



HAL
open science

Descripteurs de Fourier inspirés de la structure du cortex visuel primaire humain : Application à la reconnaissance de navires dans le cadre de la surveillance maritime

Amine Bohi

► **To cite this version:**

Amine Bohi. Descripteurs de Fourier inspirés de la structure du cortex visuel primaire humain : Application à la reconnaissance de navires dans le cadre de la surveillance maritime. Traitement des images [eess.IV]. Université de Toulon, 2017. Français. NNT : 2017TOUL0002 . tel-01611384

HAL Id: tel-01611384

<https://theses.hal.science/tel-01611384>

Submitted on 5 Oct 2017

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

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

ÉCOLE DOCTORALE MER ET SCIENCES

Laboratoire des Sciences de l'Information et des Systèmes

THÈSE

présentée par :

Amine BOHI

soutenue le : **22 mai 2017**

pour obtenir le grade de Docteur en Informatique

Spécialité : Traitement et analyse d'images

Descripteurs de Fourier inspirés de la structure du cortex visuel primaire humain

**Application à la reconnaissance de navires dans le
cadre de la surveillance maritime**

THÈSE dirigée par :

**Frédéric BOUCHARA
Jean-Paul GAUTHIER**

Maître de conférences, HDR, Université de Toulon
Professeur, Université de Toulon

JURY :

Johel MITERAN	Professeur Faculté Mirande, Dijon	(Rapporteur)
William PUECH	Professeur Université de Montpellier	(Rapporteur)
Yuliya TARABALKA	Chargée de recherche Inria Sophia Antipolis	(Examineur)
Nicole VINCENT	Professeur Université Paris Descartes	(Examineur)
Vincente GUIZ	Ingénieur de recherche Université de Toulon	(Co-Encadrant)
Dario PRANDI	Chargé de recherche CNRS Centrale Supélec	(Co-Encadrant)

Thesis title: Fourier descriptors inspired by the structure of the human primary visual cortex

Application to vessels recognition in the framework of maritime surveillance.

Abstract

In this thesis, we develop a supervised object recognition method using new global image descriptors inspired by the model of the human primary visual cortex V1. Mathematically speaking, the latter is modeled as the semi-discrete roto-translation group $SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{Z}_N$ (semi-direct product between \mathbb{R}^2 and \mathbb{Z}_N). Therefore, our technique is based on generalized and rotational Fourier descriptors defined in $SE(2, N)$, and which are invariant to natural geometric transformations (translations, and rotations). Furthermore, we show that such Fourier descriptors are weakly complete, in the sense that they allow to distinguish over an open and dense set of compactly supported functions in $L^2(SE(2, N))$, hence between real-world images. These descriptors are later used in order to feed a Support Vector Machine (SVM) classifier for object recognition purposes. We have conducted a series of experiments aiming both at evaluating and comparing the performances of our method against existing both local - and global - descriptor based state of the art techniques, using the RL, the CVL, and the ORL face databases, and the COIL-100 image database (containing various types of objects). The obtained results have demonstrated that our approach was able to compete with many existing state of the art object recognition techniques, and to outperform many others. These results have also shown that our method is robust to noise. Finally, we have applied the proposed method on vessels recognition in the framework of maritime surveillance.

Keywords: Object recognition - Primary visual cortex V1 - Geometric transformation - Local/global image descriptor - Generalized Fourier descriptor - Support Vector Machine - Vessels recognition - Maritime surveillance.

Résumé

Dans cette thèse, nous développons une approche supervisée de reconnaissance d'objets basée sur l'utilisation de nouveaux descripteurs d'images globaux inspirés du modèle du cortex visuel humain primaire V1 en tant que groupe de roto-translations semi-discrètes $SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{Z}_N$ (produit semi-direct entre \mathbb{R}^2 et \mathbb{Z}_N). La méthode proposée est basée sur des descripteurs de Fourier généralisés et rotationnels définis sur le groupe $SE(2, N)$, qui sont invariants aux transformations géométriques (translations, et rotations). De plus, nous montrons que ces descripteur de Fourier sont faiblement complets, dans le sens qu'ils permettent de discriminer sur un ensemble ouvert et dense $L^2(SE(2, N))$ de fonctions à support compact, donc distinguer entre des images réelles. Ces descripteurs sont ensuite utilisés pour alimenter un classifieur de type SVM dans le cadre de la reconnaissance d'objets. Nous avons mené une séries d'expérimentations dans le but d'évaluer notre méthode sur les bases de visages RL, CVL et ORL et sur la base d'images d'objets variés COIL-100, et de comparer ses performances à celles des méthodes basées sur des descripteurs globaux et locaux. Les résultats obtenus ont montré que notre approche est en mesure de concurrencer de nombreuses techniques de reconnaissance d'objets existantes et de surpasser de nombreuse autres. Ces résultats ont également montré que notre méthode est robuste aux bruits. Enfin, nous avons employé la technique proposée pour reconnaître des navires dans un contexte de surveillance maritime.

Mots-clefs : Reconnaissance d'objets – Cortex visuel primaire V1 – Transformations géométriques – Descripteur d'images local/global – Descripteur de Fourier généralisé – Machines à vecteurs de supports – reconnaissance de navires – Surveillance maritime.

Remerciements

Une thèse est un effort collectif et il est utopique de croire qu'un thésard solitaire puisse s'épanouir et réaliser un travail original.

Cet ouvrage est tout d'abord le fruit d'une collaboration entre le Laboratoire des Sciences de l'Information et des Systèmes (LSIS) de l'Université de Toulon et l'entreprise Opéra Ergonomie.

Je tiens donc à exprimer ma profonde gratitude et mon profond respect ainsi que mes vifs remerciements à mes directeurs de thèse : Frédéric BOUCHARA et Jean-Paul GAUTHIER, ainsi qu'à mes tuteurs au sein d'Opéra Ergonomie, pour leur patience, leurs explications et leurs critiques toujours éclairées. J'ai tout particulièrement apprécié l'autonomie qu'ils m'ont accordée dans les choix et les orientations de mon travail. J'adresse également mes remerciements à Vincente GUIZ et Dario PRANDI pour leurs précieux conseils qui m'ont grandement aidé dans mes travaux, ainsi que leur disponibilité et leur gentillesse en toute circonstance. Je veux aussi témoigner de ma gratitude à mes collègues Ikhlef BECHAR, Kheireddine AZIZ et à mon ami Badr BOUKKOURI, pour avoir partagé avec et dans la bonne humeur de nombreux problèmes, soucis administratifs et autres situations inextricables. Je remercie également toute l'équipe Signal/Image du Laboratoire LSIS avec qui j'ai eu le plaisir à travailler.

J'adresse évidemment un grand merci aux doctorants du laboratoire qui ont croisé ma route, qu'ils m'aient précédé ou suivi : Omar CHERRAK, Rémy ANDRE, Diogon SYLLA, Victor MURANDI, Cécile PESCHOUD, Vincent MARIÉ, Emilien ROYER, Vincent MARTIN, ainsi qu'à tous les étudiants ou stagiaires que j'ai côtoyés qui sont tous révélés être des collègues très sympathiques.

J'adresse également des remerciements tout particuliers aux rapporteurs, les professeurs Johel MITERAN et William PUECH qui ont pris le temps de lire et évaluer mes travaux ainsi que pour leurs remarques judicieuses.

Je remercie enfin l'ensemble des membres du jury pour l'intérêt qu'ils ont porté à mes travaux.

Je terminerai en remerciant ma famille qui m'a toujours soutenu, et ce quoi que je fasse.

Sommaire

Abstract	2
Résumé.....	3
Remerciements	4
Liste des figures.....	8
Liste des Tableaux	10
Chapitre 1 : Introduction générale.....	12
1. Contexte	12
2. Contributions.....	12
3. Organisation du manuscrit	13
Chapitre 2 : Descripteurs d'images	15
1. Introduction.....	15
2. Descripteurs locaux	16
2.1. Haar	16
2.2. SIFT	20
2.3. RIFT	27
2.4. DSIFT (SIFT denses).....	29
2.5. SURF.....	30
2.6. GLOH.....	32
2.7. DAISY	34
2.8. MOPS	36
2.9. Shape Context	36
2.10. Histogrammes de gradients orientés	39
2.11. Descripteurs binaires.....	44
3. Descripteurs globaux.....	48
3.1. Filtres de Gabor	48
3.2. GIST.....	56
3.3. Les moments géométriques	58
3.4. Les moments orthogonaux.....	59
3.5. Les descripteurs de Fourier	61
4. Conclusion	68
Chapitre 3 : Descripteurs de Fourier généralisés et rotationnels	70
1. Introduction.....	70

2.	Préliminaires.....	70
2.1.	Analyse harmonique sur les groupes abéliens localement compacts	71
2.2.	La transformée de Fourier sur les groupes non-commutatifs localement compacts	72
2.3.	Contexte de notre travail	74
2.4.	Fonctions faiblement cycliques	77
3.	Descripteurs de Fourier dans les groupes	78
3.1.	Invariants spectraux et bi-spectraux	78
3.2.	Invariants spectraux et bi-spectraux sur les groupes abéliens localement compacts	80
3.3.	Invariants spectraux et bi-spectraux sur $SE(2, N)$	81
4.	Descripteurs de Fourier basés sur la structure du cortex visuel primaire humain V1	83
4.1.	Présentation du pipeline naturel pour la reconnaissance d'objets	83
4.2.	Présentation du modèle mathématique du cortex visuel primaire V1	86
4.3.	La non-applicabilité du théorème de la complétude dans $SE(2, N)$	87
4.4.	Les invariants spectraux et bi-spectraux rotationnels et leur complétudes dans $SE(2, N)$ 88	
5.	Le calcul pratique des descripteurs de Fourier	92
6.	Conclusion	93
Chapitre 4 : Résultats et expérimentations.....		96
1.	Introduction.....	96
2.	Evaluation des descripteurs de Fourier	97
2.1.	Les bases d'images utilisées	97
2.2.	Extraction des descripteurs	102
2.3.	Expérimentations	104
3.	Application à la surveillance maritime	114
3.1.	Navires subissant des rotations 2D	114
3.2.	Navires sous différents angles de vue	115
4.	Conclusion	117
Chapitre 5 : Conclusion et perspectives		119
1.	Conclusion	119
2.	Perspectives.....	120
Annexes		123
A.	Décomposition du produit tensoriel des représentations	123
B.	Lemmes auxiliaires pour la démonstration du théorème 3.3.1	124
C.	Machines à vecteurs de support (SVM)	128

Bibliographie.....	131
Publications dans le cadre de la thèse	139

Liste des figures

Figure 1: Exemples de Descripteurs rectangulaires de Haar [43]	17
Figure 2 : Descripteurs de Haar dans une fenêtre de taille fixe: à 2, 3 et 4 rectangles [84]	18
Figure 3 : Description complète d'un descripteur de Haar [43].....	18
Figure 4 : Exemple de calcul de l'image intégrale	19
Figure 5 : Image Intégrale [85]	19
Figure 6 : Calcul de la somme du rectangle D avec l'image intégrale [84].....	20
Figure 7 : Différence de Gaussiennes [45, 44].....	22
Figure 8 : Recherche d'extrema dans <i>DoG</i> [45, 44].....	23
Figure 9 : Illustration de la construction de l'histogramme des orientations	26
Figure 10 : Construction d'un descripteur SIFT	27
Figure 11 : Construction du descripteur RIFT [42]	28
Figure 12 : Extraction de l'angle avec invariance à la rotation.....	29
Figure 13 : Opérateurs de dérivation d'ordre 2 discrétisés et dérivées d'ordre 2 de la gaussienne [3]31	
Figure 14 : Blobs détectés [3]	31
Figure 15 : La structure du descripteur SURF [4]	32
Figure 16 : Structure de la fenêtre de calcul du SIFT, Structure de la fenêtre du calcul du GLOH [36] 33	
Figure 17 : La structure du masque d'analyse du descripteur GLOH [47].....	33
Figure 18 : La structure du masque d'analyse du descripteur DAISY [80]	35
Figure 19 : Structure des descripteurs MOPS [14]	36
Figure 20: Diagramme d'histogramme log-polaire	37
Figure 21 : Comparaison des contextes de forme de deux versions différentes de la lettre "A" [5]....	38
Figure 22 : Chaîne d'extraction des descripteurs HoG proposée par Dalal et Triggs dans [17]	40
Figure 23 : (a) image d'un piéton en niveau de gris normalisée, (b) composante horizontale du gradient, (c) composante verticale du gradient, (d) la norme de gradient [18]	42
Figure 24 : Construction du HoG d'une image	42
Figure 25 : Les étapes de calcul du LBP	44
Figure 26 : Détection de contours par le filtre de Sobel [89, 90]	45
Figure 27 : Exemple d'une image transformée par la Census Transform [89, 90]	46
Figure 28 : Histogrammes des sous-régions de la Census Transform [89, 90].....	47
Figure 29 : Illustration des contraintes entre les valeurs de la CT des pixels voisins [89, 90].....	47
Figure 30 : Filtres de Gabor à différentes échelles et fréquences spatiales	49
Figure 31 : Filtre de Gabor (a) en partie réelle (b) les modules de 5 fréquences	50

Figure 32 : Transformation de Gabor (a) en module (b) en phase.....	50
Figure 33 : QBC de phase de Gabor [91]	52
Figure 34 : Patterns GGPP (a) partie réelle (b) partie imaginaire [91]	53
Figure 35 : Illustration de l'opérateur LXP [91]	54
Figure 36 : Patterns LGPP (a) partie réelle (b) partie imaginaire [91].....	54
Figure 37 : (a) 8×8 sous-régions de LGPP (b) diagramme général de HGPP [91].....	56
Figure 38 : Principe du descripteur GIST [58].....	57
Figure 39 : Les polynômes de Zernike au 5 premiers ordres	60
Figure 40 : Reconstruction des formes de quatre images de papillons en fonction du nombre de descripteurs invariants utilisés pour la reconstruction (TFMA) [33].....	62
Figure 41 : (a) l'image originale dans l'espace polaire; (b) l'image polaire de (a) dans tracée dans l'espace cartésien [92].....	64
Figure 42 : (a) un motif et son spectre de Fourier; (b) le motif pivoté de (a) et son spectre de Fourier; (c) l'image polaire de (a) et son spectre de Fourier; (d) l'image polaire de (b) et son spectre de Fourier [92]	65
Figure 43: Rotation de la tranche de camembert Sc par $\{Rk\}k \in \mathbb{Z}N$	76
Figure 44 : Log du Power Spectrum d'une image synthétique et sa version pivotée	79
Figure 45 : Pipeline naturel pour la reconnaissance d'objets	85
Figure 46 : Base des images synthétiques.....	97
Figure 47: Extrait de la base RL	98
Figure 48: Extrait de la base CVL	99
Figure 49: Trois expressions faciales extraites de la base CVL.....	99
Figure 50 : Base de visages ORL	99
Figure 51 : Exemples de visages enregistrés sous des vues différentes	100
Figure 52: Base de visages Sheffield.....	100
Figure 53: Base de visages Extended Yale B.....	101
Figure 54 : Les 100 objets de COIL-100	101
Figure 55 : Exemples d'objets vus sous des angles différents.....	102
Figure 56 : Etapes de calcul des descripteurs de Fourier	103
Figure 57: Test sur la base d'images synthétiques.....	106
Figure 58 : Exemple d'images bruitées pour un objet de la base COIL-100.....	112
Figure 59: Taux de reconnaissance pour différentes tailles de la base d'apprentissage	113
Figure 60: Exemples de bateaux qui ont subit des rotations 2D.....	114
Figure 61: Exemples de bateaux qui ont subit des transformations générales	116
Figure 62: Principe de l'architecture hexagonale en spiral	121

Liste des Tableaux

Tableau 1 : Dimension des vecteurs descripteurs de Fourier considérés	104
Tableau 2: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages RL.....	107
Tableau 3: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages CVL	108
Tableau 4: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages ORL....	109
Tableau 5: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages Sheffield	110
Tableau 6: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages Extended Yale B.....	111
Tableau 7: Le taux de reconnaissance pour chaque descripteur en utilisant la base de données COIL- 100 (cas non-bruité)	113
Tableau 8: taux de classification pour le cas bruité de la base COIL-100.....	114
Tableau 9: Résultats obtenus pour la base de navires 2D	115
Tableau 10: Résultats obtenus sur la base de navire "Marine Traffic"	117

Chapitre 1 : Introduction générale

1. Contexte

La tâche de description d'images est primordiale dans la plupart des applications en vision par ordinateur. Il s'agit entre autres de : la reconnaissance d'objets [44, 6, 74], la recherche d'images [46, 67], la reconnaissance de texture [42], la mise en correspondance d'images [82, 65].

Cette thèse est une contribution dans le domaine de la reconnaissance d'objets dans les images couleurs. Ici nous adoptons l'approche classique qui repose sur l'utilisation d'une méthode de classification supervisée prenant en entrée des vecteurs descripteurs d'images issues des bases de données considérées.

La finalité de la tâche de description d'images est de fournir une représentation d'image compacte, distinctive et invariante aux transformations géométriques d'images, changement de luminosité, occlusions et bruits. A cette fin, différentes méthodes ont été proposées, qui peuvent être classées en deux grandes catégories : locales où l'image est représentée par un ensemble de descripteurs locaux [45, 11, 42, 47, 79, 14, 17, 84, 4, 5, 53, 90], et globales qui utilisent un unique vecteur descripteur [55, 91, 39, 76, 68, 21, 92, 74].

Un descripteur d'images efficace doit discriminer différents types d'images qui ont subi différentes transformations géométriques (rotations, translations, changement d'échelle) et qui sont soumises à des altérations complexes (ajout de bruit, changement de luminosité, occlusions...).

2. Contributions

Les travaux réalisés au cours de cette thèse s'articulent autour d'un axe principal de recherche ayant pour but de proposer une nouvelle solution de reconnaissance automatique d'objets basée sur des descripteurs de Fourier.

Une première contribution sera le développement d'un ensemble de descripteurs de Fourier nommés par la suite « Rotational Fourier Descriptors » (RBS et RPS pour Rotational Bi-Spectrum invariant et Rotational Power-Spectrum invariant, respectivement), basés sur la

structure du cortex visuel primaire humain V1 qui est modélisé mathématiquement par un groupe de roto-translations semi-discrètes $SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{Z}_N$. Dans ce modèle, les stimuli corticaux sont des fonctions dans l'espace $L^2(SE(2, N))$ des fonctions de carré intégrable, par rapport à la mesure de Haar de $SE(2, N)$, et les images provenant du plan visuel sont relevées à des stimuli corticaux par l'intermédiaire d'une opération de relèvement injectif et invariant à gauche $\mathcal{L}: L^2(\mathbb{R}^2) \rightarrow L^2(SE(2, N))$.

Enfin, la deuxième contribution dans cette thèse sera l'association de ces descripteurs de Fourier à un classifieur pour résoudre le problème de reconnaissances de navires, dans le cadre de la surveillance maritime.

3. Organisation du manuscrit

Cette thèse comprend trois chapitres, les deux premiers introduisent des notions théoriques, le dernier les valide par des tests et des expérimentations.

Nous commençons dans le premier chapitre par un état de l'art des descripteurs d'images locaux et globaux utilisés pour la reconnaissance d'objet, domaine dans lequel la notion de l'invariance aux transformations géométriques, la robustesse aux bruits, aux changements de luminosité, et aux occlusions, est centrale.

Dans le deuxième chapitre, nous introduisons un nouvel ensemble de descripteurs de Fourier dits « rotationnels » basés sur la structure du cortex visuel primaire humain V1. Nous commençons d'abord par présenter quelques notions préliminaires sur l'analyse harmonique (i.e. l'étude de la transformée de Fourier et ses propriétés) sur les groupes abéliens et non-abéliens localement compacts. Ensuite, nous présentons les invariants spectraux et bi-spectraux, ainsi que leurs propriété de complétude (i.e. capacité de discrimination) dans les groupes abéliens localement compacts et dans le groupe dérivé de produit semi-direct $SE(2, N)$ défini dans la section 2.3 par $SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{Z}_N$. Dans la section d'après, nous introduisons les descripteurs de Fourier rotationnels basés sur la structure du cortex visuel primaire humain V1. Enfin, pour clore le chapitre, nous présentons quelques techniques pour le calcul pratique de ces descripteurs.

Le troisième, et dernier chapitre, a pour objectif d'évaluer les performances de notre ensemble de descripteurs de Fourier, dans un contexte de reconnaissance d'objets. A cette fin, nous proposons d'utiliser une structure classique basée sur un descripteur et un classifieur, et nous menons une série de tests et d'expérimentations sur des bases d'images différentes afin de comparer notre ensemble de descripteurs de Fourier aux principaux de la littérature, tel que les descripteurs locaux SIFT et les histogrammes de gradients orientés, et les descripteurs globaux représentés par les moments de Zernike et de Hu, et les invariants de Fourier-Mellin. Enfin, nous évaluons ces descripteurs de Fourier pour une problématique de reconnaissance de navires dans un contexte de surveillance maritime.

Chapitre 2 : Descripteurs d'images

1. Introduction

La description de l'image est souvent l'étape principale, pour plusieurs applications de traitement d'images. Celles-ci incluent, par exemple, le suivi par vidéo [81, 28, 52], la localisation et cartographie pour robots mobiles [77, 78], et la reconnaissance d'objet [44, 6].

La raison principale derrière cette étape de description d'image est de fournir une représentation de l'image, compacte, distinctive et invariante (robuste) aux transformations géométriques et aux altérations simples et complexes.

Dans ce but, de nombreux algorithmes d'extraction de descripteurs ont été proposés dans la littérature de traitement d'images, et qui peuvent être classés en deux grandes catégories, approches locales et globales :

- Locale : basée sur des méthodes locales, l'image est représentée par un ensemble de descripteurs locaux qui encodent les propriétés (e.g, distribution, variation, ...) des informations collectées au voisinage du point caractéristique considéré. Cette catégorie comprend plusieurs types de descripteurs, on peut citer:
 - Ceux basés sur le calcul des orientations du gradient comme SIFT [45], DSIFT [11], RIFT [42], GLOH [47], DAISY [79], MOPS [14] et HOG [17].
 - Ceux basés sur l'utilisation de l'image intégrale comme les caractéristiques de Haar [84] et les descripteurs SURF [4] (basés sur le même principe que SIFT).
 - Autres types de descripteurs locaux comme le Shape Context [5] utilisé pour la description de la forme et les descripteurs binaires LBP [53] et Centrist [90] (basé sur la Census Transform).

- Globale : pour les méthodes globales, une image est représentée par un unique descripteur (vecteur), qui encode les propriétés (e.g, distribution, variation, ...) des informations disponibles sur l'ensemble de l'image. Cette catégorie comprend :

- Les descripteurs basés sur les filtres de Gabor comme GIST [55] et HGPP [91] (Histogrammes de Gabor Phase Patterns) qui encodent les informations de phase de Gabor.
- Les descripteurs basés sur les moments géométriques (Hu) [39] et orthogonaux (Zernike) [76].
- Les descripteurs exploitant les propriétés de la transformée de Fourier comme les descripteurs de Fourier-Mellin (FMT) [68, 21], Fourier génériques [92] et Fourier généralisés [74].

Une méthode de description d'images efficace doit fonctionner indifféremment tout en maintenant de bonnes performances (i.e., caractère distinctif et robustesse) sur différents types d'images et sous des altérations difficiles. Ceux-ci incluent, par exemple, les problèmes qui se posent avec des images extraites de scènes homogènes et texturées, ou avec celles soumises à des transformations géométriques compliquées.

Pour contourner ces problèmes, plusieurs techniques ont été suggérées. Nous les présenterons en détails dans les deux sections suivantes.

2. Descripteurs locaux

2.1. Haar

2.1.1. Extraction des caractéristiques de Haar

Les valeurs d'un pixel nous informent sur la luminance et la couleur d'un point donné. Cependant, pour être à la fois plus efficace et plus rapide, il est plus judicieux d'utiliser des caractéristiques plus globales de l'objet, c'est-à-dire des représentations synthétique et informative, calculée à partir des valeurs des pixels. C'est le cas des descripteurs de Haar ou pseudo-Haar définis dans [84]. Les descripteurs de Haar sont des fonctions permettant de connaître la différence de contraste entre deux ou plusieurs régions rectangulaires adjacentes. La Figure 1 donne des exemples des caractéristiques rectangulaires de Haar les plus utilisées, dans lesquelles la somme de pixels sombres est soustraite de la somme des pixels blancs. La valeur du descripteur de Haar est obtenue grâce à la relation suivante :

$$f_i = \sum_i I_{i,blanche} - \sum_i I_{i,noire}$$

Où I représente l'intensité dans l'image.

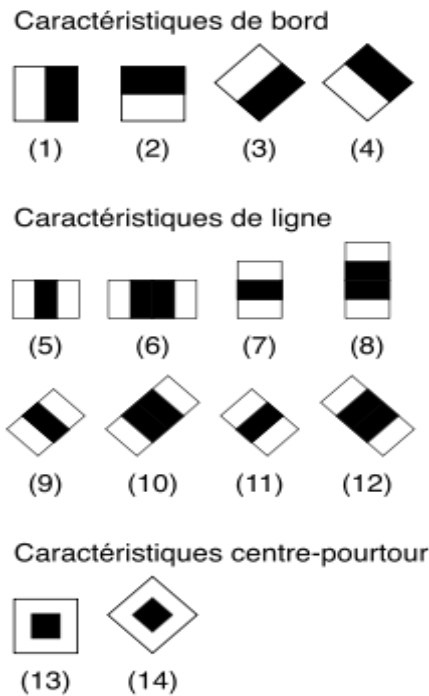


Figure 1: Exemples de Descripteurs rectangulaires de Haar [43]

Ces descripteurs sont calculés dans une fenêtre glissante de taille fixe (par exemple 24x24 pixels). Généralement, on trouve 3 sortes de descripteurs de Haar: à 2 rectangles, à 3 rectangles et à 4 rectangles (Figure 2). Les descripteurs à 2 rectangles sont utilisés horizontalement et verticalement (Figure 2: A et B). Les zones blanches ont des poids positifs et celles noires ont des poids négatifs.

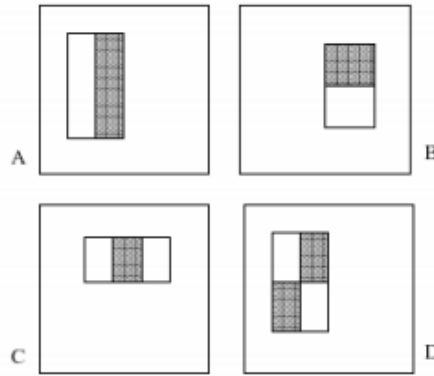


Figure 2 : Descripteurs de Haar dans une fenêtre de taille fixe: à 2, 3 et 4 rectangles [84]

Un descripteur de Haar est caractérisé par les paramètres suivants: le nombre de rectangles utilisés par filtre, la position de chaque filtre dans la fenêtre (c'est le coin supérieur gauche), la largeur w et la hauteur h de chaque filtre avec $0 < x, x + w < W; 0 < y, y + h < H$, et enfin les poids positifs ou négatifs attribué à chaque filtre. Une description complète d'un exemple de descripteur de Haar est donnée dans la Figure 3.

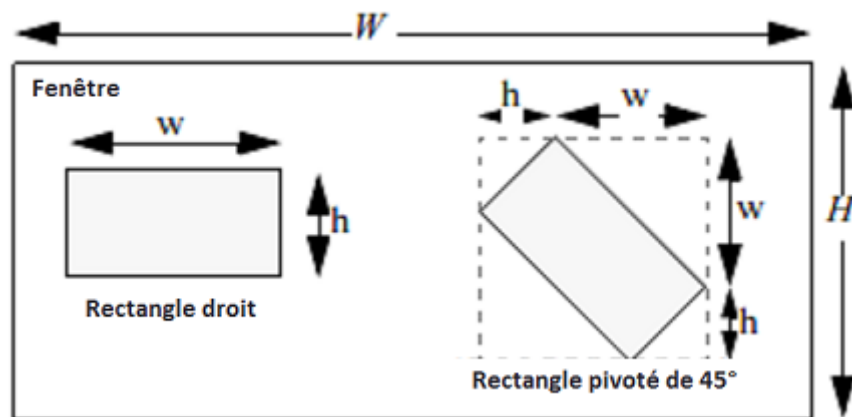


Figure 3 : Description complète d'un descripteur de Haar, H et W représentent la taille de la fenêtre [43]

Les descripteurs de Haar sont très simples à calculer mais très nombreux du fait du balayage de toute l'image avec des fenêtres de taille et de position variable. Selon [84], pour une fenêtre donnée de résolution 24x24 pixels, on peut définir environ 160 000 détecteurs possibles dans cette fenêtre.

L'opération de balayage de toute l'image par une sous-fenêtre de taille et de position variable est trop coûteuse en temps. Pour contourner ce problème, Viola et Jones ont introduit l'idée d'image intégrale afin d'optimiser le calcul.

2.1.2. ***Image intégrale***

L'image intégrale est une nouvelle représentation sous la forme d'une image, de même taille que l'image d'origine, qui va permettre de calculer plus rapidement les attributs du descripteur. L'idée est de calculer seulement une fois la somme de tous les pixels de l'image [84]. Le pixel à la position (x, y) de l'image intégrale contient la somme de tous les pixels situés au-dessus de lui, à sa gauche et lui-même (voir la Figure 4 et la Figure 5).



Figure 4 : Exemple de calcul de l'image intégrale

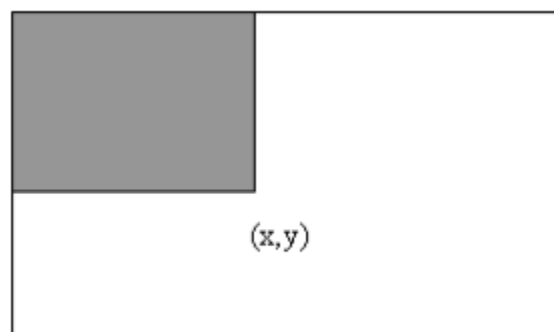


Figure 5 : Image Intégrale (la valeur de l'image intégrale au point (x, y) est égale à la somme de tous les pixels situés au-dessus et à sa gauche [85])

L'image intégrale $II(x, y)$ au point (x, y) est définie à partir de l'image d'origine $I(x, y)$ par :

$$I(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y')$$

La somme des valeurs des pixels dans le rectangle D (Figure 6) peut être calculée avec seulement quatre références. La valeur de l'image intégrale à l'endroit 1 est la somme des valeurs des pixels dans le rectangle A. la valeur à l'endroit 2 est $A + B$, à l'endroit 3 est $A + C$, et à l'endroit 4 est $A + B + C + D$. La somme dans D peut donc être calculée par : $4 + 1 - (2 + 3)$.

Grâce à cette représentation, une caractéristique à 2 rectangles peut donc être calculée en seulement 6 accès à l'image intégrale, et donc en un temps constant quelle que soit la taille de la caractéristique.

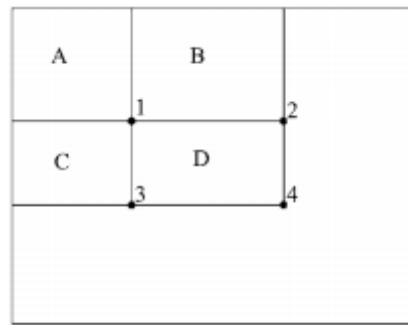


Figure 6 : Calcul de la somme du rectangle D avec l'image intégrale [84]

D'autres formes et orientations de descripteurs de Haar ont été proposées. Parmi ces travaux les exploitant, on peut citer ceux de [43] pivoté à 45 degrés.

2.2. SIFT

2.2.1. Description de la méthode SIFT

SIFT (Scale-invariant feature transform) est une approche proposée et développée par David Lowe [45, 44] pour détecter des points d'intérêt et d'extraire des caractéristiques distinctives afin de les identifier entre différentes images. La force des caractéristiques de SIFT est qu'elles sont invariantes à la rotation, à la translation, à l'échelle et partiellement invariantes aux changements d'illumination et de projection affine et 3D.

Lowe propose une approche constituée de deux parties :

- Une partie pour la détection et l'extraction de caractéristiques sur les points d'intérêt.
- Une partie pour la mise en correspondance

Cependant, c'est la première partie qui a assuré la popularité de la méthode de Lowe et elle se déroule en quatre étapes principales :

- Détection d'extrema d'espace-échelle (scale-space),
- Localisation des points d'intérêt,
- Choix de l'orientation des descripteurs,
- Calcul des Descripteurs.

Nous détaillerons pas à pas les différentes étapes de cette approche.

2.2.2. Détection d'extrema d'espace-échelle

La détection d'extrema se fait dans un espace discret appelé « espace des échelles » (scale space en anglais) qui comporte trois dimensions (les coordonnées cartésiennes x et y et le facteur d'échelle σ). Soit une image I et un noyau gaussien G de paramètre d'échelle σ , on appelle le gradient de facteur d'échelle σ (noté L) le résultat de la convolution de I par G et il est donné par:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

Avec

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

La convolution de l'image originale par le filtre gaussien permet de lisser l'image I de telle sorte que les détails trop petits sont estompés. Par conséquent, la détection des extrema dans l'espace des échelles s'effectue en utilisant la fonction Différence de Gaussiennes (DOG) définie comme la différence de deux images filtrées par un noyau gaussien, séparées par un facteur k . La fonction DOG est définie comme suit :

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

$$= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$

Où k est un paramètre multiplicatif fixe (généralement fixé à $\sqrt{2}$) qui dépend de la finesse de la discrétisation de l'espace des échelles voulue.

Afin d'optimiser le calcul des images floutées à un grand nombre d'échelles, Lowe utilise une pyramide dont la base est l'image originale et chaque niveau (octave) est obtenu à partir du niveau précédent en divisant la résolution de l'image par 2, c.à.d. doubler le facteur d'échelle. Le nombre d'intervalles par octave est défini par k (Figure 7).

Pour chaque octave de l'espace des échelles, l'image initiale est successivement convoluée par une gaussienne pour produire les différentes images à échelle. La différence entre images à échelle successive est calculée pour produire la *DoG* correspondante comme montré à droite. Après chaque octave, l'image gaussienne est sous échantillonnée d'un facteur 2, et le processus recommence.

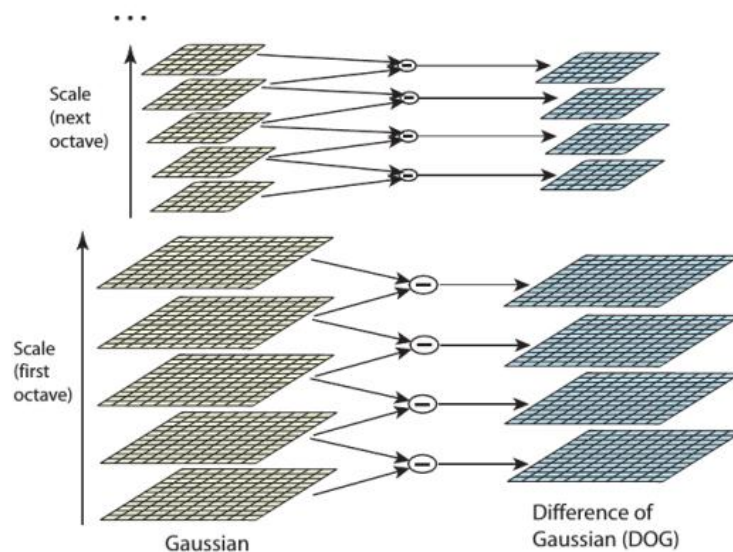


Figure 7 : Différence de Gaussiennes [45, 44]

Un point-clé candidat retenu (x, y, σ) correspond à un extremum local du *DoG*. Chaque point est comparé à ses 8 voisins immédiats et aux 9 voisins des niveaux (ou échelles) supérieurs et inférieurs comme le montre la Figure 8 (typiquement un voisinage de

type $3 \times 3 \times 3$), c'est-à-dire on a besoin d'un ensemble contenant 26 comparaisons défini par :

$$\{D(x + \Delta_x, y + \Delta_y, t\sigma), \Delta_x \in \{-1, 0, 1\}, \Delta_y \in \{-1, 0, 1\}, t \in \{k^{-1}, 1, k\}\}$$

Si la valeur du pixel est supérieure ou inférieure aux valeurs des pixels testés, on retient le point.

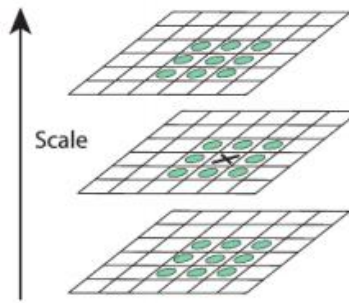


Figure 8 : Recherche d'extrema dans DoG [45, 44]

2.2.3. Localisation précise des points d'intérêt

Une interpolation de la fonction $D(x, y, \sigma)$ a pour but : d'une part, de raffiner la position des points d'intérêt pour obtenir une localisation plus précise en espace et en échelle, d'autre part, d'éliminer les points de faible contraste ou situés sur des arêtes de contour à faible courbure et donc susceptible de glisser facilement. En appliquant le développement polynomial de Taylor à l'ordre 2 sur la fonction $D(x, y, \sigma)$ autour du point-clé candidat, on obtient :

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X$$

Où $X = (x, y, \sigma)^T$ est le décalage par rapport au point d'intérêt M selon les 3 dimensions (x, y, σ) . La position précise de l'extremum local \hat{X} est obtenue en annulant la dérivée de la fonction précédente par rapport à X . On obtient donc:

$$\hat{X} = - \left(\frac{\partial^2 D}{\partial X^2} \right)^{-1} \frac{\partial D}{\partial X}$$

Si \hat{X} est supérieur à un seuil égal à 0.5 dans l'une des trois dimensions, cela signifie que l'extremum se trouve plus près d'un des points voisins dans l'espace des échelles. Dans ce cas, on recommence cette interpolation autour du point voisin le plus proche. Sinon, l'extremum \hat{X} est ajouté à la position du point candidat initial interpolé pour déterminer la position exacte de l'extrema.

Pour rejeter les points-clé de faible contraste, il suffit d'évaluer la fonction D à l'extrema \hat{X} . En cette position on a :

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x}$$

Pour les points de faible contraste, la valeur absolue de la fonction $D(\hat{X})$ est petite, il suffit donc d'appliquer un seuillage pour les éliminer.

Finalement une analyse des rapports des valeurs propres de la matrice hessienne carrée permet d'éliminer les points d'intérêt détectés sur des contours à faible courbure. Pour rappel, la matrice hessienne (ou simplement la hessienne) d'une fonction numérique f est la matrice carée, noté $H(f)$, de ses dérivées partielles secondes de la gaussienne. Cette matrice est définie par :

$$H = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix}$$

Si une des valeurs propres est très supérieure à l'autre, alors le point détecté est sur un contour (Forte variation uniquement dans un sens), sinon il s'agit d'un coin (Forte variation dans tous les sens).

Les points détectés sur un contour peuvent être filtrés facilement en analysant la trace et le déterminant de la matrice hessienne H . On peut écrire :

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\lambda_1 + \lambda_2)^2}{\lambda_1 \lambda_2} = \frac{(r\lambda_1 + r\lambda_2)^2}{r\lambda_2^2} = \frac{(r+1)^2}{r}$$

Avec $r = \frac{\lambda_1}{\lambda_2}$. Ce ratio augmente quand r augmente, on peut donc filtrer les points où ce ratio est inférieur à un seuil (Lowe utilise un seuil égal à 10).

2.2.4. Assignation d'orientation

L'étape suivante consiste à attribuer à chaque point-clé sélectionné une ou plusieurs orientations déterminées localement sur l'image et qui correspondent à la direction majoritaire des gradients spatiaux d'intensité calculés dans un voisinage autour du point d'intérêt à l'échelle préalablement déterminée. Cette étape est primordiale pour garantir l'invariance des descripteurs SIFT à la rotation.

Pour un point donné (x_0, y_0, σ_0) , le calcul est réalisé sur l'image $L(x, y, \sigma_0)$, à savoir le gradient de la pyramide dont le facteur d'échelle est le plus proche de celui du point. De cette manière, les calculs sont effectués dans un contexte indépendant de l'échelle. À chaque position dans un voisinage du point (x, y, σ) , on estime le gradient par différences finies symétriques, puis son amplitude (c.-à-d. sa norme) $m(x, y)$, et son orientation $\theta(x, y)$ dont leurs formules sont données par:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2}$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \right)$$

Un histogramme est constitué à partir de l'orientation des gradients des points sur le voisinage du point d'intérêt (Figure 9). L'histogramme est réalisé avec 36 bins (c.-à-d. intervalles), couvrant chacun 10 degrés d'angle. L'histogramme est doublement pondéré : d'une part, par une fenêtre gaussienne circulaire d'écart type de 1,5 fois l'échelle du point, d'autre part, par l'amplitude du gradient de chaque point.

L'orientation du point correspond au pic maximal de l'histogramme. Tout pic supérieur à 80% de la valeur du pic maximal est aussi pris en considération, ce qui provoque si nécessaire la génération de nouveaux points supplémentaires à la même échelle et la même position.

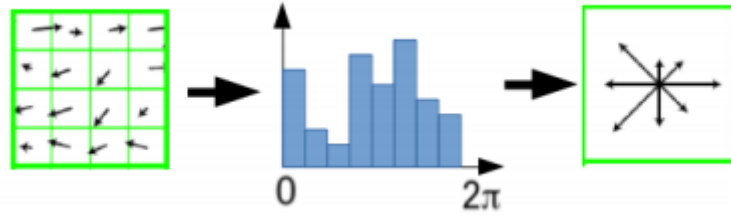


Figure 9 : Illustration de la construction de l'histogramme des orientations. A gauche : gradients des pixels définissant le voisinage du point, au centre : histogrammes des orientations à 36 bins, à droite : extraction des orientations dominantes

2.2.5. Descripteur de point d'intérêt

Une fois les points détectés et qu'on leur ait associé des facteurs d'échelle et des orientations, et assuré leur invariance aux changements d'échelles et aux rotations, on calcule un descripteur pour chaque point. À cette occasion, Lowe propose des traitements supplémentaires permettant de rendre également les descripteurs invariants à d'autres transformations telles que les changements affines de l'éclairage et de point de vue 3D, etc. Cette étape est réalisée sur l'image lissée avec le paramètre de facteur d'échelle le plus proche de celui du point considéré.

Afin de garantir l'invariance à la rotation, on modifie le système de coordonnées local autour du point d'intérêt considéré, en utilisant une rotation d'angle égal à l'orientation du point, mais de sens opposé. À partir d'une fenêtre 16×16 , toujours autour de ce point, on calcule l'histogramme des orientations par sous-fenêtres de taille 4×4 , orientés selon l'orientation principale du point calculée dans l'étape antérieure. En chaque point de ces sous-fenêtres, l'orientation et la norme du gradient sont calculés comme précédemment. L'orientation détermine l'intervalle à incrémenter dans l'histogramme, ce qui nécessite, d'une part, une pondération par la norme du gradient, et d'autre part, par un noyau gaussien centré au point d'intérêt et d'écart type 0.5 fois la taille de la fenêtre de calcul du descripteur.

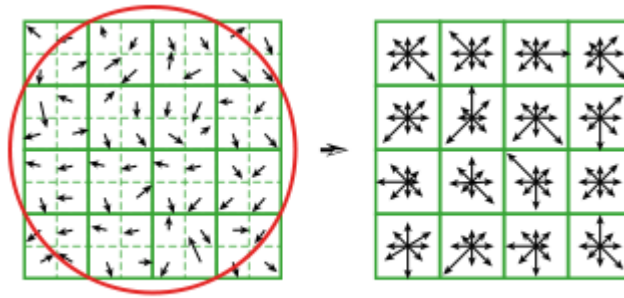


Figure 10 : Construction d'un descripteur SIFT $4 \times 4 \times 8$ à partir de gradients calculés dans une fenêtre 16×16 ¹

L'étape suivante consiste à concaténer et normaliser les seize histogrammes à huit intervalles chacun. Dans le but de limiter la sensibilité du descripteur aux changements de luminosité, les valeurs sont plafonnées à 0.2 et l'histogramme est de nouveau normalisé, le résultat est un vecteur descripteur de $4 \times 4 \times 8 = 128$ éléments.

2.3. RIFT

RIFT (Rotation invariant Feature Transform) est un descripteur invariant par rotation dérivé de SIFT, adapté aux images texturées pour lesquelles la notion d'orientation principale n'a pas vraiment de sens. Le SIFT original a été noté pour sa performance supérieure dans différentes applications, cependant, et d'après [42], il ne peut pas être utilisé directement pour la représentation de la texture, car il dépend de trouver l'orientation dominante dans un patch normalisé. Le descripteur RIFT est construit de la manière suivante : d'abord on décompose l'image en patches normalisés dans lesquels on calcule un histogramme d'orientations de gradient. La différence avec le SIFT est qu'on utilise là des patches divisés en anneaux concentriques et que les orientations du gradient en un point sont calculées par rapport à la direction du centre en ce point (Figure 11).

¹

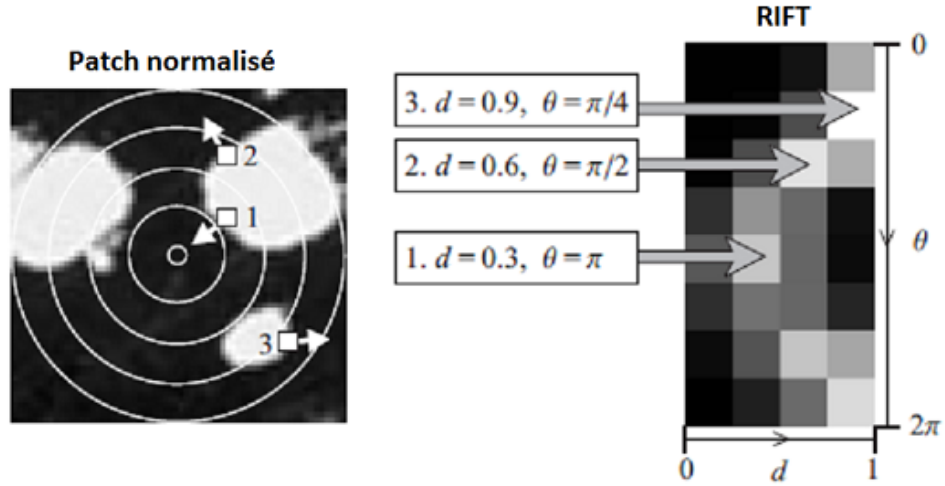


Figure 11 : Construction du descripteur RIFT [42]

Pour chaque région, on note la position centrale par (x_c, y_c) , pour une position aléatoire (x, y) dans cette région, les gradients dans les directions horizontales et perpendiculaires sont donnés par :

$$dx = \frac{I(x+1, y) - I(x-1, y)}{2}$$

$$dy = \frac{I(x, y+1) - I(x, y-1)}{2}$$

L'amplitude du gradient est $m(x, y) = \sqrt{dx^2 + dy^2}$ et son orientation $\theta(x, y) = \tan^{-1}\left(\frac{dx}{dy}\right)$.

Pour garantir l'invariance à la rotation, l'orientation du gradient du point est mesurée à chaque point par rapport à la direction pointant vers l'extérieur à partir du point central, elle est donnée par :

$$\varphi(x, y) = \theta(x, y) - \alpha(x, y)$$

Où la direction par rapport au centre est :

$$\alpha(x, y) = \tan^{-1}\left(\frac{y - y_c}{x - x_c}\right)$$

Lorsque l'image tourne, $\varphi(x, y)$ ne changera pas, d'où l'invariance à la rotation illustrée dans la Figure 12.

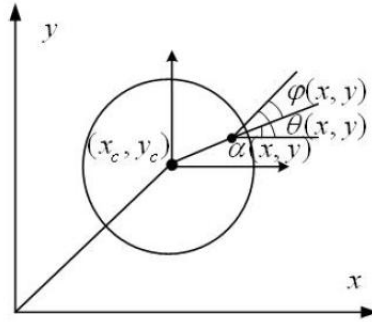


Figure 12 : extraction de l'angle avec invariance à la rotation

Une fois que l'invariance à la rotation est obtenue pour chaque point de la région locale, l'histogramme peut être facilement calculé. Le descripteur de caractéristiques RIFT est défini comme $F(d, \theta) = m(x, y)g(d, \sigma)$, avec $g(d, \sigma)$ est la fenêtre gaussienne. Finalement, le descripteur final RIFT est obtenu par :

$$RIFT(x_c, y_c) = (F(d_1, \theta_1), F(d_2, \theta_2), \dots, F(d_4, \theta_8))$$

Dans [42], les auteurs utilisent quatre anneaux et huit histogrammes d'orientations, ce qui donne des descripteurs de taille 32.

2.4. DSIFT (SIFT denses)

Le SIFT dense (DSIFT) est une variante de SIFT avec des descripteurs extraits à de multiples échelles. Ici les descripteurs SIFT sont calculés à des points sur une grille régulière avec un espacement de M pixels, une échelle et une orientation fixée. Dans [11], les auteurs choisissent $M = 5, 10$ et 15 . Au niveau de chaque point de la grille des descripteurs SIFT sont calculés sur des patchs circulaires avec des rayon $r = 4, 8, 12$ et/ou 16 pixels. En conséquence, chaque point est représenté par n descripteurs SIFT (où n est le nombre de patchs circulaires), chacun de taille 128. Lorsque $n > 1$, plusieurs descripteurs sont calculés pour permettre la variation d'échelle entre les images. Les patchs avec des rayons de 8, 12 et 16 se chevauchent. À noter que ces descripteurs sont également invariants par rotation et souvent utilisés pour la catégorisation d'objet.

2.5. SURF

Le descripteur SURF (Speeded Up Robust Features) a été introduit par Bay et al. dans [3, 4], fortement influencé par les SIFT de Lowe, puisqu'il traduit la distribution des intensités dans le voisinage du point d'intérêt. L'approche proposée par Bay utilise une approximation de la matrice hessienne afin de détecter les structures de types « blobs ». Les blobs sont localisés aux maxima du déterminant de la matrice hessienne (cf. Localisation précise des points d'intérêt). L'originalité du descripteur SURF réside dans le fait d'utiliser la notion d'image intégrale et de son calcul récurent [84] afin de diminuer fortement les temps de calculs car elles permettent le calcul rapide des convolutions avec les approximations de type « box-filters » (cf. Image intégrale).

Dans le contexte du descripteur SURF, la matrice hessienne en un point donné $X(x, y)$ et à l'échelle σ est définie par :

$$H(X, \sigma) = \begin{pmatrix} L_{xx}(X, \sigma) & L_{yx}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{pmatrix}$$

Avec $L_{xy}(X, \sigma)$, $L_{xx}(X, \sigma)$, $L_{yy}(X, \sigma)$, $L_{yx}(X, \sigma)$ sont les résultats de la convolution gaussienne de second ordre $\frac{\partial^2}{\partial x^2} g(\sigma)$, $\frac{\partial^2}{\partial x \partial y} g(\sigma)$, $\frac{\partial^2}{\partial y \partial x} g(\sigma)$ et $\frac{\partial^2}{\partial y^2} g(\sigma)$ de l'image au point $X(x, \sigma)$.

Pour pouvoir tirer parti des images intégrales, Bay et al. Proposent d'approximer les valeurs de la matrice hessienne par des filtres rectangulaires (box-filters) (Figure 13), ce qui permet de réduire le temps de calcul. Typiquement, on peut distinguer des filtres de tailles 9×9 , 15×15 , 21×21 , 27×27 , 39×39 , 51×51 , etc.

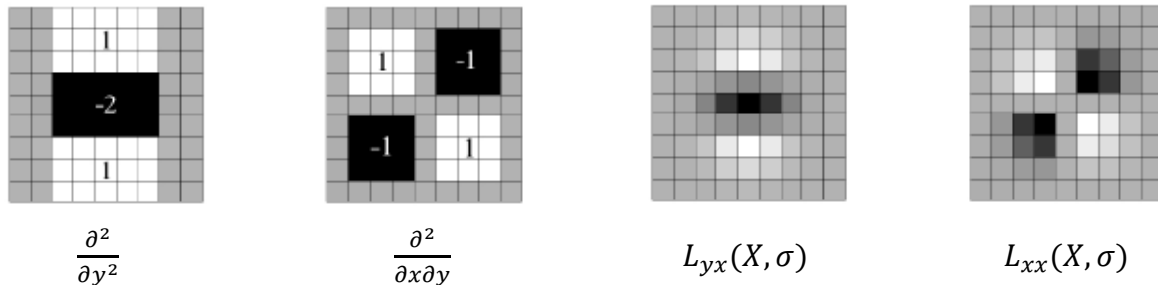


Figure 13 : Opérateurs de dérivation d'ordre 2 discrétisés et dérivées d'ordre 2 de la gaussienne [3]

Les nouvelles valeurs de la matrice Hessienne après approximation par les filtres correspondent à la somme des termes sur le voisinage des points X , et sont ensuite utilisées pour normaliser le déterminant de la hessienne. L'approximation de ce dernier est stockée dans des structures de données appelées « blob response map » (Figure 14), puis les maxima locaux sont recherchés sur ces blob pour déterminer les coordonnées et la taille des meilleurs points d'intérêt.

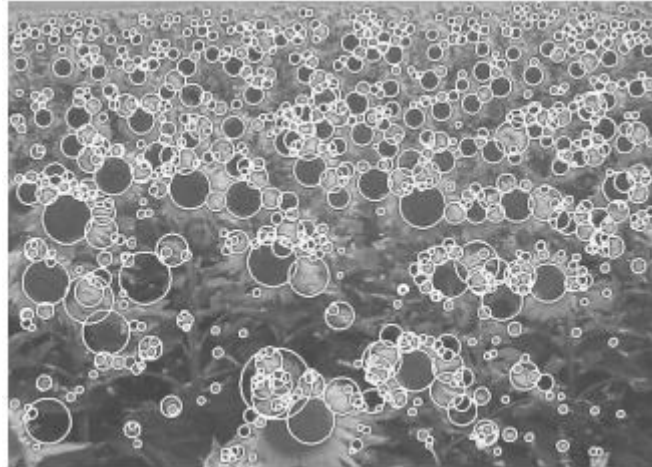


Figure 14 : Blobs détectés [3]

Le descripteur SURF est invariant aux changements d'échelle et à la rotation. Comme pour le SIFT de Lowe, l'échelle de détection σ permet de définir la taille de la fenêtre du descripteur (6σ) ainsi que la taille du box-filter utilisé. Dans la région de calcul du descripteur, les réponses aux filtres dans les directions x et y sont calculées et pondérées par une fonction gaussienne d'écart-type 2σ (Figure 15). Comme dans la construction du descripteur SIFT,

Pour chaque pixel du voisinage, un vecteur d'orientation est calculé et l'orientation maximale est considérée comme étant l'orientation dominante. Ainsi la fenêtre de calcul est orientée dans la direction de l'orientation dominante et subdivisée en sous-fenêtres de taille $4 * 4$. Pour chacune des ces sous-fenêtres, on calcule les quatre valeurs suivantes : $\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|$, où d_x et d_y sont les réponses du filtre de Haar de premier ordre (box-filter) dans les directions x et y respectivement (Figure 15).

Finalement, le vecteur descripteur SURF est obtenu par concaténation de ces quatre valeurs pour chacune des sous-fenêtres considérées. Une dernière étape de normalisation du vecteur est nécessaire afin d'assurer l'invariance au contraste.

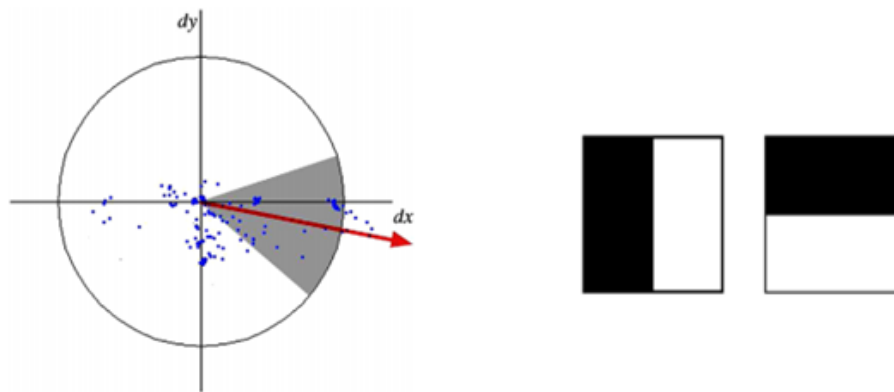


Figure 15 : La structure du descripteur SURF. A droite : les ondelettes de Haar. A gauche : calcul de l'orientation dominante [4]

2.6. GLOH

L'opérateur GLOH (Gradient Location and Orientation Histogram) [47] est une extension du descripteur SIFT destiné à accroître sa robustesse et sa spécificité. Il s'applique sur des régions concentriques à la place des grilles de patterns utilisées dans le SIFT de Lowe. Il se génère de la manière suivante : pour l'opérateur original SIFT, les positions sont établies avec trois rayons du petit au grand dans la direction radiale, sur la base des coordonnées polaires logarithmiques ($r = 6, r = 11, r = 15$), et huit orientations angulaires : l'espace des coordonnées polaires est défini par (r, θ, φ) qui représente respectivement le rayon, l'orientation polaire et l'orientation du gradient du pixel.

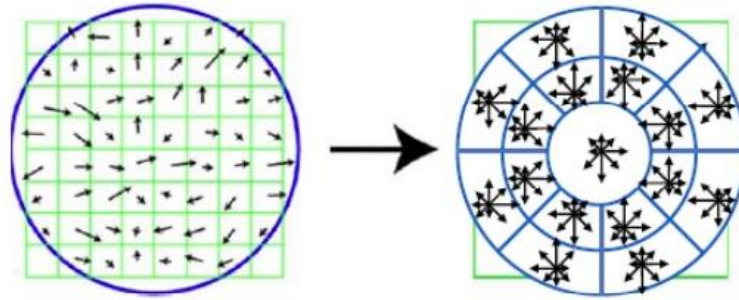


Figure 16 : (à gauche) la structure de la fenêtre de calcul du SIFT, (à droite) la structure de la fenêtre du calcul du GLOH [36]

La différence avec le descripteur SIFT réside dans la structure spatiale de la fenêtre de calcul du descripteur (Figure 16). Celle-ci est représentée par une grille polaire de 17 sous-régions (une région circulaire centrale et $8 * 2$ sous-régions à des intervalles de $\frac{\pi}{4}$ sur deux régions circulaires centrées sur le point d'intérêt). À noter que la région centrale ne se divise pas dans les directions angulaires. Les orientations du gradient sont quantifiées sur 16 sous-régions. En chacune des sous-régions, un histogramme de gradients orientés est construit suivant seize bins (intervalles de $\frac{\pi}{8}$). L'histogramme final est donc constitué de $16 \times 17 = 272$ données, seuillées et normalisées. Enfin, la taille est réduite à 128 dimensions en utilisant l'analyse en composante principale ACP [41]. La structure du masque d'analyse du descripteur GLOH est montrée dans la Figure 17.

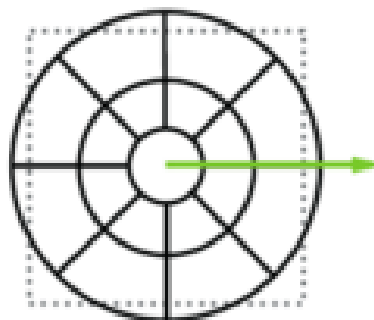


Figure 17 : La structure du masque d'analyse du descripteur GLOH [47]

2.7. **DAISY**

Le descripteur DAISY proposé par Tola et al. en 2008 et repris ensuite en 2010 [79, 80] est proche des descripteurs SIFT et GLOH. Il a été introduit pour accélérer les temps de calculs et d'améliorer l'invariance. Selon les auteurs, il est 66 fois plus rapide que le SIFT. La différence essentielle réside dans la forme de la région dans laquelle est calculé le descripteur et dans l'utilisation des filtres de dérivées gaussiennes orientées à la place des gradients exploités dans les méthodes précédemment citées. Étant donné une image I et une orientation θ , les auteurs proposent de créer huit orientations définies par :

$$M_{\theta}^{\sigma} = g_{\sigma} * \left(\frac{\partial I}{\partial \theta}\right)^{+}$$

Le terme M_{θ}^{σ} correspond à l'ensemble des gradients d'une orientation donnée, ayant une norme positive. L'image I est convertie en une série de maps d'orientations (une dans chaque direction).

Tola et al. présentent un masque d'analyse ayant une structure constituée de 25 cercles se chevauchant et centrée sur le point d'intérêt (Figure 18) : les rayons des cercles augmentent proportionnellement à leurs distance par rapport au centre du descripteur, à savoir le pixel pour lequel le descripteur est calculé, et la puissance du lissage gaussien est proportionnelle aux rayons des cercles. Pour chaque région circulaire, un histogramme d'orientation est calculé à partir de toutes les valeurs de M_{θ}^{σ} suivant une orientation particulière θ , il est défini en une position (x, y) par :

$$h_{\sigma} = [M_1^{\sigma}(x, y), M_2^{\sigma}(x, y), \dots, M_8^{\sigma}(x, y)]^T$$

Ces histogrammes sont calculés et normalisés selon les mêmes principes de SIFT et GLOH afin d'accroître l'invariance aux changements de luminosité. Ils sont notés par $\tilde{h}_{\sigma}(x, y)$.

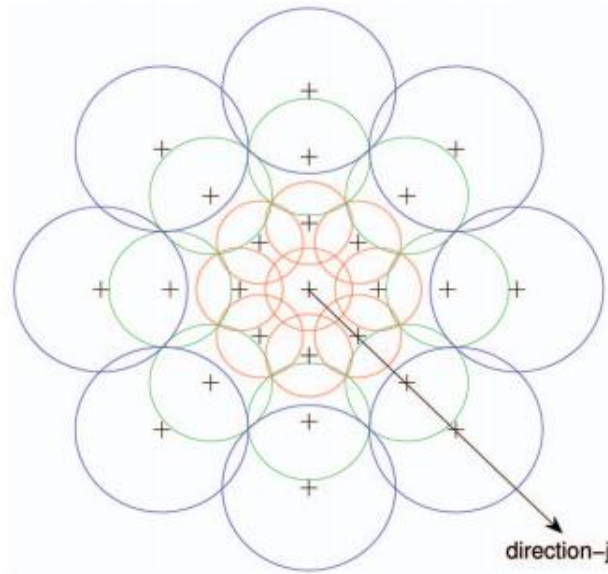


Figure 18 : La structure du masque d'analyse du descripteur DAISY [80]

Une fois que ces calculs sont faits, le vecteur descripteur final $D(x_0, y_0)$ est construit par la concaténation de tous les histogrammes d'orientation \tilde{h} de la région considérée, et se compose donc de vingt-cinq histogrammes (un par cercle) possédant chacun huit orientations. Il peut être écrit de la manière suivante :

$$D(x_0, y_0) = [\tilde{h}_{\sigma_1}^T(x_0, y_0), \\ \tilde{h}_{\sigma_1}^T(I_1(x_0, y_0, R_1)), \dots, \tilde{h}_{\sigma_1}^T(I_N(x_0, y_0, R_1)), \\ \tilde{h}_{\sigma_2}^T(I_1(x_0, y_0, R_2)), \dots, \tilde{h}_{\sigma_2}^T(I_N(x_0, y_0, R_2)), \\ \tilde{h}_{\sigma_3}^T(I_1(x_0, y_0, R_3)), \dots, \tilde{h}_{\sigma_3}^T(I_N(x_0, y_0, R_3)),]^T$$

Où $I_j(x, y, R)$ correspond à la position du cercle avoisinant à une distance R de (x, y) et suivant une orientation j . Les cercles sont ordonnés suivant trois rayons : R_1, R_2 et R_3 et suivant huit orientations : $1 \rightarrow N$ avec l'indice 1 pour l'orientation locale du cercle central (le point d'intérêt considéré).

Dans leurs tests, les auteurs recommandent d'utiliser : $N = 8$ orientations avec $R_1 = 2.5, R_2 = 3R_1$ et $R_3 = 6R_1$ ainsi que $\sigma_1 = 2.55, \sigma_2 = 3\sigma_1$ et $\sigma_3 = 5\sigma_1$. Enfin, leur descripteur est constitué de $8 + 8 \times 3 \times 8 = 200$ valeurs et extrait à partir de 25 cercles et suivant 8 orientations.

2.8. MOPS

Comme le SIFT de Lowe, les motifs orientés multi-échelles (en anglais multi-scale oriented patches) [14] utilisent les gradients du voisinage du point d'intérêt pour extraire une orientation dominante. Les MOPS sont construit en utilisant un échantillonnage des valeurs d'intensité par le biais de patches de taille 8×8 pixels calculés dans un voisinage de rayon 4.5 fois l'échelle de détection du point d'intérêt, avec un échantillonnage d'espacement 5 pixels entre chaque pixel retenu (Figure 19). Cet échantillonnage basse-fréquence donne au descripteur une certaine robustesse aux erreurs (bruit) de localisation des points d'intérêt. En conséquence, un vecteur de 64 dimensions est obtenu et normalisé. Finalement, une transformée en ondelettes de Haar est appliquée à ce vecteur pour former un vecteur descripteur contenant les 64 coefficients des ondelettes. Ce descripteur est invariant aux faibles changements d'intensité.

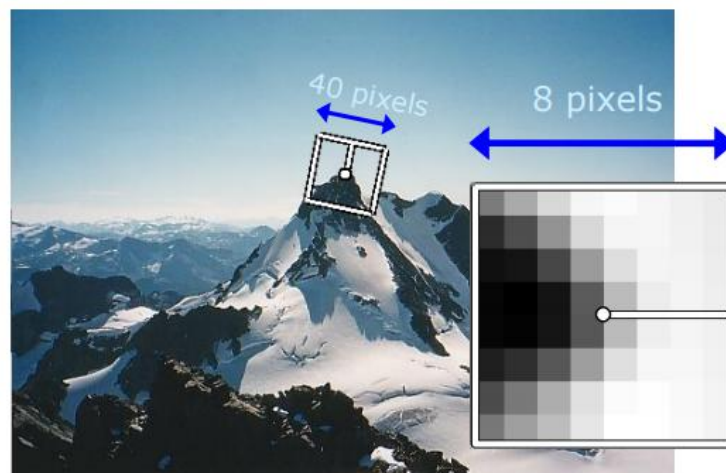


Figure 19 : Structure des descripteurs MOPS [14]

2.9. Shape Context

Shape context (SC), que l'on peut traduire par contexte de forme, est un descripteur d'images présenté pour la première fois en 2000 par Serge Belongie et Jitendra Malik [5], et repris en 2002 par les mêmes auteurs [6] afin de mesurer la similarité entre deux formes différentes et pour l'utiliser aussi dans la reconnaissance d'objets. Le descripteur SC se base sur l'idée qu'une forme est associée à un objet dont le principe est d'extraire d'une image les points p_i décrivant les contours, et d'obtenir pour chacun de ces points le contexte de forme en déterminant la distribution des points de contour dans la région au voisinage de p_i , au

moyen d'un histogramme de coordonnées relatives des points de contour par rapport à ces points p_i qu'on appelle les points de référence.

Une forme est décrite par un sous-ensemble discret de points de contours externes et internes. Supposons que $C = \{p_1, p_2, \dots, p_n\}$, $p_i \in \mathbb{R}^2$ est l'ensemble des points de contours détectés sur l'image, avec n est le nombre de points de contour :

Pour chaque point p_i du contour de la forme, on considère les $n - 1$ vecteurs obtenus en connectant le point p_i à tous les autres points. Le contexte de forme h_i du point p_i est l'histogramme des coordonnées relatives des autres $n - 1$ points, il est défini par :

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in \text{bin}(k)\}$$

Où $h_i(k)$ est le nombre de points de contour appartenant à la $k^{\text{ème}}$ classe de l'histogramme et $\text{bin}(k) = (d_q, \theta_q)$.

Avec :

$$q = (\log(d_q), \theta_q), \forall q \neq p_i \wedge q \in C$$

Où d_q est la distance entre q et p_i , θ_q est l'angle entre le vecteur $\overrightarrow{p_i q}$ et l'axe horizontal (Figure 20)

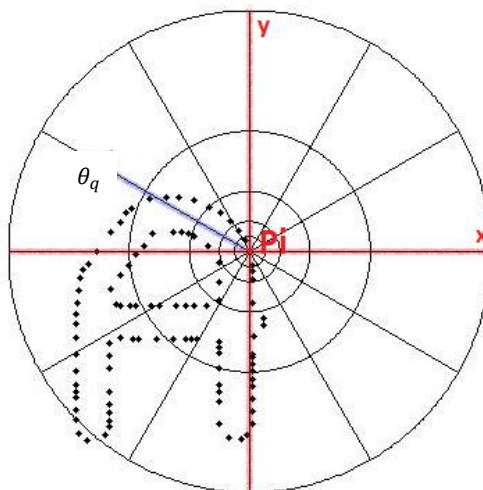


Figure 20: Diagramme d'histogramme log-polaire

Les auteurs utilisent des bins (ou régions) qui sont uniformes dans l'espace log-polaire, rendant le descripteur plus sensible aux points les plus proches qu'aux ceux les plus loin.

Une comparaison entre les descripteurs SC de deux versions différentes de la lettre « A » est montrée dans la Figure 21.

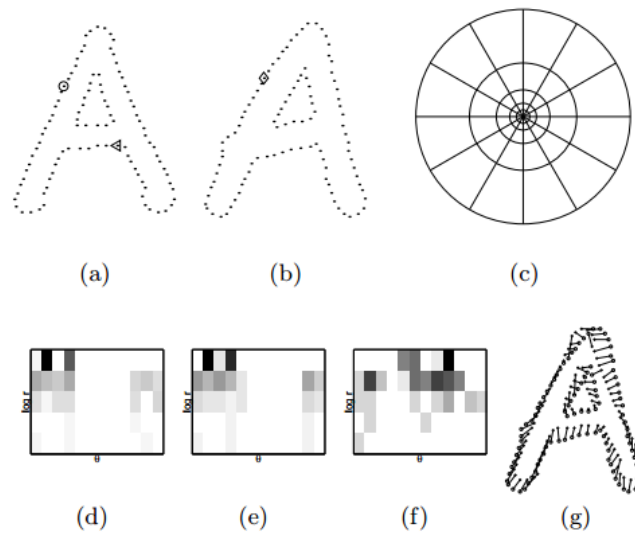


Figure 21 : Comparaison des contextes de forme de deux versions différentes de la lettre "A". (a) et (b) correspondent aux points des contours des deux formes. (c) est le diagramme d'histogramme log-polaire utilisé pour calculer le shape context. Belongie et al. utilisent 5 bins pour $\log(d)$ et 12 pour θ . (d-f) les contextes de forme pour les points de référence marqués respectivement dans (a) et (b) par un cercle, un triangle et un rectangle. (g) les correspondances trouvées [5]

Chaque contexte de forme est un histogramme log-polaire des coordonnées de l'ensemble des autres points calculés en prenant le point de référence comme origine. À noter dans la Figure 21, la similitude visuelle entre le contexte de forme de la référence marquée par un cercle et celui de la référence marquée par un rectangle, et qui ont été calculés pour des points relativement similaires sur les deux formes. En revanche, le contexte de forme de la référence marquée par un triangle est tout à fait différent. Les correspondances sont trouvées en utilisant un appariement avec des coûts définis par la distance χ^2 entre les histogrammes.

Considérons un point p_i dans la première forme et un point q_j dans la deuxième forme. Soit $C_{ij} = C(p_i, q_j)$ représente le coût de mise en correspondance de ces deux points.

Puisque les contextes de forme sont des distributions représentées sous forme d'histogrammes, il est donc naturel d'utiliser le test statistique χ^2 :

$$C_{ij} = C(p_i, q_j) = \frac{1}{2} \sum_{k=1}^K \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)}$$

Où $h_i(k)$ et $h_j(k)$ représentent le nombre de points de contour appartenant à la $k^{\text{ème}}$ classe des histogrammes aux points p_i et q_i respectivement.

Cependant le contexte de forme décrit ci-dessus n'est pas invariant à la rotation et au changement d'échelle. Pour obtenir l'invariance au changement d'échelle, les distances radiales sont normalisées. Par la distance moyenne α des n^2 paires de points de la forme.

Pour que le contexte de forme soit aussi invariant à a rotation, les auteurs ont proposé d'utiliser le vecteur tangent associé à chaque point au lieu de l'axe absolu horizontal. Dans chaque région ou *bin* k , les vecteurs t_j tangents à chacun des points q_j de contour Q appartenant à la grille sont calculés. L'orientation est donnée par la somme sur chaque *bin* :

$$\hat{h}_i(k) = \sum_{q_j \in Q} t_j, \text{ où } Q = \{q_j \neq p_i, (q_j - p_i) \in \text{bin}(k)\}$$

Ce descripteur est appelé *Generalized Shape Context* [49] et contient l'orientation dominante des contours dans chaque bin de l'histogramme.

2.10. Histogrammes de gradients orientés

Les descripteurs HOG (Histogram of Oriented Gradients) ont été introduits par Navneet Dalal et Bill Triggs [17, 18], deux chercheurs de l'Institut National de Recherche en Informatique et en Automatique (INRIA). Le but des HoG est de représenter l'apparence et la forme d'objet dans une image grâce à la distribution d'intensité des gradients ou de direction des contours. Ceci est effectué en divisant l'image en régions connectées, appelées cellules, et en calculant pour chaque cellule un histogramme des directions de gradients ou des orientations de contour pour les pixels appartenant à cette cellule. Le descripteur final est obtenu par la concaténation de ces histogrammes. La Figure 22 montre le processus d'utilisation des HOG dans un contexte de détection d'objets à base des SVM.

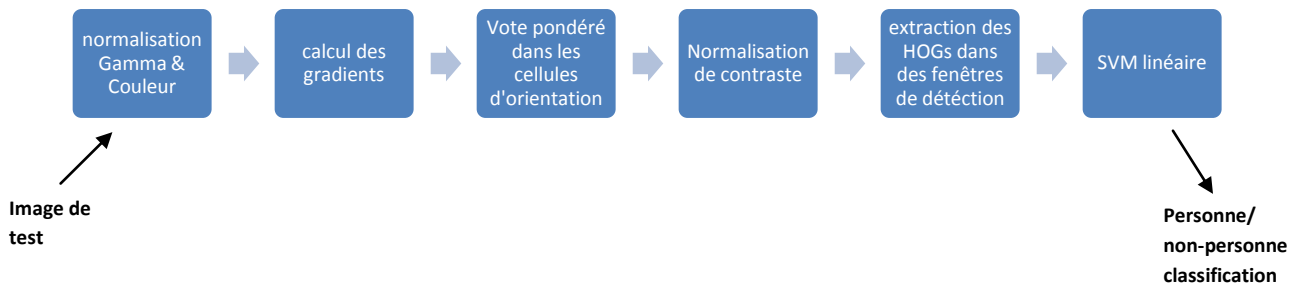


Figure 22 : Chaîne d'extraction des descripteurs HoG et de détection d'objet proposé par Dalal et Triggs dans [17]

2.10.1. Normalisation Gamma/Couleur

Avant de commencer, une étape de prétraitement facultative peut être effectuée : la normalisation des couleurs de l'image et une correction gamma correcte. Cette étape n'a qu'une faible incidence sur les performances et n'est donc pas obligatoire, la normalisation du descripteur HOG lui-même s'avèrent suffisante. Dalal et Triggs ont montré des résultats comparables pour les espaces couleurs RGB et LAB, alors que le niveau de gris réduit les performances.

2.10.2. Calcul des gradients

Le calcul du gradient est une étape critique dans la formation du descripteur, les performances de ce dernier sont sensibles à la manière dont ont été calculés les gradients. La justesse des orientations calculées, et des histogrammes, dépends de cette étape et les résultats sont donc étroitement liés à la méthode employée pour calculer le gradient dans l'image. Il s'agit de calculer le gradient pour tous les points de l'image ; généralement deux filtres dérivatifs sont appliqués sur l'image, un horizontal et un vertical, qui peuvent être des masques de dérivation simple 1-D (centré $[-1 \ 0 \ 1]$ et non centré $[-1 \ 1]$), des opérateurs 2-D de Sobel, ou encore des opérateurs récurrents de Deriche.

Dans le cas des images couleurs, le gradient est calculé séparément pour chaque canal, le gradient ayant la norme la plus grande est gardé.

Une opération de lissage gaussien 2-D est souhaitable afin de réduire le bruit et de limiter le nombre de contours. Différents filtres dérivatifs ont été testés par Dalal et Triggs:

- Filtre dérivatif 1-D non centré : $[-1 \ 1]$
- Filtre dérivatif 1-D centré : $[-1 \ 0 \ 1]$

- Filtre dérivatif 1-D ajusté cubiquement : $[1 \quad -8 \quad 0 \quad 8 \quad -1]$
- Filtre de Sobel 3×3 : $\begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -2 \end{bmatrix}$
- Filtre diagonale 2×2 : $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ et $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

Prenons l'exemple du filtre de Sobel de taille 3×3 , sa convolution avec l'image permet de calculer les approximations des dérivées horizontales et verticales. Soit I l'image source, pour chaque point de I , une approximation de la dérivée horizontale notée G_x (respectivement verticale notée G_y) du gradient est obtenue (Figure 23) :

$$G_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -2 \end{bmatrix} * I \text{ Et } G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I$$

En chaque point, les approximations horizontales et verticales sont combinées pour obtenir une approximation de la norme du gradient (voir Figure 23.d) :

$$G = \sqrt{G_x^2 + G_y^2}$$

On calcule également l'orientation du gradient comme suit :

$$\theta = \text{Arc tan} \left(\frac{G_x}{G_y} \right)$$

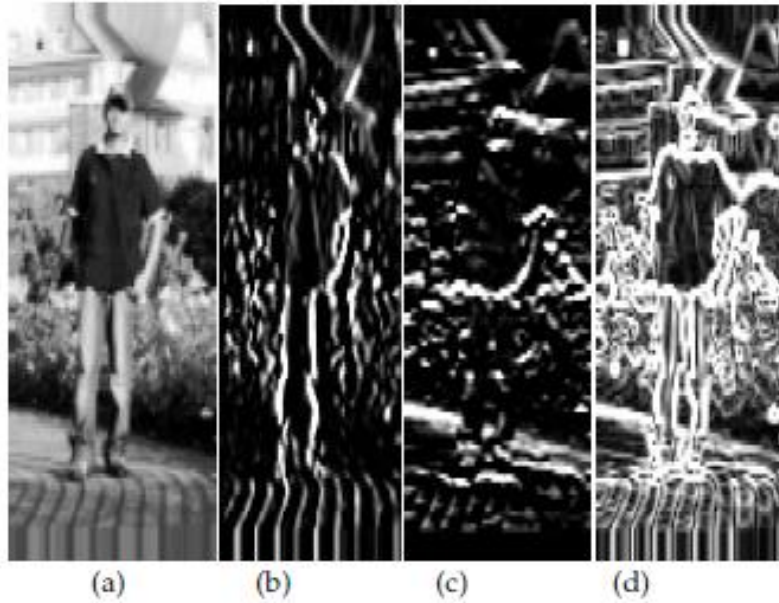


Figure 23 : (a) image d'un piéton en niveau de gris normalisée, (b) composante horizontale du gradient, (c) composante verticale du gradient, (d) la norme de gradient [18]

2.10.3. Calcul des histogrammes

L'image est découpée en blocs (ou cellules), généralement de taille 8×8 pixels, pour chacun d'entre eux on calcule l'histogramme d'orientation de gradient. Chaque pixel de blocs vote pour une orientation entre 0 et 180 dans le cas du gradient non signé (Figure 24.b), ou entre 0 et 360 dans le cas du gradient signé. Ces orientations du gradient sont représentées par les classes de l'histogramme.

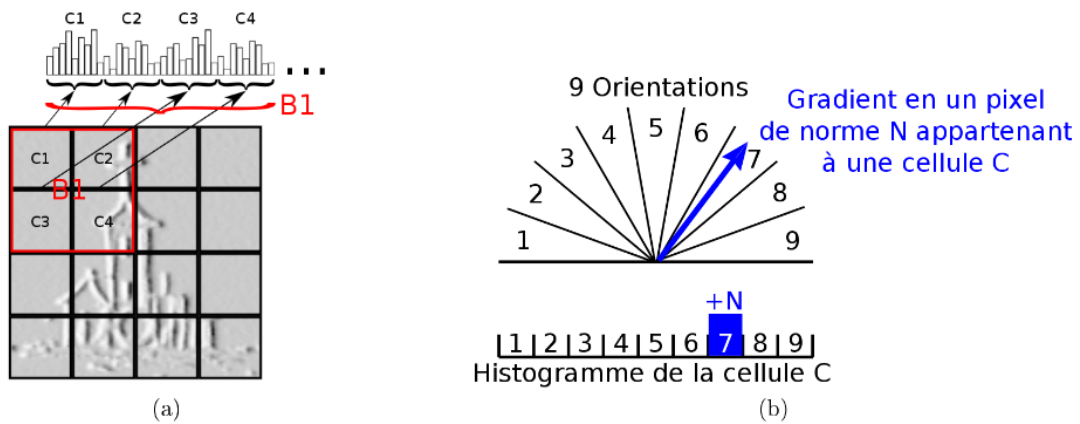


Figure 24 : Construction du HoG d'une image: (a) découpage en blocs et formation de l'histogramme des blocs adjacents, (b) vote d'un pixel selon l'orientation de ses gradients

L'étape suivante est la normalisation des histogrammes, afin d'éviter les disparités dues aux variations d'illumination, ainsi que l'introduction de redondance dans le descripteur. Pour cela, les histogrammes de blocs adjacents sont concaténés pour former un vecteur qui est ensuite normalisé.

Le descripteur HOG final est obtenu en répétant l'opération sur tous les blocs adjacents possibles.

Les normalisations possibles du descripteur final sont les suivantes (v représente l'histogramme d'un bloc) :

- L2-Norme : $v \rightarrow \frac{v}{\sqrt{\|v\|_2^2 + \epsilon^2}}$
- L1-Norme : $v \rightarrow \frac{v}{\|v\|_1 + \epsilon}$
- L1-Racine : $v \rightarrow \sqrt{\frac{v}{\|v\|_1 + \epsilon}}$

Différents types de masques dérivatifs ont été testés dans [17], permettant de différencier deux types d'architectures : les R-HOG (Rectangular-HOG) calculés avec des fenêtres rectangulaires, et les C-HOG (Circular-HOG) calculés avec des fenêtres circulaires. Les différents descripteurs utilisant les HOG peuvent donc être classés dans l'une des deux classes.

Ces dernières années, de nombreuses variantes par rapport aux HOG ont été proposées [57, 2]. Afin de choisir la meilleure version à implémenter, il est nécessaire de passer par une étape de test pour régler les différents paramètres des HOG, généralement, le type du filtre dérivatif et le nombre de bins de l'histogramme.

Ce descripteur est devenu l'un des plus performants en reconnaissance d'objet, par sa capacité d'être applicable en temps réel, notamment à travers l'utilisation des images intégrales introduites par Viola et Jones [84], l'idée est inspirée de la méthode proposée par Porikli [61] qui calcule rapidement des histogrammes d'une région rectangulaire aléatoire, nommée 'Histogramme intégral'.

2.11. Descripteurs binaires

2.11.1. Local Binary Patterns (LBP)

L'opérateur LBP (Local Binary Patterns) a été introduit en 1996 par Ojala et al. [53] puis complété dans [54]. L'idée de cet opérateur de texture est d'assigner à chaque pixel un code dépendant des valeurs des niveaux de gris des pixels situés dans le voisinage (Figure 25). La valeur du niveau de gris du pixel central n_c de coordonnées (x_c, y_c) est comparée à celle de ses pixels voisins n_i suivant l'équation :

$$LBP(x_c, y_c) = \sum_{i=0}^{N-1} s(n_c - n_i) * 2^i$$

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Avec N est le nombre de pixels voisins équi-répartis sur le cercle de centre n_c et de rayon R . Cet opérateur renvoie un mot binaire créé en comparant successivement l'intensité du pixel central n_c avec celles des pixels $n_i (i = 0, \dots, N)$. Si la valeur du niveau de gris du pixel central est supérieure ou égale à celle de son voisin, la valeur est mise à 1, sinon à 0.

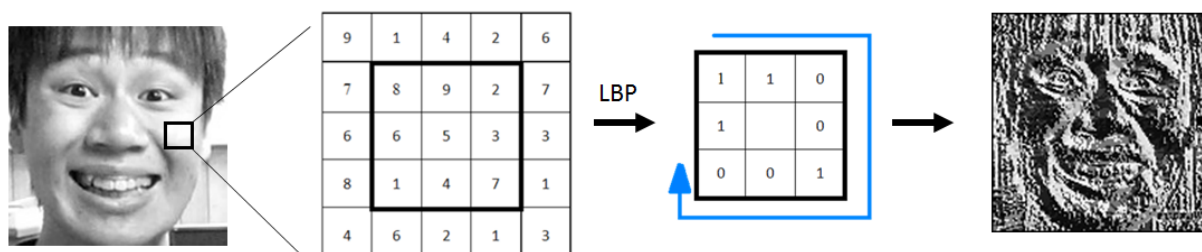


Figure 25 : Les étapes de calcul du LBP²

Généralement N est fixé à huit, c.à.d. un voisinage de 3×3 pixels. On obtient donc, dans le cas d'une image en niveaux de gris, une matrice contenant des valeurs des LBP comprises entre 0 et 255 pour chaque pixel. Finalement, un histogramme est calculé en se basant sur ces valeurs pour former le descripteur LBP.

Un inconvénient de ce descripteur est le nombre de paramètres à fixer au préalable. Dans la littérature, il existe des versions différentes de ce descripteur, suivant le paramètre fixé N

² <http://robinhsieh.com/?p=156>

désignant le nombre de pixels voisins et suivant le rayon R choisi. Dans le cas des grandes bases de données, le paramétrage optimum devient très compliqué. En plus, l'application de ce descripteur en temps réel est difficile si le nombre de pixels voisins considéré est important.

Néanmoins, l'avantage du descripteur LBP est sa représentation binaire ayant une capacité et une rapidité de traitement car il ne nécessite pas de gestion de nombres décimaux. De plus, le stockage mémoire des informations, surtout dans le cas des grandes bases de données.

2.11.2. CENTRIST

CENTRIST (CENSus TRansform hISTogram) est un descripteur visuel qui a été proposé par Jianxin Wu et James Rehg [90] pour contourner le problème de la catégorisation de scènes et de la reconnaissance de lieux. Les différentes étapes du calcul sont les suivantes:

- Extraction de contour (Figure 26). Cette étape peut-être réalisée à l'aide d'un opérateur quelconque tel que le filtre de Sobel.



Image originale

Extraction de contours

Figure 26 : Détection de contours par le filtre de Sobel [89, 90]

- Encodage de la structure locale du gradient au voisinage de chaque pixel ou « Census Transform » (CT). Il s'agit d'une transformation locale qui réalise la comparaison du niveau de gris du pixel considéré avec celui de ses 8 voisins. Dans cette

transformation chacun des pixels voisins se voit affecté la valeur « 0 » ou « 1 » suivant le résultat « vrais » ou « faux » de la comparaison. Voici un exemple :

$$\begin{pmatrix} 32 & 64 & 96 \\ 32 & 64 & 96 \\ 32 & 32 & 96 \end{pmatrix} \rightarrow \begin{matrix} 1 & 1 & 0 \\ 1 & P_{x,y} & 0 \\ 1 & 1 & 0 \end{matrix} \rightarrow (11010110)_2 \rightarrow CT = 214$$

Les huit bits générés à partir des comparaisons des valeurs d'intensités peuvent être mis ensemble dans un ordre quelconque (du gauche à droite, et du haut vers le bas). La nouvelle valeur du pixel central est donnée par la conversion en base 10 de cette suite de bits, cette valeur correspond à la Census Transform du pixel central. En tant qu'approche visuelle, une image transformée par la CT est créée en remplaçant la valeur du niveau de gris de chaque pixel par celle de sa CT (Figure 27).



Image originale

CT

Figure 27 : Exemple d'une image transformée par la Census Transform [89, 90]

- L'étape finale consiste à découper l'image en $M \times M$ régions de taille 3×3 pixels et de calculer l'historgramme dans chacune de ces sous-régions (Figure 28). L'ensemble de ces histogrammes forment le descripteur visuel CENTRIST. À noter qu'on obtient 256 valeurs possibles pour la CT car on travaille sur 8 bits.

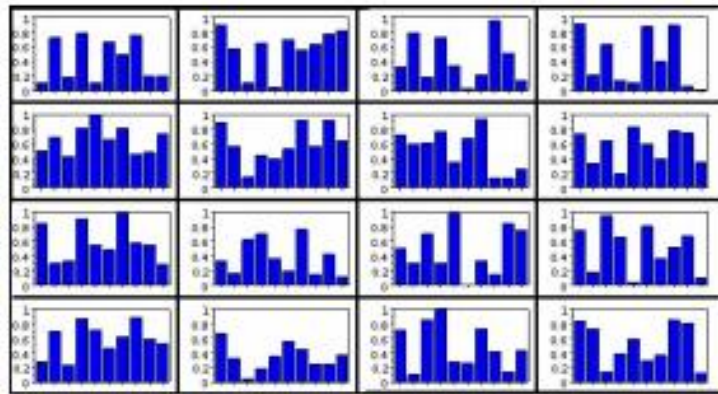


Figure 28 : Histogrammes des sous-régions de la Census Transform [89, 90]

L'inconvénient de ce descripteur est qu'il n'est calculé que pour une seule échelle. De plus, les auteurs utilisent le filtre de Sobel qui n'est pas le meilleur choix, car il existe d'autres détecteurs de contours plus efficaces dans la littérature. Enfin, une autre limitation du CENTRIST est que les valeurs de la CT des voisinages sont fortement corrélées. Un exemple est illustré dans la Figure 29 : le bit 5 de $CT(x, y)$ et le bit 4 de $CT(x + 1, y)$ sont toujours complémentaires sauf dans le cas de $I(x, y) = I(x, y + 1)$, car le premier vaut 1 si $I(x, y) \geq I(x, y + 1)$ et le deuxième vaut 1 si $I(x, y) \leq I(x, y + 1)$. En effet, Wu et Rehg mentionnent que ces contraintes (et d'autres) existent mais ils ne les exploitent pas. Pour réduire la dimensionnalité du vecteur, une analyse en composante principale (ACP) est utilisée. À noter que la Census transform est équivalente (à l'ordre de parcours près) à l'encodage LBP (Local Binary Pattern) d'OJALA [53].

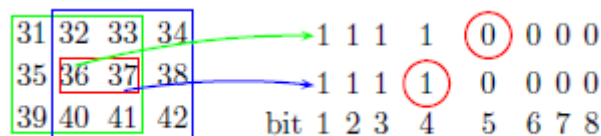


Figure 29 : Illustration des contraintes entre les valeurs de la CT des pixels voisins [89, 90]

Néanmoins, l'avantage de ce descripteur est la notion de la transformation CT qui est robuste aux changements d'illumination et aux variations gamma, etc.

3. Descripteurs globaux

3.1. Filtres de Gabor

Les filtres de Gabor sont largement utilisés en vision par ordinateur, pour la description de la texture [94]. Généralement, les paramètres de texture sont déterminés en calculant un ensemble de valeurs des moyennes et d'écart type des niveaux de gris de l'image filtrée par Gabor, ces valeurs seront regroupées dans un seul vecteur descripteur.

Les filtres de Gabor sont des ondes sinusoïdales avec une fréquence et une orientation particulière modulée par une gaussienne elliptique. Dans le cas bidimensionnel, cette onde sinusoïdale est la somme de deux fonctions sinusoïdales, la première paire et réelle, et la deuxième impaire et imaginaire. Ces filtres ont une réponse impulsionnelle de la forme:

$$h(x, y) = g(X, Y)e^{2\pi j(Ux+VY)}$$

avec :

- $(X, Y) = (x \cos \phi + y \sin \phi, -x \sin \phi + y \cos \phi)$ ou l'angle de rotation ϕ de (X, Y) par rapport à (x, y) donne l'orientation de l'enveloppe gaussienne dans le domaine spatial, et
- $g(X, Y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(\frac{X}{\lambda})^2 + Y^2}{2\sigma^2}}$ est une fonction gaussienne bidimensionnelle de rapport d'axe λ , de facteur de dilatation σ .

Un filtre de Gabor en partie réelle et imaginaire, est illustré dans la Figure 30.

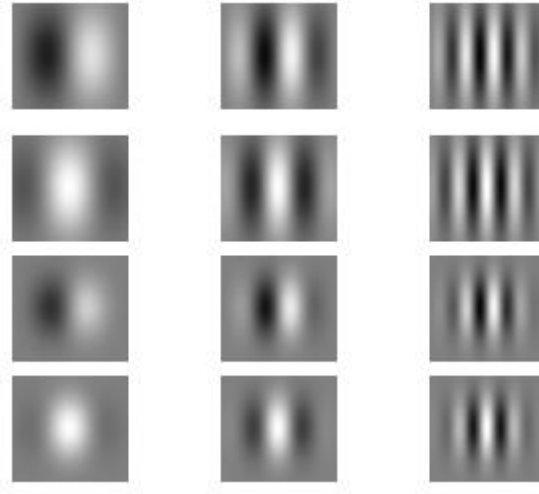


Figure 30 : Filtres de Gabor à différentes échelles et fréquences spatiales

La transformée de Fourier correspondante à $h(x, y)$ est donnée par :

$$H(u, v) = e^{-2\pi^2\sigma^2[(u'-U')^2\lambda^2+(v'-V')^2]}$$

Où

- $(u', v') = (u \cos \phi + v \sin \phi, -u \sin \phi + v \cos \phi)$, et
- (U', V') est obtenu par rotation du point (U, V) par le même angle ϕ .

$H(u, v)$ est une gaussienne passe-bande de rapport d'axes $1/\lambda$, tournée d'un angle ϕ , centrée en $F = \sqrt{U^2 + V^2}$ et d'orientation $\theta = \tan^{-1}(\frac{U}{V})$. La connaissance de ces paramètres permet de définir complètement le filtre de Gabor.

La Figure 31 présente 40 filtres de Gabor de différentes fréquences et de différentes orientations, par leur partie réelle et les modules de 5 fréquences.

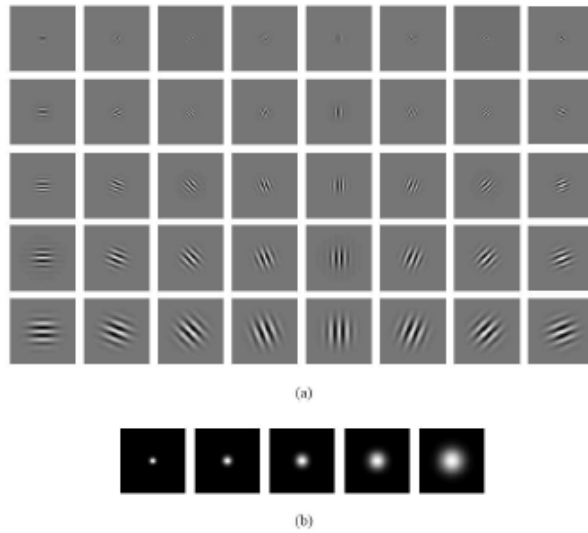


Figure 31 : Filtre de Gabor (a) en partie réelle (b) les modules de 5 fréquences

Le filtrage de Gabor conserve les aspects temporels et fréquentiels du signal. Dans le domaine spatial, ce filtrage est effectué par la convolution complexe d'une image $I(x, y)$ par un filtre de Gabor, elle est donnée par :

$$G(x, y) = I(x, y) * h(x, y)$$

Dans le domaine fréquentiel, le produit de convolution se réduit à une simple multiplication des transformées de Fourier. La Figure 32 montre les résultats de l'application de cette transformation sur une même image avec différentes fréquences et orientations.

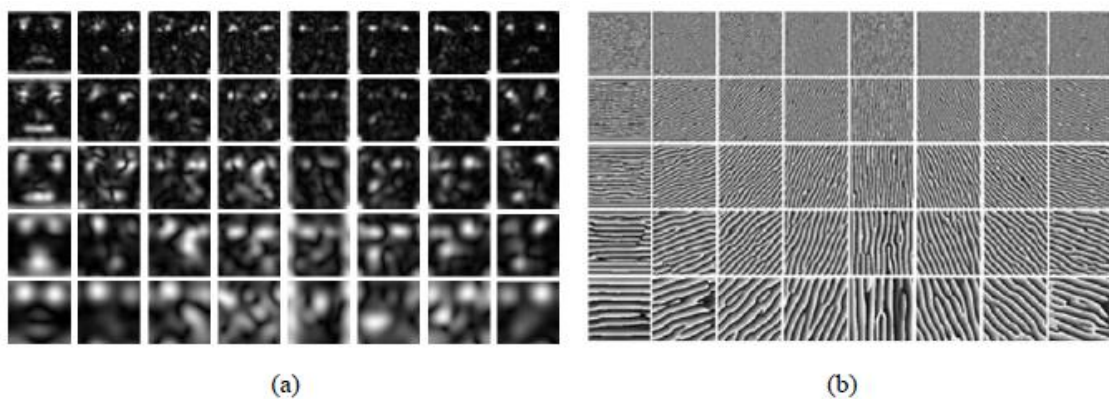


Figure 32 : Transformation de Gabor (a) en module (b) en phase

Dans la littérature, La plupart des travaux effectués n'utilisent que la partie module, comme dans [95, 88]. Les phases sont considérées comme des informations inutiles pour la description d'images, car elles varient considérablement même dans des régions locales presque identiques.

C'est en 2007 que Zhang et al. ont proposé une nouvelle méthode de représentation de la texture pour la reconnaissance de visage [91], appelée Histogramme de Gabor phase pattern (HGPP), combinant ainsi l'histogramme spatiale et les informations de phase de Gabor.

3.1.1. Histogramme de Gabor Phase Patterns (HGPP)

Pour encoder les informations de phase de Gabor, Zhang et al. ont proposé deux patterns : GGPP (Global Gabor Phase Patterns) et LGPP (Local Gabor Phase Patterns) [91], les deux sont définis en fonction du codage QBC (Quadrant-bit codes) [19] de la partie réelle et de la partie imaginaire de la transformation de Gabor. Alors que le GGPP encode les informations d'orientation à chaque échelle, le LGPP encode les variations locales au voisinage à chaque orientation et échelle. Enfin, GGPP et LGPP sont combinés avec des histogrammes spatiaux pour modéliser l'image d'origine de l'objet.

a. Encodage QBC (Quadrant-bit codes)

Selon la méthode proposée par Daugman dans [19] et appliquée dans la reconnaissance de l'iris, chaque pixel de l'image résultante sera codé sur deux bits, et donc le calcul des patterns de phase de Gabor se fera selon les formules suivantes :

$$P_{u,v}^{Re}(Z) = \begin{cases} 0, & \text{if } Re(G_{u,v}(Z)) > 0 \\ 1 & \text{if } Re(G_{u,v}(Z)) \leq 0 \end{cases}$$

$$P_{u,v}^{Im}(Z) = \begin{cases} 0, & \text{if } Im(G_{u,v}(Z)) > 0 \\ 1 & \text{if } Im(G_{u,v}(Z)) \leq 0 \end{cases}$$

Où $Re(G_{u,v}(Z))$ et $Im(G_{u,v}(Z))$ sont les parties réelles et imaginaires des coefficients de Gabor respectivement.

La méthode d'encodage de Daugman, citée ci-dessus, peut être reformulée de la manière suivante :

$$P_{u,v}^{Re}(Z) = \begin{cases} 0, & \text{if } \theta_{u,v}(Z) \in \{I, IV\} \\ 1 & \text{if } \theta_{u,v}(Z) \in \{II, III\} \end{cases}$$

$$P_{u,v}^{Im}(Z) = \begin{cases} 0, & \text{if } \theta_{u,v}(Z) \in \{I, II\} \\ 1 & \text{if } \theta_{u,v}(Z) \in \{III, IV\} \end{cases}$$

Où $\theta_{u,v}(Z)$ est l'angle de phase de Gabor pour le pixel à la position Z .

Le QBC sépare les convoluées de Gabor dans 4 quadrants dans l'espace complexe, 00 pour le quadrant I , 10 pour le quadrant II , 11 pour le quadrant III et 01 pour le quadrant IV (Figure 33).

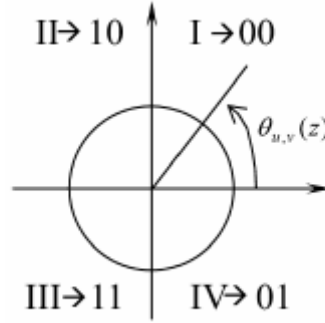


Figure 33 : QBC de phase de Gabor [91]

Maintenant à partir de QBC, on extrait les deux Patterns : GGPP (Global Gabor Phase Patterns) et LGPP (Local Gabor Phase Patterns).

b. GGPP (Global Gabor Phase Patterns)

Le pattern GGPP calcule une chaîne binaire pour chaque pixel en concaténant la partie réelle ou imaginaire du QBC de toutes les orientations pour une fréquence donnée, dans une image de même taille que l'image originale. La valeur GGPP, $GGPP_u(Z_0)$, pour la fréquence v à la position Z_0 dans une image, est définie par :

$$GGPP_v^{Re}(Z_0) = [P_{0,v}^{Re}(Z_0), P_{1,v}^{Re}(Z_0), \dots, P_{k,v}^{Re}(Z_0)]$$

$$GGPP_v^{Im}(Z_0) = [P_{0,v}^{Im}(Z_0), P_{1,v}^{Im}(Z_0), \dots, P_{k,v}^{Im}(Z_0)]$$

Où $GGPP_v^{Re}(Z_0)$ est la valeur du GGPP réel et $GGPP_v^{Im}(Z_0)$ est la valeur du GGPP imaginaire.

Dans [91] $k = 7$, et huit bits représentant huit orientations sont exploités qui forment un octet, pour représenter 256 modes d'orientation différents. Ces modes peuvent être facilement calculés par les équations suivantes :

$$GGPP_v^{Re}(Z_0) = P_{0,v}^{Re}(Z_0) * 2^k + P_{1,v}^{Re}(Z_0) * 2^{k-1} + \dots + P_{k,v}^{Re}(Z_0)$$

$$GGPP_v^{Im}(Z_0) = P_{0,v}^{Im}(Z_0) * 2^k + P_{1,v}^{Im}(Z_0) * 2^{k-1} + \dots + P_{k,v}^{Im}(Z_0)$$

Un exemple des Patterns GGPP en partie réelle et imaginaire est montré dans la Figure 34(a) et (b).

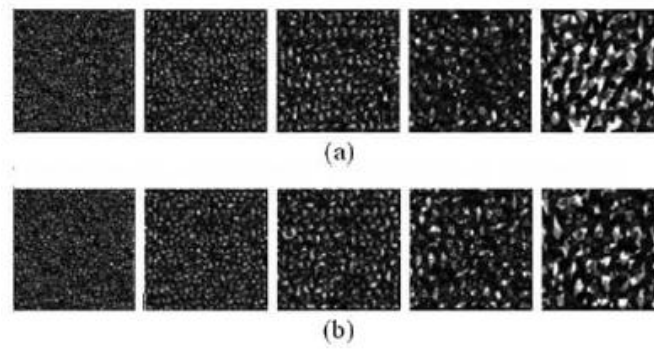


Figure 34 : Patterns GGPP (a) partie réelle (b) partie imaginaire [91]

c. LGPP (Local Gabor Phase Patterns)

Contrairement aux patterns GGPP qui encode des informations d'orientation, Zhang et al . proposent aussi un encodage des variations locales pour chaque pixel, appelé LGPP. Pour chaque orientation u et chaque fréquence v , la valeur de la partie réelle ou imaginaire de LGPP à la position Z_0 est calculée en encodant la différence de signe du pixel à la position Z_0 et ses 8 voisins par l'opérateur LXP (local XOR pattern):

$$LGPP_{u,v}^{Re}(Z_0) = [P_{u,v}^{Re}(Z_0) XOR P_{u,v}^{Re}(Z_1), P_{u,v}^{Re}(Z_0) XOR P_{u,v}^{Re}(Z_2), \dots, P_{u,v}^{Re}(Z_0) XOR P_{u,v}^{Re}(Z_8)]$$

$$LGPP_{u,v}^{Im}(Z_0) = [P_{u,v}^{Im}(Z_0) XOR P_{u,v}^{Im}(Z_1), P_{u,v}^{Im}(Z_0) XOR P_{u,v}^{Im}(Z_2), \dots, P_{u,v}^{Im}(Z_0) XOR P_{u,v}^{Im}(Z_8)]$$

Où $Z_i, i = 1, 2, \dots, 8$, sont les huit voisins du pixel Z_0 , et XOR désigne l'opérateur OU Exclusif. Cette procédure est illustrée dans la Figure 35.

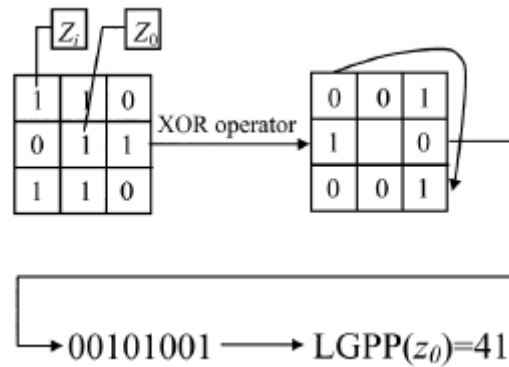


Figure 35 : Illustration de l'opérateur LXP [91]

Similaire aux patterns GGPP, huit bits qui représentent huit voisins forment un octet (255 niveaux de gris) pour chaque pixel. Un exemple de 40 images de partie réelle du pattern LGPP, ainsi pour la partie imaginaire, est illustré dans la Figure 36.

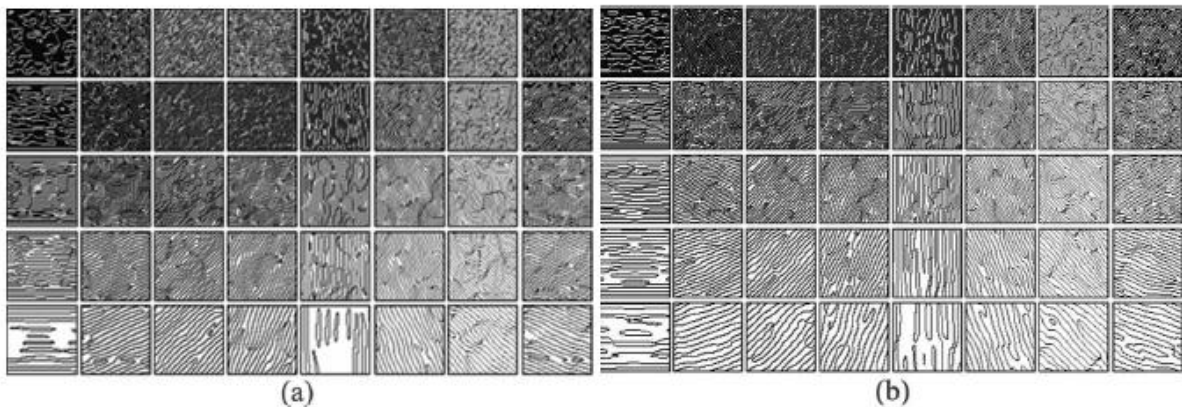


Figure 36 : Patterns LGPP (a) partie réelle (b) partie imaginaire [91]

d. HGPP (Histogramme de Gabor Phase Patterns)

Dans la méthode de Daugman pour la reconnaissance d'iris, les QBC sont utilisés directement pour former la représentation d'une image d'iris, et la classification est réalisée par une distance de Hamming. Toutefois, pour cinq fréquences et huit orientations, la méthode d'encodage GPP proposée donnera 90 images (5 GGPPs réels, 5 GGPPs imaginaires, 40 LGPPs réels et 40 LGPPs imaginaires), qui ont la même taille que l'image originale. Pour les modéliser efficacement, les auteurs exploitent les histogrammes spatiaux pour modéliser les GPPs encodés (GGPP et LGPP), appelés Histogrammes de Gabor Phase Patterns (HGPP). La raison du choix d'une méthode à base d'histogramme réside dans les observations que les

GPPs ressemblent à des images ayant une texture structurée (voir Figure 34 et Figure 36), en plus, l'histogramme spatial est un bon outil pour représenter des images de texture.

Si on calcule les histogrammes directement sur une image entière, les détails locaux seront perdus. Les auteurs proposent de diviser l'image originale en 64 sous-régions (Figure 37(a)) et de calculer les histogrammes HGPP de toutes ces sous-régions. En effet, chaque histogramme est un micro-pattern, en 16 bins.

Ces HGPPs sont formulés comme suit:

$$HGPP = (H_{GGPP}^{Re}, H_{GGPP}^{Im}, H_{LGPP}^{Re}, H_{LGPP}^{Im})$$

Où H_{GGPP}^{Re} , H_{GGPP}^{Im} , H_{LGPP}^{Re} , H_{LGPP}^{Im} sont, respectivement, les histogrammes de la sous-région de la partie réelle du GGPP, les histogrammes de la sous-région de la partie imaginaire du GGPP, les histogrammes de la sous-région de la partie réelle du LGPP et les histogrammes de la sous-région de la partie imaginaire du LGPP, et ils sont donnés par :

$$H_{GGPP}^{Re} = (H_{GGPP}^{Re}(v, l): v = 0, \dots, 4; l = 1, \dots, L)$$

$$H_{GGPP}^{Im} = (H_{GGPP}^{Im}(v, l): v = 0, \dots, 4; l = 1, \dots, L)$$

$$H_{LGPP}^{Re} = (H_{LGPP}^{Re}(u, v, l): u = 0, \dots, 7; v = 0, \dots, 4; l = 1, \dots, L)$$

$$H_{LGPP}^{Im} = (H_{LGPP}^{Im}(u, v, l): u = 0, \dots, 7; v = 0, \dots, 4; l = 1, \dots, L)$$

Où L est le nombre de sous-régions divisées pour le calcul de l'histogramme.

Le diagramme général de calcul du HGPP est présenté dans la Figure 37(b). Selon Zhang et al. les images doivent être normalisées en 128×128 pixels avant la convolution avec les filtres de Gabor 2D afin d'avoir le meilleur résultat.

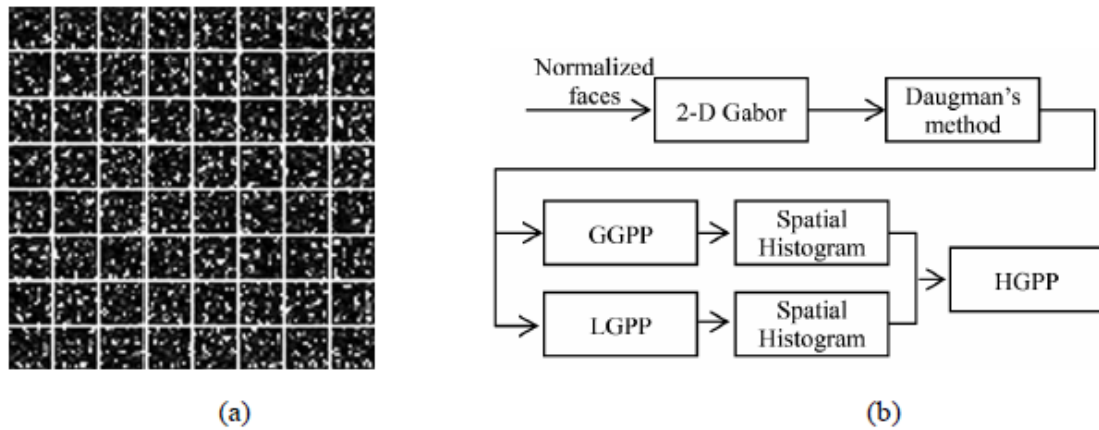


Figure 37 : (a) 8×8 sous-régions de LGPP (b) diagramme général de HGPP [91]

3.2. GIST

Dans le domaine de la vision par ordinateur, le descripteur GIST est une représentation d'une image en basse dimension qui contient suffisamment d'informations pour identifier une scène. Il a été introduit par les chercheurs Oliva et Torralba dans [55, 56] dans le but d'une classification automatique des images. Ces deux auteurs ont défini le descripteur GIST de l'image en analysant les fréquences spatiales et les orientations. Il est issu d'une suite de travaux de recherche à la fois psychologiques et informatiques sur la classification automatique de scènes. Il y a plusieurs présentations différentes de ce descripteur. Nous nous basons ici sur une étude du code source disponible sur la page web de Torralba (<http://people.csail.mit.edu/torralba/code/spatialenvelope/>).

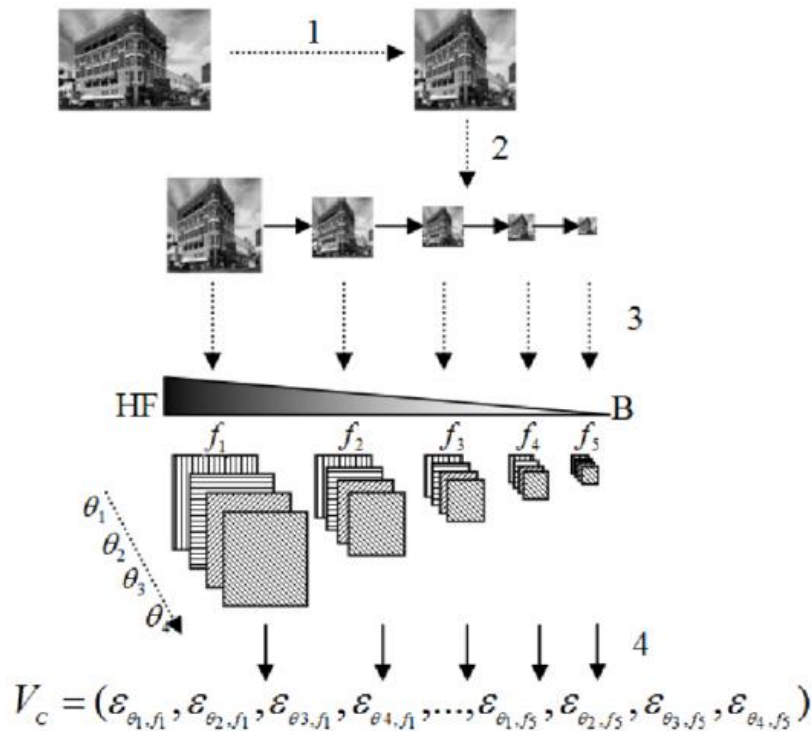


Figure 38 : Principe du descripteur GIST [58]

GIST permet d'extraire Une enveloppe spatiale globale qui correspond aux différentes fréquences et orientations contenues dans l'image, d'où vient le critère de globalité de ce descripteur. Du fait de la description globale de l'information essentielle dans l'image, il est possible de ne pas conserver les détails d'une image et donc n'identifier que les fréquences et les orientations principales de celle-ci dans l'objectif de la classer.

D'après Torralba et Oliva, le descripteur GIST peut être considéré comme une bonne solution pour des problèmes de catégorisation de scènes [56]. Le succès du descripteur GIST s'est rapidement propagé vers d'autres applications, d'abord il a été réutilisé efficacement pour résoudre des problèmes de reconnaissance de lieux [50], ensuite il a été appliqué dans le domaine de la recherche d'image à l'échelle du web [22].

Le descripteur GIST se calcule suivant les étapes montrées dans la Figure 38 :

- D'abord, les images d'entrée sont réduites en imageries carrées, d'une taille comprise entre 32×32 et 128×128 , quelles que soient leur proportions.
- Ensuite elles sont convoluées avec un banc de filtres de Gabor avec N_θ orientations et N_σ échelles.

- Les images en sortie du filtre sont divisées en une grille régulière de $M \times M$ régions sur lesquelles on calcule la moyenne.
- Enfin, le descripteur GIST de l'image est obtenu par concaténation des descripteurs des différentes régions.

Dans le cas des images couleurs, ces étapes sont répétées pour chaque canal.

La taille du vecteur descripteur final est $M^2 \times N_\theta \times N_\sigma$ (x3 dans le cas d'une image couleur), indépendante de la taille de l'image. On choisit souvent $N_\theta = 8$ orientations et $N_\sigma = 4$ échelles. Enfin, une analyse en composantes principales (ACP) est utilisée pour réduire la dimension des vecteurs.

3.3. Les moments géométriques

Les moments géométriques sont utilisés pour caractériser l'intégralité de la forme d'une région à l'aide de propriétés statistiques. Ils étaient utilisés dans un premier temps globalement, puis localement en limitant le calcul des moments au voisinage d'un point d'intérêt. Ces moments géométriques ont été introduits et appliqués pour la première fois en tant qu'approche pour la description d'images dans [39]. En général, ils sont définis par :

$$m_{p,q}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

Où $p + q$ est l'ordre du moment. $m_{0,0}$ est le moment d'ordre 0 et représente l'aire de la forme de l'objet dans l'image. Ce dernier est associé aux moments d'ordre 1 $m_{0,1}$ et $m_{1,0}$ afin de calculer le centre de gravité de l'objet dont les coordonnées sont données par :

$$x_c = \frac{m_{1,0}}{m_{0,0}} \quad \text{et} \quad y_c = \frac{m_{0,1}}{m_{0,0}}$$

D'une manière plus courante, on utilise les moments centraux qui ont la propriété importante d'être invariants par translation, ils sont calculés à base des coordonnées du centre de gravité de l'objet et sont notés par $\mu_{p,q}$:

$$\mu_{p,q}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y) dx dy$$

A partir de ces moments géométriques, et dans le but de les utiliser pour la description de forme en vue d'une classification ou d'une indexation, Hu [39] a proposé d'utiliser une succession de sept polynômes calculés à partir des moments centraux et sont donnés par les équations suivantes :

$$I_1 = \mu_{20} + \mu_{02}$$

$$I_2 = (\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2$$

$$I_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2$$

$$I_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2$$

$$I_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12}[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2]) + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03}[3(\mu_{30} + \mu_{12})^2 - (\mu_{12} + \mu_{03})^2])$$

$$I_6 = (\mu_{20} + \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})^2(3\mu_{21} + \mu_{03})$$

$$I_7 = (3\mu_{21} - \mu_{03})(\mu_{30} - 3\mu_{12})[(\mu_{30} + \mu_{12})^2 - (3\mu_{21} + \mu_{03})^2] - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$$

En plus du fait qu'ils soient invariants aux translations, rotations et changement d'échelle, ces moments géométriques sont très simples et rapides à calculer. Hu montre qu'il est suffisant d'utiliser les deux premiers invariants I_1 et I_2 pour séparer les 26 caractères alphabétiques, cela montre bien qu'il n'est pas nécessaire d'utiliser tous les moments de l'ensemble de Hu en vue d'une classification d'objets.

Par ailleurs ces descripteurs sont assez sensibles aux bruits et aux déformations. En plus, ils ne sont ni orthogonaux ni complets, ce qui fait, qu'ils souffrent d'un niveau élevé de redondance d'informations. Un autre inconvénient de ces descripteurs, c'est qu'ils doivent être dérivés à la main, au contraire des moments orthogonaux tels que les moments invariants de Zernike qui peuvent être calculés à des ordres arbitraire.

3.4. Les moments orthogonaux

A l'inverse des moments géométriques qui sont définis par rapport à une base quelconque, les moments orthogonaux sont définis dans une base orthogonale, ce qui résout le problème

de la redondance des informations portées par chacun des moments. Les moments orthogonaux les plus utilisés en vision par ordinateur sont ceux de Zernike, introduits par Teague [76] qui a proposé d'utiliser une série de polynômes complexes et orthogonaux définis sur le disque unitaire $x^2 + y^2 \leq 1$, qu'on peut écrire sous la forme :

$$V_{nm}(x, y) = V_{nm}(r \cos \theta, \sin \theta) = R_{nm}(r)e^{jm\theta}$$

$$\text{Avec } \begin{cases} n = 0, 1, 2, \dots; 0 \leq |m| \leq n \\ n - |m| \text{ est pair} \end{cases}$$

Et $R_{nm}(r)$ le polynôme orthogonal défini par :

$$R_{nm}(r) = \sum_{s=0}^{\frac{(n-|m|)}{2}} (-1)^s \frac{(n-s)!}{s! \times \left(\frac{n-2s+|m|}{2}\right)! \left(\frac{n-2s-|m|}{2}\right)!} r^{n-2s}$$

La Figure 39 ci-dessous présente une pyramide des polynômes de Zernike jusque quatrième ordre.

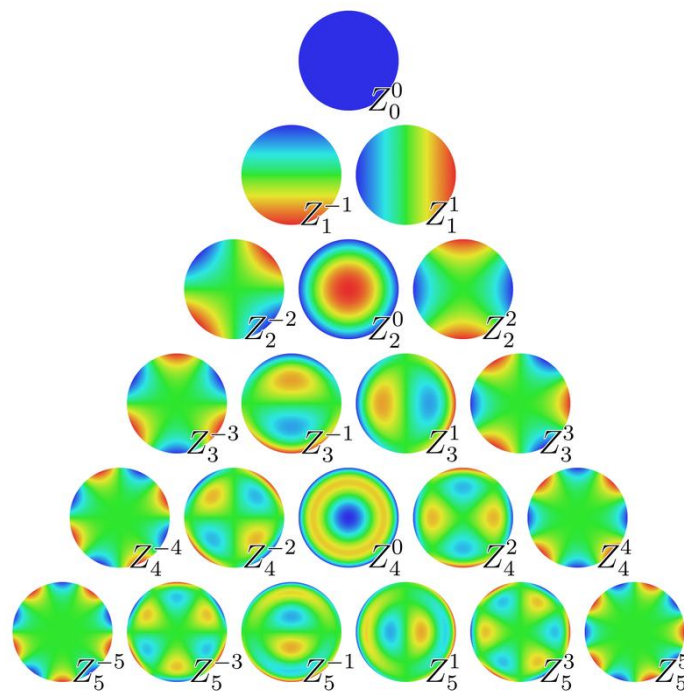


Figure 39 : Les polynômes de Zernike au 5 premiers ordres (wikipédia.org)

Les moments de Zernike $Z_f(n, m)$ d'une image sont construits par les projections de cette image f sur la base de ces polynômes $\{V_{nm}(r, \theta)\}$. Le moment de Zernike d'ordre n et de répétition m de l'image $f(x, y)$ peut être donné par :

$$Z_f(n, m) = \frac{n+1}{\pi} \sum_r \sum_{\theta} f(r \cos \theta, r \sin \theta) r \cdot R_{nm}(r) e^{jm\theta} \quad r \leq 1$$

Ces moments possèdent certains avantages dont :

- Les modules des moments de Zernike sont invariants par rotation [15, 87],
- Ils sont robustes aux altérations telles que le bruit, changement de luminosité, etc. (voir section 2.3.6),

Une des limitations de ce descripteur, c'est que la complexité de calcul des polynômes de Zernike augmente fortement dans les hautes fréquences de l'image qui sont principalement codées dans les moments d'ordres supérieurs.

3.5. Les descripteurs de Fourier

3.5.1. Les descripteurs de Fourier-Mellin

Les descripteurs de Fourier-Mellin sont construits principalement à base de la transformée de Fourier Mellin (TFM) qui est très utilisée en traitement d'images [68, 69, 20, 21]. Pour une image en niveaux de gris I de \mathbb{R}^2 dans \mathbb{C} passée dans l'espace polaire, la TFM est définie par :

$$\forall (u, v) \in \mathbb{Z} \times \mathbb{R}, M_I(u, v) = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} I(r, \theta) r^{-iv} e^{-iu\theta} d\theta \frac{dr}{r}$$

Où I est supposée intégrable sur le groupe des similitudes planes vectorielles (rotation, translation et homothétie) $G = \mathbb{R}_+^* \times S^1$, avec S^1 est le cercle unité de \mathbb{R}^2 .

La TFM a été définie dans le but de mettre en œuvre des approches invariantes aux similitudes directes, ces approches ne prennent en considération que le module de la TFM, par contre, la phase est souvent non exploitée, donc une représentation partielle des objets, ce qui est insuffisant pour discriminer des objets surtout dans le cas des grandes bases d'images. Afin de résoudre ce problème, Derrode et al, ont présenté un procédé de

normalisation de la transformée de Fourier-Mellin analytique (TFMA) permettant de mettre en œuvre des familles complètes de descripteurs invariants par similitude.

La TFMA a été introduite dans [33] pour pallier le problème de divergence des intégrales lorsqu'elles sont calculées sur des images en niveaux de gris. Celle-ci est définie par :

$$\forall (u, v) \in \mathbb{Z} \times \mathbb{R}, M_{I_\sigma}(u, v) = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} I_\sigma(r, \theta) r^{-iv} e^{-iu\theta} d\theta \frac{dr}{r}$$

Avec $\sigma > 0$ un nombre réel strictement positif et $(r, \theta) \in \mathbb{R}_+^* \times S^1$ où S^1 désigne le cercle unité. $I_\sigma(r, \theta) = r^\sigma I(r, \theta)$ et I_σ est intégrable sur le groupe de similitudes G .

La TFMA admet une transformée inverse ce qui permet de reconstruire des images depuis son spectre (Figure 40). Elle s'exprime par la formule suivante :

$$\forall (r, \theta) \in \mathbb{R}_+^* \times S^1, I(r, \theta) = \int_{-\infty}^{+\infty} \sum_{u \in \mathbb{Z}} M_{I_\sigma}(u, v) r^{-\sigma+iv} e^{iu\theta} dv$$

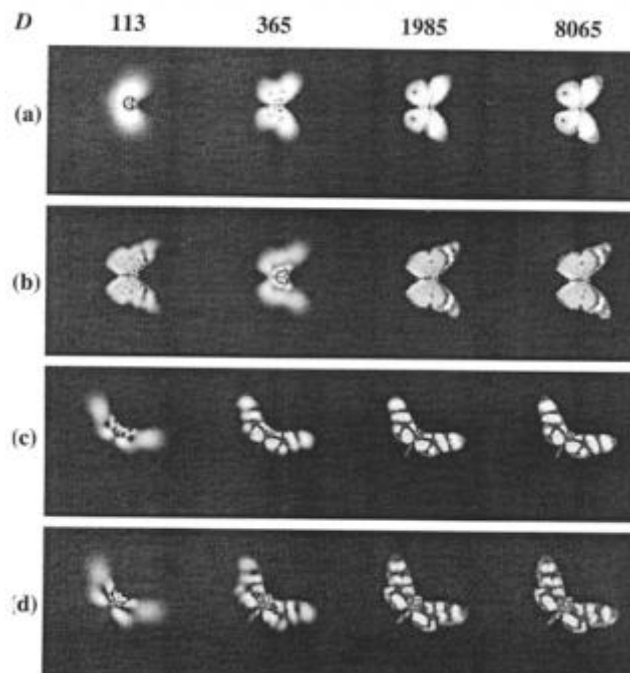


Figure 40 : Reconstruction des formes de quatre images de papillons en fonction du nombre de descripteurs invariants utilisés pour la reconstruction (TFMA) [33]

La famille complète des descripteurs invariants par similitude issue de la TFMA a été proposé dans [33], et peut être réécrite facilement et appliquée pour toute valeur strictement positive de σ :

$$\forall (u, v) \in \mathbb{Z} \times \mathbb{R}, F_{I_\sigma}(u, v) = M_{I_\sigma}(0,0) \frac{-\sigma+iv}{\sigma} e^{iu \text{Arg}(M_{I_\sigma}(1,0))} M_{I_\sigma}(u, v)$$

Partant de ces invariants, si $g(r, \theta) = I(\alpha r, \theta + \beta)$, il peut être montré facilement que $F_g(u, v) = F_I(u, v)$.

De plus, cette famille d'invariants est complète car il est possible de reconstruire la TFMA d'un objet à partir des valeurs des descripteurs invariants $F_{I_\sigma}(u, v)$ et des deux paramètres de normalisation $M_{I_\sigma}(0,0)$ et $M_{I_\sigma}(1,0)$. On peut alors reconstruire l'objet de l'image originale en utilisant la TFMA inverse. Dans [33], il est prouvé également que cette famille d'invariants est convergente au sens L^2 . Ces deux propriétés ont pour conséquence de définir une vraie distance dans l'espace des formes donnée par :

$$d_2(F_{I_\sigma}, F_{g_\sigma}) = \left(\int_{-\infty}^{+\infty} \sum_{u \in \mathbb{Z}} |F_{I_\sigma}(u, v) - F_{g_\sigma}(u, v)|^2 \right)^{\frac{1}{2}}$$

La distance $d_2 = 0$ pour des objets de même forme et elle utilisée comme critère de mesure de degré de similarité entre deux objets indépendamment de leur orientation, leur taille et leur pose.

3.5.2. Les descripteurs de Fourier Génériques

La transformée de Fourier a été largement utilisée dans le traitement et l'analyse d'images. L'avantage d'analyser une image dans le domaine fréquentiel est qu'il est facile de pallier au problème du bruit qui est fréquent dans les images numériques. Cependant, il n'est pas recommandé de calculer des descripteurs utilisant directement la transformée de Fourier, car ces derniers ne seront pas invariant par rotation. En outre, les descripteurs acquis ne sont pas compacts (Figure 41(a)(b)). Densheng Zhang et Guojun Lu [92, 93] ont contourné ce problème en proposant un descripteur de Fourier générique (GFD), celui-ci est basé sur une transformée de Fourier modifiée polairement (MPFT), c.à.d. en traitant l'image polaire dans

l'espace polaire comme une image rectangulaire à deux dimensions dans l'espace cartésien (Figure 41).

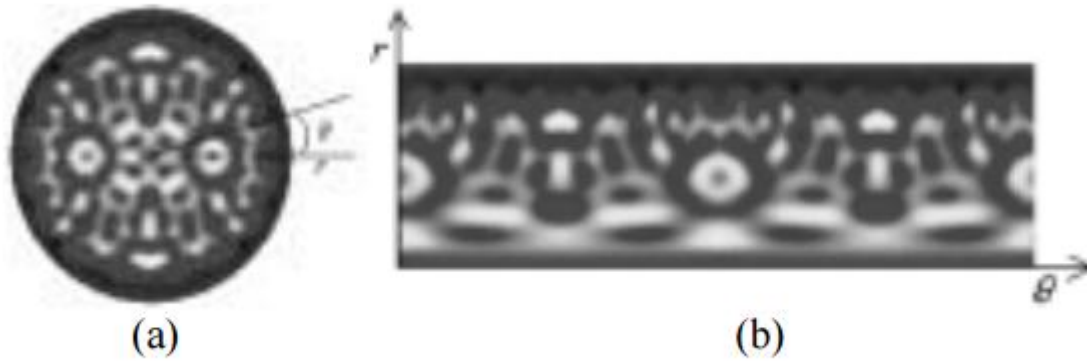


Figure 41 : (a) l'image originale dans l'espace polaire; (b) l'image polaire de (a) dans tracée dans l'espace cartésien [92]

L'image polaire de la Figure 41(b) est l'image rectangulaire. Par conséquent, si nous appliquons une transformé de Fourier 2D sur cette image rectangulaire, la FT polaire aura une forme similaire à celle d'une FT discrète classique dans l'espace cartésien. Donc, pour une image donnée $f(x, y)$, la transformé de Fourier modifiée polairement (MPFT) est définie par :

$$PF(\rho, \phi) = \sum_r \sum_i f(r, \theta_i) e^{j2\pi(\frac{r}{R}\rho + \frac{2\pi i}{T}\phi)}$$

Où $0 \leq r = \sqrt{(x - x_c)^2 + (y - y_c)^2} < R$ et $\theta_i = i(\frac{2\pi}{T})$ avec $(0 \leq i < T)$; (x_c, y_c) sont les coordonnées du centre de gravité de la forme. $(0 \leq \rho < R)$, $(0 \leq \phi < T)$ avec R et T sont les résolutions radiales et angulaires. La signification physique de ρ et ϕ est claire, ce sont respectivement la ρ^{eme} fréquence radiale et la ϕ^{eme} fréquence angulaire. La détermination du nombre de ρ et de ϕ pour la description des formes est physiquement réalisable, parce que les caractéristiques de la forme sont normalement extraites par les quelques basses fréquences. Les Figure 42(c)(d) montrent des images polaires et les spectres de Fourier polaire des deux motifs des Figure 42(a)(b). On peut observer sur la Figure 42(c)(d) que la rotation du motif dans l'espace cartésien se traduit par un shift (ou décalage) circulaire dans l'espace polaire. Le décalage circulaire ne modifie pas la répartition spectrale dans l'espace polaire. Le spectre de Fourier polaire est plus concentré autour de l'origine de l'espace polaire. Ceci est particulièrement bien adapté pour la représentation des formes, car pour

représenter une forme efficacement, le nombre de caractéristiques spectrales sélectionnées pour la décrire, ne devrait pas être grand. Le spectre est circulairement symétrique, on peut ne considérer qu'un quart des caractéristiques spectrales pour décrire une forme. Les coefficients de Fourier polaire acquis sont invariants par translation.

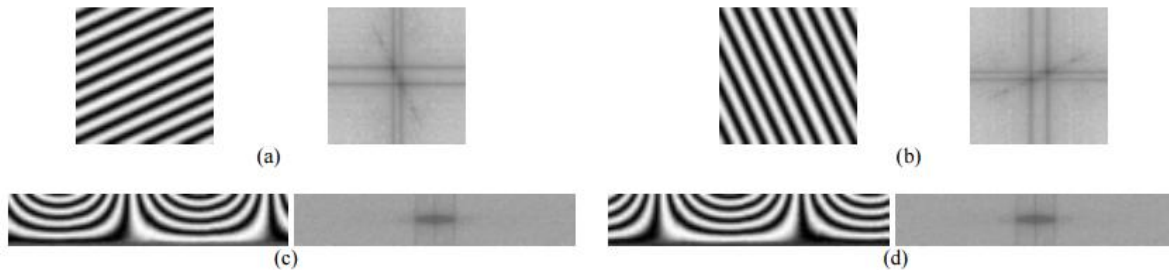


Figure 42 : (a) un motif et son spectre de Fourier; (b) le motif pivoté de (a) et son spectre de Fourier; (c) l'image polaire de (a) et son spectre de Fourier; (d) l'image polaire de (b) et son spectre de Fourier [92]

L'invariance par rotation et par changement d'échelle est atteinte par la normalisation suivante :

$$GFD = \left\{ \frac{|PF(0,0)|}{area}, \frac{|PF(0,1)|}{|PF(0,0)|}, \dots, \frac{|PF(0,n)|}{|PF(0,0)|}, \dots, \frac{|PF(m,0)|}{|PF(0,0)|}, \dots, \frac{|PF(m,n)|}{|PF(0,0)|} \right\}$$

Où *area* est la surface du cercle englobant la forme d'intérêt, *m* et *n* sont respectivement le nombre maximal des fréquences radiales et angulaires sélectionnées.

Pour une représentation efficace d'objets, seulement un petit nombre de caractéristiques GFD sont sélectionnées. Dans l'implémentation de Zhang et Lu, 36 caractéristiques GFD reflétant 4 fréquences radiales et 9 fréquences angulaires sont sélectionnées pour indexer l'objet. Les caractéristiques GFD sélectionnées forment un vecteur descripteur qui est utilisé pour indexer un objet. Pour deux formes représentées par leurs descripteurs GFD, la similarité entre ces deux formes est mesurée par la distance *City Block* entre leurs GFD. Par conséquent, la mise en correspondance en ligne est simple et efficace.

Zhang et Lu ont comparé leur descripteur GFD avec celui basé sur les moments de Zernike, ils ont pu conclure que : le GFD est plus simple à calculer, les attributs sont purement spectraux et montrent de meilleures performances de recouvrement car l'analyse multi-

résolution dans les deux directions radiale et spectrale, est possible. Enfin, les auteurs ont montré aussi que le GFD donnait de meilleurs résultats que les moments de Zernike.

3.5.3. Les descripteurs de Fourier Généralisés

Les premiers descripteurs de Fourier généralisés ont été introduit par Gauthier et al dans [29], ils sont construits à partir des actions de groupe et plus particulièrement du groupe de déplacement $SE(2)$, ce groupe est constitué des translations et des rotations dans le plan. Dans [74], les auteurs ont proposé un second ensemble de descripteurs issu également des actions du groupe de déplacement $SE(2)$, et qui peuvent être utilisés de manière souple et robuste pour la reconnaissance de forme basée sur des images couleur, en les associant avec des SVM, ils ont aussi proposé une implantation matérielle qui permet d'accélérer le calcul de leurs descripteurs.

a. Les premiers invariants

Ces invariants sont de type « densité spectrale » et sont définis comme suit : Soit f une fonction carrée sommable sur le plan, et \hat{f} sa transformée de Fourier définie par l'équation suivante :

$$\hat{f}(\xi) = \int_{\mathbb{R}^2} f(x) e^{-i\langle x, \xi \rangle} dx$$

où $\langle ., . \rangle$ est le produit scalaire sur \mathbb{R}^2 .

Si (λ, θ) sont les coordonnées polaires du point ξ , nous noterons à nouveau par $\hat{f}(\lambda, \theta)$ la transformée de Fourier de f au point (λ, θ) . Dans [29], Gauthier et al ont défini les premiers descripteurs de Fourier Généralisé (DF1) par :

$$I_1^r(f): \mathbb{R}_+ \rightarrow \mathbb{R}_+$$

$$r \rightarrow I_1^r(f)$$

$$I_1^r(f) = \int_0^{2\pi} |\hat{f}(r, \theta)|^2 d\theta$$

Dans le domaine discret, ces descripteurs sont remplacés par un ensemble fini de valeurs formant les composantes d'un vecteur qu'il est possible d'utiliser à l'entrée d'un processus de classification.

Les descripteurs de Fourier I_1^r calculés selon l'équation précédente ont des propriétés élémentaires cruciales pour la reconnaissance d'objets. Ils sont invariants par déplacement et réflexion :

- Si M est un déplacement tel que $g = f \circ M$, alors $I_1^r(f) = I_1^r(g)$, $\forall r \in \mathbb{R}_+$
- S'il existe une réflexion \mathfrak{R} tel que $g = f \circ \mathfrak{R}$, alors $I_1^r(f) = I_1^r(g)$, $\forall r \in \mathbb{R}_+$
- Ces descripteurs sont aussi covariants par homothétie :

Si k est une constante réelle tel que $g(x) = f(kx)$ pour tout $x \in \mathbb{R}^2$, $I_1^r(g) = \frac{1}{k^4} I_1^{\frac{r}{k}}(f)$, $\forall r \in \mathbb{R}_+$

b. Les invariants complets

Les premiers invariants cités précédemment et introduits par Gauthier dans [29] ne sont pas complets (i.e. deux images différentes peuvent avoir le même DF1), mais ceci n'est pas un inconvénient car ils sont suffisant pour une large classe d'applications. Dans [74], Smach et al. définissent une seconde famille d'invariants (contenant également les premiers déjà cités) qui sont complets dans les cas des groupes commutatifs et dans une large classe de fonctions, dans le cas des groupes compacts. Ces invariants sont de type « décalage de phase », notés par I^{ξ_1, ξ_2} et définis par :

$$I^{\xi_1, \xi_2}(f) = \int_{S_1} \hat{f}[R_\theta(\xi_1 + \xi_2)] \overline{\hat{f}[R_\theta(\xi_1)]} \overline{\hat{f}[R_\theta(\xi_2)]} d\theta$$

Avec $\xi_1, \xi_2 \in \mathbb{R}^2$ et $R_\theta(\xi)$ désigne l'action de S_1 dans \mathbb{R}^2 définie par :

$$R_\theta(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta),$$

À noter que :

- Il est clair que I^{ξ_1, ξ_2} est invariant par déplacement.
- Il est également clair que l'ensemble des invariants I^{ξ_1, ξ_2} est complètement déterminé par le plus petit ensemble obtenu en prenant ξ_1 sous la forme $(0, r_1)$, $r_1 \in \mathbb{R}^+$.

Par conséquent, une autre définition de I^{ξ_1, ξ_2} est donnée par :

$$I_f^\omega(\lambda_1, \lambda_2) = \int_{S_1} [\hat{f}(-\lambda_1 \sin(\theta + \omega) - \lambda_2 \sin \theta, \lambda_1 \cos(\theta + \omega) + \lambda_2 \cos \theta) \times \overline{\hat{f}}(-\lambda_1 \sin(\theta + \omega), \lambda_1 \cos(\theta + \omega)) \times \overline{\hat{f}}(-\lambda_2 \sin(\theta), \lambda_2 \cos(\theta))] d\theta$$

Où $\lambda_1, \lambda_2 \in \mathbb{R}^+$ et $\omega \in [0, 2\pi[$.

Cet ensemble d'invariants stables, nommés DF2, est adapté au groupe $SE(2)$ des déplacements dans le plan, et la complétion de cette famille est assurée dans le groupe $SE(2, N)$ des déplacements d'angle $\frac{2k\pi}{N}$. Ceci permet de garantir que deux objets différents auront des DF2 différents, et par conséquent, donner une raison solide à leur utilisation en tant que descripteurs discriminants pour la reconnaissance de formes.

4. Conclusion

Dans ce chapitre, nous avons fait l'inventaire des différentes techniques d'extraction de caractéristiques. Ainsi, les propriétés de chaque descripteur d'images ont été étudiées pour montrer leurs différentes invariances. De plus, un soin particulier a été apporté à l'étude des avantages et des inconvénients.

Si on considère tout d'abord les descripteurs locaux, on peut remarquer en premier temps que ces descripteurs sont extraits de manière fiable autour de points d'intérêts préalablement détectés. L'efficacité de ces descripteurs a été démontrée dans de multiples travaux, mais dans certains contextes, tel que la présence d'un bruit important, les descripteurs globaux sont les plus efficaces.

Si toutes ces méthodes présentent des qualités et des défauts, nous allons devoir développer de nouvelles méthodes plus adaptées à notre objectif de reconnaissance d'objet, et c'est l'objet du prochain chapitre.

Chapitre 3 : Descripteurs de Fourier généralisés et rotationnels

1. Introduction

La méthode que nous proposons est un prolongement des travaux de Smach et al. [74] qui ont permis de définir un ensemble de descripteurs de Fourier Généralisés (GFD). Comme nous le verrons dans leur définition, les descripteurs de Fourier Généralisés munis d'un relèvement cyclique dans les groupes de roto-translations discrétisées $SE(2, N)$, sont invariants en translation et en rotation. Cette approche a guidé la définition de notre nouvel ensemble de descripteurs de Fourier dits « rotationnels » dont le calcul réel sur des images relevées (en considérant plutôt un relèvement invariant à gauche) nécessite N fois moins de temps de calcul et d'espace par rapport aux descripteurs de Fourier généralisés proposés dans [74].

Ce chapitre est organisé comme suit : Dans la section suivante, nous présentons quelques notions préliminaires sur l'analyse harmonique (i.e. l'étude de la transformée de Fourier et ses propriétés) sur les groupes abéliens et non-abéliens localement compacts. La deuxième section concerne les invariants spectraux et bi-spectraux, ainsi que leurs propriété de complétude dans les groupes abéliens localement compacts et dans le groupe dérivé de produit semi-direct $SE(2, N)$ défini dans la section 2.3 par $SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{Z}_N$. Dans la troisième section, nous introduisons les descripteurs de Fourier rotationnels basés sur la structure du cortex visuel humain primaire V1. Enfin, pour clore ce chapitre, nous présentons quelques techniques pour le calcul pratique de ces descripteurs.

2. Préliminaires

Pour définir et calculer les descripteurs de Fourier spécifiques à notre étude, nous introduisons quelques généralités sur la transformée de Fourier dans les groupes non-commutatifs. Nous nous référons à [63, 37] pour une introduction générale du sujet.

2.1. Analyse harmonique sur les groupes abéliens localement compacts

Soit G un groupe abélien localement compact. Un *caractère* λ de G est un homomorphisme continu de groupes $\lambda : G \rightarrow \mathbb{C}$ tel que $|\lambda(a)| = 1$ pour tout $a \in G$. En définissant le produit de deux caractères comme une multiplication point par point, et l'inverse comme le conjugué complexe, l'ensemble

$$\hat{G} = \{\lambda \mid \lambda \text{ est un caractère de } G\},$$

muni de la topologie de la convergence uniforme sur tout compact, est un groupe abélien localement compact, appelé le *groupe dual (de Pontryagin)* de G .

Soit l'application $\Omega : G \rightarrow \hat{\hat{G}}$ définie par

$$\Omega_a(\lambda) := \lambda(a) \quad \text{pour tout } \lambda \in \hat{G} \text{ et } a \in G.$$

Ceci est un homomorphisme continu de groupes, et donc $G \subset \hat{\hat{G}}$.

Théorème 2.1.1 (Dualité de Pontryagin). *L'application Ω est un isomorphisme de groupe, et par conséquent G est canoniquement isomorphe au dual de \hat{G} .*

Grâce au théorème ci-dessus, la transformée de Fourier permet de réaliser cet isomorphisme au niveau des fonctions à valeurs complexes définies sur G et \hat{G} . A savoir, on peut munir G de sa mesure de Haar³ et pour tout $f \in L^2(G) \cap L^1(G)$ définir sa transformée de Fourier $\hat{f} \in L^2(\hat{G})$ par

$$\hat{f}(\lambda) := \int_G f(x) \bar{\lambda}(x) dx.$$

Avec :

$$L^2(G) = \{f : G \rightarrow \mathbb{C} \text{ (ou } \mathbb{R}) \mid \int |f(g)|^2 dg < \infty\},$$

$$L^1(G) = \{f : G \rightarrow \mathbb{C} \text{ (ou } \mathbb{R}) \mid \int |f(g)| dg < \infty\}.$$

Notons en particulier que, si l'on définit $\text{avg } f = \int_G f(x) dx$, alors on a : $\text{avg } f = \hat{f}(\hat{\delta})$, où $\hat{\delta}(x) = 1$ est l'identité de \hat{G} . Nous avons le théorème suivant.

³ Une mesure de Haar sur un groupe localement compact G est une mesure de Borel quasi-régulière non-nulle λ invariante par translation à gauche. En particulier, pour tout ensemble Borélien B de G et pour tout g dans G , on a : $\lambda(gB) = \lambda(B)$.

Théorème 2.1.2 (Théorème de Plancherel). *Il existe une mesure unique $d\lambda$ sur \hat{G} , appelée mesure de Plancherel, de sorte que la transformée de Fourier définie ci-dessus peut être étendue à une isométrie $F : L^2(G) \rightarrow L^2(\hat{G})$. En particulier, lorsque $f \in L^2(G) \cap L^1(G)$ et $\hat{f} \in L^2(\hat{G}) \cap L^1(\hat{G})$, on a :*

$$F^{-1}(\hat{f})(x) = \int_{\hat{G}} \hat{f}(\lambda) \lambda(x) d\lambda.$$

Remarque 2.1.3. *Lorsque $G = \mathbb{R}$, la procédure ci-dessus nous donne la transformée de Fourier classique.*

$$\hat{f}(\lambda) = \int_{\mathbb{R}} f(x) e^{-2\pi i \lambda x} dx \quad f \in L^2(\mathbb{R}).$$

En effet, la mesure de Haar sur \mathbb{R} est la mesure de Lebesgue, $\hat{\mathbb{R}} \cong \mathbb{R}$ peut être réalisée comme l'ensemble des $x \rightarrow e^{2\pi i \lambda x}$ pour $\lambda \in \mathbb{R}$, et la mesure de Plancherel sur \mathbb{R} devient la mesure de Lebesgue, i.e. la transformée de Fourier inverse peut être réécrite comme suit :

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(\lambda) e^{2\pi i \lambda x} dx \quad \hat{f} \in L^2(\hat{\mathbb{R}}).$$

Considérons la représentation régulière à gauche de G , i.e. l'opérateur de translation $x \rightarrow \tau_x \in \mathcal{U}(L^2(G))$ définie comme $\tau_x f(y) = f(y - x)$, la propriété fondamentale de la transformée de Fourier pour nos besoins est la suivante.

Théorème 2.1.4. *Pour tout $f, g \in L^2(G)$ et tout $x \in G$, on a :*

$$f = \tau_x g \Leftrightarrow \hat{f}(\lambda) = \bar{\lambda}(x) \hat{g}(\lambda) \quad \forall \lambda \in \hat{G}.$$

2.2. La transformée de Fourier sur les groupes non-commutatifs localement compacts

Soit maintenant G un groupe localement compact unimodulaire⁴, non nécessairement abélien. Une *représentation unitaire* T de G est un homomorphisme continu $T : G \rightarrow \mathcal{U}(\mathcal{H}_T)$, où \mathcal{H}_T est un espace de Hilbert complexe (éventuellement de dimension infinie)

⁴ Un groupe est unimodulaire si et seulement si sa mesure de Haar est invariante non seulement à gauche mais aussi à droite. En effet, sa fonction modulaire (qui mesure le défaut d'invariance à droite d'une mesure de Haar) est le morphisme constant $g \rightarrow 1$.

et $\mathcal{U}(\mathcal{H}_T)$ est l'ensemble des opérateurs unitaires sur \mathcal{H}_T . Une représentation T est *irréductible* s'il n'y a aucun sous-espace fermé non triviale de \mathcal{H}_T invariant pour toute $T(a)$, $a \in G$. Deux représentations T, T' sont *équivalentes* s'il existe un opérateur linéaire inversible $A : \mathcal{H}_T \rightarrow \mathcal{H}_{T'}$ tel que $A \circ T = T' \circ A$. Dans ce cas nous écrivons $T \simeq T'$.

L'ensemble dual de G est l'ensemble \hat{G} de toutes les classes d'équivalence des représentations irréductibles unitaires de G . Nous remarquons que \hat{G} , en général, n'a pas une structure de groupe. Cependant, cela est suffisant pour généraliser la transformée de Fourier à cet égard. Soit $f \in L^2(G) \cap L^1(G)$, alors sa transformée de Fourier est définie par :

$$\hat{f}(T) = \int_G f(a)T(a)^{-1}da, \quad \forall T \in \hat{G}. \quad (1)$$

Observons que $\hat{f}(T)$ est un opérateur de Hilbert-Schmidt sur \mathcal{H}_T . Nous avons la généralisation suivante du Théorème 2.1.2.

Théorème 2.2.1 (Théorème de Plancherel non-commutatif unimodulaire). *Soit G un groupe localement compact unimodulaire. Alors, il existe une (unique) mesure de Plancherel $\hat{\mu}_G$ sur \hat{G} telle que la définition ci-dessus peut être étendue à une isométrie $F : L^2(G) \rightarrow L^2(\hat{G}, \mu_G)$. En particulier, la formule d'inversion suivante est vérifiée :*

$$f(a) = \int_{\hat{G}} \text{Trace}(\hat{f}(T) \circ T(a)) d\hat{\mu}_G(T).$$

De façon plus générale, si T est une représentation unitaire de G – pas nécessairement irréductible – on peut définir la transformée de Fourier $\hat{f}(T)$ par la même formule (1).

De la même façon que dans le cas abélien, la transformée de Fourier satisfait le théorème (2.1.2) en diagonalisant l'action de la représentation régulière à gauche $\Lambda : G \rightarrow \mathcal{U}(L^2(G))$, définie par $\Lambda(a)f(b) := f(a^{-1}b)$.

Théorème 2.2.2. *Pour tout $f, g \in L^2(G)$ et tout $a \in G$, il en est que :*

$$f = \Lambda(a)g \Leftrightarrow \hat{f}(T) = g(T) \circ T^{-1}(a) \quad \forall T \in \hat{G}. \quad (2)$$

2.3. Contexte de notre travail

Dans le reste de ce chapitre, nous considérons le cas particulier de produit semi-direct de deux groupes $SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{Z}_N$. On rappelle que :

- \mathbb{R}^2 est un groupe abélien localement compact, connexe et séparable.
- \mathbb{Z}_N est un groupe abélien fini de cardinalité N .
- L'action de rotation $k \in \mathbb{Z}_N \rightarrow R_k \in \text{Aut}(\mathbb{R}^2)$ de \mathbb{Z}_N sur \mathbb{R}^2 est libre et la mesure de Haar sur \mathbb{R}^2 (i.e. la mesure de Lebesgue) est invariante sous les R_k 's.
- l'opération (ou action) du groupe (non-commutatif) de $SE(2, N)$ est :

$$(x, k)(y, r) = (x + R_k y, k + r).$$

Ici, nous identifions implicitement $k + r$ par $k + r \bmod N$.

Les hypothèses ci-dessus garantissent que $SE(2, N)$ est unimodulaire. Plus tard, nous allons calculer explicitement les représentations irréductibles unitaires de $SE(2, N)$, qui seront de dimension finie, démontrant ainsi que $SE(2, N)$ est un groupe de Moore⁵.

Une notation additive est utilisée pour les deux groupes \mathbb{R}^2 et \mathbb{Z}_N . Nous notons :

- l'identité de \mathbb{R}^2 par o et celle de \mathbb{Z}_N par e .
- Les lettres x, y, z sont réservées aux éléments de \mathbb{R}^2 , alors que k, h, l, α, β sont des éléments de \mathbb{Z}_N .
- Les éléments des duals de Pontryagin $\widehat{\mathbb{R}^2}$ et $\widehat{\mathbb{Z}_N}$ sont notés, respectivement, $\lambda, \mu \in \widehat{\mathbb{R}^2}$ et $\hat{k}, \hat{h}, \dots \in \widehat{\mathbb{Z}_N}$.
- Les identités des duals de Pontryagin sont \hat{o} et \hat{e} .
- Les éléments de $SE(2, N)$ sont notés soit par $a, b \in SE(2, N)$ ou par les couples $(x, k), (y, h)$.
- L'action de rotation de \mathbb{Z}_N sur \mathbb{R}^2 induit une action de rotation de \mathbb{Z}_N sur $\widehat{\mathbb{R}^2}$, toujours notée $k \rightarrow R_k$ et définie par $R_k \lambda(x) = \lambda(R_{-k} x)$.
- Les représentations régulières à gauches de \mathbb{R}^2 et \mathbb{Z}_N sont appelées *translation* et *opérateurs de décalage*, et notées, respectivement, $x \rightarrow \tau_x \in \mathcal{U}(L^2(\mathbb{R}^2))$

⁵ Un groupe de Moore est un groupe localement compact dont les représentations unitaires irréductibles sont de dimension finie.

et $k \rightarrow S^k \in \mathcal{U}(L^2(\mathbb{Z}_N))$. Leur action de rotation sur $f \in L^2(\mathbb{R}^2)$ et $v \in L^2(\mathbb{Z}_N) \simeq \mathbb{C}^N$ est donnée par :

$$\tau_x f(y) := f(y - x) \quad \text{et} \quad (S^k v)_j = v_{j-k}.$$

Observons que, \mathbb{Z}_N étant cyclique, l'opérateur de décalage est complètement déterminé par $S = S^e$ via $S^k v = S \circ \dots \circ S v$.

La représentation régulière à gauche de $SE(2, N)$ est notée par $\Lambda : SE(2, N) \rightarrow \mathcal{U}(L^2(SE(2, N)))$, et son action sur $f \in L^2(SE(2, N))$ est $\Lambda(a)f(b) = f(a^{-1}b)$. En exploitant la structure du produit semi-direct de $SE(2, N)$, nous pouvons considérer la représentation quasi-régulière de $SE(2, N)$, notée $\pi : SE(2, N) \rightarrow \mathcal{U}(L^2(\mathbb{R}^2))$, dont l'action sur $f \in L^2(\mathbb{R}^2)$ est $\pi(x, k)f(y) = f(R_{-k}(y - x))$.

Nous allons aussi considérer la représentation $\hat{\pi}$ obtenue par conjugaison de π avec la transformée de Fourier sur \mathbb{R}^2 , celle-ci est facilement vue opérer sur $L^2(\widehat{\mathbb{R}^2})$ via :

$$\hat{\pi}(x, k)\hat{f}(\lambda) = F(\pi(x, k)f)(\lambda) = \overline{\lambda(x)}\hat{f}(R_{-k}\lambda).$$

Une description complète des représentations irréductibles unitaires de $SE(2, N)$ peut être obtenue au moyen de la théorie de Mackey (voir, e.g., [63], Ch. 17.1, Théorème 4 et 5). Nous la rappelons dans ce qui suit.

Théorème 2.3.1 (Représentations des $SE(2, N)$). *Le dual $\widehat{SE(2, N)}$ est paramétré par les orbites de l'action des rotations $\{R_k\}_{k \in \mathbb{Z}_N}$ sur \mathbb{R}^2 , i.e., par la tranche de camembert $Sc \subset \mathbb{R}^2$ qui en coordonnées polaires est $(\rho, \theta) \in (0, +\infty) \times [0, 2\pi/N)$ (Figure 43). En outre, correspondant à l'origine, il y a les caractères de \mathbb{Z}_N . A savoir, à chaque $\lambda \in Sc$ correspond la représentation T^λ agissant sur \mathbb{C}^N via :*

$$T^\lambda(x, k)v = \text{diag}_h(e^{i\langle \lambda, R_h x \rangle}) \circ S^{-k}v = (e^{i\langle \lambda, R_h x \rangle} v_{h+k})_{h=0}^{N-1}, \quad (3)$$

où l'on désigne par $\text{diag}_h v_h$ la matrice diagonale de diagonale $v \in \mathbb{C}^N$. D'autre part, à chaque $k \in \mathbb{Z}_N$ correspond la représentation sur \mathbb{C} donnée par : $T^{\hat{\delta} \times \hat{k}}(x, n)z = e^{i\frac{2\pi nk}{N}}z$.

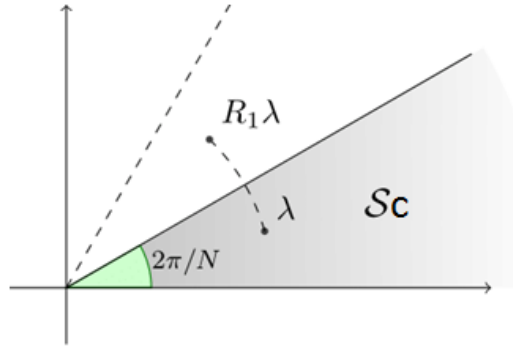


Figure 43: Rotation de la tranche de camembert Sc par $\{R_k\}_{k \in \mathbb{Z}_N}$

Proposition 2.3.2. Soit $f \in L^2(SE(2, N))$. Alors, pour tout $\lambda \in \widehat{\mathbb{R}^2} \setminus \{\hat{\delta}\}$, on peut vérifier que :

$$\hat{f}(T^\lambda)v(h) = \sum_{k \in \mathbb{Z}_N} F(f(\cdot, h - k))(R_{-k}\lambda)v(k) \quad \forall v \in \mathbb{C}^N. \quad (4)$$

De plus, pour tout $f \in L^1(SE(2, N)) \cap L^2(SE(2, N))$ et pour tout $\hat{k} \in \widehat{\mathbb{Z}_N}$, on a :

$$\hat{f}(T^{\hat{\delta} \times \hat{k}}) = \widehat{avg}_f(\hat{k}).$$

Démonstration. D'après le théorème 2.3.1, des calculs simples nous donnent :

$$\begin{aligned} \hat{f}(T^\lambda)v(h) &= \sum_{l \in \mathbb{Z}_N} \int_{\mathbb{R}^2} f(x, l) e^{-i\langle \lambda, R_{h-l}x \rangle} v(h-l) dx \\ &= \sum_{k \in \mathbb{Z}_N} v(k) \int_{\mathbb{R}^2} f(x, h-k) e^{-i\langle R_k \lambda, x \rangle} dx, \end{aligned}$$

Ce qui implique le premier énoncé. D'autre part, pour démontrer le second énoncé, il suffit de calculer :

$$\hat{f}(T^{\hat{\delta} \times \hat{k}}) = \sum_{l \in \mathbb{Z}_N} \int_{\mathbb{R}^2} f(x, l) e^{-i\frac{2\pi kl}{N}} dx = \sum_{l \in \mathbb{Z}_N} avg_f(l) e^{-i\frac{2\pi kl}{N}} = \widehat{avg}_f(\hat{k}).$$

■

2.4. Fonctions faiblement cycliques

On vérifie que les fonctions utilisées pour nos descripteurs sont faiblement cycliques, une condition nécessaire pour démontrer que les invariants rotationnels, introduits dans la section 4.4, soit faiblement complets.

Un vecteur $v \in \mathbb{C}^N$ est *cyclique* si $\{S^k v\}_{k \in \mathbb{Z}_N}$ est une base de \mathbb{C}^N . Ceci est équivalent à ce que la matrice circulante suivante soit inversible :

$$\text{Circ } v = (v, Sv, \dots, S^{N-1}v).$$

Fixons $f \in L^2(\mathbb{R}^2)$. Pour $\lambda \in \widehat{\mathbb{R}^2}$, nous considérons le vecteur $\omega_f(\lambda) \in \mathbb{C}^N$ tel que :

$$\omega_f(\lambda)_k = \hat{f}(R_{-k}\lambda) \quad \forall k \in \mathbb{Z}_N. \quad (5)$$

Ici, nous désignons par \hat{f} la transformée de Fourier abélienne sur \mathbb{R}^2 . Observons que $S^k \omega_f(\lambda) = \omega_f(R_k \lambda)$.

Etant donné que $\omega_f(o) = (\hat{f}(o), \dots, \hat{f}(o))$, le vecteur $\omega_f(\lambda)$ ne peut pas être cyclique pour tout $\lambda \in \widehat{\mathbb{R}^2}$, motivant ainsi la définition suivante.

Définition 2.4.1. Une fonction $f \in L^2(\mathbb{R}^2)$ est *faiblement cyclique* si $\omega_f(\lambda)$ est cyclique pour presque tout $\lambda \in \widehat{\mathbb{R}^2}$. Nous notons $\mathcal{C} \subset L^2(\mathbb{R}^2)$ l'ensemble des fonctions faiblement cycliques.

a. Fonctions à valeurs réelles

Nos arguments, dans ce qui suit, sont fortement basés sur l'exploitation de la propriété de la cyclicité faible des fonctions. Toutefois, pour les fonctions à valeurs réelles cela se révèle en général impossible.

Proposition 2.4.2. Soit N un nombre paire et $f \in L^2(\mathbb{R}^2)$. Alors, $\omega_f(\lambda) \in \mathcal{X}$ pour tout $\lambda \in \widehat{\mathbb{R}^2}$, où \mathcal{X} est le sous-espace \mathbb{R} -vectoriel propre de \mathbb{C}^N défini par :

$$\mathcal{X} = \{v \in \mathbb{C}^N \mid v(h) = \overline{v(h + \frac{N}{2})} \quad \forall h \in \mathbb{Z}_N\}. \quad (6)$$

En particulier, $\omega_f(\lambda)$ n'est jamais cyclique.

Démonstration : A partir de la parité de N , il en résulte que $R_{-h}\lambda = -R_{-h-N/2}\lambda$ pour tout $\lambda \in \widehat{\mathbb{R}^2}$ et $h \in \mathbb{Z}_N$. Etant donné que $\hat{f}(\lambda) = \overline{\hat{f}(-\lambda)}$, cela implique que :

$$\omega_f(\lambda)_h = \hat{f}(R_{-h}\lambda) = \hat{f}(-R_{-h-N/2}\lambda) = \overline{\hat{f}(R_{-h-N/2}\lambda)} = \overline{\omega_f(\lambda)_{h+N/2}}, \quad \forall h \in \mathbb{Z}_N,$$

Ce qui démontre l'énoncé. ■

Observons que \mathcal{X} est invariant sous l'action de l'opérateur de décalage. Nous disons alors que $\omega \in \mathcal{X}$ est \mathbb{R} -cyclique si $\text{span}\{S^k\omega\}_{\forall k \in \mathbb{Z}_N} = \mathcal{X}$, et posons la définition suivante.

Définition 2.4.3. Si N est paire, une fonction à valeurs réelles $f \in L^2_{\mathbb{R}}(\mathbb{R}^2)$ est faiblement \mathbb{R} -cyclique si $\omega_f(\lambda)$ est \mathbb{R} -cyclique pour presque tout $\lambda \in \widehat{\mathbb{R}^2}$. D'autre part, si N est impaire, $f \in L^2_{\mathbb{R}}(\mathbb{R}^2)$ est faiblement \mathbb{R} -cyclique si et seulement si elle est faiblement cyclique dans le sens de la définition 2.4.1.

Nous notons $C_{\mathbb{R}} \subset L^2_{\mathbb{R}}(\mathbb{R}^2)$ l'ensemble des fonctions faiblement \mathbb{R} -cycliques.

3. Descripteurs de Fourier dans les groupes

3.1. Invariants spectraux et bi-spectraux

Des applications $f \rightarrow I_f$ sont appelées des *invariants* pour G si $I_f = I_{\Lambda(a)f}$ pour tout $a \in G$. Un choix d'invariants est *complet* s'il vérifie la propriété suivante : pour tout $f, g \in L^2(G)$, nous avons :

$$\exists a \in G \text{ tel que } I_f = I_g \Leftrightarrow f = \Lambda(a)g.$$

Un choix d'invariants est *faiblement complet* si l'énoncé ci-dessus est vrai uniquement sur un sous ensemble résiduel⁶ de $L^2(G)$.

Comme il a été montré dans les sous-sections 2.1 et 2.2 qui présentent les propriétés associées à la représentation régulière à gauche dans le cas abélien (Th 1.1.4) et non abélien

⁶ Intersection dénombrable d'ensembles ouverts et denses.

(Th 1.2.2), nous utilisons ces deux théorèmes afin de mettre en évidence le caractère invariant des descripteurs spectraux et bi-spectraux.

Les premiers invariants que l'on peut envisager sont les suivants.

Définition 3.1.1. Les invariants spectraux (Power Spectrum invariant ou PS) de $f \in L^2(G)$ sont l'ensemble $PS_f = \{PS_f(T) \mid T \in \text{supp } \mu_{\hat{G}}\}$, où :

$$PS_f(T) := \hat{f}(T) \circ \hat{f}(T)^*.$$

Les invariants spectraux ne sont pas faiblement complets, même dans le cas simple de $G = \mathbb{R}$. Dans ce cas $PS_f(\lambda) = |\hat{f}(\lambda)|^2$ pour tout $\lambda \in \text{supp } \mu_{\hat{G}} = \hat{\mathbb{R}}$, et il est facile de construire un contre-exemple. En effet, il suffit de fixer un certain $\phi : \hat{\mathbb{R}} \rightarrow \mathbb{S}^1$ et considérer la fonction $g = F^{-1}(e^{i\phi(\lambda)} \hat{f}(\lambda))$. De toute évidence, g est tel que $PS_f = PS_g$ mais $f = \Lambda(a)g$ si et seulement si $\phi(\lambda) = a\lambda$.

Ainsi, nous devons considérer des ensembles plus riches d'invariants, comme les suivants.

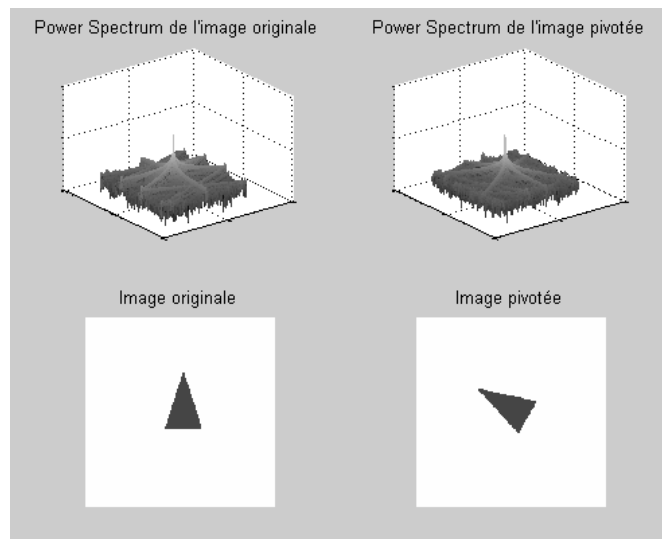


Figure 44 : Log du Power Spectrum d'une image synthétique et sa version pivotée

Définition 3.1.2. Les invariants bi-spectraux (Bi-Spectrum invariant ou BS) de $f \in L^2(G)$ sont l'ensemble $BS_f = \{BS_f(T_1, T_2) \mid T_1, T_2 \in \text{supp } \mu_{\hat{G}}\}$, où :

$$BS_f(T_1, T_2) := \hat{f}(T_1) \otimes \hat{f}(T_2) \circ \hat{f}(T_1 \otimes T_2)^*.$$

A priori, on a besoin d'utiliser les deux invariants, spectraux et bi-spectraux, bien que nous allons voir que dans la plupart des cas, et en particulier, dans le cas du groupe $SE(2, N)$, nous avons $BS_f \supset PS_f$.

3.2. Invariants spectraux et bi-spectraux sur les groupes abéliens localement compacts

Dans cette section, nous allons démontrer la faible complétude des invariants bi-spectraux dans le cas où G est un groupe abélien localement compact. Dans ce cas, toutes ses représentations sont unidimensionnelles et la mesure de Plancherel est la mesure de Haar sur le caractère du groupe \hat{G} . En effet on va montrer que les invariants sont faiblement complets sur l'ensemble :

$$\mathcal{G} := \{f \in L^2(G) \mid f \text{ a un support compact et } \hat{f}(\lambda) = 0 \text{ pour au plus une quantité discrète des } \lambda \in \hat{G}\}$$

Des calculs simples montrent que:

$$PS_f(\lambda) = |\hat{f}(\lambda)|^2 \quad \text{et} \quad BS_f(\lambda_1, \lambda_2) = \hat{f}(\lambda_1)\hat{f}(\lambda_2)\bar{\hat{f}}(\lambda_1 + \lambda_2). \quad (7)$$

Dans ce cas, nous avons que $BS_f \supset PS_f$ pour tout $f, g \in L^1(G) \cap L^2(G)$. En effet, observons que le fait de choisir $\lambda_1 = \lambda_2 = \hat{\delta}$ dans les invariants bi-spectraux donne $\text{avg}(f)|\text{avg}(f)|^2 = \text{avg}(g)|\text{avg}(g)|^2$, ce qui implique que $\text{avg}(f) = \text{avg}(g)$. Ceci montre que $BS_f(\lambda_1, 0) = PS_f(\lambda_1)$.

Théorème 3.2.1. *Les invariants bi-spectraux sont complets sur l'ensemble \mathcal{G} .*

Démonstration. Soit $f, g \in \mathcal{G}$ tel que $BS_f = BS_g$. Etant donné que ceci implique $PS_f = PS_g$, nous avons que $|\hat{f}| = |\hat{g}|$. Ainsi, \hat{f} et \hat{g} s'annulent sur le même ensemble I . En outre, observons que étant donné f et g sont à support compact, leurs transformées de Fourier \hat{f} et \hat{g} sont continues.

Soit $u(\lambda) = \hat{g}(\lambda)/\hat{f}(\lambda)$ pour tout $\lambda \in I$. Puisque u est le ratio de deux fonctions continues qui s'annulent uniquement sur un ensemble discret, elle est mesurable. Alors, par l'égalité des invariants bi-spectraux et (7), il en résulte que u satisfait :

$$u(\lambda_1 + \lambda_2) = u(\lambda_1)u(\lambda_2).$$

Ceci implique que u est un caractère mesurable de \hat{G} et donc, par le biais d'un résultat connu ([37], Théorème 22.17), il doit être continu. Par la dualité de Pontryagin, ceci démontre l'existence de $a \in G$ tel que $u(\lambda) = \lambda(a)$. Ainsi, nous avons démontré que $\hat{f}(\lambda) = \lambda(a)\hat{g}(\lambda)$, ce qui d'après le théorème 2.2.2 implique que $f = \tau_{a^{-1}}g$, complétant ainsi la démonstration.

■

Dans le cas où $G = \mathbb{R}^n$ le résultat ci-dessus peut être renforcé.

Corollaire 3.2.2. *Les invariants bi-spectraux sur \mathbb{R}^n sont complets sur les fonctions à support compact de $L^2(\mathbb{R}^2)$.*

Démonstration. Il suffit d'observer que par le théorème de Paley-Wiener, les transformées de Fourier des fonctions à support compact sont analytiques. Etant donné que les fonctions non nulles analytiques ont un ensemble discret de zéros isolés, ceci implique que l'ensemble \mathcal{G} du théorème 3.2.1 coïncide avec toutes les fonctions considérées.

■

3.3. Invariants spectraux et bi-spectraux sur $SE(2, N)$

Dans cette sous-section, nous allons étudier la faible-complétude des invariants spectraux et bi-spectraux dans le cas où $G = SE(2, N)$.

Soit $K \subset \mathbb{R}^2$ un ensemble compact. Dans ce qui suit, nous serons principalement concernés par les fonctions qui sont à support compact soit dans K ou dans $K \times \mathbb{Z}_N \subset SE(2, N)$.

Le résultat suivant généralise, par une démonstration simplifiée, le résultat présenté dans [74]. Notons que ce résultat est vrai dans un cadre plus général, comme il sera montré dans un article à paraître par Prandi et Gauthier.

Théorème 3.3.1. *Soit $K \subset \mathbb{R}^2$ un ensemble compact. Le PS et BS généralisés sont faiblement complets sur $L^2(\mathbb{Z}_N \times K)$. En particulier, ils discriminent sur l'ensemble ouvert et dense \mathcal{G} de*

fonctions f à support dans $\mathbb{Z}_N \times K$ et dont la transformée de Fourier est inversible pour un ensemble ouvert et dense des λ s.

Autrement dit, $f, g \in \mathcal{G}$ sont tel que $PS_f = PS_g$ et $BS_f = BS_g$ si et seulement si $f = \Lambda(x, k)g$ pour certains $(x, k) \in SE(2, N)$.

Démonstration. Le fait que \mathcal{G} est ouvert et dense est démontré dans le lemme 1 (Annexes : Lemmes auxiliaires pour la démonstration du théorème 3.3.1). Soit $f, g \in \mathcal{G}$ tel que $PS_f = PS_g$ et $BS_f = BS_g$. L'égalité du BS généralisé implique que l'ensemble des λ s pour lequel $\hat{f}(\lambda)$ et $\hat{g}(\lambda)$ ne peuvent pas être inversibles, est le même. Nous le notons I et considérons

$$U(T^\lambda) := \hat{f}(T^\lambda)^{-1} \hat{g}(T^\lambda) \quad \forall \lambda \in I.$$

Afin de compléter la démonstration de l'énoncé, nous allons démontrer que $U(T^\lambda)$ peut être définie pour tous λ s dans \mathbb{R}^2 et, en outre, que $\exists a \in SE(2, N)$ tel que $U(T^\lambda) = T^\lambda(a)$.

En effet, par (2) cela implique donc que $f = \Lambda(a)g$.

De plus $U(T^\lambda)$ est unitaire pour tout $\lambda \in I$. En effet, par l'égalité du PS généralisé, nous avons :

$$U(T^\lambda)^* U(T^\lambda) = g(T^\lambda)^* PS_f(\lambda) g(T^\lambda) = PS_g(\lambda)^* PS_f(\lambda) = I.$$

Observons que l'égalité du BS généralisé et la définition de U , impliquent que pour tout $\lambda_1, \lambda_2 \in I$, on a :

$$BS_f(\lambda_1, \lambda_2) = \hat{f}(T^{\lambda_1}) \otimes \hat{f}(T^{\lambda_2}) \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \circ \hat{g}(T^{\lambda_1} \otimes T^{\lambda_2})^*.$$

Par l'inversibilité de $\hat{f}(T^{\lambda_1}) \otimes \hat{f}(T^{\lambda_2})$ et l'unitarité de U , on obtient :

$$\hat{f}(T^{\lambda_1} \otimes T^{\lambda_2}) \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) = \hat{g}(T^{\lambda_1} \otimes T^{\lambda_2}). \quad (8)$$

Le résultat annoncé est alors une conséquence des trois faits suivants, qui sont démontrés dans l'annexe (Lemmes auxiliaires pour la démonstration du théorème 3.3.1) :

1. Lemme 2 : La fonction $\lambda \rightarrow U(T^\lambda)$ est continue sur I .
2. Lemme 3 : La fonction $\lambda \rightarrow U(T^\lambda)$ peut être étendue à une fonction continue sur \mathbb{R}^2 , pour laquelle (8) est toujours vraie.
3. Lemme 4 : Il existe $a \in SE(2, N)$ tel que $U(T^\lambda) = T^\lambda(a)$.

■

Un corollaire immédiat est le suivant.

Corollaire 3.3.2. *Soit $\tilde{\mathcal{L}} : L^2(\mathbb{R}^2) \rightarrow L^2(SE(2, N))$ un opérateur de relèvement injectif. Supposons qu'il existe un ensemble résiduel $R \subset L^2(\mathbb{R}^2)$ tel que $\tilde{\mathcal{L}}(R) \cap \mathcal{G}$ est résiduelle. Alors, les PS et BS généralisés sont faiblement complets sur $L^2(\mathbb{R}^2)$. A savoir, pour tout $f, g \in R$ on admet que $BS_{\tilde{\mathcal{L}}f} = BS_{\tilde{\mathcal{L}}g}$ si et seulement si $f = \pi(x, k)g$ pour certains $(x, k) \in SE(2, N)$.*

4. Descripteurs de Fourier basés sur la structure du cortex visuel primaire humain V1

4.1. Présentation du pipeline naturel pour la reconnaissance d'objets

Dans cette partie, suivant une ligne de recherche commencée dans [74], nous présentons un cadre théorique qui nous permet de construire des descripteurs de Fourier généralisés invariants par rapport aux roto-translations (semi-discrètes) d'images.

En fait, les contributions de certains des auteurs à un modèle assez récent du cortex visuel primaire humain V1 [62, 9, 8, 10], ont montré que celui-ci peut être modélisé comme un groupe de roto-translations semi-discrètes $SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{Z}_N$. Dans ce modèle, les stimuli corticaux sont des fonctions dans $L^2(SE(2, N))$, par rapport à la mesure de Haar de $SE(2, N)$, et les images provenant du plan visuel sont relevées à des stimuli corticaux via une opération naturelle de relèvement injectif et invariant à gauche $\mathcal{L} : L^2(\mathbb{R}^2) \rightarrow L^2(SE(2, N))$. Un tel relèvement est défini comme la transformée en ondelettes par rapport à une ondelette mère (voir: section 4.2).

A partir de ces faits, un pipeline naturel pour la reconnaissance d'objets est le suivant (Figure 45) :

1. Etant donné une image $f \in L^2(\mathbb{R}^2)$, relevons-la à un stimulus cortical $\mathcal{L}f \in L^2(SE(2, N))$.
2. Calculer les descripteurs de Fourier généralisés de $\mathcal{L}f$ sur le groupe non-commutatif $SE(2, N)$.
3. Si le relèvement d'une autre image $g \in L^2(\mathbb{R}^2)$ a les mêmes descripteurs de Fourier que $\mathcal{L}f$, on déduit que $\mathcal{L}f \approx \mathcal{L}g$ à une action d'un élément de $SE(2, N)$ près.
4. Grâce à l'invariance à gauche et à l'injectivité du relèvement \mathcal{L} , on obtient aussi que $f \approx g$ à une action d'un élément de $SE(2, N)$ près.

Ce pipeline a été déjà étudié dans [74], où les auteurs ont considéré un relèvement non-invariant à gauche. Pour ce relèvement, ils ont démontré un résultat de faible complétude du BS pour des images, représentées comme des fonctions de $L^2(\mathbb{R}^2)$ à support à l'intérieur d'un ensemble compact et fixe.

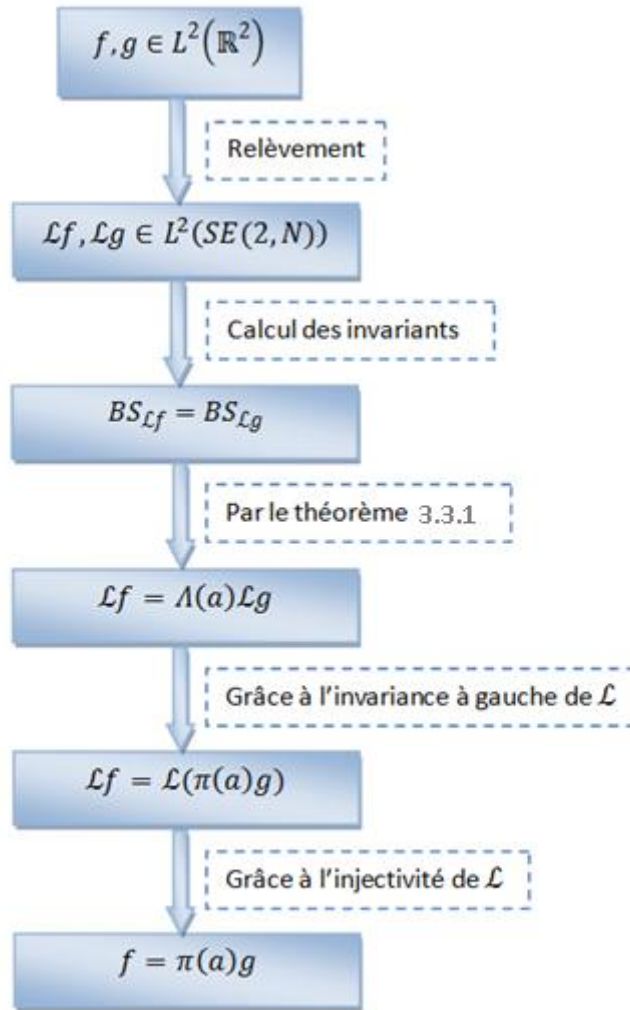


Figure 45 : Pipeline naturel pour la reconnaissance d'objets

Dans la suite, nous considérons la même question pour les relèvements invariants à gauche, où la situation se révèle être plus compliquée. En particulier, comme il sera expliqué dans la section 4.4, afin d'assurer la faible complétude, nous sommes amenés à considérer des invariants « plus forts » que le BS généralisé. Toutefois, comme il sera observé dans la section 4.4, le calcul réel de ces plus forts invariants sur des images relevées nécessite N fois moins de temps de calcul et d'espace par rapport au calcul des invariants proposés dans [74].

4.2. Présentation du modèle mathématique du cortex visuel primaire V1

Comme mentionné précédemment, la principale nouveauté de notre approche est son lien avec un modèle assez récent du cortex visuel primaire humain V1, grâce à Petitot et Citti-Sarti [16, 60] et les contributions récentes de certains des auteurs [8, 9, 10, 62, 7]. La théorie « Orientations score » introduite dans [24, 25], est aussi fortement connectée avec ce travail, en particulier pour son exploitation du relèvement invariant à gauche. Nous mentionnons aussi [73], où les invariants d'images basés sur la structure du groupe de roto-translation $SE(2)$ ont été introduits pour les textures. Dans cette section, nous présentons les caractéristiques de ce modèle qui sont essentielles pour notre approche.

Etant donné qu'il est bien connu [40] que les neurones dans V1 sont sensibles non seulement aux différentes positions dans le champ visuel, mais aussi aux orientations locales, et qu'il est raisonnable de supposer que ces orientations soient finies, dans [8] V1 a été modélisé comme un groupe de roto-translations semi-discrètes $SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{Z}_N$ pour $N \in \mathbb{N}$ paires.

Les stimuli visuels $f \in L^2(\mathbb{R}^2)$ sont supposés être relevés à des motifs (patterns) d'activation dans $L^2(SE(2, N))$ par un opérateur de relèvement $\mathcal{L} : L^2(\mathbb{R}^2) \rightarrow L^2(SE(2, N))$. Motivés par des évidences neurophysiologiques, nous supposons donc que :

(H) l'opérateur de relèvement \mathcal{L} est linéaire et défini par

$$\mathcal{L}f(x, k) = \int_{\mathbb{R}^2} f(y) \bar{\psi}(R_{-k}(y - x)) dy, \quad (9)$$

pour une ondelette mère donnée $\psi \in L^2(\mathbb{R}^2)$ tel que \mathcal{L} est injectif et borné.

Remarque 4.2.1. Cette hypothèse signifie que l'opérateur de relèvement sous considération est la transformation en ondelettes par rapport à ψ (voir, e.g. [27]). Le fait que \mathcal{L} soit injectif et borné est alors équivalent au fait que l'ondelette mère est *faiblement admissible*, i.e., est tel que l'application $\lambda \in \mathbb{R}^2 \rightarrow \sum_{k \in \mathbb{Z}_N} |\hat{\psi}(R_{-k}\lambda)|^2$ est strictement positive et essentiellement bornée.

Comme conséquence de l'hypothèse ci-dessus, l'opération de relèvement \mathcal{L} est invariante à gauche par rapport à l'action de $SE(2, N)$. A savoir :

$$\Lambda(x, k) \circ \mathcal{L} = \mathcal{L} \circ \pi(x, k). \quad (10)$$

Ici Λ et π sont les actions de $SE(2, N)$ sur $L^2(SE(2, N))$ et $L^2(\mathbb{R}^2)$ respectivement. C'est-à-dire :

$$[\Lambda(x, k)\varphi](y, r) = \varphi((x, k)^{-1}(y, r)) = \varphi(R_{-k}(y - x), k + r),$$

$$[\pi(x, k)f](y) = f((x, k)^{-1}y) = f(R_{-k}(y - x)).$$

La formule (10) peut être vue comme une version semi-discrète de la symétrie *shift-twist* [13].

L'observation principale pour nos besoins est que (10) signifie que deux images $f, g \in L^2(\mathbb{R}^2)$ peuvent être déduites par roto-translation (i.e., $f = \pi(x, k)g$ pour certains $(x, k) \in SE(2, N)$) si et seulement si leurs relèvements peuvent être déduits par $\Lambda(x, k)$. C.-à-d. que,

$$\mathcal{L}f = \Lambda(a) \circ \mathcal{L}g \Leftrightarrow f = \pi(a)g.$$

4.3. La non-applicabilité du théorème de la complétude dans $SE(2, N)$

Le corollaire 3.3.2 ne peut jamais être appliqué aux relèvements de la forme (9). En fait, on a le résultat suivant.

Proposition 4.3.1. Considérons $\omega_f(\lambda) := (\hat{f}(R_{-k}\lambda))_{k=0}^{N-1} \in \mathbb{C}^N$, nous avons :

$$\widehat{\mathcal{L}f}(T^\lambda) = \omega_\psi(\lambda)^* \otimes \omega_f(\lambda)^*, \quad (11)$$

Où pour $v, w \in \mathbb{C}^N$, nous considérons $v^* = (\overline{v_k})_k$ et $(v \otimes w)_{k,h} = v_k \overline{w_h}$, de sorte que $(v \otimes w)u = \langle w, u \rangle v$ pour tout $u \in \mathbb{C}^N$. En particulier, on a que $\text{rang } \widehat{\mathcal{L}f}(T^\lambda) \leq 1$ et donc $\mathcal{L}(L^2(\mathbb{R}^2)) \cap \mathcal{G} = \emptyset$ lorsque $N > 1$.

Démonstration. Soit $\lambda \in Sc$ et considérons $v \in \mathbb{C}^N$. Observons que $(x, k)^{-1} = (-R_{-k}x, -k)$.

Ensuite, par (1), (4) et (3), pour tout $h \in \mathbb{Z}_N$, nous avons

$$\begin{aligned}
(\hat{\mathcal{L}}(T^\lambda).v)_h &= \sum_{k=0}^{N-1} \int_{\mathbb{R}^2} \mathcal{L}(x, k) e^{-i\langle \lambda, R_{h-k}x \rangle} v_{h-k} dx \\
&= \sum_{k=0}^{N-1} v_{h-k} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f(y) \bar{\psi}(R_{-k}(y-x)) e^{-i\langle \lambda, R_{h-k}x \rangle} dy dx \\
&= \sum_{k=0}^{N-1} v_{h-k} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \bar{\psi}(z) f(y) e^{-i\langle R_{h-k}\lambda, y-z \rangle} dy dz \\
&= \bar{\psi}(R_h\lambda) \sum_{k=0}^{N-1} v_{h-k} \hat{f}(R_{h-k}\lambda) \\
&= \overline{\omega_\psi(\lambda)_h} \langle \overline{\omega_f(\lambda)}, v \rangle.
\end{aligned}$$

Par définition de $\omega_\psi(\lambda)^* \otimes \omega_f(\lambda)^*$ ceci complète la démonstration. ■

4.4. Les invariants spectraux et bi-spectraux rotationnels et leur complétudes dans $SE(2, N)$

Pour contourner la difficulté posée par la non-inversibilité de la transformée de Fourier pour les fonctions relevées, nous sommes amenés à considérer les descripteurs plus forts suivants.

Définition 4.4.1. *Les invariants spectraux et bi-spectraux rotationnels (RPS et RBS respectivement) de $f \in L^2(SE(2, N))$ sont les collections de matrices, pour tout $\lambda, \lambda_1, \lambda_2 \in S$ et $h \in \mathbb{Z}_N$,*

$$RPS_f(\lambda, h) := \hat{f}(T^{R_h\lambda}) \circ \hat{f}(T^\lambda)^*.$$

$$RBS_f(\lambda_1, \lambda_2, h) := \hat{f}(T^{R_h\lambda_1}) \otimes \hat{f}(T^{\lambda_2}) \circ \hat{f}(T^{\lambda_1} \otimes T^{\lambda_2})^*.$$

Les descripteurs rotationnels sont invariants uniquement sous l'action de rotations de $\mathbb{Z}_N \subset SE(2, N)$ mais pas sous l'action de translations. Pour éviter ce problème, nous fixons un $K \subset \mathbb{R}^2$ compact et considérons l'ensemble $\mathcal{A} \subset L^2(\mathbb{R}^2)$ des fonctions à support

compact dans K , avec une moyenne non-nulle. Observons que ceci est un sous-ensemble ouvert et dense de $L^2(K)$. On peut donc définir le barycentre $c_f \in \mathbb{R}^2$ de $f \in \mathcal{A}$ comme :

$$c_f = \frac{1}{\text{avg } f} \left(\int_{\mathbb{R}^2} x_1 f(x) dx, \int_{\mathbb{R}^2} x_2 f(x) dx \right),$$

Et l'opérateur de centrage $\phi : \mathcal{A} \rightarrow \mathcal{A}$ comme :

$$\phi f(x) := f(x - c_f). \quad (12)$$

Ensuite, en considérant le relèvement centré $\mathcal{L}_c = \mathcal{L} \circ \phi$, nous avons $\mathcal{L}_c f = \mathcal{L}_c g$ si et seulement si g est une translatée de f . En particulier,

$$\mathcal{L}_c f = \Lambda(0, k) \mathcal{L}_c g \Leftrightarrow f = \pi(x, k) g \quad \text{Pour certains } x \in \mathbb{R}^2.$$

Définition 4.4.2. Soit $R \subset L^2(\mathbb{R}^2)$ l'ensemble des fonctions à valeurs réelles f à support dans K , tel que $\hat{f}(\lambda) \neq 0$ pour presque tout $\lambda \in \mathbb{R}^2$ et la famille $\Omega_f = \{S^k \omega_f(\lambda)\}_{k=0}^{N-1}$ est une base pour \mathbb{C}^N , si N est impair, ou, si N est pair, pour

$$\chi = \left\{ v \in \mathbb{C}^N \mid v(h) = \overline{v\left(h + \frac{N}{2}\right)} \quad \forall h \in \mathbb{Z}_N \right\}.$$

La dépendance de cette définition de la parité de N provient du fait que $\hat{f}(\lambda) = \overline{\hat{f}(-\lambda)}$.

En effet, pour N paire, cela implique que $S^k \omega_f(\lambda) \in \chi$ pour tout $k \in \mathbb{Z}_N$. A ce titre, il n'y a aucun espoir pour que la famille Ω_f génère la totalité de \mathbb{C}^N .

Finalement, nous avons le théorème suivant.

Théorème 4.4.3. *Pour tout compact $K \subset \mathbb{R}^2$, si l'ondelette mère $\psi \in R$, les RPS et RBS sont faiblement complets sur $L^2(K) \cap \mathcal{A}$. A savoir, l'ensemble R est ouvert et dense dans $L^2(K)$ et pour tout $f, g \in R \cap \mathcal{A}$, on admet que $RPS_{\mathcal{L}_c f} = RPS_{\mathcal{L}_c g}$ et $RBS_{\mathcal{L}_c f} = RBS_{\mathcal{L}_c g}$ si et seulement si $f = \pi(x, k)g$ pour certain $(x, k) \in SE(2, N)$.*

Ici, nous nous contentons de présenter uniquement une esquisse de la démonstration de ce résultat pour le cas N impaire. La parité de N ne présente pas les problèmes essentiels, il

suffit d'exploiter le fait que $\widehat{\mathcal{L}f}(T^\lambda) \subset \chi$ pour tout $f \in R$ et $\lambda \in S$, où χ est défini dans (6) et que l'équivalence A du théorème d'induction-réduction se réduit tout simplement à une équivalence entre $\chi \times \chi$ et $\bigoplus_{k \in \mathbb{Z}_N} \chi$. Cependant, afin de démontrer le point-clé technique (13), nous avons besoin d'une étude plus raffinée des propriétés des opérateurs circulants, ce qui est en dehors du cadre de ce travail et nous nous en reportons à un article par Prandi et Gauthier, à paraître.

Démonstration. (Esquisse dans le cas N est impair). Le fait que R est ouvert et dense dans $L^2(K)$ résulte des mêmes arguments dans le lemme 1 (Annexe : Lemmes auxiliaires pour la démonstration du théorème 3.3.1).

Soit $Circ v$ la matrice circulante associée à v , donnée par, $Circ v = [v, Sv, \dots, S^{N-1}v]$. Alors la condition sur Ω_f pour $f \in R$ est équivalente à l'inversibilité de $Circ \omega_f(\lambda)$ pour un ensemble ouvert et dense de λ s. Selon les propriétés de la transformée de Fourier sur \mathbb{R}^2 par rapport aux translations, il en résulte :

$$\omega_{\phi_f}(\lambda) = \text{diag}_k(e^{-i(\lambda, R_k c_f)}) \omega_f(\lambda).$$

Ceci implique que $Circ \omega_f(\lambda)$ est inversible si et seulement si $Circ \omega_{\phi_f}(\lambda)$ l'est. Par conséquent, l'énoncé est équivalente au fait que pour tout couple $f, g \in R$ nous avons $RBS_{\mathcal{L}f} = RBS_{\mathcal{L}g}$ si et seulement si $f = R_k g$ pour certain $k \in \mathbb{Z}_N$.

La démonstration est similaire à celle du théorème 3.3.1 mais avec des difficultés techniques supplémentaires. Soit I l'ensemble où $Circ \omega_f(\lambda)$ et $Circ \omega_g(\lambda)$ sont inversibles. Par hypothèse I est ouvert et dense. Pour surmonter la non-inversibilité de $\widehat{\mathcal{L}f}$ dans la définition U candidate, nous exploitons l'inversibilité des matrices circulantes $Circ \omega_f(\lambda)$ et $Circ \omega_g(\lambda)$ sur un ensemble ouvert et dense. A savoir, pour tout $\lambda \in I$, nous considérons

$$U(T^\lambda)^* := Circ \omega_g(\lambda) (Circ \omega_f(\lambda))^{-1}.$$

Par définition, $U(T^\lambda)$ est circulant et $U(T^\lambda)^* S^k \omega_f(\lambda) = S^k \omega_g(\lambda)$ pour tout $k \in \mathbb{Z}_N$. En outre, par (11), ceci est équivalent à :

$$\widehat{\mathcal{L}f}(T^{R_k \lambda}) U(T^\lambda) = \widehat{\mathcal{L}g}(T^{R_k \lambda}), \quad \forall k \in \mathbb{Z}_N.$$

En particulier, $\lambda \rightarrow U(T^\lambda)$ est constante sur les orbites $\{R_k\lambda\}_{k \in \mathbb{Z}_N}$. Finalement, $U(T^\lambda)$ est unitaire, en conséquence, e.g. du théorème 5.1.

La difficulté principale dans la démonstration maintenant est d'obtenir l'équivalent de l'identité (8), qui est, que pour un ensemble ouvert et dense de couples (λ_1, λ_2) nous avons

$$\widehat{\mathcal{L}}f(T^{R_k\lambda_1} \otimes T^{R_k\lambda_2})U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) = \widehat{\mathcal{L}}g(T^{R_k\lambda_1} \otimes T^{R_k\lambda_2}), \quad (13)$$

$$\forall k \in \mathbb{Z}_N.$$

Comme déjà mentionnée, la démonstration de cette identité exige une utilisation profonde des propriétés des opérateurs circulants, ce qui est en dehors du cadre de ce travail. Nous reportons donc à un article à paraître.

Une fois (13) est connue, l'énoncé suit en appliquant les mêmes arguments que ceux dans le théorème 3.3.1. A savoir :

1. La fonction $\lambda \rightarrow U(T^\lambda)$ est continue sur I . Cela peut être fait par les mêmes arguments que dans le lemme 2 (Annexe : Lemmes auxiliaires pour la démonstration du théorème 3.3.1).
2. La fonction $\lambda \rightarrow U(T^\lambda)$ peut être étendue à une fonction continue sur S satisfaisant (13). Cela peut être fait exactement comme dans le lemme 3 (Annexe : Lemmes auxiliaires pour la démonstration du théorème 3.3.1).
3. Il existe $k \in \mathbb{Z}_N$ tel que $U(T^\lambda) = T^\lambda(0, k)$. Cela est démontré suivant le lemme 4 (Annexe : Lemmes auxiliaires pour la démonstration du théorème 3.3.1). En effet, le fait que $\lambda \rightarrow U(T^\lambda)$ est maintenant constante sur les orbites $\{R_k\lambda\}_{k \in \mathbb{Z}_N}$, implique que les f_k s y obtenues doivent être indépendantes de k . Comme $f_k(\lambda) = e^{i\langle R_k x_0, \lambda \rangle}$ pour certains $x_0 \in \mathbb{R}^2$, ceci implique que $x_0 = 0$ et donc $f_k \equiv 0$. Evidemment, cela démontre que $U(T^\lambda) = S^k = T^\lambda(0, k)$, pour certains $k \in \mathbb{Z}_N$.

■

5. Le calcul pratique des descripteurs de Fourier

Ici, nous présentons des formules explicites pour le calcul des descripteurs de Fourier présentés dans les sections précédentes.

Dans ce qui suit, nous montrons que, sous certaines hypothèses sur l'ondelette mère ψ , le calcul concret des PS et BS généralisés et de leurs homologues rotationnels RPS et RBS, ne dépend que de la transformée de Fourier 2D de f . Dans la suite, étant donné deux vecteurs $v, w \in \mathbb{C}^N$, on note \odot , le produit élément par élément de deux vecteurs.

Théorème 5.1. *Nous supposons que l'ondelette mère $\psi \in R$. Alors :*

- Pour tout $f \in R$, les PS et BS généralisés de $\mathcal{L}f$ sont déterminés respectivement par les quantités, pour presque tout $\lambda, \lambda_1, \lambda_2 \in S$:

$$I_1^\lambda(f) = \|\omega_f(\lambda)\|^2 = \sum_{k=0}^{N-1} |\hat{f}(R_{-k}\lambda)|^2$$

$$I_1^{\lambda_1, \lambda_2}(f) = \langle \omega_f(\lambda_1) \odot \omega_f(\lambda_2), \omega_f(\lambda_1 + \lambda_2) \rangle$$

$$= \sum_{k=0}^{N-1} \hat{f}(R_{-k}\lambda_1) \hat{f}(R_{-k}\lambda_2) \overline{\hat{f}(R_{-k}(\lambda_1 + \lambda_2))}.$$

- Pour tout $f \in \mathcal{A} \cap R$, les RPS et RBS de $\mathcal{L}f$ sont déterminés respectivement par les quantités, pour presque tout $\lambda, \lambda_1, \lambda_2 \in S$ et $h \in \mathbb{Z}_N$:

$$I_2^{\lambda, h}(f) = \langle \omega_{\phi f}(R_h\lambda), \omega_{\phi f}(\lambda) \rangle$$

$$= \sum_{k=0}^{N-1} \overline{\hat{f}(R_{-k+h}\lambda)} \hat{f}(R_{-k}\lambda)$$

$$I_2^{\lambda_1, \lambda_2, h}(f) = \langle \omega_{\phi f}(R_h\lambda_1) \odot \omega_{\phi f}(\lambda_2), \omega_{\phi f}(\lambda_1 + \lambda_2) \rangle$$

$$= \sum_{k=0}^{N-1} \hat{f}(R_{-k+h}\lambda_1) \hat{f}(R_{-k}\lambda_2) \overline{\hat{f}(R_{-k}(\lambda_1 + \lambda_2))}.$$

Ici, $\phi : \mathcal{A} \rightarrow \mathcal{A}$ est l'opérateur de centrage défini dans (12).

Remarque 5.2. Le théorème 5.1 montre en particulier que le résultat du théorème 4.4.3 est en effet plus fort que le résultat de complétude pour le BS généralisé du relèvement cyclique obtenu dans [74]. En effet, dans ce travail, il est démontré que celui-ci (pour N

impaire) est déterminé exactement par les quantités, pour presque tout $\lambda_1, \lambda_2 \in S$ et $h, k \in \mathbb{Z}_N$:

$$\tilde{I}_2^{\lambda_1, \lambda_2, k, h} = \langle \omega_{\phi_f}(R_h \lambda_1) \odot \omega_{\phi_f}(R_k \lambda_2), \omega_{\phi_f}(\lambda_1 + R_{h+k} \lambda_2) \rangle.$$

En particulier, pour chaque $\lambda_1, \lambda_2 \in S$, on doit calculer N fois plus de quantités que celles pour les RBS.

Comme corollaire du théorème 5.1, nous montrons que, dans le but de comparer le PS et BS, il est généralement suffisant de comparer que le dernier.

Corollaire 5.3. *Soit $\psi \in R$ et $f, g \in R \cap \mathcal{A}$. Alors, si $\mathcal{L}f$ et $\mathcal{L}g$ ont le même BS généralisé (resp. rotationnel), elles ont aussi le même PS généralisé (resp. rotationnel).*

Démonstration. Nous démontrons uniquement le résultat pour les descripteurs rotationnels. Dans le but de démontrer celui des descripteurs généralisés, il sera suffisant de fixer $h = 0$ dans ce qui suit. D'après le théorème 5.1, il est suffisant de montrer qu'à chaque fois que $I_2^{\lambda_1, \lambda_2, h}(f) = I_2^{\lambda_1, \lambda_2, h}(g)$ pour presque tout $\lambda_1, \lambda_2 \in S$ et tout $h \in \mathbb{Z}_N$, alors $I_2^{\lambda, h}(f) = I_2^{\lambda, h}(g)$ pour presque tout $\lambda \in S$ et tout $h \in \mathbb{Z}_N$. Nous commençons par observer que par le théorème de Paley-Wiener, toutes ces quantités sont analytiques, étant donné que f et g sont à support compact. De plus,

$$\lim_{\lambda_1, \lambda_2 \downarrow 0} I_2^{\lambda_1, \lambda_2, h}(f) = N \hat{f}(0) |\hat{f}(0)|^2 = N \text{avg}(f)^3,$$

Et la même chose est vraie pour g . Ainsi, $\text{avg}(f) = \text{avg}(g)$.

Finalement, le résultat suit en observant que :

$$\lim_{\lambda_2 \downarrow 0} I_2^{\lambda_1, \lambda_2, h}(f) = \text{avg}(f) I_2^{\lambda_1, h}(f).$$

■

6. Conclusion

Dans ce chapitre, nous avons présenté un ensemble de descripteurs de Fourier sur le groupe de roto-translations semi-discrètes $SE(2, N)$. Ensuite, nous avons démontré que les

descripteurs PS et BS généralisés – et leurs homologues rotationnels RPS et RBS – sont faiblement complets, dans le sens où ils permettent de discriminer sur un ensemble ouvert et dense de fonctions à support compact $f \in L^2(S(2, N))$ à une action d'un élément de $SE(2, N)$ près.

Les aspects techniques, les tests comparatifs effectués et les résultats obtenus seront présentés dans le chapitre suivant.

Chapitre 4 : Résultats et expérimentations

1. Introduction

Les chapitres précédents ont abordé un des points clés essentiels pour réaliser la tâche de reconnaissance d'objets : l'extraction des caractéristiques. Le but de ce chapitre est d'évaluer les performances, dans un contexte de reconnaissance d'objets, des descripteurs proposés dans ce document.

De nombreuses applications de reconnaissance automatique d'objets sont proposées dans la littérature. Certaines sont spécifiquement destinées à des traitements temps réel tel que les travaux de Gould et al. [35] et Shotton et al. [72]. Les structures des systèmes de reconnaissance les plus communes utilisent un seul descripteur suivi d'un seul classifieur [18, 59, 86], d'autres systèmes font appel à des combinaisons associant plusieurs descripteurs et/ou classifieurs [66, 32, 30].

- Dans cette thèse, nous proposons d'utiliser la structure classique basée sur un classifieur. En ce qui concerne les descripteurs, nous mettons en place ceux dans le chapitre précédent, à savoir : les invariants spectraux (PS) et bi-spectraux (BS) généralisés et leurs homologues rotationnels (RPS) et (RBS), nous considérons aussi combinaison des deux descripteurs BS et RPS. En effet, combiner ces deux semble être un bon compromis entre le résultat théorique de la complétude donné théorème 4.4.3 (qui ne tient que pour le RBS) et les calculs, comme le montreront les résultats sur les différentes bases d'images. Pour les classifieurs, nous aurons recours méthode de classification supervisée à noyau : les machines à vecteurs de support

Annexe :

Machines à vecteurs de support (SVM)).

Dans l'objectif de comparer nos ensembles de descripteurs de Fourier aux principaux de la littérature, une série d'expérimentations complexes est menée. Cette dernière testera et comparera des descripteurs globaux tels que les moments de Zernike et de Hu, les invariants de Fourier-Mellin, des descripteurs locaux tels que les SIFT et les histogrammes de gradients orientés, et notre ensemble de descripteurs de Fourier {PS, BS, RPS, RBS, RPS & BS}. Les

performances de reconnaissance de ces différents descripteurs par rapport à l'invariance par rotation, la capacité de discrimination et la robustesse aux bruits, seront calculées.

2. Evaluation des descripteurs de Fourier

Pour évaluer les descripteurs de Fourier que nous avons définis dans la partie théorique (chapitre 3), nous mettons en place différentes expérimentations sur des bases d'images standards. Nous commencerons par étudier la capacité de discrimination chez ces descripteurs en les évaluant sur une base d'images synthétiques, ensuite nous comparerons leurs performances avec celles des descripteurs locaux et globaux en les testant sur des bases de visages comme RL⁷, CVL [75], ORL [64], Shieffield⁸ et Extended Yale b[31], et sur une base d'images d'objets variés larges comme COIL-100 [51]. Enfin, nous évaluerons ces descripteurs de Fourier pour une problématique de reconnaissance de navires dans un contexte de surveillance maritime.

2.1. Les bases d'images utilisées

a. Base d'images synthétiques

La première base d'images de test est composée des images de quelques formes géométriques simples (triangles, rectangles, étoiles, ellipses) et qui ont subi des rotations entre 0 à 2π (Figure 46)

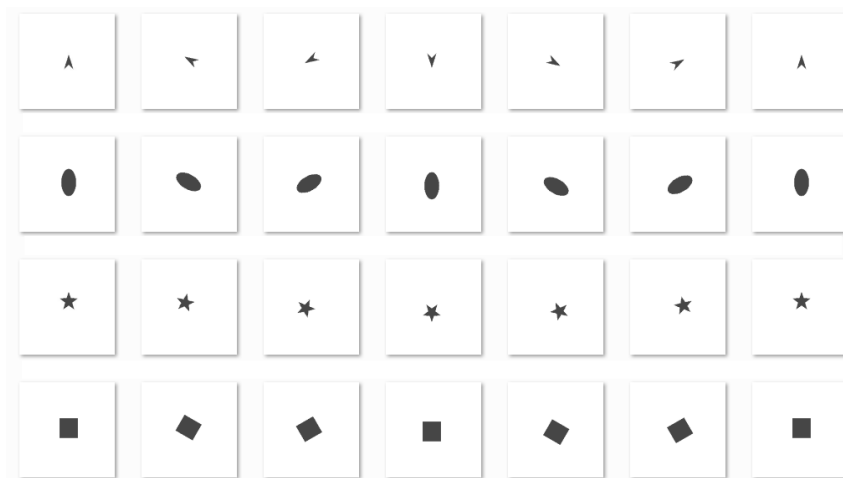


Figure 46 : Base des images synthétiques

⁷ http://robotics.csie.ncku.edu.tw/Databases/FaceDetect_PoseEstimate.htm#Our_Database_

⁸ <https://www.sheffield.ac.uk/eee/research/iel/research/face>

b. Base de visages RL

La base de visages RL a été construite par le laboratoire RL (Robotics Laboratory, Department of Computer Science and Information Engineering), basé à l'université nationale de Cheng Kung, Taiwan.

Elle contient 6660 images de 90 sujets. Chaque sujet a 74 images, où 37 images ont été prises tous les 5 degrés du profil droit (défini par +90°) au profil gauche (défini par -90°) dans le plateau de rotation (Figure 47). Les 37 images restantes sont générées (synthétisées) par les 37 images existantes à l'aide d'un logiciel commercial de traitement d'images afin d'obtenir les images symétriques. Les images suivantes constituent un exemple de cette base de données, avec des rotations de 0° (frontal) -> -90° (profil de gauche) (Figure 47) :



Figure 47: Extrait de la base RL. Pour chacune des 90 personnes enregistrées, on dispose de 74 vues avec des poses différentes

c. Base de visages CVL

La base CVL a été collectée par le chercheur Peter Peer, membre du Computer Vision Laboratory (CVL) à l'université de Ljubljana, Slovénie. Elle contient 114 personnes dont 90% sont de sexe masculin et avec un âge moyen de 18 ans, chacune étant enregistrée sous 7 vues différentes avec des expressions faciales différentes (Figure 48 et Figure 49). Les images sont en couleur de taille 640 × 480 pixels. Pour tous les sujets, les images ont été collectées dans des conditions d'éclairage uniformes avec des variations de poses et d'expressions faciales. Parmi ces vues, trois seulement sont de face.



Figure 48: Extrait de la base CVL. Ensembles de vues collectées pour un individu de la base



Figure 49: Trois expressions faciales extraites de la base CVL

d. Base de visages ORL

Cette base de données de visage a été créée au laboratoire AT&T, basé à Cambridge (Figure 50). Elle contient 40 personnes de sexe différent. Les images sont de taille 112×92 pixels. 10 vues différentes de chaque sujet ont été collectées (Figure 51). Ces vues présentent différentes poses et expressions faciales (expression neutre, sourire et yeux fermés) et des occlusions partielles par des lunettes, sous des conditions de luminosité variables.



Figure 50 : Base de visages ORL



Figure 51 : Exemples de visages enregistrés sous des vues différentes

e. Base de visages Sheffield

La base de visages Sheffield (auparavant UMIST) est composée de 564 images de 20 individus (race/genre/apparence mixtes), chacun représenté dans un intervalle de poses, allant des vues de profil aux vues frontales (Figure 52). Les images des visages sont sous le format PGM et sont de taille 200×200 pixels.



Figure 52: Base de visages Sheffield

f. Base de visages Extended Yale B

La base extended Yale B est une version étendue de la base de visages Yale B construite par l'université de Yale. Cette base est considérée comme la base standard pour l'évaluation de la robustesse des systèmes de biométrie faciale en cas de conditions d'illumination variables (Figure 53). Elle est composée de 16128 images faciales de 10 personnes, chacune enregistrées sous 9 poses et 64 conditions différentes d'éclairages (en total 576 images par

individu). Le format des images de cette base est le même que celui des images de la base d'origine Yale B : format GIF en niveaux de gris.



Figure 53: Base de visages Extended Yale B

g. Base d'images COIL-100

Coil-100 (Columbia Object Image Library) est composée de 7200 images couleur de taille 128×128 de 100 objets différents (Figure 54). Chaque image est composée d'un fond noir et chaque objet est pris sous 72 angles de vue différents (Figure 55). Cette base, utilisée dans des travaux similaires [74], peut être qualifiée de « facile » car le seul contenu fréquentiel est celui de l'objet d'intérêt. De plus, les objets ont souvent un contenu très géométrique.



Figure 54 : Les 100 objets de COIL-100

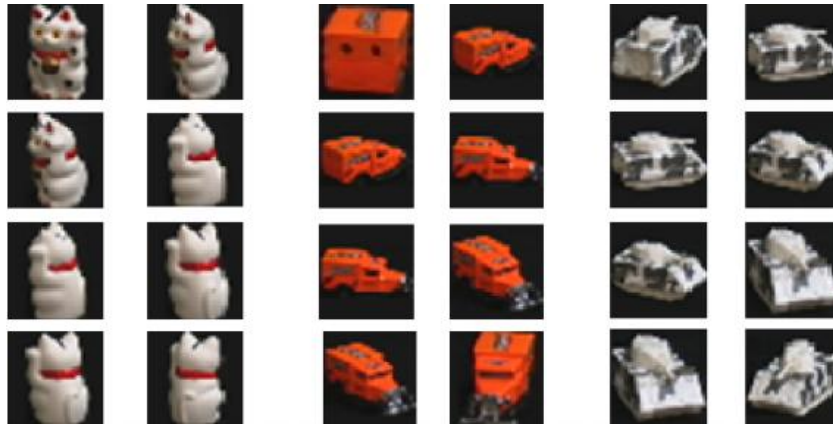


Figure 55 : Exemples d'objets vus sous des angles différents

2.2. Extraction des descripteurs

Comme démontré dans le théorème 5.1, l'égalité des descripteurs de Fourier que nous avons introduite, ne dépend pas du choix de l'ondelette mère ψ . Par conséquent, dans notre implémentation, nous avons calculé uniquement les quantités introduites dans le théorème 5.1, dont la complexité est réduite au calcul efficace du vecteur $\omega_f(\lambda)$, pour $\lambda \in S$ donnée. Nous rappelons que ce vecteur est obtenu par l'évaluation de la transformée de Fourier de f sur l'orbite de λ sous l'action des rotations discrètes R_{-k} pour $k \in \mathbb{Z}_N$.

D'un point de vue théorique, l'utilisation d'une ondelette mère spécifique permet d'améliorer les performances en termes de discrimination du descripteur. Les tests préliminaires que nous avons réalisés en utilisant une ondelette de Gabor, ont effectivement montré une amélioration sensible mais pour des coûts de calcul plus importants.

Pour l'implémentation, nous choisissons de considérer $N = 6$ et de travailler avec des images composées de pixels hexagonaux. Il y a deux raisons pour ce choix :

- Il est bien connu que les cellules rétinienne sont réparties sur une grille hexagonale, et il est donc raisonnable de supposer que les activations corticales reflètent ce fait.
- Les grilles hexagonales sont invariantes sous l'action de \mathbb{Z}_6 et les translations discrétisées. En effet, à part le réseau hexagonal, les seules autres grilles sur \mathbb{R}^2 qui sont invariantes par certaines \mathbb{Z}_N et par des translations discrètes appropriées, sont obtenues avec $N = 2,3,4$.

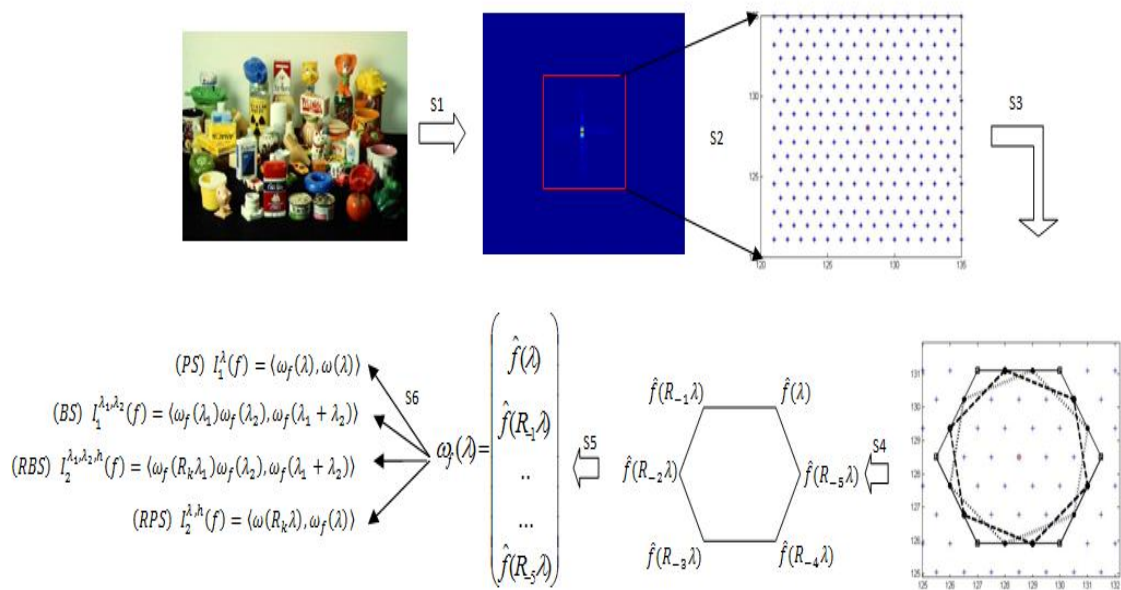


Figure 56 : Etapes de calcul des descripteurs de Fourier. (S1) calcul de la FFTSHIFT de l'image f , (S2) génération du maillage hexagonal, (S3) extraction des hexagons, (S4) évaluation de la FFT de f sur chaque hexagone extrait, (S5) génération du vecteur $\omega_f(\lambda)$ et (S6) calcul des quatres invariants

Les différentes étapes de calcul des descripteurs⁹ sont décrites dans la Figure 56 et données de la façon suivante :

1. L'image d'entrée est convertie en mode niveau de gris, la transformée de Fourier est calculée, et la composante de fréquence zéro est décalée au centre du spectre. (Figure 56 S1).
2. Pour des raisons de coût de calcul et puisque nous traitons des images naturelles, dont les fréquences concernées sont les basses, nous extrayons une grille de $16 * 16$ pixels autour de l'origine (Figure 56 S2).
3. Les invariants du théorème 5.1 sont calculés à partir des valeurs de la transformée de Fourier décalée (FFTSHIFT), sur toutes les fréquences dans un maillage hexagonal à l'intérieur de cette grille de taille $16 * 16$ pixels. Une interpolation bilinéaire est appliquée pour obtenir les valeurs correctes de $\omega_f(\lambda)$ (Figure 56 S3, S4, S5, S6). La dimension finale du vecteur descripteur est donnée dans le Tableau 1.

⁹ Exemple de code Matlab pour l'implémentation des invariants bi-spectraux rotationnels : https://nbviewer.jupyter.org/github/dprn/bispectral-invariant-svm/blob/master/Invariant_computation_matlab.ipynb

Descripteurs	Dimension
PS	136
BS	717
RPS	816
RBS	4417
RPS + BS	1533

Tableau 1 : Dimension des vecteurs descripteurs de Fourier considérés

2.3. Expérimentations

2.3.1. Protocole de test

Nous utilisons nos descripteurs de Fourier pour alimenter un classifieur de type SVM, en l'appliquant sur les différentes bases d'images. Finalement, nous comparons les résultats obtenus avec ceux obtenus via des descripteurs traditionnels.

Le résultat de l'étape d'apprentissage consiste en l'ensemble des vecteurs supports SVM. Durant l'étape de décision, le classifieur calcule les descripteurs de Fourier et le modèle déterminé durant l'étape d'apprentissage est utilisé pour réaliser la décision du SVM. La sortie est la classe image.

Nous évaluons séparément le taux de reconnaissance obtenu en utilisant les quatre descripteurs cités précédemment et la combinaison des deux invariants RPS et BS afin de tester leur complémentarité. Ensuite, nous comparons leur performance avec les moments de Hu (HM), les moments de Zernike (ZM), les descripteurs de Fourier-Mellin (FM), décrits dans l'état de l'art, nous avons comparé aussi avec des descripteurs locaux comme SIFT et HOG.

Le classifieur SVM que nous avons utilisé est basé sur un noyau Gaussien dont la taille σ est fixée empiriquement afin de maximiser le taux de reconnaissance.

Les performances des différents descripteurs invariants sont analysées en fonction du taux de reconnaissance. Par conséquent, pour un ratio donné, les ensembles d'apprentissage et de test ont été construits en divisant aléatoirement tous les exemples. Puis, en raison du caractère aléatoire de cette procédure, de multiples essais ont été effectués avec différents tirages au hasard de l'ensemble d'apprentissage et de test. Dans le cas où un bruit est ajouté, puisque comme mentionné précédemment l'ensemble d'apprentissage comprend toutes les images, cette procédure est appliquée uniquement à l'ensemble de test.

Les paramètres fixés pour nos expérimentations sont les suivant :

- L'ensemble d'apprentissage c_i correspondant aux valeurs d'un descripteur invariant calculé sur une image de la base de données.
- Les classes $\hat{c}_i \in \{1, 100\}$ correspondant à la classe d'objet (100 classes pour le cas de la base COIL-100).
- Nombre d'essais aléatoires : fixé à 5.
- Noyau K : un noyau gaussien de bande passante σ est choisi

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

x et y correspondent aux vecteurs descripteurs des objets.

Pour résoudre un problème multi-classes, les deux approches les plus populaires sont la méthode One-Against-All (OAA) et la méthode One-Against-One (OAO). Pour notre propos, nous avons choisi un multi-classifieur SVM basé sur la méthode OAO, car cette dernière est beaucoup plus rapide pour l'étape d'apprentissage et semble préférable pour les problèmes avec un très grand nombre de classes.

2.3.2. Tests sur la base des images synthétiques

Les tests sur la base d'images synthétiques ont été réalisés dans le but de valider la propriété de faible-complétude (i.e. capacité de discrimination) des descripteurs de Fourier définis dans le chapitre précédent {PS, BS, RPS, RBS, BS+RPS}. Tout d'abord, nous commençons par calculer ces différents descripteurs pour chaque image synthétique de la base, ainsi pour d'autres images réelles. Dans le but de comparer ces descripteurs, nous avons eu recours à la norme euclidienne classique qui calcule la distance entre deux vecteurs descripteurs.

Dans Figure 57, les notations {D-D60, D-D120, D-D180, D-D240, D-D300, D-D360, D-rose, etc.} signifient, respectivement, les distances euclidiennes entre le descripteur de l'image de l'objet original {Diamant, Triangle, Ellipse...} et ceux des images du même objet subissant des rotations de 0 à 2π , ainsi que les descripteurs des images réelles.

La Figure 57 montre bien que les normes euclidiennes entre le descripteur de l'image de l'objet original et ceux des images du même objet pivoté, sont presque égales et convergent vers 0, au contraire de celle entre le même descripteur et ceux des images réelle, qui divergent. En analysant les différents graphes de la figure, nous constatons aussi que les descripteurs BS et RBS sont légèrement plus faiblement complets que les descripteurs PS, RPS et la combinaison (BS,RPS).

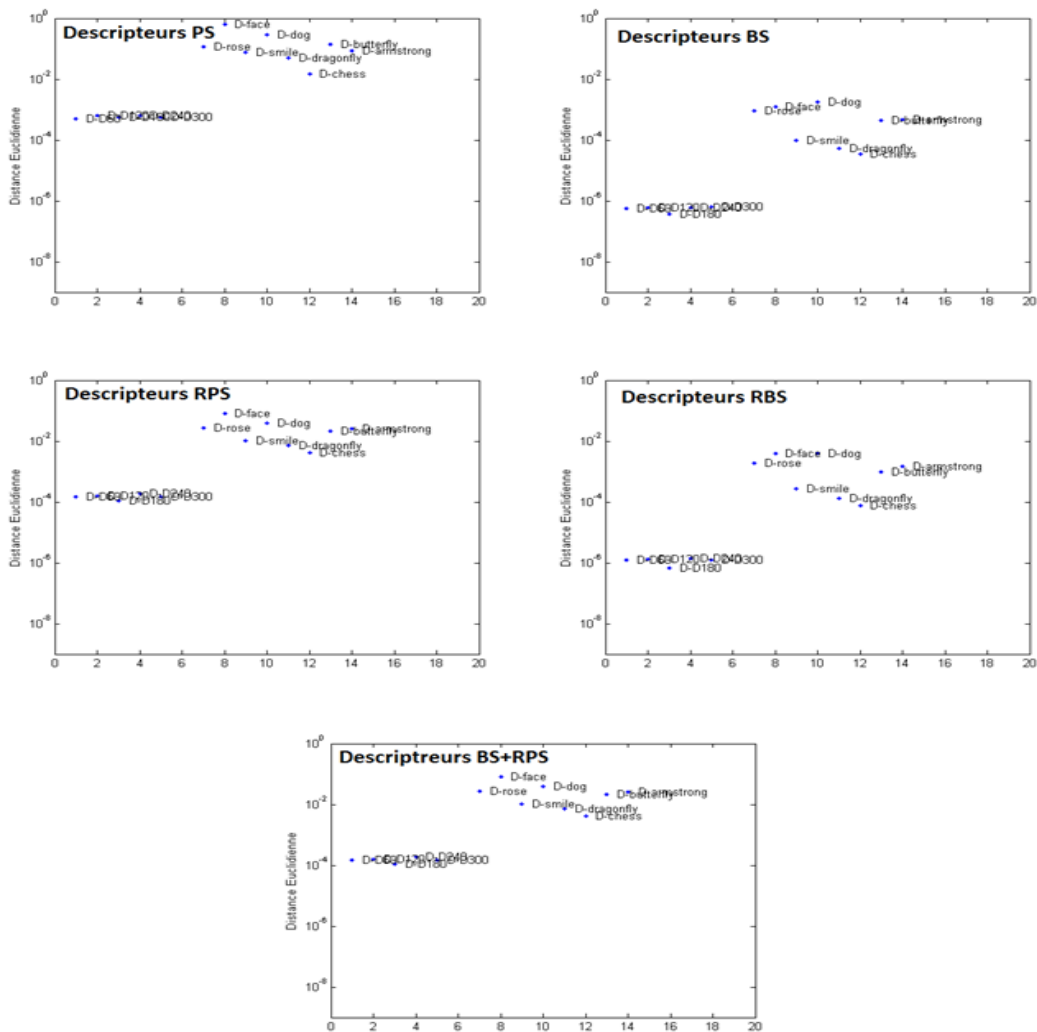


Figure 57: Test sur la base d'images synthétiques

2.3.3. Tests sur la base des visages RL

Pour la base de visages RL, les tests ont été réalisés en utilisant 75% d'images de visages pour l'apprentissage (environ 55 images par visage) et 25% pour le test (environ 19 images par visage).

Nous effectuons des expérimentations sur la base de données RL en utilisant l'ensemble des descripteurs globaux RBS, BS, PS, RPS, ZM, HU, FM, la combinaison du RPS & BS, et les deux fameux descripteurs locaux SIFT et HOG.

Le Tableau 2 présente les résultats obtenus en testant notre méthode de reconnaissance d'objets sur la base de visages RL. Le meilleur résultat a été obtenu en utilisant le descripteur RBS confirmant ainsi son invariance aux rotations, suivi des descripteurs SIFT et HOG et de la combinaison du RPS + BS.

Descripteurs	Taux de reconnaissance (%)
RBS	99.5
BS	91.4
PS	53
RPS	96.4
RPS + BS	97.6
ZM	95
HM	42.8
FM	51.5
HOG	98.8
SIFT	98.7

Tableau 2: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages RL

2.3.4. Tests sur la base des visages CVL

La base de visages CVL a été utilisée dans plusieurs travaux, sous des conditions d'apprentissage et de test différents. Elle est notamment utilisée par Albiol et al. dans [1] qui présentent un nouvel algorithme de reconnaissance de visages basé sur une combinaison des EBGM (Elastic Bunch Graph Matchnig) et des descripteurs HOG (Histogram of Oriented Gradients), et par Goal et al. dans [34], où les auteurs ont mené une étude de la faisabilité des RP (Random Projection) pour la reconnaissance de visages.

Les tests ont été réalisés en utilisant les mêmes descripteurs d'images exploités dans la base de visages RL. Le Tableau 3 montre que le descripteur RBS et la combinaison du RPS et BS, approchent les performances des descripteurs locaux SIFT et HOG, et donnent un meilleur résultat par rapport aux autres descripteurs globaux.

Descripteurs	Taux de reconnaissance (%)
RBS	97.5
BS	95
PS	93.5
RPS	96.7
RPS + BS	96
ZM	80.4
HM	65.2
FM	69
HOG	98.2
SIFT	98.4

Tableau 3: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages CVL

2.3.5. Tests sur la base des visages ORL

Dans la littérature, le protocole utilisé pour l'apprentissage et le test est différent d'un papier à l'autre. Dans [64], une approche basée sur le modèle de Markov caché (HMM) est utilisée, et le meilleur modèle a abouti à un taux de reconnaissance de 95%, avec un coût de calcul élevé. Dans [38], Hjelmas a atteint un taux de reconnaissance de 85% en utilisant la base de visages ORL et des vecteurs descripteurs constitués de coefficients de Gabor.

Les résultats sont présentés dans le Tableau 4, où l'on voit clairement que le descripteur RBS approche les descripteurs locaux et surpasse ceux globaux en terme de performances.

Descripteurs	Taux de reconnaissance (%)
RBS	92.5
BS	70
PS	50
RPS	83.5
RPS + BS	86.5
ZM	82
HM	50
FM	54
HOG	95
SIFT	92.5

Tableau 4: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages ORL

2.3.6. Tests sur la base de visages Sheffield

Dans le cas de la base de visages Sheffield, les expérimentations ont été réalisées en respectant le même protocole de test (c.à.d. 75% d'images pour l'apprentissage et 25% pour le test) et en utilisant les mêmes descripteurs testés dans le cas de la base de visages RL.

Le Tableau 5 montre que le descripteur RBS et la combinaison du RPS et BS égalent les performances du descripteur local SIFT, et surpassent celle du HOG et des autres descripteurs globaux.

Descripteurs	Taux de reconnaissance (%)
RBS	100
BS	90
PS	83
RPS	95
RPS + BS	100
ZM	92.5
HM	81.2
FM	85
HOG	90
SIFT	100

Tableau 5: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages Sheffield

2.3.7. Tests sur la base de visages Extended Yale B

Pour la base de visages Extended Yale B, nous refaisons les mêmes tests et nous réutilisons l'ensemble de descripteurs exploité dans les cas des bases de visages RL et Sheffield.

Le Tableau 6 présente les résultats obtenus en testant notre méthode de reconnaissance d'objets sur cette base de visages. Le meilleur résultat a été obtenu en utilisant le

descripteur local HOG, qui surpassent nettement les autres descripteurs, suivi des descripteurs SIFT et RBS qui donnent aussi un taux de reconnaissance acceptable.

Descripteurs	Taux de reconnaissance (%)
RBS	89.9
BS	70.3
PS	66.4
RPS	84.3
RPS + BS	87.4
ZM	82
HM	63.5
FM	68.1
HOG	98
SIFT	92.1

Tableau 6: Taux de reconnaissance pour chaque descripteur en utilisant la base de visages Extended Yale B

2.3.8. Tests sur la base des images COIL-100

Pour la base d'images COIL-100, deux cas ont été étudiés : un cas avec bruit et un autre sans bruit. Dans le premier cas, les tests ont été réalisés en utilisant 75% de la base d'images pour l'apprentissage (54 images par objet) et 25% pour le test (18 images par objet). Dans le second cas, nous avons utilisé un ensemble de données d'apprentissage composé de 6700 images (100 objets avec 67 vues pour chacun) non bruitées, et un ensemble de données de test composé de 15 vues par objet sélectionnées aléatoirement pour lesquelles un bruit gaussien avec $Std = 5, 10$ et 20 est ajouté (Figure 58).

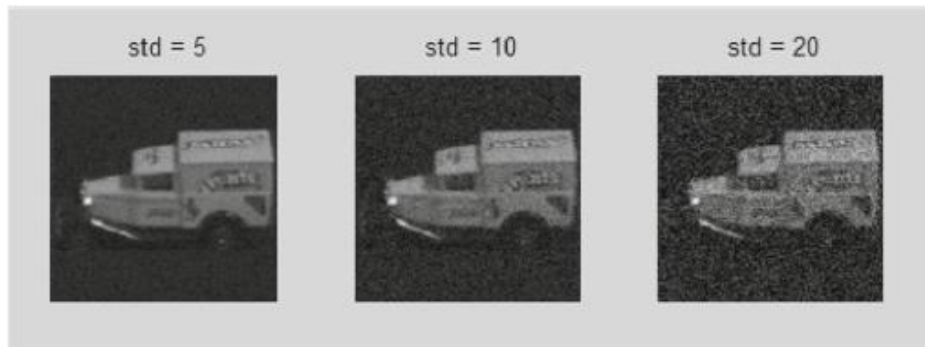


Figure 58 : Exemple d'images bruitées pour un objet de la base COIL-100

a. Cas simple

Le Tableau 7 présente les résultats obtenus en testant notre méthode de reconnaissance d'objets avec la base d'images COIL-100. Les meilleurs résultats ont été obtenus en utilisant le descripteur RBS. Le descripteur local SIFT vient en deuxième place et le HOG troisième. Alors qu'il a été démontré dans la littérature, que les méthodes locales donnent actuellement les meilleurs résultats [15]. De plus, si le bruit est ajouté sur l'image, l'utilisation d'une approche globale semble plus efficace que l'approche locale. La raison principale est que le détecteur de points-clés utilisé dans les méthodes locales produit dans ce cas de nombreux points-clés qui ne sont pas pertinents pour la reconnaissance d'objets. Cela sera montré dans la sous-section suivante.

Descripteurs	Taux de reconnaissance (%)
RBS	95.5
BS	88
PS	84.3
RPS	89.8
RPS + BS	92.8
ZM	91.9
HM	80.2
FM	89.6

HOG	95.3
SIFT	93.4

Tableau 7: Le taux de reconnaissance pour chaque descripteur en utilisant la base de données COIL-100 (cas non-bruité)

Dans la Figure 59, nous présentons le taux de reconnaissance en fonction de la taille de l'ensemble d'apprentissage. Comme prévu, il s'agit d'une fonction croissante et nous remarquons que les descripteurs HOG et SIFT, le RBS et la combinaison RPS + BS nécessitent moins de données d'apprentissage afin de donner un bon taux de reconnaissance, au contraire des descripteurs qui ont besoin d'une grande taille de données.

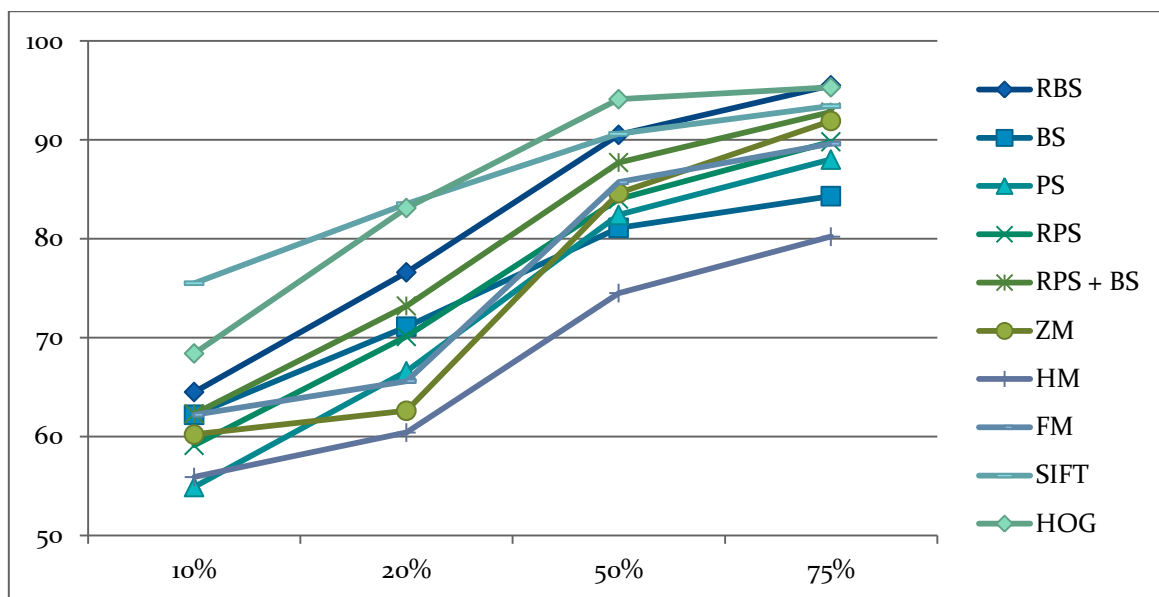


Figure 59: Taux de reconnaissance pour différentes tailles de la base d'apprentissage

b. Cas bruité

Les résultats présentés dans le Tableau 8 montrent que le bruit a peu d'influence sur la performance de classification quand on utilise un descripteur global tel que le RBS, BS, la combinaison du BS & RPS, ZM, HM et FM. Il a cependant une grande influence sur les descripteurs locaux SIFT et HOG.

Std	RBS (%)	BS (%)	PS (%)	RPS (%)	RPS+BS (%)	ZM (%)	HM (%)	FM (%)	SIFT (%)	HOG (%)	SIFT (%)
5	98.2	93.4	85.4	91.8	94.2	93.7	82	90.2	89.27	4	1.4
10	98.2	92.8	85	91.4	94	93.7	82	90.2	88.89	1.2	0.6
20	98.2	92	85	90.8	93.6	92.8	80.8	89	85.46	1	0.3

Tableau 8: Taux de classification pour le cas bruité de la base COIL-100

3. Application à la surveillance maritime

3.1. Navires subissant des rotations 2D

3.1.1. Présentation de la base de données

Cette base d'images de navires a été construite par les membres de l'équipe SIIM (Signal Image) du LSIS.

Elle contient 875 images de 125 navires. Chaque navire a 7 images, qui représentent des rotations sur le plan 2D (Figure 60) :



Figure 60: Exemples de bateaux qui ont subi des rotations 2D

3.1.2. Résultats obtenus

Vu le nombre petit d'images par navire (7 images par navire), nous avons choisi d'alimenter le classifieur SVM par un ensemble d'apprentissage composé de 6 images, et se contenter d'une seule image par navire pour le test. Les tests ont été réalisés en utilisant les mêmes descripteurs d'images exploités dans les cas des bases de visages RL, CVL et ORL.

Le Tableau 9 présente les résultats obtenus sur cette base et montre que les descripteurs SIFT, HOG, RBS et RPS+BS donnent les meilleurs résultats et sont par la suite les plus discriminants.

Descripteurs	Taux de reconnaissance (%)
RBS	98.6
BS	94
PS	91.2
RPS	97.1
RPS + BS	98.2
ZM	95
HM	84.3
FM	89.5
HOG	98.2
SIFT	99

Tableau 9: Résultats obtenus pour la base de navires 2D

3.2. Navires sous différents angles de vue

3.2.1. Présentation de la base de données

Les images de cette base ont été téléchargées depuis le site <https://www.marinetraffic.com/>. Nous avons pu collecter 2200 images de 44 navires, où

chaqu'un a été pris sous 50 vues différentes avec des arrière-plans complexes et des conditions d'éclairage qui changent d'une image à une autre (Figure 61) :

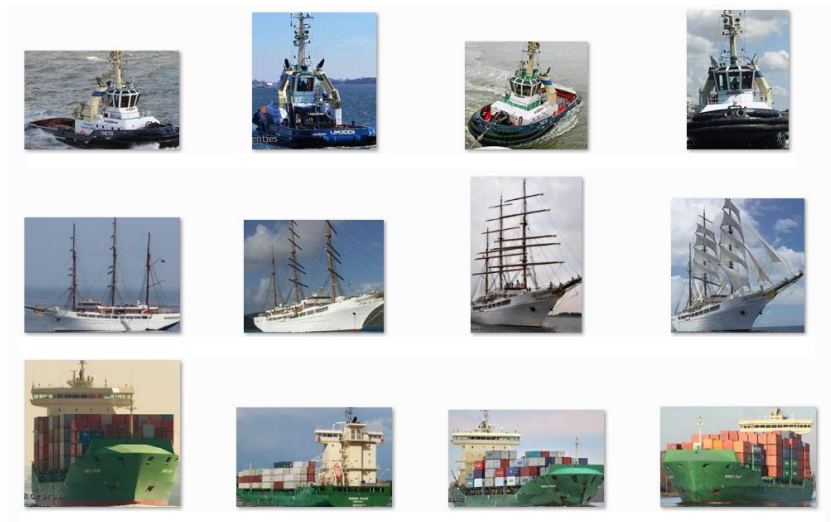


Figure 61: Exemples de bateaux qui ont subi des transformations générales

3.2.2. Résultats obtenus

Pour la base de navires « Marine Traffic », les tests ont été réalisés de la même manière que dans le cas de la base précédente, c.-à-d., nous avons pris une image parmi les 50 images par navire de façon aléatoire pour le test, et nous avons utilisé les 49 images restantes pour l'apprentissage. Les tests ont été réalisés en utilisant les mêmes descripteurs d'images exploités dans les cas des bases de visages RL, CVL et ORL.

Les résultats des expérimentations sur cette base sont donnés par le Tableau 10, qui montre que le descripteur RBS approche les performances du descripteur SIFT et discrimine mieux que le descripteur local HOG et les autres descripteurs globaux. La difficulté principale rencontrée lors de ces expérimentations était constituée par des arrière-plans complexes, qui comportent, par exemple, d'autres objets que celui que nous cherchons à reconnaître (arbres, bâtiments, autres navires ...), ou encore, des conditions météorologiques et de l'état de la mer variables, qui influencent légèrement sur la qualité de la description d'images dans un espace fréquentiel.

Descripteurs	Taux de reconnaissance (%)
RBS	83.5
BS	68.2
PS	61.2
RPS	72.7
RPS + BS	75
ZM	71
HM	37.8
FM	41.2
HOG	83.4
SIFT	88

Tableau 10: Résultats obtenus sur la base de navire "Marine Traffic"

4. Conclusion

Dans ce chapitre, nous avons exposé les performances obtenues par notre ensemble de descripteurs et nous les avons comparées à celle des descripteurs principaux (locaux et globaux) de la littérature. Tout d'abord, nous avons validé et confirmé la propriété de la faible-complétude de nos descripteurs de Fourier démontrée théoriquement dans le chapitre 2, par une série de tests que nous avons effectués sur des images synthétiques comportant des formes géométriques simples (triangles, ellipses, étoiles...). Ensuite, nous avons testé ces descripteurs et les avons comparés à d'autres descripteurs locaux et globaux, dans le cadre de la reconnaissance d'objet, en utilisant des bases de visages telles que RL, CVL, ORL, Sheffield, Extended Yale B, et des bases d'images d'objets variés comme COIL-100. Pour la base d'images COIL-100, deux cas ont été traités : un cas simple de reconnaissance d'objet et un cas bruité, où différentes images de la base ont subi des altérations différentes telles que l'ajout de bruit gaussien. Cette série d'expérimentations que nous avons menée, nous a permis de constater que la théorie était validée. Ainsi, nos descripteurs

de Fourier rotationnels fournissent des résultats supérieurs en moyenne à tous les autres descripteurs.

Dans un second temps nous avons testé nos descripteurs de Fourier dans un contexte de surveillance maritime, en les utilisant pour reconnaître des navires. Une fois encore nous avons pu montrer que nos descripteurs de Fourier rotationnels mis au point permettent d'obtenir de bons résultats de reconnaissance sur des bases d'images de navires ayant un arrière-plan complexe et pour des conditions d'acquisition variables.

Chapitre 5 : Conclusion et perspectives

1. Conclusion

Les travaux décrits dans cette thèse portent sur la définition et la mise en œuvre de nouveaux descripteurs de Fourier sur le groupe de roto-translations semi-discrètes $SE(2, N)$.

Nous avons commencé par présenter un état de l'art des différentes approches de description d'images. Les propriétés de chaque méthode ont été étudiées pour montrer leurs invariances aux différentes transformations géométriques et leurs robustesses aux altérations. Nous avons également présenté les avantages et inconvénients de ces différentes approches.

La deuxième partie concerne plus précisément notre contribution. Celle-ci débute par des rappels sur l'analyse harmonique dans les groupes abéliens et non-abéliens localement compacts. Ensuite, nous avons démontré que les invariants spectraux (PS) et bi-spectraux (BS) généralisés, et leurs homologues rotationnels (RBS) et (RPS), sont faiblement complets, c'est-à-dire qu'ils permettent de discriminer deux fonctions définies sur un ensemble ouvert et dense de fonctions à support compact $\varphi \in L^2(SE(2, N))$ à une action d'un élément de $SE(2, N)$ près. Cela généralise un résultat de [74]. Par la suite, nous avons utilisé ces descripteurs dans le cadre de la reconnaissance d'objets invariante aux roto-translations, inspiré par certaines propriétés neurophysiologiques du cortex visuel humain primaire V1. Dans ce cadre, nous avons montré que les invariants bi-spectraux rotationnels sont en effet des invariants aux roto-translations faiblement complets pour les images planaires. De plus, bien que les descripteurs de Fourier proposés soient définis en termes d'objets mathématiques complexes, nous avons montré qu'ils peuvent être implémentés de façon linéaire en tant que combinaisons linéaires des valeurs de la transformée de Fourier 2D de l'image.

Dans la deuxième partie de cette thèse, nous avons proposé une évaluation des performances de ces descripteurs de Fourier dans un contexte de reconnaissance d'objets et nous avons présenté les résultats obtenus sur différents bases de données : les bases de visages RL, CVL, ORL, Sheffield et Extended Yale B, sur lesquelles différents visages sont soumis à plusieurs types de variations ; la base d'images COIL-100, composées de plusieurs

objets soumis à des changements de rotations 3D et d'échelles. Pour toutes ces bases de données, les descripteurs de Fourier globaux introduits dans cette thèse sont, en moyenne, les descripteurs testés les plus efficaces. Bien que pour les images non altérées, les descripteurs de Fourier rotationnels approchent souvent les performances des descripteurs locaux SIFT et HOG et parfois donnent un meilleur taux de reconnaissance, l'addition de bruit et de différentes altérations donne toujours l'avantage aux descripteurs globaux.

Ces résultats montrent ainsi que l'invariant bi-spectral rotationnel (RBS) est un très bon descripteur de Fourier pour la reconnaissance d'objets, en cohérence avec le résultat théorique de la faible-complétude. Lorsque la dimension du vecteur descripteur de Fourier pose un problème, le RBS peut être remplacé par la combinaison des descripteurs bi-spectraux généralisés BS et des descripteurs spectraux rotationnels RPS, qui donne des résultats légèrement moins bons avec un vecteur descripteur de taille égale au tiers de celle du RBS.

2. Perspectives

Les travaux réalisés au cours de cette thèse nous ouvrent un large champ de perspectives.

D'une part, l'implémentation d'une architecture hexagonale en spirale (Figure 62) introduite par Sheridan [71, 70], à la place de la grille hexagonale classique utilisée dans nos expérimentations, peut améliorer le temps de calcul et réduire la taille de nos descripteurs de Fourier. L'architecture en spirale est un moyen d'indexer des hexagones de la grille avec un seul index qui permet d'introduire une opération, multiplication en spirale, qui, avec la même complexité d'une multiplication normale, calcule des rotations avec des multiples $\pi/3$.

Il existe des méthodes efficaces [23] pour simuler des pixels hexagonaux par un sur-échantillonnage de l'image par un ratio de 7 puis en utilisant ce que nous appelons des « hyperpels », composés de 56 pixels pour approximer un pixel hexagonal.

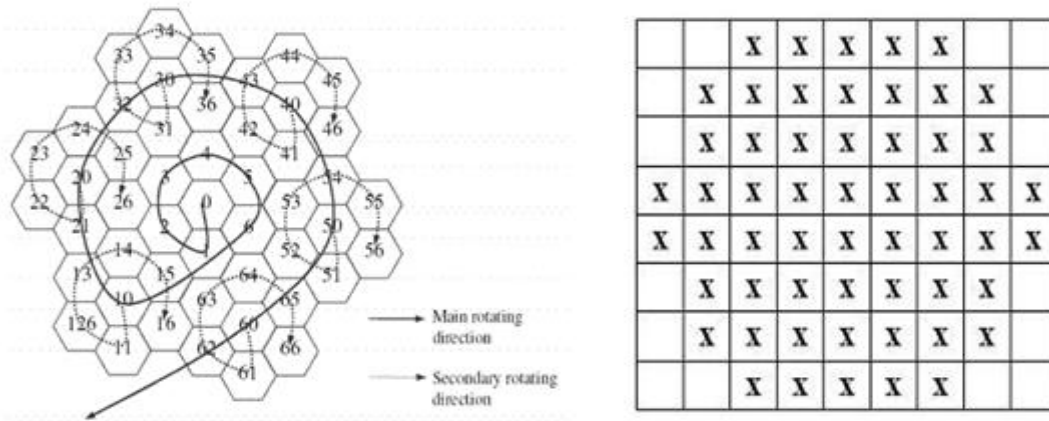


Figure 62: Principe de l'architecture hexagonale en spirale

Enfin, en s'inspirant de la méthode de détection d'objets, proposée par les chercheurs Paul Viola et Michael Jones en 2001 [84], nous envisageons d'associer ce type de descripteurs à un classifieur de type Adaboost (ou Adaptive Boosting) [26] qui semble très prometteur et bien adapté au problème de détection d'objets.

Annexes

A. Décomposition du produit tensoriel des représentations

Théorème 1.2. (Théorème d'Induction-Reduction). *Pout tout $\lambda_1, \lambda_2 \in \mathbb{R}^2 \setminus \{\hat{0}\}$, il en est que :*

$$T^{\lambda_1} \otimes T^{\lambda_2} \simeq \bigoplus_{k \in \mathbb{Z}_N} T^{\lambda_1 + R_k \lambda_2} \quad (14)$$

L'équivalence $A : L^2(\mathbb{Z}_N \times \mathbb{Z}_N) \rightarrow \bigoplus_{k \in \mathbb{Z}_N} L^2(\mathbb{Z}_N)$ dans le théorème (14) peut être explicitement calculée comme :

$$(AV)_k(h) = A_k V(h) = V(h, h - k), \quad \forall V \in L^2(\mathbb{Z}_N \times \mathbb{Z}_N) \quad (15)$$

Ici, $A_k : \bigoplus_{k \in \mathbb{Z}_N} L^2(\mathbb{Z}_N) \rightarrow L^2(\mathbb{Z}_N)$ est la composition de A avec la projection π_k sur le k -ième composant. L'inverse de A est son adjoint $A^* : \bigoplus_{k \in \mathbb{Z}_N} L^2(\mathbb{Z}_N) \rightarrow L^2(\mathbb{Z}_N \times \mathbb{Z}_N)$, donné par :

$$(A^*\psi)(i, j) = \psi_{i-j}(i) \quad \forall \psi = (\psi_h)_{h \in \mathbb{Z}_N} \in \bigoplus_{h \in \mathbb{Z}_N} L^2(\mathbb{Z}_N).$$

Soit $p_k : L^2(\mathbb{Z}_N) \rightarrow \bigoplus_{h \in \mathbb{Z}_N} L^2(\mathbb{Z}_N)$ l'inverse à droite de π_k , défini par :

$$(p_k v)_h = \begin{cases} v & \text{si } k = h, \\ 0 & \text{sinon.} \end{cases}$$

Par ailleurs, l'action d'un opérateur linéaire $B : \bigoplus_{k \in \mathbb{Z}_N} L^2(\mathbb{Z}_N) \rightarrow \bigoplus_{k \in \mathbb{Z}_N} L^2(\mathbb{Z}_N)$ est donnée par :

$$(B\psi)_k = \sum_{l \in \mathbb{Z}_N} B_{k,l} \psi_l \quad \forall \psi = (\psi_h)_{h \in \mathbb{Z}_N} \in \bigoplus_{h \in \mathbb{Z}_N} L^2(\mathbb{Z}_N),$$

Où $B_{k,l} = \pi_k \circ B \circ p_l$ sont les composants du bloc de B .

Dans ce qui suit, nous recueillons quelques faits utiles.

Proposition 1.3. *Soit A l'équivalence dans (15). Alors, on a ce qui suit :*

- *Pour tout opérateur linéaire $T : L^2(\mathbb{Z}_N \times \mathbb{Z}_N) \rightarrow L^2(\mathbb{Z}_N \times \mathbb{Z}_N)$ avec les composantes $T = (T_{i,j,r,s})_{i,j,r,s}$, l'opérateur $A \circ T \circ A^*$ a k, l composant de bloc :*

$$(A \circ T \circ A^*)_{k,l} = A_k \circ T \circ A^* \circ p_l = (T_{i,i-k,j,j-l})_{i,j \in \mathbb{Z}_N}.$$

En particulier, pour un couple d'opérateurs linéaires $B, C : L^2(\mathbb{Z}_N) \rightarrow L^2(\mathbb{Z}_N)$, il en est que :

$$(A \circ (B \otimes C) \circ A^*)_{k,l} = (B_{i,j} C_{i-k,j-l})_{i,j \in \mathbb{Z}_N}.$$

- Soit $\tilde{S}^j : \bigoplus_{k \in \mathbb{Z}_N} L^2(\mathbb{Z}_N) \rightarrow \bigoplus_{k \in \mathbb{Z}_N} L^2(\mathbb{Z}_N)$ définie par $(S^j \psi)_l = \psi_{l-j}$ pour tout $\psi = (\psi_h)_{h \in \mathbb{Z}_N}$ et $l \in \mathbb{Z}_N$. Alors :

$$A \circ S^j \otimes S^j \circ A^* = \tilde{S}^{j-j} \circ \bigoplus_{k \in \mathbb{Z}_N} S^j.$$

B. Lemmes auxiliaires pour la démonstration du théorème 3.3.1

Lemme 1. *L'ensemble \mathcal{G} introduit dans le théorème 3.3.1 est ouvert et dense dans $L^2(\mathbb{Z}_N \times K)$.*

Démonstration. Nous commençons par montrer que $\mathcal{G} \neq \emptyset$. Pour ce faire, il suffit de considérer f tel que $f(\cdot, k) \equiv 0$ pour tout $k \in \mathbb{Z}_N \setminus \{0\}$ et $f(\cdot, 0) \neq 0$ tel que $\text{supp} F(f(\cdot, 0)) = \mathbb{R}^2$. Par (4), nous avons alors $f \in \mathcal{G}$, comme

$$\det f(T^\lambda) = \prod_{k \in \mathbb{Z}_N} F(f(\cdot, 0))(R_{-k}\lambda) \neq 0 \quad \forall \lambda \in S.$$

Pour tout $f \in \mathcal{G}$ et $k \in \mathbb{Z}_N$, le théorème de Paley-Wiener implique que $F(f(\cdot, k))$ est analytique. En particulier, par (4), $\lambda \rightarrow \det f(T^\lambda)$ est analytique. Ainsi, $f \in \mathcal{G}$ si et seulement si $\exists \lambda_0 \in S$ tel que $f(T^{\lambda_0})$ est inversible.

Nous affirmons que l'ensemble \mathcal{G} est dense. En effet, soit $f \notin \mathcal{G}$ et fixons un certain $g \in \mathcal{G}$ et $\lambda_0 \in S$ tel que $\hat{g}(T^{\lambda_0})$ est inversible. Par l'analyticité de $\varepsilon \rightarrow \det(\hat{f}(T^{\lambda_0}) + \varepsilon \hat{g}(T^{\lambda_0}))$, il en résulte que $f + \varepsilon g \in \mathcal{G}$ pour un $\varepsilon > 0$ suffisamment petite, ce qui implique que $g \in \bar{\mathcal{G}}$, démontrant ainsi ce que nous avons affirmé.

Démontrons maintenant que \mathcal{G} est ouvert dans $L^2(\mathbb{Z}_N \times K)$. Pour ce faire, fixons $f \in \mathcal{G}$ et $f_n \rightarrow f$ dans $L^2(\mathbb{Z}_N \times K)$. Ceci implique que $\hat{f}_n \rightarrow \hat{f}$ dans $L^2(\widehat{SE(2, N)})$, et en particulier que $\hat{f}_n \rightarrow \hat{f}$ est en mesure. Par la définition de la convergence en mesure, ceci implique que

pour n suffisamment grand on doit admettre que $\hat{f}_n(T^{\lambda_0}) \neq 0$. D'où, $f_n \in \mathcal{G}$ pour n suffisamment grand et \mathcal{G} est ouvert. ■

Avant de plonger dans les démonstrations des autres lemmes auxiliaires, nous faisons l'observation suivante. Soit $\lambda_1, \lambda_2 \in I$ tel que $\lambda_1 + R_k \lambda_2 \in I$ pour tout $k \in \mathbb{Z}_N$. L'application du théorème d'Induction-Reduction (14) sur (8) donne :

$$\begin{aligned} A \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \circ A^{-1} \\ &= \bigoplus_{k \in \mathbb{Z}_N} f(T^{\lambda_1 + R_k \lambda_2})^{-1} g(T^{\lambda_1 + R_k \lambda_2}) \\ &= \bigoplus_{k \in \mathbb{Z}_N} U(T^{\lambda_1 + R_k \lambda_2}) \end{aligned} \quad (16)$$

Lemme 2. La fonction $\lambda \rightarrow U(T^\lambda)$ est continue sur I .

Démonstration. Fixons $\lambda_0 \in I$ et un ensemble ouvert $V \subset \mathbb{R}^2$ tel que

$$\int_V U(T^{\lambda_2})_{i,j}^* d\lambda_2 > 0 \quad \forall i, j \in \mathbb{Z}_N.$$

Ceci est possible puisque $U \not\equiv 0$. Puisque l'ensemble I est ouvert et dense, jusqu'à la réduction de V nous pouvons supposer qu'il existe un voisinage W de λ_0 tel que $V + \lambda \subset I$ pour tout $\lambda \in W$. Ensuite, (16) est valable pour $\lambda_1 \in W$ et $\lambda_2 \in V$. En calculant explicitement le bloc 0,0 de (16), nous avons

$$U(T^{\lambda_1})_{i,j} U(T^{\lambda_2})_{i,j}^* = U(T^{\lambda_1 + \lambda_2})_{i,j} \quad \forall i, j \in \mathbb{Z}_N.$$

Ensuite, en l'intégrant sur V w.r.t. λ_2 , ça donne

$$U(T^{\lambda_1})_{i,j} = \frac{\int_{V+\lambda} U(T^{\lambda_2})_{i,j} d\lambda_2}{\int_V U(T^{\lambda_2})_{i,j}^* d\lambda_2} \quad \forall \lambda_1 \in W, \forall i, j \in \mathbb{Z}_N.$$

Puisque la fonction sur le côté droit est clairement continue sur W , ceci prouve la continuité à λ_0 de $\lambda \rightarrow U(T^\lambda)$, complétant ainsi la démonstration. ■

Lemme 3. La fonction $\lambda \rightarrow U(T^\lambda)$ peut être étendue à une fonction continue sur \mathbb{R}^2 , pour laquelle (8) est toujours vraie.

Démonstration. Soit $\lambda_0 \notin I$. Puisque I est un ensemble ouvert et dense, ceci implique que λ_0 est dans sa fermeture et que nous pouvons choisir $\lambda_1, \lambda_2 \in I$ tel que $\lambda_0 = \lambda_1 + R_{k_0}\lambda_2$ pour un certain $k_0 \in \mathbb{Z}_N$ et $\lambda_1 + R_k\lambda_2 \in I$ pour tout $k \neq k_0$. Nous considérons ensuite

$$U(T^{\lambda_0}) := (A \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \circ A^*)^{k_0, k_0} \quad (17)$$

Nous montrons maintenant que la définition ci-dessus ne dépend pas du choix de λ_1, λ_2 et k_0 . Par l'ouverture de I , il existe un voisinage V de λ_2 entièrement contenu dans I . Puis, jusqu'à prendre un petit V , on admet que $\lambda_1 + R_{k_0}\lambda'_2 \in I$ pour tout $\lambda'_2 \in I \setminus \{\lambda_2\}$. Par (16), ceci implique que pour tout $\mu_1 + R_l\mu_2 = \lambda_0$, il en est que

$$(A \circ U(T^{\lambda_1}) \otimes U(T^{\lambda'_2}) \circ A^*)^{k_0, k_0} = (A \circ U(T^{\mu_1}) \otimes U(T^{\mu'_2}) \circ A^*)^{l, l}.$$

Pour λ'_2 et μ'_2 assez près, mais différents, de λ_2 et μ_2 , respectivement. Par la continuité de U sur I , démontrée dans le lemme 2, ceci implique que cette équation est valable aussi pour $\lambda'_2 = \lambda_2$ et $\mu'_2 = \mu_2$. D'où, (17) ne dépend pas du choix de λ_1, λ_2 et k_0 .

Finalement, le fait que $\hat{f}(T^{\lambda_1} \otimes T^{\lambda_2}) \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) = \hat{g}(T^{\lambda_1} \otimes T^{\lambda_2})$ pour tout λ_1, λ_2 résulte de (17) et (16).

■

Lemme 4. Il existe $a \in SE(2, N)$ tel que $U(T^\lambda) = T^\lambda(a)$.

Démonstration. Par la définition de U on admet que

$$\bigoplus_{k \in \mathbb{Z}_N} U(T^{\lambda_1 + R_k\lambda_2}) \circ A = A \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \quad \forall \lambda_1, \lambda_2 \neq 0.$$

Ensuite, pour tout i, j, l, k ,

$$U(T^{\lambda_1})_{l,i} U(T^{\lambda_1})_{l-k,j} = \begin{cases} U(T^{\lambda_1 + R_k\lambda_2})_{l,i} & \text{si } j = i - k \\ 0 & \text{sinon.} \end{cases} \quad (18)$$

Par l'inversibilité de $U(T^{\lambda_1})$, il existe $i_0 \in \mathbb{Z}_N$ tel que $U(T^{\lambda_1})_{0,i_0} \neq 0$. En utilisant (18) ceci implique que $U(T^{\lambda_2})_{-k,j} = 0$ pour tout $j \neq i_0 - k$. A savoir, nous avons démontré qu'il existe une famille de fonctions $f_{-k}: S \rightarrow \mathbb{C}$ tel que $U(T^{\lambda_1})_{-k,\cdot} = f_{-k}(\lambda_1)\delta_{i_0-k}$ ou, de manière équivalente, que

$$U(T^\lambda) = \text{diag}_k f_k(\lambda) S^{i_0}.$$

Par l'expression explicite (3) de T^λ , dans le but de compléter la démonstration, il suffit de démontrer que $\exists x_0 \in \mathbb{R}^2$ tel que $f_k(\lambda) = e^{i\langle x_0, R_{-k}\lambda \rangle}$.

Par la continuité et l'unitarité de U , les f_k 's sont continues et satisfont $|f_h(\lambda)| = 1$. En utilisant encore une fois (18) avec $j = i_0 - k$, nous obtenons

$$f_l(\lambda_1 + R_k \lambda_2) = f_l(\lambda_1) f_{l-k}(\lambda_2) \quad (19)$$

Pour tout $\lambda_1, \lambda_2 \neq 0$ et $l, k \in \mathbb{Z}_N$.

Nous affirmons que les f_l 's sont les caractères de \mathbb{R}^2 . En effet, fixons $k = 0$ dans (19) :

$$f_l(\lambda_1 + \lambda_2) = f_l(\lambda_1) f_l(\lambda_2) \quad (20)$$

Choisir $\lambda_2 = -\lambda_1$ dans ce qui est ci-dessus montre que f_l peut être étendue à 0. De plus, considérer $\lambda_1 = 0$ et prendre la limite $\lambda_2 \rightarrow 0$ montre que cette extension est continue. Puisque les caractères de \mathbb{R}^2 sont exactement les fonctions continues satisfaisant (20), l'affirmation est démontrée.

Par la dualité de Pontryagin, il existe $x_l \in \mathbb{R}^2$ tel que $f_l(\lambda) = e^{i\langle \lambda, x_l \rangle}$. Finalement, par (19) avec $k \in \mathbb{Z}_N$ on obtient $R_{-k} x_l = x_{l-k}$, qui démontre qu'il existe $x_0 \in \mathbb{R}^2$ tel que $f_l(\lambda) = e^{i\langle x_0, R_{-k}\lambda \rangle}$. Ceci complète la démonstration de l'énoncé.

■

C. Machines à vecteurs de support (SVM)

La plupart des systèmes de reconnaissance d'objet incluent une étape de classification. Nous avons choisi ici la fameuse et efficace approche SVM.

SVM est une machine d'apprentissage universelle (développée en particulier par Vladimir Vapnik [83, 12]). Une revue des principes de base suit, en considérant un problème de deux classes (quelque soit le nombre de classes, il peut être réduit à un problème de deux classes, via une approche One-Against-All ou One-Against-One [48]).

L'objectif des machines à vecteurs de support est de déterminer un classifieur ou une fonction de régression qui minimise le risque empirique (c'est-à-dire, l'erreur de l'ensemble d'apprentissage) et l'intervalle de confiance (qui correspond à l'erreur de généralisation ou de l'ensemble de test).

Vapnik et al. ont introduit des classifieurs SVM pour lesquels les surfaces de décision sont des hyperplans dans un espace K :

$$f(x) = \text{sgn}(\langle w, x \rangle + b) = 0, \text{ where } w \in K, b \in \mathfrak{R}$$

Etant donné un ensemble d'apprentissage de N exemples linéairement séparables $S = \{x_i \in \mathbb{R}^n \mid i = 1, 2, \dots, N\}$, où chaque exemple appartient à l'une des deux classes, représenté par $y_i \in \{+1, -1\}$, la méthode d'apprentissage SVM cherche l'hyperplan optimal $w \cdot x + b = 0$, comme surface de décision, qui sépare les exemples positifs et négatifs avec une plus grande marge. La fonction de décision pour classifier les données linéairement séparables est :

$$f(x) = \text{sign}(\langle w, x \rangle + b)$$

Où w et b sont obtenues à partir de l'ensemble d'apprentissage en résolvant un problème d'optimisation quadratique contraint. La fonction de décision finale est :

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i (x_i \cdot x) + b \right)$$

La fonction dépend des exemples d'apprentissage pour lesquels α_i est non-nulle. Ces exemples sont appelés les vecteurs de support. Souvent, le nombre de vecteurs de support n'est qu'une fraction de l'ensemble de données d'origine. La formulation SVM de base peut être étendue au cas non-linéaire en utilisant des noyaux non-linéaires qui mappent l'espace d'entrée à un espace caractéristique de grande dimension. Dans cet espace caractéristique de grande dimension, une classification linéaire peut être réalisée.

Les SVM ont été mis en place avec succès pour les tâches de détection et de reconnaissance de visages.

Bibliographie

- [1] Alberto Albiol, David Monzo, Antoine Martin, Jorge Sastre, and Antonio Albiol. Face recognition using hog+ebgm. *Pattern Recognition Letters*, 29(10):1537–1543, 2008.
- [2] Mohamed ATRI, Taoufik SAIDANI, and Rached TOURKI. Détection d'individu d'histogramme intégral.
- [3] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.
- [4] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer vision—ECCV 2006*, pages 404–417. Springer, 2006.
- [5] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *NIPS*, volume 2, page 3, 2000.
- [6] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.
- [7] Amine Bohi, Dario Prandi, Vincente Guis, Frédéric Bouchara, and Jean-Paul Gauthier. Fourier descriptors based on the structure of the human primary visual cortex with applications to object recognition. *Journal of Mathematical Imaging and Vision*, 57(1):117–133, 2017.
- [8] Ugo Boscain, Roman A Chertovskih, Jean-Paul Gauthier, and AO Remizov. Hypoelliptic diffusion and human vision: a semidiscrete new twist. *SIAM Journal on Imaging Sciences*, 7(2):669–695, 2014.
- [9] Ugo Boscain, Jean Duplaix, Jean-Paul Gauthier, and Francesco Rossi. Anthropomorphic image reconstruction via hypoelliptic diffusion. *SIAM Journal on Control and Optimization*, 50(3):1309–1336, 2012.
- [10] Ugo Boscain, Jean-Paul Gauthier, Dario Prandi, and Alexey Remizov. Image reconstruction via non-isotropic diffusion in dubins/reed-shepp-like control systems. In *53rd IEEE Conference on Decision and Control*, pages 4278–4283. IEEE, 2014.
- [11] Anna Bosch, Andrew Zisserman, and Xavier Muñoz. Scene classification via plsa. In *Computer Vision—ECCV 2006*, pages 517–530. Springer, 2006.
- [12] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.
- [13] Paul C Bressloff, Jack D Cowan, Martin Golubitsky, Peter J Thomas, and Matthew C Wiener. Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 356(1407):299–330, 2001.

- [14] Matthew Brown, Richard Szeliski, and Simon Winder. Multi-image matching using multi-scale oriented patches. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pages 510–517, Washington, DC, USA, 2005. IEEE Computer Society.
- [15] Anant Choksuriwong, Bruno Emile, Helene Laurent, and Christophe Rosenberger. Comparative study of global invariant descriptors for object recognition. *Journal of Electronic imaging*, 17(2):023015–023015, 2008.
- [16] Giovanna Citti and Alessandro Sarti. A cortical based model of perceptual completion in the roto-translation space. *Journal of Mathematical Imaging and Vision*, 24(3):307–326, 2006.
- [17] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
- [18] Navneet Dalal, Bill Triggs, and Cordelia Schmid. Human detection using oriented histograms of flow and appearance. In *Computer Vision—ECCV 2006*, pages 428–441. Springer, 2006.
- [19] John G Daugman. High confidence visual recognition of persons by a test of statistical independence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1148–1161, 1993.
- [20] Stéphane Derrode. *Représentation de formes planes à niveaux de gris par différentes approximations de Fourier-Mellin analytique en vue d'indexation de bases d'images*. PhD thesis, 1999.
- [21] Stéphane Derrode and Faouzi Ghorbel. Robust and efficient fourier–mellin transform approximations for gray-level image reconstruction and complete invariant description. *Computer Vision and Image Understanding*, 83(1):57–78, 2001.
- [22] Matthijs Douze, Hervé Jégou, Harsimrat Sandhawalia, Laurent Amsaleg, and Cordelia Schmid. Evaluation of gist descriptors for web-scale image search. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, page 19. ACM, 2009.
- [23] Shlomo Dubnov, Naftali Tishby, and Dalia Cohen. Polyspectra as measures of sound texture and timbre. *Journal of New Music Research*, 26(4):277–314, 1997.
- [24] Remco Duits and Erik Franken. Left-invariant parabolic evolutions on $se(2)$ and contour enhancement via invertible orientation scores part i: Linear left-invariant diffusion equations on $se(2)$. *Quarterly of Applied Mathematics*, pages 255–292, 2010.
- [25] Remco Duits and Erik Franken. Left-invariant parabolic evolutions on $se(2)$ and contour enhancement via invertible orientation scores part ii: Nonlinear left-invariant diffusions on invertible orientation scores. *Quarterly of applied mathematics*, pages 293–331, 2010.
- [26] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory*, pages 23–37. Springer, 1995.

- [27] Hartmut Führ and Matthias Mayer. Continuous wavelet transforms from semidirect products: Cyclic representations and plancherel measure. *Journal of Fourier Analysis and Applications*, 8(4):375–398, 2002.
- [28] Pierre F Gabriel, Jacques G Verly, Justus H Piater, and André Genon. The state of the art in multiple object tracking under occlusion in video sequences. In *Advanced Concepts for Intelligent Vision Systems*, pages 166–173. Citeseer, 2003.
- [29] Jean-Paul Gauthier, Guy Bornard, and Martine Silberman. Motions and pattern analysis: harmonic analysis on motion groups and their homogeneous spaces. *Systems, Man and Cybernetics, IEEE Transactions on*, 21(1):159–172, 1991.
- [30] Philip Geismann and Georg Schneider. A two-staged approach to vision-based pedestrian recognition using haar and hog features. In *Intelligent Vehicles Symposium, 2008 IEEE*, pages 554–559. IEEE, 2008.
- [31] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE transactions on pattern analysis and machine intelligence*, 23(6):643–660, 2001.
- [32] David Gerónimo, Antonio López, Daniel Ponsa, and Angel D Sappa. Haar wavelets and edge orientation histograms for on-board pedestrian detection. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 418–425. Springer, 2007.
- [33] Faouzi Ghorbel. A complete invariant description for gray-level images by the harmonic analysis approach. *Pattern recognition letters*, 15(10):1043–1051, 1994.
- [34] N Goal, George Bebis, and Ara Nefian. Face recognition experiments with random projection. In *Proceedings SPIE Vol*, volume 5779, pages 426–437, 2005.
- [35] Stephen Gould, Joakim Arfvidsson, Adrian Kaehler, Benjamin Sapp, Marius Messner, Gary R Bradski, Paul Baumstarck, Sukwon Chung, Andrew Y Ng, et al. Peripheral-foveal vision for real-time object recognition and tracking in video. In *IJCAI*, volume 7, pages 2115–2121, 2007.
- [36] M Hassaballah, Aly Amin Abdelmgeid, and Hammam A Alshazly. Image features detection, description and matching. In *Image Feature Detectors and Descriptors*, pages 11–45. Springer, 2016.
- [37] Edwin Hewitt and Kenneth A Ross. Preliminaries. In *Abstract harmonic analysis*, pages 1–15. Springer, 1963.
- [38] Erik Hjelmås and Boon Kee Low. Face detection: A survey. *Computer vision and image understanding*, 83(3):236–274, 2001.
- [39] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *information Theory, IRE Transactions on*, 8(2):179–187, 1962.
- [40] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat’s striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.
- [41] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

- [42] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. A sparse texture representation using local affine regions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1265–1278, 2005.
- [43] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–900. IEEE, 2002.
- [44] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.
- [45] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [46] Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 525–531. IEEE, 2001.
- [47] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10):1615–1630, 2005.
- [48] Jonathan Milgram, Mohamed Cheriet, and Robert Sabourin. “one against one” or “one against all”: Which one is better for handwriting recognition with svms? In *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft, 2006.
- [49] Greg Mori, Serge Belongie, and Jitendra Malik. Efficient shape matching using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11):1832–1837, 2005.
- [50] Ana Cris Murillo and J Kosecka. Experiments in place recognition using gist panoramas. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 2196–2203. IEEE, 2009.
- [51] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al. Columbia object image library (coil-20). 1996.
- [52] Lucas PJJ Noldus, Andrew J Spink, and Ruud AJ Tegelenbosch. Ethovision: a versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers*, 33(3):398–414, 2001.
- [53] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.
- [54] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.
- [55] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.

- [56] Aude Oliva and Antonio Torralba. Building the gist of a scene: The role of global image features in recognition. *Progress in brain research*, 155:23–36, 2006.
- [57] Patrick Ott and Mark Everingham. Implicit color segmentation features for pedestrian and object detection. In *ICCV*, pages 723–730, 2009.
- [58] M Oujaoura, B Minaoui, M Fakir, R El Ayachi, and O Bencharef. Recognition of isolated printed tiffinagh characters. *International Journal of Computer Applications*, 85(1), 2014.
- [59] Constantine P Papageorgiou, Michael Oren, and Tomaso Poggio. A general framework for object detection. In *Computer vision, 1998. sixth international conference on*, pages 555–562. IEEE, 1998.
- [60] Jean Petitot. *Neurogéométrie de la vision: modeles mathematiques et physiques des architectures fonctionnelles*. Editions Ecole Polytechnique, 2008.
- [61] Fatih Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 829–836. IEEE, 2005.
- [62] Dario Prandi, Ugo Boscain, and Jean-Paul Gauthier. Image processing in the semidiscrete group of rototranslations. In *International Conference on Networked Geometric Science of Information*, pages 627–634. Springer, 2015.
- [63] Ryszard Raczka. *Theory of group representations and applications*. Polish scientific Publishers, 1977.
- [64] Ferdinando S Samaria and Andy C Harter. Parameterisation of a stochastic model for human face identification. In *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pages 138–142. IEEE, 1994.
- [65] Frederik Schaffalitzky and Andrew Zisserman. Multi-view matching for unordered image sets, or “how do i organize my holiday snaps?”. In *European conference on computer vision*, pages 414–431. Springer, 2002.
- [66] Sam Schauland, Anton Kummert, Su-Birm Park, Uri Iurgel, and Yan Zhang. Vision-based pedestrian detection–improvement and verification of feature extraction methods and svm-based classification. In *2006 IEEE Intelligent Transportation Systems Conference*, pages 97–102. IEEE, 2006.
- [67] Cordelia Schmid and Roger Mohr. Local grayvalue invariants for image retrieval. *IEEE transactions on pattern analysis and machine intelligence*, 19(5):530–535, 1997.
- [68] Yunlong Sheng and Henri H Arsenault. Experiments on pattern recognition using invariant fourier–mellin descriptors. *JOSA A*, 3(6):771–776, 1986.
- [69] Yunlong Sheng and Jacques Duvernoy. Circular-fourier–radial-mellin transform descriptors for pattern recognition. *JOSA A*, 3(6):885–888, 1986.
- [70] Phil Sheridan, Tom Hintz, and David Alexander. Pseudo-invariant image transformations on a hexagonal lattice. *Image and Vision Computing*, 18(11):907–917, 2000.

- [71] Phillip Sheridan. *Spiral Architecture for machine vision*. PhD thesis, 1996.
- [72] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.
- [73] Laurent Sifre and Stéphane Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1233–1240, 2013.
- [74] Fethi Smach, Cedric Lematre, Jean-Paul Gauthier, Johel Miteran, and Mohamed Atri. Generalized fourier descriptors with applications to objects recognition in svm context. *Journal of Mathematical Imaging and Vision*, 30(1):43–71, 2008.
- [75] Franc Solina, Peter Peer, Borut Batagelj, Samo Juvan, and Jure Kovac. Color-based face detection in the " 15 seconds of fame" art installation. 2003.
- [76] Michael Reed Teague. Image analysis via the general theory of moments*. *JOSA*, 70(8):920–930, 1980.
- [77] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. A real-time algorithm for mobile robot mapping with applications to multi-robot and 3d mapping. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 1, pages 321–328. IEEE, 2000.
- [78] Sebastian Thrun et al. Robotic mapping: A survey. *Exploring artificial intelligence in the new millennium*, 1:1–35, 2002.
- [79] Engin Tola, Vincent Lepetit, and Pascal Fua. A fast local descriptor for dense matching. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [80] Engin Tola, Vincent Lepetit, and Pascal Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(5):815–830, 2010.
- [81] Emanuele Trucco and Konstantinos Plakas. Video tracking: a concise survey. *IEEE Journal of Oceanic Engineering*, 31(2):520–529, 2006.
- [82] Tinne Tuytelaars and Luc Van Gool. Matching widely separated views based on affine invariant regions. *International journal of computer vision*, 59(1):61–85, 2004.
- [83] Vladimir Naumovich Vapnik and Vlamimir Vapnik. *Statistical learning theory*, volume 1. Wiley New York, 1998.
- [84] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.
- [85] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

- [86] Paul Viola, Michael J Jones, and Daniel Snow. Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, 63(2):153–161, 2005.
- [87] Åke Wallin and Olaf Kübler. Complete sets of complex zernike moment invariants and the role of the pseudoinvariants. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (11):1106–1110, 1995.
- [88] Laurenz Wiskott, Jean-Marc Fellous, N Kuiger, and Christoph Von Der Malsburg. Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):775–779, 1997.
- [89] Jianxin Wu. *Visual place categorization*. PhD thesis, Georgia Institute of Technology, 2009.
- [90] Jianxin Wu and James M Rehg. Centrist: A visual descriptor for scene categorization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(8):1489–1501, 2011.
- [91] Baochang Zhang, Shiguang Shan, Xilin Chen, and Wen Gao. Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition. *Image Processing, IEEE Transactions on*, 16(1):57–68, 2007.
- [92] Dengsheng Zhang and Guojun Lu. Generic fourier descriptor for shape-based image retrieval. In *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, volume 1, pages 425–428. IEEE, 2002.
- [93] Dengsheng Zhang and Guojun Lu. Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, 17(10):825–848, 2002.
- [94] Dengsheng Zhang, Aylwin Wong, Maria Indrawan, and Guojun Lu. Content-based image retrieval using gabor texture features. In *IEEE Pacific-Rim Conference on Multimedia, University of Sydney, Australia*, pages 91–110, 2000.
- [95] Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang. Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 786–791. IEEE, 2005.

Publications dans le cadre de la thèse

Fourier Descriptors Based on the Structure of the Human Primary Visual Cortex with Applications to Object Recognition

Amine Bohi, Dario Prandi, Vincente Guis, Frédéric Bouchara & Jean-Paul Gauthier

Journal of Mathematical Imaging and Vision

ISSN 0924-9907
Volume 57
Number 1

J Math Imaging Vis (2017) 57:117-133
DOI 10.1007/s10851-016-0669-1



 Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Fourier Descriptors Based on the Structure of the Human Primary Visual Cortex with Applications to Object Recognition

Amine Bohi¹ · Dario Prandi² · Vincente Guis¹ · Frédéric Bouchara¹ · Jean-Paul Gauthier¹

Received: 8 September 2015 / Accepted: 14 June 2016 / Published online: 4 July 2016
© Springer Science+Business Media New York 2016

Abstract In this paper, we propose a supervised object recognition method using new global features and inspired by the model of the human primary visual cortex V1 as the semidiscrete roto-translation group $SE(2, N) = \mathbb{Z}_N \times \mathbb{R}^2$. The proposed technique is based on generalized Fourier descriptors on the latter group, which are invariant to natural geometric transformations (rotations, translations). These descriptors are then used to feed an SVM classifier. We have tested our method against the COIL-100 image database and the ORL face database, and compared it with other techniques based on traditional descriptors, global and local. The obtained results have shown that our approach looks extremely efficient and stable to noise, in presence of which it outperforms the other techniques analyzed in the paper.

Keywords Descriptor · Fourier transform · Hexagonal grid · Geometric transformations · Support vector machine · Object recognition

✉ Amine Bohi
amine.bohi@lsis.org
Dario Prandi
prandi@ceremade.dauphine.fr
Vincente Guis
guis@univ-tln.fr
Frédéric Bouchara
bouchara@univ-tln.fr
Jean-Paul Gauthier
gauthier@univ-tln.fr

¹ LISIS Laboratory, University of South Toulon - Var, B.P. 20132 83957, La Garde Cedex, France

² CEREMADE Laboratory, University of Paris-Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16, France

1 Introduction

Object recognition is a fundamental problem in computer vision and keeps attracting more and more attention nowadays. Its concepts have been applied in multiple fields, such as manufacturing, surveillance system, optical character recognition, face recognition, etc.

Almost every object recognition algorithm proposed in the literature is based on the computation of certain features of the image, which allow to characterize the object depicted and to discriminate it from others. In particular, since objects can appear at different locations and with different sizes, it is desirable for such features to be invariant by translation, rotation and scale. These invariant features can be global, i.e., computed taking into account the whole image, or local, i.e., computed considering only neighborhoods of key-points in the image.

In this paper we focus on *Fourier descriptors*, an important class of global invariant features used since the seventies [19, 40] based on algebraic properties of the Fourier transform. In particular, inspired by some neurophysiological facts on the structure of the human primary visual cortex, we extend this theory to define features invariant to translation and rotation and we apply them for invariant object recognition in SVM context. These results are then compared with those obtained with another important class of global invariant features, the *moment invariants* (see Appendix 1), used e.g., in [10, 34, 35], and with two different local invariants. For more information on object recognition via local features we refer to [2, 12, 25, 27, 28, 30].

Our choice of a global approach is motivated by the better results obtained by these methods in presence of noise, luminance changes and other different alterations, with respect to algorithms based on local invariant features [7]. Indeed, under these conditions, key-points detectors used in the local

approach produce key-points that are not relevant for the object recognition.

In the following, we briefly introduce the theory of Fourier descriptors, before discussing the framework used in this paper and our contributions.

1.1 Fourier Descriptors

The basic idea behind Fourier descriptors is that the action of an abelian locally compact group G on functions in $L^2(G)$ is much easier to treat at the level of their Fourier transforms. In the specific case of images, $f, g \in L^2(\mathbb{R}^2)$, this is expressed by the well-known equivalence for the translation of $a \in \mathbb{R}^2$:

$$\begin{aligned} f(x) &= g(x - a) \quad \forall x \in \mathbb{R}^2 \\ \iff \hat{f}(\lambda) &= e^{i\langle a, \lambda \rangle} \hat{g}(\lambda) \quad \forall \lambda \in \mathbb{R}^2, \end{aligned} \quad (1)$$

where the Fourier transform is defined¹ by

$$\hat{f}(\lambda) = \int_{\mathbb{R}^2} f(x) e^{-i\langle \lambda, x \rangle} dx, \quad \forall \lambda \in \mathbb{R}^2.$$

In this setting, Fourier descriptors are quantities associated with functions of $L^2(\mathbb{R}^2)$ that can be easily computed starting from their Fourier representation and that are invariant under the action of translations. Ideally, a Fourier descriptor should be **complete**, meaning that for any couple of images $f, g \in L^2(\mathbb{R}^2)$ the equality of the Fourier descriptor is equivalent to the equality of f and g up to translations. Indeed, the lack of completeness could yield to problems in applications, notably to false positives in the classification.

However, a result as strong as completeness is usually out of reach and unnecessary for practical applications. In this case, one looks for Fourier descriptors that are at least **weakly complete**, meaning that they are complete on a sufficiently big subset of $L^2(\mathbb{R}^2)$, usually either open and dense or at least residual, i.e., the intersection of countably many open and dense sets. This guarantees that the Fourier descriptor will correctly classify a sufficiently large class of images.

Various Fourier descriptors have been defined in the literature [19, 24, 26, 38, 40]. In this work we are mainly interested in the following two, whose invariance w.r.t. translations can be checked via (1).

- **Power-spectrum:** the quantity $PS_f(\lambda) := |\hat{f}(\lambda)|^2$ for $\lambda \in \mathbb{R}^2$, which is the Fourier transform of the auto-correlation function

$$a_f(x) := \int_{\mathbb{R}^2} \overline{\hat{f}(y)} \hat{f}(y + x) dy.$$

¹ Here we use a non-unitary definition of the Fourier transform for future convenience in computations.

It is easy to show that the power-spectrum is not weakly complete, and indeed it is used in texture synthesis to identify the translation invariant Gaussian distribution of textures [18].

- **Bispectrum:** an extension of the power-spectrum, it is the quantity $BS_f(\lambda_1, \lambda_2) := \hat{f}(\lambda_1) \hat{f}(\lambda_2) \overline{\hat{f}(\lambda_1 + \lambda_2)}$, or equivalently the Fourier transform of the triple correlation, defined as

$$a_{3,f}(x_1, x_2) := \int_{\mathbb{R}^2} \overline{\hat{f}(y)} \hat{f}(y + x_1) \hat{f}(y + x_2) dy.$$

These descriptors are complete on compactly supported functions of $L^2(\mathbb{R}^2)$ and are well established in statistical signal processing. See e.g., [14], where they are applied to sound texture recognition.

These two Fourier descriptors can be easily generalized to functions on $L^2(G)$ of a locally compact abelian group G to obtain invariants under the action of G . This can be applied, for example, to 2D shape recognition. However, when working with images, these descriptors are unsatisfying. Indeed, they are invariant only under translations, and so cannot be used to classify images under the action of rotations.

1.2 Framework of the Paper

In this paper, following a line of research started in [38], we present a theoretical framework that allows us to build generalized Fourier descriptors which are invariant w.r.t. (semidiscrete) roto-translations of images. We exploit the following two facts:

- It is possible to define a natural generalization of the power-spectrum and the bispectrum on non-commutative groups, as it has been done in [24, 38].
- Contributions of some of the authors to a fairly recent model of the human primary visual cortex V1 [4, 5] have shown that the latter can be modeled as the semidiscrete group of roto-translations $SE(2, N) = \mathbb{Z}_N \times \mathbb{R}^2$. In this model, cortical stimuli are functions in $L^2(SE(2, N))$, w.r.t. the Haar measure of $SE(2, N)$, and images from the visual plane are lifted to cortical stimuli via a natural injective and left-invariant lift operation $\mathcal{L} : L^2(\mathbb{R}^2) \rightarrow L^2(SE(2, N))$. Such lift is defined as the wavelet transform w.r.t. a mother wavelet Ψ , see Sect. 2.

From these facts, a natural pipeline for invariant object recognition is the following:

1. Given an image $f \in L^2(\mathbb{R}^2)$ lift it to a cortical stimulus $\mathcal{L}f \in L^2(SE(2, N))$.

2. Compute the generalized Fourier descriptors of $\mathcal{L}f$ on the non-commutative group $SE(2, N)$.
3. If the lift of another image $g \in L^2(\mathbb{R}^2)$ has the same Fourier descriptors as $\mathcal{L}f$, deduce that $\mathcal{L}f \approx \mathcal{L}g$ up to the action of $SE(2, N)$.
4. Thanks to the left-invariance and injectivity of the lift \mathcal{L} , obtain that also $f \approx g$ up to the action of $SE(2, N)$.

This pipeline was already investigated in [38], where the authors considered a non-left-invariant lift, the *cyclic lift*. For this lift they then proved a weak completeness result of the generalized bispectrum for images, represented as $L^2(\mathbb{R}^2)$ functions with support inside a fixed compact set.

In this paper, we consider the same question for left-invariant lifts, where the situation turns out to be more complicated. In particular, as explained in the following section, to ensure the weak completeness we are led to consider “stronger” invariants than the generalized bispectrum. However, as observed in Remark 4, the actual computation of these stronger invariants on lifted images requires N times less computational time and space w.r.t. the computation of the generalized bispectrum of cyclically lifted images.

1.3 Contributions of the Paper

Let $K \subset \mathbb{R}^2$ be any compact set, representing the size of the images under consideration. According to the pipeline for object recognition introduced above, the weak completeness of the generalized Fourier descriptors on images can be proved in two steps:

1. Prove the completeness of the generalized Fourier descriptors on some residual set $\mathcal{G} \subset L^2(\mathbb{Z}_N \times K)$ of cortical stimuli;
2. Prove that for some residual set $\mathcal{R} \subset L^2(\mathbb{R}^2)$ of images with support in K we have $\mathcal{L}(\mathcal{R}) \subset \mathcal{G}$.

The first point is addressed in Theorem 1, where is identified an open and dense set $\mathcal{G} \subset L^2(\mathbb{Z}_N \times K)$ on which the combination of the generalized power-spectrum and bispectrum holds. This generalizes the result in [38], where the same result was proved for a residual subset of the range of the cyclic lift.

Unfortunately, it turns out that for this set \mathcal{G} and a left-invariant lift \mathcal{L} there is no hope of finding a set $\mathcal{R} \subset L^2(\mathbb{R}^2)$ satisfying the second point above. We are then led to introduce stronger Fourier descriptors, the rotational power-spectrum and bispectrum, which are invariant only w.r.t. rotations. To solve this problem we preprocess images by centering them at their barycenter, a procedure that is essential also in [38]. Theorem 2 then shows that the resulting invariants are complete for an open and dense set of functions in $L^2(K)$, for any compact $K \subset \mathbb{R}^2$. The proof of this com-

pleteness requires fine technical tools from harmonic analysis and the theory of circulant operators, and for this reason we only present a sketch of it, evidencing the technical difficulties. A complete proof will be presented in a forthcoming paper by the second and last authors.

Finally, in Theorem 3 we show that, under mild assumptions on the mother wavelet Ψ , to check the equality of all these Fourier descriptors it is sufficient to compute simple quantities computed from the 2D Fourier transform of the image. This allows for an efficient implementation on regular hexagonal grids. After using these descriptors to feed an SVM-based classifier, we compare their performances with those of Hu and Zernike moments, the Fourier–Mellin transform and some well-known local descriptors. To this purpose, we test them on two large databases: the COIL-100² object recognition database, composed of 7200 objects presenting rotation and scale changes [31], and the ORL³ face database, on which different human faces are subject to several kinds of variations.

1.4 Structure of the Paper

The remainder of the paper is organized as follows. In Sect. 2, we present the features of a mathematical model of the primary visual cortex V1 that are essential to our approach. In Sect. 3, we introduce some generalities on the Fourier Transform on the semidiscrete group of roto-translation $SE(2, N)$. In Sect. 4, we describe the natural generalization of the power-spectrum and the bispectrum on \mathbb{R}^2 to $SE(2, N)$. We then prove the weak completeness result (Theorem 1) and show that under the chosen lift operator this does not imply weak completeness for images. Finally, we introduce the rotational power-spectrum and bispectrum and sketch the proof of the corresponding weak completeness result (Theorem 2) for images. We end this section with some results on the practical computation of these descriptors. In Sect. 5, we illustrate some numerical results where these descriptors are compared with those obtained via global descriptors such as Zernike moments, Hu moments, Fourier–Mellin transform, and local ones like the SIFT and HoG descriptors. Finally, we conclude with some practical suggestions in Sect. 6.

2 A Mathematical Model of the Primary Visual Cortex V1

As mentioned in the introduction, the main novelty of our approach is its connection with a fairly recent model of the human primary visual cortex V1 due to Petitot and Citti-Sarti [11, 32] and our recent contributions [3–5, 33]. The theory

² <http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php>.

³ <http://www.cl.cam.ac.uk/research/dtp/attarchive/facedatabase.html>.

of orientation scores introduced in [15, 16] is also strongly connected with this work, in particular for its exploitation of left-invariant lift operators. We also mention [37], where image invariants based on the structure of the roto-translation group $SE(2)$ are introduced for textures. In this section, we present the features of this model that are essential to our approach.

Since it is well-known [23] that neurons in V1 are sensitive not only to positions in the visual field, but also to local orientations and that it is reasonable to assume these orientations to be finite, in [4] V1 has been modeled as the semidiscrete group of roto-translations $SE(2, N) = \mathbb{Z}_N \times \mathbb{R}^2$ for some even $N \in \mathbb{N}$. Letting R_k be the rotation of $2\pi/k$, the (non-commutative) group operation of $SE(2, N)$ is

$$(x, k)(y, r) = (x + R_k y, k + r).$$

Here, we are implicitly identifying $k + r$ with $k + r \pmod N$.

Visual stimuli $f \in L^2(\mathbb{R}^2)$ are assumed to be lifted to activation patterns in $L^2(SE(2, N))$ by a lift operator $\mathcal{L} : L^2(\mathbb{R}^2) \rightarrow L^2(SE(2, N))$. Motivated by neurophysiological evidence, we then assume that

(H) The lift operator \mathcal{L} is linear and is defined as

$$\mathcal{L}f(x, k) := \int_{\mathbb{R}^2} f(y)\bar{\psi}(R_{-k}(y-x))dy, \quad (2)$$

for a given mother wavelet $\psi \in L^2(\mathbb{R}^2)$ such that \mathcal{L} is injective and bounded.

Remark 1 This assumption means that the lift operator under consideration is the wavelet transform w.r.t. ψ (see e.g., [17]). The fact that \mathcal{L} be injective and bounded is then equivalent to the fact that the mother wavelet ψ is weakly admissible, i.e., is such that the map

$$\lambda \in \mathbb{R}^2 \mapsto \sum_{k \in \mathbb{Z}_N} |\hat{\psi}(R_{-k}\lambda)|^2$$

is strictly positive and essentially bounded.

As a consequence of the above assumption, the lift operation \mathcal{L} is left-invariant w.r.t. to the action of $SE(2, N)$. Namely,

$$\Lambda(x, k) \circ \mathcal{L} = \mathcal{L} \circ \pi(x, k). \quad (3)$$

Here Λ and π are the actions of $SE(2, N)$ on $L^2(SE(2, N))$ and $L^2(\mathbb{R}^2)$, respectively. That is,

$$\begin{aligned} [\Lambda(x, k)\varphi](y, r) &= \varphi\left((x, k)^{-1}(y, r)\right) \\ &= \varphi(R_{-k}(y-x), r-k), \\ [\pi(x, k)f](y) &= f\left((x, k)^{-1}y\right) = f(R_{-k}(y-x)). \end{aligned}$$

Formula (3) can be seen as a semidiscrete version of the shift-twist symmetry [6].

The main observation for our purposes is that (3) means that two images f and $g \in L^2(\mathbb{R}^2)$ can be deduced via roto-translation (i.e., $f = \pi(x, k)g$ for some $(x, k) \in SE(2, N)$) if and only if their lifts can be deduced via $\Lambda(x, k)$.

3 Preliminaries on Non-commutative Harmonic Analysis

In this section we introduce some generalities on the (non-commutative) Fourier transform on $SE(2, N)$, an essential tool to define and compute the Fourier descriptors we are interested in. We refer to [1, 20] for a general introduction to the topic.

Since $SE(2, N)$ is a non-commutative unimodular group, the Fourier transform of $\varphi \in L^2(SE(2, N))$ is an operator associating to each (continuous) irreducible unitary representations T^λ of $SE(2, N)$ some Hilbert-Schmidt operator on the Hilbert space where T^λ acts. Here, λ is an index taking values in the dual object of $SE(2, N)$, which is denoted by $\widehat{SE(2, N)}$ and is the set of equivalence classes of irreducible unitary representations.

The set of irreducible representations of a semi-direct product group can be obtained via Mackey's machinery (see e.g., [1, Ch. 17.1, Theorems 4 and 5]). Accordingly, $\widehat{SE(2, N)}$ is parametrized by the orbits of the (contragredient) action of rotations $\{R_k\}_{k \in \mathbb{Z}_N}$ on \mathbb{R}^2 , i.e., by the slice $S \subset \mathbb{R}^2$ which in polar coordinates is $(0, +\infty) \times \{0, 2\pi/N\}$. Additionally, corresponding to the origin, there are the characters of \mathbb{Z}_N . Namely, to each $\lambda \in S$ corresponds the representation T^λ acting on \mathbb{C}^N via

$$\begin{aligned} T^\lambda(x, k)v &= \text{diag}_{\mathbb{R}^2}(e^{i(\lambda, R_k x)}) \circ S^k v \\ &= \left(e^{i(\lambda, R_k x)} v_{h+k} \right)_{h=0}^{N-1}. \end{aligned} \quad (4)$$

where we denoted by $\text{diag}_{\mathbb{R}^2} v_h$ the diagonal matrix of diagonal $v \in \mathbb{C}^N$ and by S the shift operator $(Sv)_j = v_{j+1}$, so that $(S^k v)_j = v_{j+k}$. On the other hand, to each $k \in \mathbb{Z}_N$ corresponds the representation on \mathbb{C} given by $z \mapsto e^{i\frac{2\pi k}{N}} z$. Since it is possible to show that to invert the Fourier transform it is enough to consider only the representations parametrized by S , we will henceforth ignore the \mathbb{Z}_N part of the dual.

Finally, the matrix-valued Fourier coefficient of a function $\varphi \in L^2(SE(2, N)) \cap L^1(SE(2, N))$ for $\lambda \in \widehat{SE(2, N)}$ is

$$\widehat{\varphi}(T^\lambda) = \int_{SE(2,N)} \varphi(a) T^\lambda(a^{-1}) da, \tag{5}$$

where da is the Haar measure⁴ of $SE(2, N)$. This is essentially the same formula for the Fourier transform on \mathbb{R}^2 , which is a scalar and is obtained using the representations $\lambda(x) = e^{i(x,\lambda)}$ acting on \mathbb{C} .

Straightforward computations yield

$$\widehat{\varphi}(T^\lambda)_{i,j} = \mathcal{F}(\varphi(\cdot, i - j))(R_{-j}\lambda), \tag{6}$$

where we let \mathcal{F} denote the Fourier transform on \mathbb{R}^2 .

As usual, the definition of the Fourier transform can be extended to the whole $L^2(SE(2, N))$ by density arguments. Then, there exists a unique measure on $\widehat{SE(2, N)}$, the Plancherel measure, supported on \mathcal{S} where it coincides with the restriction of the Lebesgue measure of \mathbb{R}^2 , such that the Fourier transform is an isometry between $L^2(SE(2, N))$ and $L^2(\widehat{SE(2, N)})$. In particular, the following inversion formula holds

$$\varphi(x, k) = \int_{\mathcal{S}} \text{Tr}(\widehat{\varphi}(T^\lambda) \circ T^\lambda(x, k)) d\lambda.$$

The fundamental property of the non-commutative Fourier transform, generalizing (1), is that for all $\varphi, \eta \in L^2(SE(2, N))$ and $a \in SE(2, N)$ it holds

$$\begin{aligned} \varphi(x, k) &= [A(a)\eta](x, k) \quad \forall (x, k) \in SE(2, N) \iff \\ \widehat{\varphi}(T) &= \widehat{\eta}(T) \circ T^{-1}(a) \quad \forall T \in \widehat{SE(2, N)}. \end{aligned} \tag{7}$$

Namely, φ, η can be deduced via the action of $SE(2, N)$ if and only if their Fourier transforms at a representation T can be deduced via multiplication by $T(a)$.

Remark 2 The fact that the Fourier transform in (5) be matrix-valued is a direct consequence of $SE(2, N)$ being a Moore group, that is, that all the T^λ act on finite-dimensional spaces. This is not true for the roto-translation group $SE(2)$. As a consequence, the Fourier transform on $SE(2)$, takes values not in the finite dimensional space of complex $N \times N$ matrices, but in the infinite dimensional space of operators over $L^2(\mathbb{S}^1)$. This is indeed the main theoretical advantage of considering $SE(2, N)$.

3.1 Decomposition of Tensor Product Representations

Proofs of Sect. 4, will use a well-known fact on tensor product representations: the Induction–Reduction Theorem (see [1]). This theorem allows to decompose the tensor products of

⁴ That is, up to a multiplicative constant, the only left and right invariant measure on $SE(2, N)$. One can check that $\int_{SE(2,N)} \varphi(a) da = \sum_{k=0}^{N-1} \int_{\mathbb{H}^2} \varphi(x, k) dx$.

representations $T^{\lambda_1} \otimes T^{\lambda_2}$, acting on $\mathbb{C}^N \otimes \mathbb{C}^N \cong \mathbb{C}^{N \times N}$, to an equivalent representation acting on $\bigoplus_{k \in \mathbb{Z}_N} \mathbb{C}^N$, which is a block-diagonal operator whose block elements are of the form $T^{\lambda_1 + R_k \lambda_2}$. Moreover, the linear transformation realizing the equivalence is explicit.

To avoid confusion, we will henceforth denote components of vectors $v \in \mathbb{C}^N$ as $v(0), \dots, v(N - 1)$, elements of $\bigoplus_{k \in \mathbb{Z}_N} \mathbb{C}^N$ as $(w_k)_{k \in \mathbb{Z}_N}$ where $w_k \in \mathbb{C}^N$, and the components of vectors $\mathbf{v} \in \mathbb{C}^N \otimes \mathbb{C}^N$ as $\mathbf{v}(k, h)$ for $k, h = 0, \dots, N - 1$. We also remark that linear operators B on $\bigoplus_{k \in \mathbb{Z}_N} \mathbb{C}^N$ can be decomposed as $B = (B^{k,h})_{k,h \in \mathbb{Z}_N}$, where each $B^{k,h}$ is an $N \times N$ complex matrix. Namely, we have

$$B(w_k)_{k \in \mathbb{Z}_N} = \left(\sum_{h=0}^{N-1} B^{k,h} w_h \right)_{k \in \mathbb{Z}_N}. \tag{8}$$

Then, exploiting the commutation of the Fourier transform with equivalences of representation, the Induction–Reduction Theorem implies that for every $\varphi \in L^2(SE(2, N))$ and any $\lambda_1, \lambda_2 \in \mathcal{S}$ it holds

$$A \circ \widehat{\varphi}(T^{\lambda_1} \otimes T^{\lambda_2}) \circ A^{-1} = \bigoplus_{k \in \mathbb{Z}_N} \widehat{\varphi}(T^{\lambda_1 + R_k \lambda_2}). \tag{9}$$

Here, $A : \mathbb{C}^N \otimes \mathbb{C}^N \rightarrow \bigoplus_{k \in \mathbb{Z}_N} \mathbb{C}^N$ is given by

$$(A\mathbf{v})_k(h) = (A_k \mathbf{v})(h) = \mathbf{v}(h, h - k), \quad \forall \mathbf{v} \in \mathbb{C}^N \otimes \mathbb{C}^N. \tag{10}$$

4 Fourier Descriptors on $SE(2, N)$

In the following sections, we introduce and study the Fourier descriptors on the group $SE(2, N)$. As already mentioned, proving a general completeness result is essentially hopeless, and we will content ourselves to prove the weak completeness.

Let $K \subset \mathbb{R}^2$ be a compact set. In the following, we will be mainly concerned with functions that are compactly supported either in K or in $K \times \mathbb{Z}_N \subset SE(2, N)$.

4.1 Generalized Fourier Descriptors

Following [38], the power-spectrum and the bispectrum on \mathbb{R}^2 can be generalized to $SE(2, N)$ as follows.

Definition 1 The *generalized power-spectrum* and *bispectrum* of $\varphi \in L^2(SE(2, N))$ are the collections of matrices for any $\lambda, \lambda_1, \lambda_2 \in \mathcal{S}$,

$$\begin{aligned} \text{PS}_\varphi(\lambda) &:= \widehat{\varphi}(T^\lambda) \circ \widehat{\varphi}(T^\lambda)^* \\ \text{BS}_\varphi(\lambda_1, \lambda_2) &:= \widehat{\varphi}(T^{\lambda_1}) \otimes \widehat{\varphi}(T^{\lambda_2}) \circ \widehat{\varphi}(T^{\lambda_1} \otimes T^{\lambda_2})^*. \end{aligned}$$

The next result generalizes, with a simplified proof, the result presented in [38]. Let us mention that this result is indeed true in a more general setting, as it will be shown in a forthcoming paper by Prandi and Gauthier.

Theorem 1 *Let $K \subset \mathbb{R}^2$ be a compact. The generalized power-spectrum and bispectrum are weakly complete on $L^2(\mathbb{Z}_N \times K)$. In particular, they discriminate on the open and dense set $\mathcal{G} \subset L^2(\mathbb{Z}_N \times K)$ of functions φ supported in $\mathbb{Z}_N \times K$ and whose Fourier transform $\hat{\varphi}(T^\lambda)$ is invertible for an open and dense set of λ s. That is, $\varphi_1, \varphi_2 \in \mathcal{G}$ are such that $PS_{\varphi_1} = PS_{\varphi_2}$ and $BS_{\varphi_1} = BS_{\varphi_2}$ if and only if $\varphi_1 = A(x, k)\varphi_2$ for some $(x, k) \in SE(2, N)$.*

Proof The fact that \mathcal{G} is open and dense is proved in Lemma 1 in Appendix 2. Let $\varphi, \eta \in \mathcal{G}$ be such that $PS_{\varphi_1} = PS_{\varphi_2}$ and $BS_{\varphi} = BS_{\eta}$. The equality of the generalized bispectrum implies that the set of λ s where $\hat{\varphi}(T^\lambda)$ and $\hat{\eta}(T^\lambda)$ fail to be invertible is the same. We will denote it by I and let

$$U(T^\lambda) := \hat{\varphi}(T^\lambda)^{-1}\hat{\eta}(T^\lambda) \quad \forall \lambda \in I.$$

In order to complete the proof of the statement, we will prove that $U(T^\lambda)$ can be defined for all λ s in \mathbb{R}^2 and, moreover, that $U(T^\lambda) = T^\lambda(a)$ for some $a \in SE(2, N)$. Indeed, by (7) this will readily imply that $\varphi = A(a)\eta$ as announced.

We claim that $U(T^\lambda)$ is unitary for all $\lambda \in I$. Indeed, by the equality of the generalized power-spectrum we have

$$U(T^\lambda)^*U(T^\lambda) = \eta(T^\lambda)^*PS_{\varphi}(\lambda)\eta(T^\lambda) = I.$$

Observe that the equality of the generalized bispectrum and the definition of U , imply that for all $\lambda_1, \lambda_2 \in I$ it holds

$$BS_{\varphi}(\lambda_1, \lambda_2) = \hat{\varphi}(T^{\lambda_1}) \otimes \hat{\varphi}(T^{\lambda_2}) \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \circ \hat{\eta}(T^{\lambda_1} \otimes T^{\lambda_2})^*.$$

By the invertibility of $\hat{\varphi}(T^{\lambda_1}) \otimes \hat{\varphi}(T^{\lambda_2})$ and the unitarity of U , this yields

$$\hat{\varphi}(T^{\lambda_1} \otimes T^{\lambda_2}) \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) = \hat{\eta}(T^{\lambda_1} \otimes T^{\lambda_2}). \quad (11)$$

The announced result is then a consequence of the following three facts, which are proved in Appendix 2.

1. Lemma 2: The function $\lambda \mapsto U(T^\lambda)$ is continuous on I .
2. Lemma 3: The function $\lambda \mapsto U(T^\lambda)$ can be extended to a continuous function on \mathbb{R}^2 for which (11) is still true.
3. Lemma 4: There exists $a \in SE(2, N)$ such that $U(T^\lambda) = T^\lambda(a)$. \square

An immediate corollary is the following.

 Springer

Corollary 1 *Let $\tilde{\mathcal{L}} : L^2(\mathbb{R}^2) \rightarrow L^2(SE(2, N))$ be an injective lift operator (not necessarily satisfying (2)). Assume that there exists a residual set $\mathcal{R} \subset L^2(\mathbb{R}^2)$ such that $\tilde{\mathcal{L}}(\mathcal{R}) \cap \mathcal{G}$ is residual. Then, the generalized power-spectrum and bispectrum are weakly complete on $L^2(\mathbb{R}^2)$. Namely, for any $f, g \in \mathcal{R}$ it holds that $BS_{\tilde{\mathcal{L}}f} = BS_{\tilde{\mathcal{L}}g}$ if and only if $f = \pi(x, k)g$ for some $(x, k) \in SE(2, N)$.*

Remark 3 In [38], the authors applied their version of Theorem 1 to a non-left-invariant lift $\tilde{\mathcal{L}}$, called cyclic lift. Indeed, for this cyclic lift, when N is odd, it is possible to prove that, for any compact $K \subset \mathbb{R}^2$, there exists a residual set $\mathcal{R} \subset L^2(K)$ satisfying the assumptions of Corollary 1.

Unfortunately, Corollary 1 can never be applied to lifts of the form (2). In fact, as proved in Appendix 3, letting $\omega_f(\lambda) := (\hat{f}(R_{-\lambda}))_{k=0}^{N-1} \in \mathbb{C}^N$, we have that

$$\widehat{\mathcal{L}}f(T^\lambda) = \omega_\psi(\lambda)^* \otimes \omega_f(\lambda)^*, \quad (12)$$

where for $v, w \in \mathbb{C}^N$, we let $v^* = (\overline{v_k})_k$ and $(v \otimes w)_{k,h} = v_k \overline{w_h}$, so that $(v \otimes w)u = \langle w, u \rangle v$ for all $u \in \mathbb{C}^N$. This immediately implies that $\text{rank } \widehat{\mathcal{L}}f(T^\lambda) \leq 1$ and hence that $\text{range } \mathcal{L} \cap \mathcal{G} = \emptyset$ whenever $N > 1$.

4.2 Rotational Fourier Descriptors

To bypass the difficulty posed by the non-invertibility of the Fourier transform for lifted functions, we are led to consider the following stronger descriptors.

Definition 2 The *rotational power-spectrum* and *bispectrum* of $\varphi \in L^2(SE(2, N))$ are the collections of matrices, for any $\lambda, \lambda_1, \lambda_2 \in \mathcal{S}$ and $h \in \mathbb{Z}_N$,

$$\begin{aligned} \text{RPS } \varphi(\lambda, h) &:= \hat{\varphi}(T^{R_{\lambda} h}) \circ \hat{\varphi}(T^\lambda)^* \\ \text{RBS}_\varphi(\lambda_1, \lambda_2, h) &:= \hat{\varphi}(T^{R_{\lambda_1} h}) \otimes \hat{\varphi}(T^{\lambda_2}) \circ \hat{\varphi}(T^{\lambda_1} \otimes T^{\lambda_2})^*. \end{aligned}$$

As already mentioned in the introduction, the rotational descriptors are invariant only under the action of $\mathbb{Z}_N \subset SE(2, N)$ but not under translations. To avoid this problem, let us fix a compact $K \subset \mathbb{R}^2$ and consider the set $\mathcal{A} \subset L^2(\mathbb{R}^2)$ of functions compactly supported in K , with non-zero average⁵. Observe that this is an open and dense subset of $L^2(K)$. We can then define the barycenter $c_f \in \mathbb{R}^2$ of $f \in \mathcal{A}$ as

$$c_f = \frac{1}{\text{avg } f} \left(\int_{\mathbb{R}^2} x_1 f(x) dx, \int_{\mathbb{R}^2} x_2 f(x) dx \right),$$

and the centering operator $\Phi : \mathcal{A} \rightarrow \mathcal{A}$ as

$$\Phi f(x) := f(x - c_f). \quad (13)$$

⁵ Recall that the average of $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is $\text{avg } f = \int_{\mathbb{R}^2} f(x) dx$, which is always well-defined for $L^2(\mathbb{R}^2)$ functions with compact support.

Then, considering the centered lift $\mathcal{L}_c = \mathcal{L} \circ \Phi$, we have that $\mathcal{L}_c f = \mathcal{L}_c g$ if and only if g is a translate of f . In particular,

$$\mathcal{L}_c f = A(0, k)\mathcal{L}_c g \iff f = \pi(x, k)g \text{ for some } x \in \mathbb{R}^2.$$

Let us consider the following set of functions.

Definition 3 Let $\mathcal{R} \subset L^2(\mathbb{R}^2)$ be the set of real-valued functions f supported in K , such that $\hat{f}(\lambda) \neq 0$ for a.e. $\lambda \in \mathbb{R}^2$ and the family $\Omega_f = \{S^k \omega_f(\lambda)\}_{k=0}^{N-1}$ is a basis for \mathbb{C}^N , if N is odd, or, if N is even, for

$$\mathcal{X} = \{v \in \mathbb{C}^N \mid v(h) = \overline{v(h + N/2)} \quad \forall h \in \mathbb{Z}_N\}.$$

The dependence of this definition on the parity of N comes from the well-known fact that $\hat{f}(\lambda) = \overline{\hat{f}(-\lambda)}$. Indeed, for N even, this implies that $S^k \omega_f(\lambda) \in \mathcal{X}$ for any $k \in \mathbb{Z}_N$. As such, there is no hope for the family Ω_f to generate the whole \mathbb{C}^N .

Finally, we have the following theorem.

Theorem 2 For any compact $K \subset \mathbb{R}^2$, if the mother wavelet $\Psi \in \mathcal{R}$, the rotational power-spectrum and bispectrum are weakly complete on $L^2(K) \cap \mathcal{A}$. Namely, the set \mathcal{R} is open and dense in $L^2(K)$ and for any $f, g \in \mathcal{R} \cap \mathcal{A}$ it holds that $\text{RPS}_{\mathcal{L}_c f} = \text{RPS}_{\mathcal{L}_c g}$ and $\text{RBS}_{\mathcal{L}_c f} = \text{RBS}_{\mathcal{L}_c g}$ if and only if $f = \pi(x, k)g$ for some $(x, k) \in SE(2, N)$.

Here, we content ourselves to present only a sketch of the proof of this result for the case N odd. The parity of N does not introduce essential problems, up to exploit the fact that $\text{range } \widehat{\mathcal{L}}f(T^\lambda) \subset \mathcal{X}$ for all $f \in \mathcal{R}$ and $\lambda \in \mathcal{S}$ and that the equivalence A of the Induction–Reduction Theorem quotients nicely to an equivalence between $\mathcal{X} \times \mathcal{X}$ and $\bigoplus_{k \in \mathbb{Z}_N} \mathcal{X}$. However, in order to prove the key technical point (14) we need a much finer study of the properties of circulant operators, which is outside the scope of this work and we defer to a forthcoming paper by Prandi and Gauthier.

Proof (Sketch in the case N odd) The fact that \mathcal{R} is open and dense in $L^2(K)$ follows from the same arguments in Lemma 1.

Let $\text{Circ } v$ be the circulant matrix associated with v , that is, $\text{Circ } v = [v, Sv, \dots, S^{N-1}v]$. Then the condition on Ω_f for $f \in \mathcal{R}$ is equivalent to the invertibility of $\text{Circ } \omega_f(\lambda)$ for an open and dense set of λ s. By the properties of the Fourier transform on \mathbb{R}^2 w.r.t. translations it follows that

$$\omega_{\Phi f}(\lambda) = \text{diag}_{\mathbb{Z}_k} \left(e^{-i\langle \lambda, R_k c_f \rangle} \right) \omega_f(\lambda).$$

This entails that $\text{Circ } \omega_f(\lambda)$ is invertible if and only if $\text{Circ } \omega_{\Phi f}(\lambda)$ is. Hence, the statement is equivalent to the fact

that for any couple $f, g \in \mathcal{R}$ we have $\text{RBS}_{\mathcal{L}_c f} = \text{RBS}_{\mathcal{L}_c g}$ if and only if $f = R_k g$ for some $k \in \mathbb{Z}_N$.

The proof is similar to the one of Theorem 1, but with additional technical difficulties. Let I be the set where $\text{Circ } \omega_f(\lambda)$ and $\text{Circ } \omega_g(\lambda)$ are invertible. By assumption I is open and dense. To overcome the non-invertibility of $\widehat{\mathcal{L}}f$ in the definition the candidate intertwining representation U , we exploit the invertibility of the circulant matrices $\text{Circ } \omega_f(\lambda)$ and $\text{Circ } \omega_g(\lambda)$ on an open and dense set. Namely, for any $\lambda \in I$ we let

$$U(T^\lambda)^* := \text{Circ } \omega_g(\lambda) (\text{Circ } \omega_f(\lambda))^{-1}.$$

By definition, $U(T^\lambda)$ is circulant and $U(T^\lambda)^* S^k \omega_f(\lambda) = S^k \omega_g(\lambda)$ for any $k \in \mathbb{Z}_N$. Moreover, by (12), this is equivalent to

$$\widehat{\mathcal{L}}f(T^{R_k \lambda}) U(T^\lambda) = \widehat{\mathcal{L}}g(T^{R_k \lambda}), \quad \forall k \in \mathbb{Z}_N.$$

In particular, $\lambda \mapsto U(T^\lambda)$ is constant on orbits $\{R_k \lambda\}_{k \in \mathbb{Z}_N}$. Finally, $U(T^\lambda)$ is unitary as a consequence, e.g., of Theorem 3.

The main difficulty in the proof is now to derive the equivalent of identity (11), that is, that for an open and dense set of couples (λ_1, λ_2) we have

$$\begin{aligned} \widehat{\mathcal{L}}f(T^{R_{k_1} \lambda_1} \otimes T^{R_{k_2} \lambda_2}) U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \\ = \widehat{\mathcal{L}}g(T^{R_{k_1} \lambda_1} \otimes T^{R_{k_2} \lambda_2}), \quad \forall k \in \mathbb{Z}_N. \end{aligned} \tag{14}$$

As already mentioned, the proof of this identity requires a deep use of properties of circulant operators, which is outside the scope of this paper. We thus defer it to a forthcoming paper.

Once (14) is known, the statement follows applying the same arguments as those in Theorem 1. Namely,

1. The function $\lambda \mapsto U(T^\lambda)$ is continuous on I . This can be done via the same arguments as in Lemma 2.
2. The function $\lambda \mapsto U(T^\lambda)$ can be extended to a continuous function on \mathcal{S} still satisfying (14). This can be done exactly as in Lemma 3.
3. There exists $k \in \mathbb{Z}_N$ such that $U(T^\lambda) = T^\lambda(0, k)$. This is proved following Lemma 4. Indeed, the fact that now $\lambda \mapsto U(T^\lambda)$ is constant on the orbits $\{R_k \lambda\}_{k \in \mathbb{Z}_N}$ implies that the φ_k s obtained there have to be independent of k . Since $\varphi_k(\lambda) = e^{i\langle R_k x_0, \lambda \rangle}$ for some $x_0 \in \mathbb{R}^2$, this implies that $x_0 = 0$ and hence $\varphi_k = 0$. Obviously, this proves that $U(T^\lambda) = S^k = T^\lambda(0, k)$, for some $k \in \mathbb{Z}_N$. \square

4.3 Practical Computation of the Fourier Descriptors

Here, we present some explicit formulae for the computation of the Fourier descriptors presented in this section.

In the following, we show that, under some assumptions on the mother wavelet Ψ , the concrete computation of the generalized power-spectrum and bispectrum and of their rotational counterparts, depend only on the 2D Fourier transform of f .

Theorem 3 Assume that the mother wavelet $\Psi \in \mathcal{R}$. Then:

- For any $f \in \mathcal{R}$, the generalized power-spectrum and bispectrum of $\mathcal{L}f$ are, respectively, determined by the quantities, for a.e. $\lambda, \lambda_1, \lambda_2 \in \mathcal{S}$,

$$I_1^\lambda(f) = \|\omega_f(\lambda)\|^2 = \sum_{k=0}^{N-1} |\hat{f}(R_{-k}\lambda)|^2$$

$$I_1^{\lambda_1, \lambda_2}(f) = \langle \omega_f(\lambda_1) \odot \omega_f(\lambda_2), \omega_f(\lambda_1 + \lambda_2) \rangle$$

$$= \sum_{k=0}^{N-1} \hat{f}(R_{-k}\lambda_1) \hat{f}(R_{-k}\lambda_2) \overline{\hat{f}(R_{-k}(\lambda_1 + \lambda_2))}.$$

- For any $f \in \mathcal{A} \cap \mathcal{R}$, the rotational power-spectrum and bispectrum of $\mathcal{L}_c f$ are, respectively, determined by the quantities, for a.e. $\lambda, \lambda_1, \lambda_2 \in \mathcal{S}$ and $h \in \mathbb{Z}_N$,

$$I_2^{\lambda, h}(f) = \langle \omega_{\Phi f}(R_h \lambda), \omega_{\Phi f}(\lambda) \rangle$$

$$= \sum_{k=0}^{N-1} \overline{\hat{f}(R_{-k+h}\lambda)} \hat{f}(R_{-k}\lambda),$$

$$I_2^{\lambda_1, \lambda_2, h}(f) = \langle \omega_{\Phi f}(R_h \lambda_1) \odot \omega_{\Phi f}(\lambda_2), \omega_{\Phi f}(\lambda_1 + \lambda_2) \rangle$$

$$= \sum_{k=0}^{N-1} \hat{f}(R_{-k+h}\lambda_1) \hat{f}(R_{-k}\lambda_2) \overline{\hat{f}(R_{-k}(\lambda_1 + \lambda_2))}.$$

Here, $\Phi : \mathcal{A} \rightarrow \mathcal{A}$ is the centering operator defined in (13).

Remark 4 Theorem 3 shows in particular that the result of Theorem 2 is indeed stronger than the completeness result for the generalized bispectrum of the cyclic lift obtained in [38]. Indeed, in that work is proved that the latter (for odd N) is determined exactly by the quantities, for a.e. $\lambda_1, \lambda_2 \in \mathcal{S}$ and $h, k \in \mathbb{Z}_N$,

$$\tilde{I}_2^{\lambda_1, \lambda_2, k, h} = \langle \omega_{\Phi f}(R_h \lambda_1) \odot \omega_{\Phi f}(R_k \lambda_2), \omega_{\Phi f}(\lambda_1 + R_{h+k}\lambda_2) \rangle.$$

In particular, for each $\lambda_1, \lambda_2 \in \mathcal{S}$ one has to compute N times more quantities than those for the rotational bispectrum.

As a corollary of Theorem 3 we show that, in order to compare the power-spectra and bispectra, it is usually enough to compare only the latter.

Corollary 2 Let $\Psi \in \mathcal{R}$ and $f, g \in \mathcal{R} \cap \mathcal{A}$. Then, if $\mathcal{L}f$ and $\mathcal{L}g$ have the same generalized (resp. rotational) bispectrum, they have also the same generalized (resp. rotational) power-spectrum.

Proof We only prove the result for the rotational descriptors.

In order to prove the one for the generalized descriptors, it will be enough to fix $h = 0$ in the following. By Theorem 3 it is enough to show that whenever $I_2^{\lambda_1, \lambda_2, h}(f) = I_2^{\lambda_1, \lambda_2, h}(g)$ for a.e. $\lambda_1, \lambda_2 \in \mathcal{S}$ and any $h \in \mathbb{Z}_N$, then $I_2^{\lambda, h}(f) = I_2^{\lambda, h}(g)$ for a.e. $\lambda \in \mathcal{S}$ and any $h \in \mathbb{Z}_N$. We start by observing that by the Paley–Wiener Theorem all these quantities are analytic, since f and g are compactly supported. Moreover,

$$\lim_{\lambda_1, \lambda_2 \downarrow 0} I_2^{\lambda_1, \lambda_2, h}(f) = N \hat{f}(0) |\hat{f}(0)|^2 = N \text{avg}(f)^3,$$

and the same is true for g . Thus, $\text{avg}(f) = \text{avg}(g)$. Finally, the result follows observing that

$$\lim_{\lambda_2 \downarrow 0} I_2^{\lambda_1, \lambda_2, h}(f) = \text{avg}(f) I_2^{\lambda_1, h}(f).$$

□

5 Experimental Results

The goal of this section is to evaluate the performance of the invariant Fourier descriptors defined in the previous section on a large image database for object recognition. In addition to the generalized power-spectrum (PS) and bispectrum (BS) and the rotational power-spectrum (RPS) and bispectrum (RBS), we also consider the combination of the RPS and BS descriptors. Indeed, combining these two descriptors seems to be a good compromise between the theoretical result of completeness given by Theorem 2, which only holds for the RBS, and computational demands, as the results on the COIL-100 database will show.

After showing how to efficiently compute these descriptors and presenting the image data-set, we analyze some experimental results. In order to estimate the features capabilities, we use a support vector machine (SVM) [39] as supervised classification method. The recognition performances of the different descriptors regarding invariance to rotation, discrimination capability and robustness against noise are compared.

5.1 Implementation

As proved in Theorem 3, the equality of the Fourier descriptors we introduced does not depend on the choice of the mother wavelet Ψ . Accordingly, in our implementation we only computed the quantities introduced in Theorem 3,

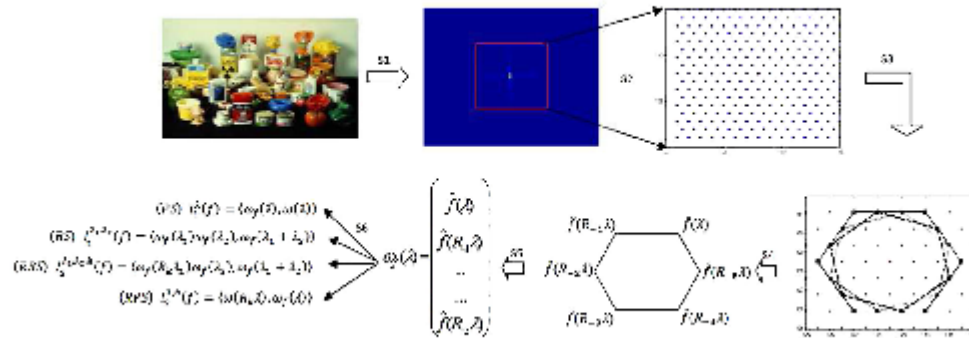


Fig. 1 Steps of computing the invariant descriptors. (S1) computation of the shifted *FFT* of the image *f*, (S2) generation of the hexagonal grid, (S3) extraction of different hexagons, (S4) evaluation of the *FFT*

of *f* on each extracted hexagon, (S5) generation of the vector $\omega_f(\lambda)$ and (S6) computation of the four invariants

whose complexity is reduced to the efficient computation of the vector $\omega_f(\lambda)$, for a given $\lambda \in \mathcal{S}$. We recall that this vector is obtained by evaluating the Fourier transform of *f* on the orbit of λ under the action of discrete rotations R_{-k} for $k \in \mathbb{Z}_N$.

Let us remark that, although in our implementation we chose this approach, in principle fixing a specific mother wavelet could be useful to appropriately weight descriptors depending on the associated frequencies. Indeed, preliminary tests with a Gabor mother wavelet (which can be easily shown to be in \mathcal{R}) showed slightly better results at a bigger computational cost.

For the implementation we chose to consider $N = 6$ and to work with images composed of hexagonal pixels. There are two reasons for this choice:

- It is well-known that retinal cells are distributed in a hexagonal grid, and thus it is reasonable to assume that cortical activations reflect this fact.
- Hexagonal grids are invariant under the action of \mathbb{Z}_6 and discretized translations, which is the most we can get in the line of the invariance w.r.t. $SE(2, 6)$. Indeed, apart from the hexagonal lattice, the only other lattices on \mathbb{R}^2 which are invariant by some \mathbb{Z}_N and appropriate discrete translations are obtained with $N = 2, 3, 4$.

The different steps of computation of the descriptors⁶ are described in Fig. 1 and given as follows:

⁶ MATLAB sample code for the implementation of the rotational bispectral invariants can be found at https://nbviewer.jupyter.org/github/dprn/bispectral-invariant-svm/blob/master/Invariant_computation_matlab.ipynb

Table 1 Dimension of the feature vectors for the Fourier descriptors under consideration

Descr.	Dim.
PS	136
BS	717
RPS	816
RBS	4417
RPS + BS	1533

1. The input image is converted to grayscale mode, the Fourier transform is computed via FFT, and the zero-frequency component is shifted to the center of the spectrum (Fig. 1, S1).
2. For cost computational reasons and since we are dealing with natural images, for which the relevant frequencies are the low ones, we extract a grid of 16×16 pixels around the origin (Fig. 1, S2).
3. The invariants of Theorem 3 are computed from the shifted Fourier transform values, on all frequencies in an hexagonal grid inside this 16×16 pixels square. A bilinear interpolation is applied to obtain the correct values of $\omega_f(\lambda)$ (Fig. 1, S3, S4, S5, S6). The final dimension of the feature vector is given in Table 1.

5.2 Test Protocol

We use the Fourier descriptors to feed an SVM classifier, via the MATLAB Statistics and Machine Learning Toolbox, applying it on a database of 7200 objects extracted from the Columbia Object Image Library (COIL-100) and a database of 400 faces extracted from ORL face database. Finally, we

compare the results obtained with those obtained using traditional descriptors.

The result of the training step consists of the set of support vectors determined by the SVM-based method. During the decision step, the classifier computes the Fourier descriptors and the model determined during the training step is used to perform the SVM decision. The output is the image class.

For COIL-100 database, two cases are studied: a case without noise and another with noise. In the first one, tests have been performed using 75% of the COIL-100 database images for training and 25% for testing. In the second one, we have used a learning data-set composed of all the 7200 images (100 objects with 72 views) without noise and a testing data-set composed of 15 randomly selected views per object to which an additive Gaussian noise with Δ_d of 5, 10 and 20 was added (see Fig. 4).

We evaluate separately the recognition rate obtained using the four previous invariant descriptors and the combination of the RPS & BS invariants to test their complementarity. Then, we compare their performance with the Hu's moments (HM), the Zernike's moments (ZM), the Fourier–Mellin transform (FM), described in Appendix 1, and the local SIFT and HOG descriptors [12] whose performance under the same conditions has been tested in [7].

Since we use the RBF kernel in the SVM classification process, this depends on the kernel size σ . The results presented here are obtained by choosing empirically the value σ_{opt} that provided maximum recognition rate.

5.3 Experiments

The performances of the different invariant descriptors are analyzed with respect to the recognition rate given a learning set. Hence, for a given ratio, the learning and testing sets have been built by splitting randomly all examples. Then, due to randomness of this procedure, multiple trials have been performed with different random draws of the learning and testing set. In the case of an added noise, since as mentioned before the learning set comprised all images, this procedure is applied only to the testing set.

The parameters of our experiments are the following:

1. The learning set c_l corresponding to the values of an invariant descriptor computed on an image from the database;
2. The classes $\hat{c}_l \in [1, 100]$ corresponding to the object class.
3. Algorithm performance: the efficiency is given through a percentage of the well recognized objects composing the testing set.
4. Number of random trials: fixed to 5.



Fig. 2 Sample objects of COIL-100 database

5. Kernel K: a Gaussian kernel of bandwidth σ is chosen

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

x and y correspond to the descriptors vectors of objects.

For solving a multi-class problem, the two most popular approaches are the one-against-all (OAA) method and the one-against-one (OAO) method [29]. For our purpose, we chose an OAO SVM, because it is substantially faster to train and seems preferable for problems with a very large number of classes.

5.3.1 COIL-100 Databases

The Columbia Object Image Library (COIL-100, Fig. 2) is a database of color images of 100 different objects, where 72 images of each object were taken at pose intervals of 5° .

5.3.2 Classification Performance

Table 2 presents results obtained testing our object recognition method with the COIL-100 database. The best results were achieved using the local SIFT descriptor. The RBS comes in the second place and the local HOG features come third. Indeed it has been demonstrated in the literature, these local methods currently give the best results. However, if noise is added on the image, the use of global approach is better than the use of local ones. The main reason is that the key-points detector used in the local method produce in these cases many key-points that are not relevant for object recognition. This will be shown in the next subsection.

In Fig. 3 we present the recognition rate as a function of the size of the training set. As expected, this is an increasing function and we remark that the RBS and the combination of the RPS and the BS give better results than the other global invariant descriptors.

Table 2 Recognition rate for each descriptor using the COIL-100 database. The test results for ZM, HM, FM, and SIFT are taken from [7]

Descriptors	Recognition rates (%)
RBS	95.5
BS	88
PS	84.3
RPS	89.8
RPS+BS	92.8
ZM	91.9
HM	80.2
FM	89.6
HOG	95.3
SIFT	100

The best results for each trial are given in bold

5.3.3 Robustness Against Noise

Also in this case, test results for ZM, HM, FM, and SIFT are taken from [7].

Results presented in Table 3 show that noise has little influence on classification performance when we use a global descriptor such as RBS, BS, the combination of BS & RPS,

Fig. 3 Classification rate for different size of the training database. The test results for ZM, HM, FM, and SIFT are taken from [7]

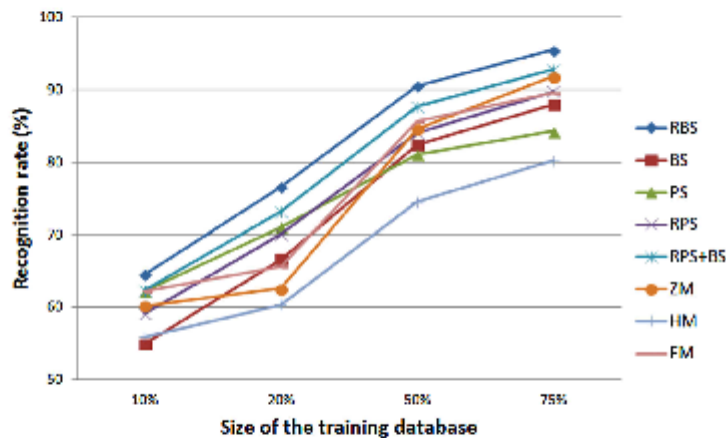


Table 3 Classification rate on COIL-100 noisy database. The test results for ZM, HM, FM, and SIFT are taken from [7]

S_d	RBS (%)	BS (%)	PS (%)	RPS (%)	RPS+BS (%)	ZM (%)	HM (%)	FM (%)	SIFT (%)	HOG (%)
5	100	100	71.5	99.8	100	100	95.2	98.6	89.27	4
10	100	100	71.2	99.8	100	100	95.2	95.2	88.89	1.2
20	100	100	67.8	99.8	100	100	91.4	90.2	85.46	1

The best results for each trial are given in bold

ZM, HM and FM. It has however a sensible effect on the SIFT local descriptor, and a big one on the HOG local descriptor.

5.3.4 The ORL Database

The Cambridge University ORL face database (Fig. 5) is composed of 400 gray level images of ten different patterns for each of 40 persons. The variations of the images are across time, size, pose and facial expression (open/closed eyes, smiling/not smiling), and facial details (glasses/no glasses).

In the literature, the protocol used for training and testing is different from one paper to another. In [36], a hidden Markov model (HMM) based approach is used, and the best model resulted in recognition rate of 95%, with high computational cost. In [21], Hjelmas reached a 85% recognition rate using the ORL database and feature vector consisting of Gabor coefficients.

We perform experiments on the ORL database using the RBS, BS, PS, RPS, ZM, HU, FM, and the combination of the RPS & BS descriptors. Since the local descriptors SIFT and HOG obtained, predictably, almost perfect scores, we do not present them. The results are shown in Table 4, where we clearly see that the RBS invariant descriptor gives the best recognition rate $c = 89.8%$, faring far better than before w.r.t. the combination of RPS and BS descriptors.



Fig. 4 Sample of COIL-100 noisy object



Fig. 5 Face samples from the ORL database

Table 4 Recognition rate for each descriptor using the ORL database

Descriptors	Recognition rates (%)
RBS	89.8
BS	67.9
PS	49.2
RPS	76.9
RPS+BS	79.8
ZM	75
HM	43.5
FM	47.6

The best result for each trial is given in bold

6 Conclusion and Perspectives

In this paper, we presented four Fourier descriptors over the semidiscrete roto-translation group $SE(2, N)$. Then, we proved that the generalized power-spectrum (PS) and bispectrum (BS)—and thus the rotational power-spectrum (RPS) and bispectrum (RBS)—are weakly complete, in the sense that they allow to distinguish over an open and dense set of compactly supported functions $\varphi \in L^2(SE(2, N))$ up to the $SE(2, N)$ action. This generalizes a result of [38]. We then considered a framework for the application of these Fourier descriptors to roto-translation invariant object recognition, inspired by some neurophysiological facts on the human primary visual cortex. In this framework, we showed that the rotational bispectrum is indeed a weakly complete roto-translation invariant for planar images. Moreover, although the proposed Fourier descriptors are given in terms of com-

plex mathematical objects, we showed that they can be implemented in a straightforward way as linear combinations of the values of the 2D Fourier transform of the image.

In the second part of the paper, we proposed an evaluation of the performances of these Fourier descriptors in object recognition and we presented the results obtained on different databases: the COIL-100 database, composed of several objects undergoing 3D rotation and scales changes, and the ORL-database, on which different human faces are subject to several kinds of variations. For both these databases, the global Fourier descriptors introduced in this paper are the most efficient global descriptors tested, equalled only, for noisy images, by the Zernike Moments. Although for unperturbed images the local SIFT descriptor gives better recognition rate, the addition of noise leads to the global descriptors outperforming the local ones. These results thus show the rotational bispectrum (RBS) to be a very good Fourier descriptor for object recognition, consistently with the theoretical weak completeness result. When the dimension of the feature vector is an issue, the RBS can be replaced by a combination of the generalized bispectrum (BS) and the rotational power-spectrum (RPS), which yields slightly worse results with a feature vector of length almost one-third.

An extension of the object recognition method presented in this paper to an AdaBoost framework for the problem of object detection is currently ongoing.

Acknowledgements This research has been supported by the European Research Council, ERC StG 2009 “GeCoMethods”, contract n. 239748. The second and last authors were partially supported by the Grant ANR-15-CE40-0018 of the ANR.

Appendix 1: Moment Invariants and Fourier–Mellin Transform

In this section, following [7], we review the two most used classes of moment invariants, Hu and Zernike, and Fourier–Mellin descriptors, that we use as a comparison for our generalized Fourier descriptors.

Moment invariants were first introduced to the pattern recognition and image processing community in 1962 by Hu [22], with the introduction of the seven Hu moments which are invariants under translation, rotation and scaling. These are derived from a scaling and translation invariant modification of the standard moments of an image $I : \mathbb{R}^2 \rightarrow \mathbb{R}$. Namely,

$$v_{p,q} = \frac{u_{p,q}}{u_{0,0}^{\frac{p+q}{2}}},$$

where

$$u_{p,q} = \int_{\mathbb{R}^2} (x - x_0)^p (y - y_0)^q I(x, y) dx dy,$$

and $x_0 = \frac{m_{1,0}}{m_{0,0}}$ and $y_0 = \frac{m_{0,1}}{m_{0,0}}$ are the coordinates of the barycenter computed via the standard $(p + q)$ th order moments of I :

$$m_{p,q} = \int_{\mathbb{R}^2} x^p y^q I(x, y) dx dy.$$

Another important class of moments are the Zernike ones, introduced in [9] and computed via orthogonal Zernike polynomials. The Zernike moment of order (m, n) is:

$$Z_{mn} = \frac{m+1}{n} \sum_x \sum_y I(x, y) [V_{mn}(x, y)],$$

where $x^2 + y^2 < 1$ and $V_{mn}(x, y)$ are the Zernike polynomials defined in polar coordinates as $V_{mn}(r, \theta) = R_{mn}(r)e^{jn\theta}$, where

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} \frac{(-1)^s (m-s)! r^{m-2s}}{s! \left(\frac{m+|n|}{2} - s\right)! \left(\frac{m-|n|}{2} - s\right)!}$$

These moments present several advantages. Indeed, beside a rotation and translation invariance they have nice orthogonality properties and are considered to be robust against image noise. In particular, the orthogonality property helps in achieving a near zero value of redundancy measure in a set of moments functions [8].

Finally, strictly related to Fourier descriptors are the descriptors obtained via Fourier–Mellin transform (FMT), presented in [13]. The FMT of an image I , that we assume to be given in polar coordinates, is defined as:

$$M_I(u, v) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty I(r, \theta) r^{-iv} e^{-i u \theta} \frac{dr}{r} d\theta.$$

Following [13], we will indeed compute the analytical Fourier–Mellin transform (AFMT). That is, we replace I in the above definition with its regularized version $I_\sigma(r, \theta) = r^\sigma I(r, \theta)$, where $\sigma > 0$. Finally, each feature $M_{I_\sigma}(u, v)$ is modified in order to compensate for the rotation, translation and size changes of the object.

Appendix 2: Auxiliary Lemmata for the Proof of Theorem 1

Lemma 1 *The set \mathcal{G} introduced in Theorem 1 is open and dense in $L^2(K \times \mathbb{Z}_N)$.*

Proof We start by showing that $\mathcal{G} \neq \emptyset$. To this aim, it suffices to consider φ such that $\varphi(\cdot, k) = 0$ for all $k \in \mathbb{Z}_N \setminus \{0\}$ and $\varphi(\cdot, 0) \neq 0$ such that $\text{supp } \mathcal{F}(\varphi(\cdot, 0)) = \mathbb{R}^2$. By (6), we then have $\varphi \in \mathcal{G}$, since

$$\det \hat{\varphi}(T^\lambda) = \prod_{k \in \mathbb{Z}_N} \mathcal{F}(\varphi(\cdot, 0))(R_{-k}\lambda) \neq 0 \quad \forall \lambda \in \mathcal{S}.$$

For any $\varphi \in \mathcal{G}$ and $k \in \mathbb{Z}_N$, the Paley–Wiener Theorem implies that $\mathcal{F}(\varphi(\cdot, k))$ is analytic. In particular, by (6), $\lambda \mapsto \det \hat{\varphi}(T^\lambda)$ is analytic. Thus, $\varphi \in \mathcal{G}$ if and only if $\varphi(T^{\lambda_0})$ is invertible for some $\lambda_0 \in \mathcal{S}$.

We claim that the set \mathcal{G} is dense. Indeed, let $\varphi \notin \mathcal{G}$ and fix some $\eta \in \mathcal{G}$ and $\lambda_0 \in \mathcal{S}$ such that $\hat{\eta}(T^{\lambda_0})$ is invertible. By analyticity of $\varepsilon \mapsto \det(\hat{\varphi}(T^{\lambda_0}) + \varepsilon \hat{\eta}(T^{\lambda_0}))$ follows that $\varphi + \varepsilon \eta \in \mathcal{G}$ for sufficiently small $\varepsilon > 0$, which entails that $\varphi \in \mathcal{G}$, proving the claim.

Let us prove that \mathcal{G} is open in $L^2(K \times \mathbb{Z}_N)$. To this aim, fix $\varphi \in \mathcal{G}$ and $\varphi_n \rightarrow \varphi$ in $L^2(K \times \mathbb{Z}_N)$. This implies that $\hat{\varphi}_n \rightarrow \hat{\varphi}$ in $L^2(\widehat{SE}(2, \mathbb{N}))$, and in particular that $\hat{\varphi}_n \rightarrow \hat{\varphi}$ in measure. By definition of convergence in measure, this implies that for sufficiently big n it has to hold $\det \hat{\varphi}_n(T^{\lambda_0}) \neq 0$. Hence $\varphi_n \in \mathcal{G}$ for n sufficiently big and \mathcal{G} is open. \square

Before diving into the proofs of the other auxiliary lemmata, we make the following observation. Let $\lambda_1, \lambda_2 \in I$ be such that $\lambda_1 + R_k \lambda_2 \in I$ for all $k \in \mathbb{Z}_N$. Applying the Induction–Reduction theorem (9) to (11) yields

$$\begin{aligned} A \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \circ A^{-1} &= \bigoplus_{k \in \mathbb{Z}_N} \varphi(T^{\lambda_1 + R_k \lambda_2})^{-1} \eta(T^{\lambda_1 + R_k \lambda_2}) \\ &= \bigoplus_{k \in \mathbb{Z}_N} U(T^{\lambda_1 + R_k \lambda_2}). \end{aligned} \tag{15}$$

Lemma 2 *The function $\lambda \mapsto U(T^\lambda)$ is continuous on I .*

Proof Fix $\lambda_0 \in I$ and an open set $V \subset \mathbb{R}^2$ such that

$$\int_V U(T^{\lambda_2})_{i,j}^* d\lambda_2 > 0 \quad \forall i, j \in \mathbb{Z}_N.$$

This is possible since $U \neq 0$. Since the set I is open dense, up to reducing V we can assume that there exists a neighborhood W of λ_0 such that $V + \lambda \subset I$ for any $\lambda \in W$. Then, (15) holds for $\lambda_1 \in W$ and $\lambda_2 \in V$. Explicitly computing the 0, 0 block of (15), we have

$$U(T^{\lambda_1})_{i,j} U(T^{\lambda_2})_{i,j}^* = U(T^{\lambda_1 + \lambda_2})_{i,j} \quad \forall i, j \in \mathbb{Z}_N.$$

Then, integrating it over V w.r.t. λ_2 yields

$$U(T^{\lambda_1})_{i,j} = \frac{\int_{V+\lambda} U(T^{\lambda_2})_{i,j} d\lambda_2}{\int_V U(T^{\lambda_2})_{i,j}^* d\lambda_2} \quad \forall \lambda_1 \in W, \forall i, j \in \mathbb{Z}_N$$

Since the function on the r.h.s. is clearly continuous on W this proves the continuity at λ_0 of $\lambda \mapsto U(T^\lambda)$, completing the proof. \square

Lemma 3 The function $\lambda \mapsto U(T^\lambda)$ can be extended to a continuous function on \mathbb{R}^2 for which (11) is still true.

Proof Let $\lambda_0 \notin I$. Since I is an open and dense set, this implies that λ_0 is in its closure and that we can choose $\lambda_1, \lambda_2 \in I$ such that $\lambda_0 = \lambda_1 + R_{k_0}\lambda_2$ for some $k_0 \in \mathbb{Z}_N$ and $\lambda_1 + R_k\lambda_2 \in I$ for any $k \neq k_0$. We then let

$$U(T^{\lambda_0}) := (A \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \circ A^*)^{k_0, k_0}. \tag{16}$$

We now prove that the above definition does not depend on the choice of λ_1, λ_2 and k_0 . By openness of I , there exists a neighborhood V of λ_2 entirely contained in I . Then, up to taking a smaller V , it holds that $\lambda_1 + R_{k_0}\lambda'_2 \in I$ for any $\lambda'_2 \in V \setminus \{\lambda_2\}$. By (15), this implies that for any $\mu_1 + R_\ell\mu_2 = \lambda_0$ it holds

$$\begin{aligned} (A \circ U(T^{\lambda_1}) \otimes U(T^{\lambda'_2}) \circ A^*)^{k_0, k_0} &= \\ (A \circ U(T^{\mu_1}) \otimes U(T^{\mu_2}) \circ A^*)^{\ell, \ell}. \end{aligned}$$

for λ'_2 and μ'_2 sufficiently near, but different, to λ_2 and μ_2 , respectively. By the continuity of U on I , proved in Lemma 2, this implies that this equation has to hold also for $\lambda'_2 = \lambda_2$ and $\mu'_2 = \mu_2$. Hence, (16) does not depend on the choice of λ_1, λ_2 and k_0 .

Finally, the fact that $\hat{f}(T^{\lambda_1} \otimes T^{\lambda_2}) \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) = \hat{g}(T^{\lambda_1} \otimes T^{\lambda_2})$ for any λ_1, λ_2 follows from (16) and (15). \square

Lemma 4 There exists $a \in SE(2, N)$ such that $U(T^\lambda) = T^\lambda(a)$.

Proof By definition of U it holds that

$$\bigoplus_{k \in \mathbb{Z}_N} U(T^{\lambda_1 + R_k\lambda_2}) \circ A = A \circ U(T^{\lambda_1}) \otimes U(T^{\lambda_2}) \quad \forall \lambda_1, \lambda_2 \neq 0.$$

Then, for any i, j, ℓ, k ,

$$U(T^{\lambda_1})_{\ell, i} U(T^{\lambda_2})_{\ell-k, j} = \begin{cases} U(T^{\lambda_1 + R_k\lambda_2})_{\ell, i} & \text{if } j = i - k, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

By invertibility of $U(T^{\lambda_1})$, there exists $i_0 \in \mathbb{Z}_N$ such that $U(T^{\lambda_1})_{0, i_0} \neq 0$. Using (17) this implies that $U(T^{\lambda_2})_{-k, j} = 0$ for any $j \neq i_0 - k$. Namely, we have proved that there exists a family of functions $\varphi_{-k} : \mathcal{S} \rightarrow \mathbb{C}$ such that $U(T^{\lambda_1})_{-k, \cdot} = \varphi_{-k}(\lambda_1) \delta_{i_0-k}$ or, equivalently, that

$$U(T^\lambda) = \text{diag}_k \varphi_k(\lambda) S^{i_0}.$$

By the explicit expression (4) of T^λ , in order to complete the proof it suffices to prove that $\varphi_k(\lambda) = e^{i(x_0, R_{-k}\lambda)}$ for some $x_0 \in \mathbb{R}^2$.

 Springer

By continuity and unitarity of U , the φ_k s are continuous and satisfy $|\varphi_k(\lambda)| = 1$. Using again (17) with $j = i_0 - k$, we obtain

$$\varphi_\ell(\lambda_1 + R_\ell\lambda_2) = \varphi_\ell(\lambda_1)\varphi_{\ell-k}(\lambda_2), \tag{18}$$

for any $\lambda_1, \lambda_2 \neq 0$ and $\ell, k \in \mathbb{Z}_N$.

We claim that the φ_ℓ 's are characters of \mathbb{R}^2 . Indeed, let us fix $k = 0$ in (18):

$$\varphi_\ell(\lambda_1 + \lambda_2) = \varphi_\ell(\lambda_1)\varphi_\ell(\lambda_2). \tag{19}$$

Choosing $\lambda_2 = -\lambda_1$ in the above shows that φ_ℓ can be extended at 0. Moreover, letting $\lambda_1 = 0$ and taking the limit $\lambda_2 \rightarrow 0$ shows that this extension is continuous. Since characters of \mathbb{R}^2 are exactly the continuous functions satisfying (19), the claim is proved.

By Pontryagin duality, there exists $x_\ell \in \mathbb{R}^2$ such that $\varphi_\ell(\lambda) = e^{i(x_\ell, \lambda)}$. Finally, by (18) with $k \in \mathbb{Z}_N$ one obtains that $R_{-k}x_\ell = x_{\ell-k}$, which proves that there exists $x_0 \in \mathbb{R}^2$ such that $\varphi_\ell(\lambda) = e^{i(x_0, R_{-k}\lambda)}$. This completes the proof of the statement. \square

Appendix 3: Proofs

Proof (Formula (12)) Let $\lambda \in \mathcal{S}$ and consider $v \in \mathbb{C}^N$. Observe that $(x, k)^{-1} = (-R_{-k}x, -k)$. Then, by (5), (2), and (4), for any $h \in \mathbb{Z}_N$ we have

$$\begin{aligned} (\tilde{\mathcal{L}}(T^\lambda, v))_h &= \sum_{k=0}^{N-1} \int_{\mathbb{R}^2} \mathcal{L}(x, k) e^{-i(\lambda, R_{k-k}x)} v_{h-k} dx \\ &= \sum_{k=0}^{N-1} v_{h-k} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f(y) \tilde{\psi}(R_{-k}(y-x)) e^{-i(\lambda, R_{k-k}x)} dy dx \\ &= \sum_{k=0}^{N-1} v_{h-k} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \tilde{\psi}(z) f(y) e^{-i(R_{k-k}\lambda, y-z)} dy dz \\ &= \tilde{\psi}(R_h\lambda) \sum_{k=0}^{N-1} v_{h-k} f(R_{k-h}\lambda) \\ &= \overline{\omega_\varphi(\lambda)_h} \overline{\omega_f(\lambda)_v}. \end{aligned}$$

By definition of $\omega_\varphi(\lambda)^* \otimes \omega_f(\lambda)^*$; this completes the proof. \square

In order to prove Theorem 3, we need the following explicit description of the equivalence in the Induction-Reduction theorem of Sect. 3.1.

Lemma 5 For any $M, N \in \mathbb{C}^{N \times N}$ we have

$$(A \circ (M \otimes N) \circ A^{-1})^{k, h} = (M_{\ell, j} N_{\ell-k, j-h})_{\ell, j \in \mathbb{Z}_N}.$$

Proof Observe that for any $v \in \mathbb{C}^N \otimes \mathbb{C}^N$ it holds $(M \otimes N) \cdot v = M \circ v \circ N^T$. Thus,

$$[A \circ (M \otimes N).\mathbf{v}]_{\mathbf{k}(i)} = \sum_{j,\ell=0}^{N-1} M_{i,j} N_{i-k,\ell} \mathbf{v}(j, \ell).$$

Since is straightforward to check that $A^{-1} : \bigoplus_{\mathbf{k} \in \mathbb{Z}_N} \mathbb{C}^N \rightarrow \mathbb{C}^{N \times N}$ is given by $[A^{-1}(w_\ell)_{\ell \in \mathbb{Z}_N}]_{\mathbf{k}, h} = w_{\mathbf{k}-h}(k)$, we then have

$$\begin{aligned} [A \circ (M \otimes N) \circ A^{-1}.(w_h)_{h \in \mathbb{Z}_N}]_{\mathbf{k}(i)} & \\ &= \sum_{j,\ell=0}^{N-1} M_{i,j} N_{i-k,\ell} w_{j-\ell}(j) \\ &= \sum_{j,h=0}^{N-1} M_{i,j} N_{i-k,j-h} w_h(j). \end{aligned}$$

By (8), the proof is completed. \square

Proof (Proof of Theorem 3) Without loss of generality we can restrict ourselves to consider functions such that $\phi f = f$ and $\mathcal{L}_c f = \mathcal{L} f$. We start by the trivial remark that the result on the rotational descriptors contains the one on the generalized ones.

Let us consider

$$\begin{aligned} I_2^{\lambda,h}(f) &:= (\omega_f(R_h \lambda), \omega_f(\lambda)), \\ I_2^{\lambda_1, \lambda_2, h}(f) &:= (\omega_f(R_h \lambda_1) \odot \omega_f(\lambda_2), \omega_f(\lambda_1 + \lambda_2)). \end{aligned}$$

Since f is assumed to be compactly supported, its Fourier transform is analytic, and so are the functions $\lambda \mapsto I_2^{\lambda,h}(f)$ and $(\lambda_1, \lambda_2) \mapsto I_2^{\lambda_1, \lambda_2, h}(f)$ for any $h \in \mathbb{Z}_N$. Thus, the statement of the proposition reduces to show that $\text{RPS}_{\mathcal{L} f} = \text{RPS}_{\mathcal{L}_g}$ (resp. $\text{RB}_{\mathcal{L} f} = \text{RB}_{\mathcal{L}_g}$) if and only if $I_2^{\lambda,h}(f) = I_2^{\lambda,h}(g)$ (resp. $I_2^{\lambda_1, \lambda_2, h}(f) = I_2^{\lambda_1, \lambda_2, h}(g)$) for a.e. $\lambda, \lambda_1, \lambda_2 \in \mathcal{S}$ and all $h \in \mathbb{Z}_N$.

Let us recall the following properties of the tensor product, valid for all $v, v_1, v_2, w, w_1, w_2 \in \mathbb{C}^N$:

1. $(v \otimes w)^* = v \otimes w$,
2. $(v_1 \otimes w_1) \odot (v_2 \otimes w_2) = (w_1, v_2) v_1 \otimes w_2$,

By these and (12), we immediately have

$$\begin{aligned} \text{RPS}_{\mathcal{L}_c f}(\lambda, h) &= (\omega_{\phi f}(R_h \lambda)^*, \omega_{\phi f}(\lambda)^*) \omega_\psi(R_h \lambda)^* \otimes \omega_\psi(\lambda)^* \\ &= I_2^{\lambda,h}(f) \omega_\psi(R_h \lambda)^* \otimes \omega_\psi(\lambda)^*. \end{aligned}$$

Hence, whenever $\omega_\psi(R_h \lambda)^* \otimes \omega_\psi(\lambda)^* \neq 0$, $\text{RPS}_{\mathcal{L}_c f}(\lambda, h) = \text{RPS}_{\mathcal{L}_c g}(\lambda, h)$ if and only if $I_2^{\lambda,h}(f) = I_2^{\lambda,h}(g)$. Since $\omega_\psi(R_h \lambda)^* \otimes \omega_\psi(\lambda)^* \neq 0$ if and only if $\omega_\psi(\lambda) \neq 0$, by the fact that $\Psi \in \mathcal{R}$ this is true for a.e. $\lambda \in \mathcal{S}$. This completes the proof of the part of the statement regarding the rotational power-spectrum.

To prove the statement regarding the rotational bispectrum, let $B_f = A \circ \text{RBS}_{\mathcal{L} f}(\lambda_1, \lambda_2, h) \circ A^{-1}$, where A is the equivalence given by the Induction–Reduction Theorem and defined in (10). Since A is invertible, determining $\text{RBS}_{\mathcal{L} f}(\lambda_1, \lambda_2, h)$ is equivalent to determining B_f . Exploiting the fact that the r.h.s. of (9) is a diagonal matrix, we have

$$B_f^{k,\ell} = \left(A \circ \widehat{\mathcal{L} f}(T^{R_h \lambda_1}) \otimes \widehat{\mathcal{L} f}(T^{\lambda_1}) \circ A^{-1} \right)^{k,\ell} \circ \widehat{\mathcal{L} f}(T^{\lambda_1 + R_\ell \lambda_2})^*.$$

By Lemma 5, formula (12), and explicit computations, we then get

$$\begin{aligned} B_f^{k,\ell} &= (\omega_f(R_h \lambda_1) \odot \omega_f(R_\ell \lambda_2), \omega_f(\lambda_1 + R_\ell \lambda_2)) \\ &\quad \times (\omega_\psi(R_h \lambda_1)^* \odot \omega_\psi(R_{h+\ell} \lambda_2)^*) \otimes \omega_\psi(\lambda_1 + R_\ell \lambda_2). \end{aligned}$$

Similarly to before, $(\omega_\psi(R_h \lambda_1)^* \odot \omega_\psi(R_{h+\ell} \lambda_2)^*) \otimes \omega_\psi(\lambda_1 + R_\ell \lambda_2) \neq 0$ for a.e. $\lambda_1, \lambda_2 \in \mathcal{S}$ since $\Psi \in \mathcal{R}$. For these couples, $B_f = B_g$ if and only if

$$\begin{aligned} &(\omega_f(R_h \lambda_1) \odot \omega_f(R_\ell \lambda_2), \omega_f(\lambda_1 + R_\ell \lambda_2)) \\ &= (\omega_g(R_h \lambda_1) \odot \omega_g(R_\ell \lambda_2), \omega_g(\lambda_1 + R_\ell \lambda_2)). \end{aligned}$$

Finally, making the change of variables $R_\ell \lambda_2 \mapsto \lambda_2$ completes the proof of the theorem. \square

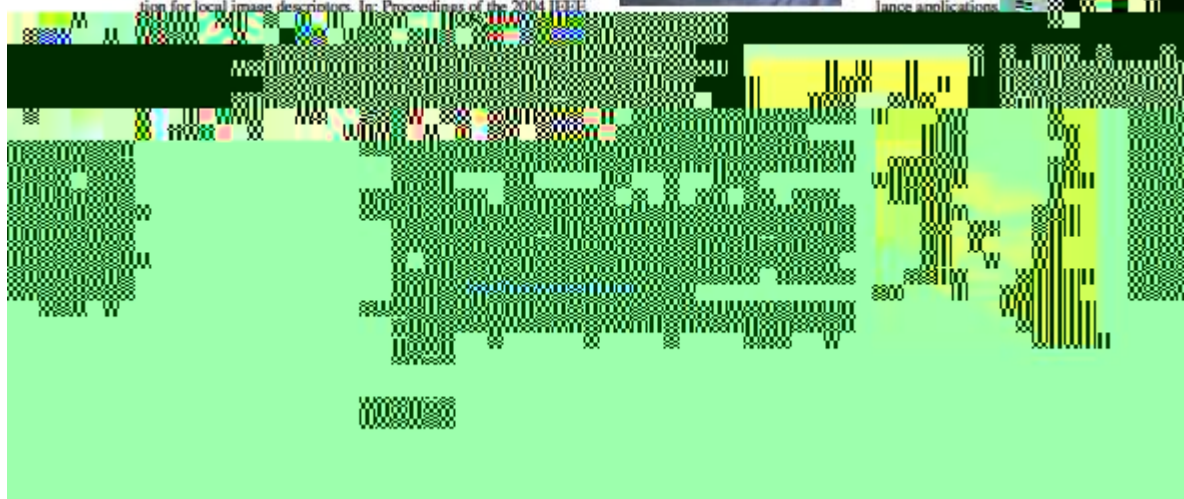
References

1. Barut, A., Raçzka, R.: Theory of Group Representations and Applications. World Scientific, Singapore (1977)
2. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: speeded up robust features. Computer Vision—ECCV 2006, pp. 404–417. Springer, Heidelberg (2006)
3. Boscaïn, U., Duplaix, J., Gauthier, J.P., Rossi, F.: Anthropomorphic image reconstruction via hypoelliptic diffusion. SIAM J. Control Optim. 50, 1–25 (2012)
4. Boscaïn, U., Chertovskih, R., Gauthier, J.P., Remizov, A.: Hypoelliptic diffusion and human vision: a semi-discrete new twist on the petiot theory. SIAM J. Imaging Sci. 7(2), 669–695 (2014a)
5. Boscaïn, U., Gauthier, J.P., Prandi, D., Remizov, A.: Image reconstruction via non-isotropic diffusion in Dubins/Reed–Shepp-like control systems. In: 53rd IEEE Conference on Decision and Control, pp. 4278–4283 (2014b)
6. Bressloff, P., Cowan, J., Golubitsky, M., Thomas, P., Wiener, M.: Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 356, 299–330 (2001)
7. Choksurivong, A., Emile, B., Rosenberger, C., Lauren, H.: Comparative study of global invariant descriptors for object recognition. J. Electron. Imaging 17, 1–35 (2008)
8. Chong, C., Raveendran, P., Mukundan, R.: A comparative analysis of algorithms for fast computation of Zernike moments. Pattern Recogn. 36(3), 731–742 (2003a)
9. Chong, C.W., Raveendran, P., Mukundan, R.: Translation invariants of zernike moments. Pattern Recogn. 36(8), 1765–1773 (2003b)

10. Chou, J., O'Neill, W., Cheng, H.: Pavement distress classification using neural networks. In: 1994 IEEE International Conference on Systems, Man, and Cybernetics, 1994. Humans, Information and Technology, vol. 1, pp. 397–401. doi:10.1109/ICSMC.1994.399871 (1994)
11. Citti, G., Sarti, A.: A cortical based model of perceptual completion in the roto-translation space. *Pattern Recogn.* **24**(3), 307–326 (2006)
12. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, IEEE, vol. 1, pp. 886–893 (2005)
13. Derrode, S., Ghorbel, F.: Robust and efficient fourier-mellin transform approximations for invariant grey-level image description and reconstruction. *Comput. Vis. Image Underst.* **83**(1), 57–78 (2001)
14. Dubnov, S., Tishby, N., Cohen, D.: Polyspectra as measures of sound texture and timbre. *Comput. Vis. Image Underst.* **26**(4), 277–314 (1997)
15. Duits, R., Franken, E.: Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores. Part I: Linear left-invariant diffusion equations on $SE(2)$. *Q. Appl. Math.* **68**, 255–292 (2010)
16. Duits, R., Franken, E.: Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores. Part II: Nonlinear left-invariant diffusions on invertible orientation scores. *Q. Appl. Math.* **68**, 1–38 (2010)
17. Führ, H., Mayer, M.: Continuous wavelet transforms from semi-direct products: Cyclic representations and Plancherel measure. *J. Fourier Anal. Appl.* **8**(4):1–23. <http://www.springerlink.com/index/G7TC4AANGTUC4HXW.pdf>, 0102002v1 (2002)
18. Galerne, B., Gousseau, Y., Morel, J.: Random phase textures: theory and synthesis. *IEEE Trans. Image Process.* **20**(1), 257–267 (2011)
19. Granlund, G.H.: Fourier preprocessing for hand print character recognition. *IEEE Trans. Image Process.* **C-21**(2), 195–201 (1972)
20. Hewitt, E., Ross, K.: *Abstract Harmonic Analysis—Volume 1*. Springer, Berlin/New York (1963)
21. Hjeltnas, E., Low, B.: Face detection: a survey. *Comput. Vis. Image Underst.* **83**(3), 236–274 (2001)
22. Hu, M.: Visual pattern recognition by moment invariants. *Comput. Vis. Image Underst.* **8**(2), 179–187 (1962)
23. Hubel, D., Wiesel, T.: Receptive fields of single neurones in the cat's striate cortex. *Comput. Vis. Image Underst.* **14**8, 574–591 (1959)
24. Kakarala, R.: The bispectrum as a source of phase-sensitive invariants for fourier descriptors: a group-theoretic approach. *Comput. Vis. Image Underst.* **44**(3), 341–353 (2012)
25. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE
31. Nene, S.A., Nayar, S.K., Murase, H., et al.: Columbia object image library (coil-20). Tech. rep., technical report CUCS-005-96 (1996)
32. Petitot, J.: *Neurogéométrie de la vision - Modèles mathématiques et physiques des architectures fonctionnelles*. Les Éditions de l'École Polytechnique (2008)
33. Prandi, D., Boscaïn, U., Gauthier, J.P.: Image processing in the semidiscrete group of roto-translations. In: To appear in the Proceedings of the 2nd Conference on Geometric Science of Information (2015)
34. Raja, D.M.S., Shanmugam, A.: Artificial neural networks based war scene classification using invariant moments and glcm features: A comparative study. *Int. J. Eng. Sci. Technol.* **3**(2), 1189–1195 (2011)
35. Rajasekaran, S., Pai, G.V.: Image recognition using simplified fuzzy artmap augmented with a moment based feature extractor. *Int. J. Pattern Recognit. Artif. Intell.* **14**(08), 1081–1095 (2000)
36. Samaria, F., Harter, A.: Parameterisation of a stochastic model for human face identification. In: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142 (1994)
37. Sifre, L., Mallat, S.: Rotation, scaling and deformation invariant scattering for texture discrimination. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1233–1240. doi:10.1109/CVPR.2013.163 (2013)
38. Smach, F., Lemaître, C., Gauthier, J.P., Miteran, J., Atri, M.: Generalized Fourier descriptors with applications to objects recognition in SVM context. *Int. J. Pattern Recognit. Artif. Intell.* **30**, 43–71 (2008)
39. Vapnik, V.N., Vapnik, V.: *Statistical Learning Theory*, vol. 1. Wiley, New York (1998)
40. Zahn, C.T., Roskies, R.Z.: Fourier descriptors for plane closed curves. *IEEE Trans. Comput.* **C-21**(3), 269–281 (1972)



Amine Bohi was born in Taza, Morocco, in 1987. He received the Master degree in computer science, computer vision and imaging from the University Sidi Mohammed Ben Abdellah - Faculty of science, Fes, Morocco, in 2011. He currently is a Ph.D. student at the University of South Toulon-Var (France). He belongs to the Laboratory for Systems and Information Science. His research interests concern object detection for multi-scale image applications.





Vincente Guis She received her Ph.D. degrees in signal and image processing applied to quality control of mammographic devices and to computer-aided diagnosis for breast cancer from the Aix-Marseille of University with ARCADES, Marseille, France, in 2002. From 2002 to 2009, she worked as a signal and image processing expert for research and development teams of different companies specializing in Astrophysics (Corot Project-CNES-LAM),

Geophysics (Magnitude joint venture of CGG Veritas Baker Hughes) and embedding image systems (Shaktiware). In 2010, she joined the LISIS Laboratory (SIIM team) at the University of Toulon, France, as a Research Engineer. Her current fields of interest are Physics, Informatics, Signal and Image Processing, and applied systems and she is an active member of the "Amicale Photon Matière" since 1999.



Jean-Paul Gauthier was born in 1952. He is currently a Professor at the University of Toulon (France) in the Dpt of Electrical Engineering. He got his Ph.D. in physics in 1982. He got the medal of "Institut Universitaire de France" in 1992. He was a member of this institute from 1992 to 97, and is now honorary member. He got a Featured review of the American Mathematical Society in 2002, for his work on the subanalyticity of Carnot-Cartheodory distances.

His current fields of interest are Automatic Control, Robotics, Signal and Image Processing, and Deterministic Observation Theory (he wrote a reference book at Cambridge University Press in 2001 on this last topic). Besides his academic activities, he is a member of "Federation Française de Go", "Ligue des Libres penseurs" and a strong supporter of "Toulon Rugby Club".



Frédéric Bouchara received the DEA (M.Sc degree) from the University of Marseille (France) and the Ph.D. degree from the University of Toulon (France) in 1992 and 1998, respectively. He is currently an associate professor of computer sciences at the University of Toulon. His research focuses on computer vision and image processing, mainly for applications in maritime surveillance and document image processing.