedings of the 3rd European Conference on Computer Vision, 1994.
- [66] D. Reisfeld, H. J. Wolfson et Y. Yeshurun. *Context-free attentional operators : the generalized symmetry transform*. International Journal of Computer Vision, 14: 119-130, 1995.
- [67] K. Rohr. *Über die Modellierung und Identifikation charakteristischer Grauwertverläufe in Realwertbildern*. Dans 12. DAGM-Symposium Mustererkennung, 1990.
- [68] K. Rohr. *Recognizing corners by fitting parametric*. International Journal of Computer Vision, 9(3): 213-230, 1992.
- [69] R. Deriche et T. Blaszk. *Recovering and characterizing image features using an efficient model based approach*. Dans Proceedings of the Conference on Computer Vision and Pattern Recognition, pages 530-535, 1993.
- [70] P. Brand et R. Mohr. *Accuracy in image measure*. Dans Proceedings of the Spie Conference on Videometrics III, volume 2350, pages 218-228, 1994.
- [71] H. Wang et J. M. Brady. *Corner detection with subpixel accuracy*. Rapport technique OUEL 1925/92, Dept. Engineering Science, Oxford University, 1992.

- [72] R. Deriche et G. Giraudon. *A computational approach for corner and vertex detection*. International Journal of Computer Vision, 10(2): 101-124, 1993.
- [73] J. Canny. *A computational approach to edge detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6): 679-698, 1986.
- [74] D. Gabor. *Theory of communication*. Proceedings of Inst. Elec. Eng., 93(26): 429-441, 1946.
- [75] C. J. Westelius, H. Knutsson et J. Wiklund. *Robust vergence control using scale-space phase information*. Rapport technique LiTH-ISY-I-1363, Linköpings tekniska högskola, Department of Electrical Engineering, 1992.
- [76] D. J. Fleet, A. D. Jepson et M. R. M. Jenkin. *Phase-base disparity measurement*. Computer Vision Graphics and Image Processing, 53(2): 198-210, 1991.
- [77] T. Sanger. *Stereo disparity computation using gabor filters*. Biological Cybernetics, 2(59): 405-418, 1988.
- [78] X. Wu et B. Bhanu. *Gabor wavelets for 3D object recognition*. Dans Proceedings of the 5th International Conference on Computer Vision, pages 537-542, 1995.
- [79] Y. Meyer. *Ondelettes et fonctions splines*. Dans Sem. Equations aux Dérivées Partielles, Ecole Polytechnique, Paris, 1986.
- [80] S. G. Mallat. *A theory for multiresolution signal decomposition: the wavelet representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7): 674-693, 1989.
- [81] M. K. Hu. *Visual pattern recognition by moment invariants*. IEEE Transactions on Information Theory, 8: 179-187, 1962.
- [82] M. R. Teague. *Image analysis via the general theory of moments*. Journal of the Optical Society of America, 70: 920-930, 1980.
- [83] C.-H. Teh et R. T. Chin. *On image analysis by the methods of moments*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(4): 496-513, 1988.
- [84] W.-Y. Kim et P. Yuan. *A practical pattern recognition system for translation, scale and rotation invariance*. Dans Proceedings of the Conference on Computer Vision and Pattern Recognition, pages 391-396, 1994.
- [85] L. Van Gool, T. Moons et D. Ungureanu. *Affine / photometric invariants for planar intensity patterns*. Dans Proceedings of the 4th European Conference on Computer Vision, pages 642-651, 1996.
- [86] J. Bigün. *Pattern recognition in images by symmetries and coordinate transformations*. Computer Vision and Image Understanding, 1995. Soumis.

-
- [87] J. Bigün et M.H. du Buf. N-folded symmetries by complex moments in gabor space and their application to unsupervised texture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1): 80-87, 1994.
 - [88] I. Weiss. *Noise resistant projective and affine invariants*. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 115-121, 1992.
 - [89] M. J. Swain and D. H. Ballard. *Color indexing*. *IEEE Transactions Pattern Analysis Machine Intelligence*, 7(1) :11–32, 1991.
 - [90] B. Schiele and J. L. Crowley. *Object recognition using multidimensional receptive field histograms*. In *European Conference on Computer Vision*, pages 610–619, 1996.
 - [91] D. Lowe. *Object recognition from local scale-invariant features*. In *IEEE International Conference on Computer Vision*, pages 1150–1157, 1999.
 - [92] D. Lowe. *Distinctive image features from scale-invariant keypoints*. *IEEE Transactions Pattern Analysis Machine Intelligence*, 20 :91–110, 2004.
 - [93] K. Mikolajczyk and C. Schmid. *A performance evaluation of local descriptors*. *IEEE Transactions Pattern Analysis Machine Intelligence*, 27(10) :1615–1630, 2005.
 - [94] J. L. Mundy et A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, 1992.
 - [95] P. Gros et L. Quan. *Projective invariants for vision*. Rapport technique RT 90 IMAG - 15 LIFIA, Grenoble, 1992.
 - [96] H. Schulz-Mirbach. *Anwendung von Invarianzprinzipien zur Merkmalgewinnung in der Mustererkennung*. Thèse de doctorat, Technische Universität Hamburg, 1995.
 - [97] Y. Moses et S. Ullman. *Limitations of non model-based recognition*. Dans *Proceedings of the 2nd European Conference on Computer Vision*, pages 820-828, 1992.
 - [98] D. J. Clemens et D. W. Jacobs. *Model-group indexing for recognition*. Dans *Proceedings of Darpa Image Understanding Workshop*, pages 604-613, 1990.
 - [99] A. Zisserman, D. A. Forsyth, J. L. Mundy, C. Rothwell, J. Liu et N. Pillow. *3D object recognition using invariance*. *Artificial Intelligence*, 78(1-2): 239-288, 1995.
 - [100] T.O. Binford et T.S. Levitt. *Quasi-invariants : theory and exploitation*. Dans *Proceedings of Darpa Image Understanding Workshop*, pages 819-829, 1993.
 - [101] A. E. Grace et M. Spann. *A comparison between Fourier-Mellin descriptors and moment based features for invariant object recognition using neural networks*. *Pattern Recognition Letters*, 12: 635-643, 1991.
 - [102] J. Rubinstein, J. Segman et Y. Zeevi. *Recognition of distorted patterns by invariance kernels*. *Pattern Recognition*, 24(10): 959-967, 1991.

- [103] M. K. Hu. *Visual pattern recognition by moment invariants*. IEEE Transactions on Information Theory, 8: 179-187, 1962.
- [104] J. J. Koenderink et A. J. van Doorn. *Representation of local geometry in the visual system*. Biological Cybernetics, 55: 367-375, 1987.
- [105] A. H. Salden, B. M. ter Haar Romeny, L. M. J. Florack, M. A. Viergever et J. J. Koenderink. *A complete and irreducible set of local orthogonally invariant features of 2-dimensional images*. Dans Proceedings of the 11th International Conference on Pattern Recognition, pages 180-184, 1992.
- [106] L. M. J. Florack. *The syntactical structure of scalare images*. Thèse de doctorat, Universiteit Utrecht, 1993.
- [107] B. M. ter Haar Romeny. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, 1994.
- [108] D. Hilbert. *Ueber die vollen Invarianten systemen*. Math. Annalen, 42: 313-373, 1893.
- [109] W. T. Freeman et E. H. Adelson. *The design and use of steerable filters*. IEEE Transactions on Pattern and Machine Intelligence, 13(9): 891-906, 1991.
- [110] D. J. Fleet, A. D. Jepson et M. R. M. Jenkin. *Phase-base disparity measurement*. Computer Vision Graphics and Image Processing, 53(2): 198-210, 1991.
- [111] P. J. Burt. *Fast filter transforms for image processing*. Computer Graphics and Image Processing, 16: 20-51, 1981.
- [112] J. L. Crowley. *A representation for visual information*. Thèse de doctorat, Carnegie-Mellon University, 1981.
- [113] J. L. Crowley et A. C. Parker. *A representation for shape based on peaks and ridges in the difference of low pass transform*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(2): 156-170, 1984.
- [114] A. P. Witkin. *Scale-space filtering*. Dans Proceedings of the 8th International Joint Conference on Artificial Intelligence, pages 1019-1023, 1983.
- [115] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.
- [116] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink et M. A. Viergever. *General intensity transformations and differential invariants*. Journal of Mathematical Imaging and Vision, 4: 171-187, 1994.
- [117] M. Berthod et G. Giraudon. *About grey-level invariants*. Version préliminaire non publiée, 1994.

- [118] J. B. Burns, R. Weiss et E. M. Riseman. *View variation of point set and line segment features*. Dans Proceedings of Darpa Image Understanding Workshop, pages 650-659, 1990.
- [119] B. M. ter Haar Romeny, L. M. J. Florack, A. H. Salden et M. A. Viergever. *Higher order differential structure of images*. Image and Vision Computing, 12(6): 317-325, 1994

Résumé

Pour obtenir des images omnidirectionnelles, le système le plus répandu est le capteur catadioptrique. Un capteur catadioptrique composé d'un miroir de révolution dans lequel se reflète une scène qui est filmée par une caméra. Le problème principal avec les images omnidirectionnelles obtenues par le capteur catadioptrique est leur différence de résolution qui est élevée à la périphérie et est faible au centre de l'image. Appliquer directement les traitements classiques sur les images catadioptriques aboutit à des erreurs, puisque la géométrie du miroir n'est pas prise en compte. Pour palier à ces problèmes, plusieurs méthodes ont été proposées. La projection de l'image sur une sphère virtuelle est l'une des méthodes proposées. Son principe est de projeter l'image sur la sphère d'équivalence pour appliquer des outils adaptés.

Nous présentons dans ce travail la détection de points d'intérêts sur les images omnidirectionnelles par le détecteur de Harris Laplace. La méthode de Harris Laplace extrait des caractéristiques invariantes à la rotation et à l'échelle. Dans un premier temps, l'algorithme détecte pour chaque échelle, les points d'intérêt 2D au sens de Harris en utilisant la matrice de structure et en exhibant les maxima locaux. Parmi les points d'intérêt détectés, ceux qui correspondent à un extremum local en échelle au sens de Lindeberg sont retenus comme étant les points d'intérêt de Harris-Laplace.

Mots-clés

Capteur catadioptrique, vision omnidirectionnelle, points d'intérêts, Harris Laplace.