Contributions à l'analyse et à l'interprétation des images : Extraction et représentation de caractéristiques

Désiré Sidibé

Habilitation à Diriger des Recherches
Discipline *Sciences et Techniques*
Section CNU : 61

Contributions à l'analyse et à l'interprétation des images :
Extraction et représentation de caractéristiques

Présentée par
Désiré Sidibé

Maître de Conférences
Université de Bourgogne Franche-Comté
Laboratoire LE2I, UMR CNRS 6306

Soutenue le 1er Décembre 2016 devant le jury composé de :

Christian Daul	Prof. Université de Lorraine	Rapporteur
Denis Friboulet	Prof. INSA de Lyon	Rapporteur
Ludovic Macaire	Prof. Université de Lille	Rapporteur
David Fofi	Prof. Université de Bourgogne Franche-Comté	Examinateur
Fabrice Mériaudeau	Prof. Université de Bourgogne Franche-Comté	Examinateur
Alain Trémeau	Prof. Université Jean Monnet, St-Etienne	Examinateur
William Puech	Prof. Université de Montpellier	Président

SOMMAIRE

I

Introduction et Présentation

1

INTRODUCTION

Ce mémoire, rédigé en vue de l'obtention de l'Habilitation à Diriger des Recherches (HDR), offre un aperçu des travaux de recherche et d'encadrement que j'ai pu mener depuis l'obtention de mon doctorat. Il montre la diversité des champs d'application et de recherche (en vision et en imagerie médicale) que j'ai pu couvrir , ainsi que mon implication dans l'encadrement doctoral. Mais il n'y sera pas fait état des autres activités telles que l'investissement dans l'administration de la recherche (écriture de projets, coordination de projet, participation à des jury de thèse), et l'investissement dans l'enseignement (notamment les responsabilités pédagogiques dans deux masters internationaux).

1.1/ THÉMATIQUES DE RECHERCHE

Le but affirmé de la vision par ordinateur est de doter les machines de systèmes de vision, simulant ou dépassant la vision humaine (vision nocturne, panoramique, ou sous-marine par exemple), pour *faire voir les ordinateurs* (ou selon la terminologie anglaise, *make computers see*). Ce domaine de recherche s'est beaucoup développé ces dernières décénies et trouve aujourd'hui des applications dans de nombreux secteurs d'activités : la santé (systèmes d'imageries et de diagnostic), la sécurité (systèmes de vidéo-surveillance), l'industrie (systèmes de production automatisés), la communication (photographie et vidéo numériqe), les loisirs (les jeux vidéos, la réalité augmentée), etc.

Dans tous ces domaines, une fois les données (images et/ou vidéos) acquises, une étape importante du processus d'analyse concerne l'extraction de caractéristiques pertinentes qui facilitent l'interprétation des images et de la scène acquise. L'extraction et la représentation de caractéristiques/primitives dans les images et les vidéos sont au cœur de nos travaux de recherche. En effet, depuis mon recrutement à l'université de Bourgogne en septembre 2009, je travaille principalement dans l'équipe « Vision pour la Robotique » du Le2i sur les problématiques d'analyse de scènes dynamiques. Pour se déplacer et se localiser dans un environnement complexe, un robot mobile doit être capable de détecter et identifier les objets présents dans la scène. Dans le cas d'une scène dynamique, le robot doit également pouvoir prédire les positions des objets dans le temps pour planifier sa trajectoire. Dans ce contexte, nos travaux portent sur la détection de regions d'intérêt dans des séquences d'images, pour réduire la taile des données à traiter, et sur la détection et le suivi d'objets mobiles à l'aide de caméras atypiques.

D'autre part, depuis environ 5 ans, je travaille également avec les collègues de l'axe « Imagerie Médicale » pour l'analyse et l'interprétation d'images médicales, en particu-

lier les images rétiniennes pour le diagnostic de la rétinopathie diabétique qui est une complication oculaire du diabète se manifestant par l'apparition de lésions sur la rétine du patient diabétique. En France, 35 à 40 % des personnes diabétiques sont atteintes d'une rétinopathie, soit environ 800 000 personnes et celle-ci est la première cause de cécité chez les personnes de moins de 65 ans. Nos travaux dans ce domaine portent sur la détection des lésions rétiniennes, leur discrimination et l'identification automatique de patients malades par rapport à des patients sains.

Mes activités de recherche se divisent donc en deux grandes parties.

1. **Analyse de scènes dynamiques :** Afin d'analyser de manière efficace le flot de données visuelles pour en extraire les objets de la scène, nous adoptons une approche de type « bottom-up », dans laquelle, il faut dans un premier temps détecter des régions d'intérêt dans l'image, ces régions pouvant potentiellement contenir un objet. Ensuite, il faut extraire de ces régions d'intérêt des primitives (features) qui servent à caractériser l'objet. Cette approche est intéressante dans la mesure où l'on se focalise sur quelques régions particulières de l'image, réduisant ainsi, d'une part, la taille des données à traiter, et, d'autre part, le temps de calcul particulièrement limité pour une application robotique.
   - La détection de régions d'intérêt
     Si la détection de régions d'intérêt (ROI) a fait l'objet de nombreux travaux, notamment par la détection de régions saillantes, ceux-ci étaient limités à l'analyse d'images fixes. Nous nous sommes donc intéressés à la détection de régions saillantes dans des vidéos complexes, i.e. avec des arrières plan dynamiques, et avons montré l'intérêt d'une approche spatio-temporelle incluant la texture et le mouvement.
   - Le suivi d'objets mobiles
     Le suivi d'objets dans une séquence d'images peut-être formulé comme un problème de recherche de motifs dans une séquence d'images en utilisant un modèle de déplacement pour le(s) motif(s). Les primitives généralement employées sont la couleur et la texture. Nous avons montré que la prise en compte de la saillance visuelle permet d'obtenir des résultats plus robustes aux occultations et aux changements d'apparence des objets. D'autre part, nous avons également proposé une approche permettant d'appliquer les algorithmes de suivi visuel à des images omnidirectionnelles. Celle-ci est basée sur une représentation sphérique de l'image qui permet de prendre en compte les distorsions et la résolution non-uniforme de ce type d'images.
   - La reconnaissance d'objets 3D
     La reconnaissance des objets, 2D ou 3D, repose sur des descripteurs qui représentent l'objet de manière unique et facilitent son identification. Parmi les nombreux descripteurs de surface proposés dans la littérature, ceux basés sur l'orientation des points 3D, sont les plus couramment utilisés. Mais ils sont généralement représentés par des histogrammes de grandes tailles dont de nombreuses cellules sont vides. En outre, les descripteurs existants sont très sensibles au bruit et aux variations de point de vue, ce qui limite leur application pour la reconnaissance d'objets dans les systèmes embarqués, par exemple, les smartphones et les robots mobiles. Nous avons proposé un nouveau descripteur de nuage de points 3D combinant à la fois des propriétés géométriques et de texture, mais de taille très réduite grâce à une ACP. Ce descripteur est à la fois plus performant en terme de précision et de robustesse au bruit, et en

terme d'occupation de la mémoire pour le stockage.

2. **Analyse d'images de la rétine :** Le dépistage de la rétinopathie diabétique (RD) par rétinographie, i.e. par photographie du fond d'œil, est basé sur la détection de différentes lésions rétiniennes telles que les microanévrismes rétiniens (premiers signes ophtalmoscopiques notables), les hémorragies rétiniennes punctiformes, les nodules cotonneux, les exsudats profonds (exsudats secs) ou les druses. Nos travaux portent sur la détection de ces lésions, leur discrimination et l'identification automatique de patients malades par rapport à des patients sains.
   — La détection de lésions rétiniennes
   Parmi les lésions rétiniennes, les microanévrismes sont les plus difficiles à détecter de par leur nature (petite forme quasi-invisible à l'œil nu). Nous avons proposé une méthode de détection inspirée de la détection de point d'intérêt en vision, basée sur la détection de « blobs » et une approche multi-échelle. D'autre part, pour pallier la difficulté d'obtention d'un grand nombre d'exemples manuellement annotés pour l'apprentissage, nous avons proposé une méthode d'apprentissage semi-supervisée. Notre méthode a obtenu d'excellents résultats dans le challenge ROC (retinopathy online challenge). Nous nous sommes également intéressés à la détection des exsudats et avons proposé une méthode basée sur la création d'un atlas.
   — La discrimination de lésions
   La plupart des méthodes de détection dans la littérature sont spécifiques à un type de lésion, alors que plusieurs lésions différentes peuvent être présentes chez un même patient. De plus, les deux lésions exsudats et druses, bien que similaires par leur apparence sont les signes de deux pathologies très différentes. Respectivement, l'œdème maculaire diabétique (OMD) et la dégénérescence maculaire liée à l'âge (AMD). Nous avons donc proposé une méthode efficace de discrimination automatique d'images contenant ces deux types de lésions basée sur l'utilisation de représentations parcimonieuses.
   — La classification automatique d'images OCT
   Un outil de dépistage complémentaire de la rétinographie, est la tomographie par cohérence optique (OCT) qui permet d'obtenir une image tridimensionnelle de l'œil. Cependant, l'analyse des images OCT présente de nombreux challenges liés à la faible résolution spatiale, au bruit et à la difficulté de la segmentation des régions d'intérêt. Depuis 2015, nous travaillons dans le cadre d'un projet PHC que j'ai initié avec le SERI (Singapore Eye Research Institute) sur la classification automatique des images OCT et la détection de signes de l'OMD. Nous avons déjà proposé plusieurs méthodes d'extraction de caractéristiques et de classification des images OCT.

1.2/ STRUCTURE DU MÉMOIRE

Dans le chapitre 2, nous présentons brièvement notre parcours ainsi qu'un bilan de nos activités d'enseignement et de recherche.

Ensuite, le reste de ce manuscript est divisé en deux parties principales, chaque partie correspondant à l'une des deux principales thématiques de recherche.

La partie II traite de la détection et la reconnaissance d'objets dans des scènes dy-

namiques et est organisée en deux chapitres. Dans le chapitre 3, nous aborbons le problème de la détection de régions saillantes dans une séquence d'images, et dans le chapitre 4 nous abordons la détection et le suivi d'objets avec des caméras omnidirectionnelles et de profondeur.

La partie III est consacrée à l'analyse d'images de la rétine pour le dépistage de la rétinopathie diabétique. Dans le chapitre 5, nous présentons le problème ainsi que les méthodes proposées pour la détection de la rétinopathie diabétique en utilisant des images de fond d'œil (fundus images). Le chapitre 6 aborde le problème du diagnostic à l'aide de la tomographie par cohérence optique (OCT) et présente nos contributions.

Enfin, le dernier chapitre, le chapitre 7, présente les conclusions de nos travaux ainsi que les perspectives de recherches pour les années à venir.

2

Curriculum Vitæ

Dans ce chapitre je présente mes principales activités scientifiques, administratives et pédagogiques. Je commencerai par présenter mon parcours professionnel et universitaire, puis je décrirai mes activités d'enseignement, de recherche et d'encadrement. Enfin, ce chapitre se termine par une liste de mes publications scientifiques.

2.1/ Présentation

2.1.1/ Données personnelles

Nom : Sidibé
Prénom : Dro Désiré
Date de naissance : 28/01/1981
Grade : Maître de conférences, classe Normale
Etablissement : Université de Bourgogne-Franche Comté
Section CNU : 61
Unité de recherche : Le2i, UMR 6306 CNRS

2.1.2/ Parcours professionnel et universitaire

Depuis 2009	Maître de conférences Université de Bourgogne - IUT Le Creusot
2008-2009	Attaché Temporaire d'Enseignement et de Recherche Télécom Saint-Etienne
2007-2008	Chercheur post-doctorant LIRMM (Equipe ICAR), UMR 5506 CNRS, Montpellier
2007	Doctorat Ecole de Mines d'Alès et Université de Montpellier II
2004	Ingénieur et DEA Ecole Centrale de Nantes
2001	CPGE (MPSI & MP*) Lycée Buffon, Paris

2.2/ ACTIVITÉS D'ENSEIGNEMENT ET RESPONSABILITÉS PÉDAGOGIQUES

Depuis 2009, je suis affecté au département GEII (Génie Electrique et Informatique Industrielle) de l'IUT du Creusot où je suis responsable du cours de physique en première année. J'assure également une partie du cours d'automatique numérique pour les étudiants de seconde année.

D'autre part, j'effectue une partie importante de mon service dans les formations internationales du site universitaire Condorcet, au Creusot, où sont dispensés les enseignements des masters internationaux Vibot (Vision & Robotics) et Computer Vision. Dans ces formations, j'assure la reponsabilité des modules *Applied Maths*, *Probabilistic Robotics* et *Visual Tracking*. Ces cours sonts dispensés en Anglais.

2.2.1/ ENSEIGNEMENTS

Mes activités d'enseignement peuvent donc se résumer comme suit :

A l'IUT

- Physique, 1ère année (CM, TD)
  Mécanique générale, thermodynamique et transfert thermique, électrostatique et magnétisme.
- Systèmes numériques, 2nd année (CM, TD)
  Systèmes échantillonnés, Transformée en Z, fonction de transfert et stabilité.

Dans les masters Vibot et Computer Vision

- Applied Maths, M1 (CM, TD)
  Algèbre linéaire, probabilité et statistiques.
- Visual Perception, M1 (CM, TD, TP)
  Modèles de caméras, calibrage, géométrie épipolaire, mise en correspondance.
- Probabilistic Robotics, M1 (CM, TD, TP)
  Filtrage bayésien, filtre de Kalman, estimation d'état, SLAM.
- Visual Tracking, M2 (CM, TD, TP)
  Suivi d'objet, mean-shift, filtres particulaires.

Le volume horaire des mes enseignements varie de 210 à 240 heures par an, dont 2/5 de CM.

2.2.2/ RESPONSABILITÉS

Depuis ma nommination en 2009, je me suis fortement investi dans les formations internationales, dont le master Erasmus Mundus Vibot. J'assure plus de la moitié de mon service dans ces formations.

Je suis depuis 2015, directeur des études de la première année des masters internationaux, et coordinateur local du nouveau master Erasmus+ en imagerie médicale MAIA

(Medical Imaging and Applications) dont la première promotion a fait sa rentrée en septembre 2016.

D'autre part, j'ai été pendant 3 ans, de 2010 à 2013, responsable des relations internationales à l'IUT du Creusot (envoie et suivi d'environ 40 étudiants en stage à l'étranger chaque année).

2.3/ ACTIVITÉS DE RECHERCHES

2.3.1/ RÉSUMÉ DES TRAVAUX

Mes activités de recherches depuis la fin de ma thèse de doctorat concernent principalement l'extraction et la représentation de caractéristiques/primitives dans les images et les vidéos, avec deux principales applications : l'imagerie médicale et la détection et le suivi d'objets dans le domaine de la robotique mobile. Si ces deux domaines d'applications peuvent sembler distincts *a priori*, les techniques d'analyse d'images (ou de vidéos) mises en œuvre ne le sont pas, et les méthodes développées dans un domaine peuvent être appliquées à l'autre. En effet, dans chacune de ces applications, la première étape d'analyse consiste à détecter des régions d'intérêt dans l'image ; celles-ci pouvant soit correspondre à des objets (dans le cadre de la détection et du suivi avec un robot mobile), soit à des lésions potentielles (dans le cadre de l'analyse d'images rétiniennes).

Dans le premier domaine d'application, l'imagerie médicale, je me suis intéressé plus particulièrement à la classification automatique d'images rétiniennes. Du fait de la difficulté d'obtention d'un grand nombre d'exemples manuellement annotés dans le domaine médical, je me suis attaché à proposer des méthodes nécessitant peu d'images annotées pour l'apprentissage, ainsi que des approches d'extraction automatiques de caractéristiques dans ces images.

Dans le domaine de l'analyse de scènes dynamiques, je m'intéresse à des méthodes de détection et de suivi pouvant s'appliquer à différents types de caméras : perspectives, catadioptriques ou caméras de profondeur de type Kinect.

2.3.2/ ENCADREMENTS

J'ai, depuis 2010, eu l'opportunité de participer à l'encadrement de plusieurs thèses de doctorat, ainsi que de travailler avec de nombreux étudiants de M2.

Thèses

- Mazen Hittawee (2013-2015) (Financement ANR) : Détection et classification de défauts sur des planches de bois en défilement. Co-encadrement avec Fabrice Mériaudeau. Actuellement en post-doc au LM2S, UTT, France.
  Publications : [C38, C35, C34].
- Yasir Salih (2011-2015) (Co-tutelle avec UTP Malaisie) : 3D descriptors for robust objects detection. Co-encadrement avec Fabrice Mériaudeau et Aamir Malik (UTP). Actuellement chercheur à Umm Al-Qura University, Makkah, Saudi Arabia.
  Publications : [C30, C27].
- François Rameau (2011-2014) (Financement DGA-Région Bourgogne) : Surveillance aérienne à l'aide d'un système de vision hybride. Co-encadrement avec

Cédric Demonceaux et David Fofi. Actuellement Post-doc au KAIST, Corée du Sud.
Publications : [R2, C39, C28, C21, W1, N2].

— Staya Muddamsetty (2010-2014) (Financement région Bourgogne) : Visual saliency applied to complex dynamic scenes. Co-encadrement avec Fabrice Mériaudeau et Alain Trémeau. Maintenant Ingénieur en Suède.
Publications : [C32, C24].

Masters

— Anas Mahna (2016) (co-encadrement avec C. Demonceaux) : Visual-based localization in large-scale environment. Actuellement ingénieur chez Adasens Automotive GmbH, Allemagne.
— Ibrahim Sadek (2014) (co-encadrement avec F. Meriaudeau) : Automatic discrimanition of retinal images. Actuellement en thèse à IPAL Singapour. Publication [C37].
— Jilliam Diaz Barros (2014) (co-encadrement avec F. Garcia) : Human pose estimation from 3D point cloud. Actuellement en thèse à IEE, Luxembourg. Publication [C36].
— Alberto Quintero Delgado (2014) (co-encadrement avec Y. Benezeth) : Automatic spatial and temporal organization of long range video sequences. Actuellement ingénieur à Paris. Publication [W5].
— Andru P. Tiwanda (2013) (co-encadrement avec A. Comport) : Life-long localization and map learning. Actuellement en thèse à ICube, Strasbourg. Publication [W3].
— Kedir Adal (2012) (co-encadrement avec F. Mériaudeau) : Microaneurysm detection in retinal images. Actuellement en thèse à Delft University of Technology, Pays-Bas. Publications [R8, C23].
— Danda Pani Paudel (2012) (co-encadrement avec A. Habed) : Camera auto-calibration. Actuellement en post-doc à l'ETH Zurich après une thèse au Le2i.
— Sharib Ali (2012) (co-encadrement avec F. Mériaudeau) : Retinal images atlas creation. Actuellement post-doc en Allemagne après une thèse au CRAN, Nancy. Publications [R7, C25, C22].
— Darshan Venkatrayappa (2012) : Online feature selection for visual tracking. Actuellement en post-doc à Clermont Ferrand après une thèse à Nîmes. Publication [C31].
— Abhilash Srikantha (2011) : Ghost detection in HDR images. Actuellement en thèse à Max-Planck Institute, Germany. Publications [R5, C20].
— François Rameau (2011) (co-encadrement avec C. Demonceaux) : Visual tracking with omnidirectional cameras. Actuellement en post-doc au KAIST, après un thèse au Le2i. Publications [R2, W1].
— Valentine Vega (2011) (co-encadrement ave Y. Fougerolle) : Road signs detection with Gielis curves. Publication [C11].

2.3.3/ COLLABORATIONS

Au cours de mes activités scientifiques, j'ai pu développer des collaborations avec plusieurs collègues du Le2i, ainsi qu'avec des collègues d'autres laboratoires français et étrangers.

Collaborations au sein du Le2i

— Fabrice Mériaudeau (Professeur) : co-encadrement de plusieurs thèses (Yasir Salih, Satya Muddamsetty, Mazen Hittawee), de stagiaires de Master, et participation à divers projets de recherche (ANR CLAMEB, PHC Merlion).
— Cédric Demonceaux (Professeur) : co-encadrement de la thèse de François Rameau et participation au projet ANR PLATINUM.
— David Fofi (Professeur) : co-encadrement de la thèse de François Rameau.
— Yannick Benezeth (Maître de conférences) : co-encadrement de stagiaires de Master, travaux sur la détection d'objets mobiles.

Collaborations avec d'autres laboratoires français
— Alain Trémeau (Professeur) du Laboratoire Hubert Curien, Saint-Etienne. Co-encadrement de la thèse de Satya Muddamsetty et participation aux travaux de thèse du doctorant de St-Etienne M. Nawaf. Publications [C32, C29, C24, C6].
— William Puech (Professeur) et Olivier Strauss (Maître de Conférences) du LIRMM, Montpellier. Travaux sur le de-ghosting d'images HDR et la détection d'arrière plan. Publications [C5, C4].

Collaborations internationales
— Aamir Malik (Maître de conférences) de l'UTP (Universiti Teknologi Petronas). Co-encadrement de la thèse de Yasir Salih. Publications [C30, C27].
— Thomas Karnowski (Chercheur) à Oak Ridge National Laboratory, USA. Travaux sur la rétinopathie. Publications [R9, R8, C23, C22].
— Carol Y. Cheung (Assistant Professor) à Chineese University of Hong-Kong. Projet de recherche sur la rétinopathie (PHC Merlion). Publications [R12, R10, C43, C42, C41].

2.3.4/ PROJETS DE RECHERCHE

J'ai eu l'occasion de participer à différents projets de recherche au cours des dernières années.

- 2016 - 2019 : Projet ANR PLATINUM (Cartographie Long-Terme pour la Navigation Urbaine)
  Partenaires : LITIS Rouen (coordinateur), MATIS (IGN), Le2i, Lagdic (INRIA).
  Implication : co-encadrement d'une thèse (à partir de 2016).

- 2016 - 2017 : Projet ANR VIPER (VIsion PolarimétriquE pour la navigation de Robots)
  Partenaires : Le2i (O. Morel est le porteur de cette ANR JCJC).
  Implication : Responsable d'un work-package.

- 2015 - 2016 : PHC Merlion (Automatic tools for diabetic macular edema detection form SD-OCT)
  Partenaires : Le2i, SERI (Singapore Eye Research Institute).
  Implication : coordinateur.

- 2012 - 2015 : Projet ANR CLAMEB (Classification mécanique non destructif du bois)
  Partenaires : LaBoMap (Arst et Métiers), Le2i, FCBA, et 3 industriels du bois.

Implication : co-encadrement de la thèse de Mazen Hittawe.

- 2011 - 2012 : Analyse d'images rétiniennes pour le diagniostic de la rétinopathie diabétique
  Partenaires : Oak Ridge National Lab (USA), University of Tenessee (USA), Le2i.
  Implication : co-encadrement de 2 stages de master.

2.3.5/ RAYONNEMENT SCIENTIFIQUE ET RESPONSABILITÉS

Prix et distinctions

- Prime d'encadrement doctoral et de recherche (PEDR), depuis 2014.
- Prix du meilleur papier étudiant, "*Winner of Robert F. Wagner Student Best Paper Award*", à la conférence SPIE Medical Imaging 2015, Orlando, USA.
- Best papers of the year 2012 du journal Computer Methods and Programs in Biomedicine (pour l'article [R3]).

Sociétés scientifiques

- Membre de IEEE, IAPR.
- Membre associé du comité technique IVMSP de IEEE Signal Processing Society.
- Membre du GDR ISIS dans le thème B : Image et Vision.

Activités éditoriales

- Membre du comité de programme de ICIRA 2011.
- Relecteur pour les conférences internationales suivantes :
  ICRA (2016), IROS (2013), EUSIPCO (2009, 2010, 2011, 2013, 2015), ICDAR (2013), ACIVS (2011), ICIRA (2010, 2011), CCIW (2009), IPTA (2008).
- Relecteur pour les revues internationales suivantes (environ 10 articles par an) :
  — IEEE Trans. Medical Imaging, IEEE Signal Processing Letters
  — Medical Image Analysis, Computers in Biology and Medicine, Computerized Medical Imaging and Graphics
  — Signal Image and Video Processing, Signal Processing :Image Communication, IET Computer Vision, Journal of Electronic Imaging, Signal Processing, Sensors

Organisations diverses

- Organisation, avec Fabrice Mériaudeau, de l'école d'été européenne COMVICS (Erasmus Intensive Programm in Computer Vision and Intelligent Systems). 40 étudiants participants de 8 pays.
- Membre du comité d'organisation de CCIW 2009.

Séminaires invités

- "*Matrix decomposition techniques in computer vision and image analysis*", Ecole doctorale I2S, Univ. de Montpellier, 27 Avril 2016.
- "*Sparse coded feature for bright lesions discrimination in retinal images*", Journée du GDR-ISIS, 19 Novembre 2014, Lyon.
- "*Target representation for visual tracking : an adaptation to catadioptric imaging systems*", Journée de travail du projet MOSCA, 27 Janvier 2012, Clermont-Ferrand.
- "*Particle Filters and Applications in Computer Vision*", Ecole doctorale I2S, Univ. de Montpellier, 6 Avril 2011.

- "*Matching local features : application to object recognition*", Séminaire invité, Nanjing University of Science and Technology, 23 Novembre 2007, Chine.

Jury de thèse

- Membre du jury (examinateur) pour la thèse de Mohamad Motasem Nawaf, Université Jean Monnet, St-Etienne, 2014.
- Membre du jury (examinateur) pour la thèse de Jean-Louis Palomares, Université de Montpellier II, 2012.
- Membre du jury (examinateur) pour la thèse de Jhimli Mitra, Université de Bourgogne, 2012.

2.4/ PRODUCTION SCIENTIFIQUE

Bilan des publications :

— **Revues internationales :** 12
— **Conférences internationales :** 44
— **Conférences nationales :** 4
— **Workshop internationaux :** 6
— **Brevet :** 1
— **En révision :** 2 revues

Revues Internationales avec comité de lecture

[R12] D. Sidibé, S. Sankar, G. Lemaître, M. Rastgoo, J. Massich, C. Y. Cheung, T. Y. Wong, G. S. W. Tan, E. Lamoureux, D. Milea, F. Meriaudeau. "An anomaly detection approach for the identification of DME patients using spectral domain optical coherence tomography images", Computer Methods and Programs in Biomedicine, 139, pp. 109-117, 2017. $IF = 1.86$

[R11] D. Sidibé, F. Mériaudeau, "Visual saliency detection in colour images based on density estimation", Electronics Letters, 53(1), pp. 24-25, 2016. $IF = 0.93$

[R10] G. Lemaître, M. Rastgoo, J. Massich, C. Y. Cheung, T. Y. Wong, E. Lamoureux, D. Milea, F. Meriaudeau, D. Sidibé, "Classification of SD-OCT volumes using local binary patterns : experimental validation for DME detection" Journal of Ophthalmology, 2016. $IF = 1.46$

[R9] D. Sidibé, I. Sadek, F. Mérieaudeau, "Discrimination of retinal images containing bright lesions using sparse coded features and SVM", Computers in Biology and Medicine, 62, pp. 175-184, 2015. $IF = 1.52$

[R8] K. Adal, D. Sidibé, S. Ali, E. Chaum, T. Karnowski, F. Mériaudeau. "Automated Detection of Microaneurysms Using Scale-Adapted Blob Analysis and Semi-Supervised Learning", Computer Methods and Programs in Biomedicine, 114(1), pp. 1-10, 2014. $IF = 1.86$

[R7] S. Ali, D. Sidibé, K. Adal, L. Giancardo, E. Chaum, T. Karnowski, F. Mériaudeau, "Statistical Atlas based Exudate Segmentation", Computerized Medical Imaging and Graphics, 37(5-6), pp. 358-368, 2013. $IF = 1.38$

[R6] S. Ghose, A. Oliver, J. Mitra, R. Marti, X. Llado, J. Freixenet, D. Sidibé, J. Vilanova, J. Comet, F. Mériaudeau. "A Supervised Learning Framework of Statistical Shape and Probability Priors for Automatic Prostate Segmentation in Ultrasound Images", Medical

Image Analysis, 17(6), pp. 587-600, 2013. $IF = 4.56$

[R5] A. Srikantha, D. Sidibé, "Ghost Detection and Removal for High Dynamic Range Images : Recent Advances", Signal Processing : Image Communication, vol. 27(6), pp. 650-662, 2012. $IF = 1.60$

[R4] J. Mitra, Z. Kato, R. Marti, A. Oliver, X. Llado, D. Sidibé, S. Ghose, J. C. Vilanova, J. Comet, F. Meriaudeau, "A Spline-Based Non-linear Diffeomorphism for Multimodal Prostate Registration", Medical Image Analysis, vol. 16(6), pp. 1259-1279, 2012. $IF = 4.56$

[R3] S. Ghose, A. Oliver, R. Marti, X. Llado, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibé, F. Meriaudeau, "A Survey of Prostate Segmentation Methodologies in Ultrasound, Magnetic Resonance and Computed Tomography Images", Computer Methods and Programs in Biomedicine, vol. 108(1), pp. 262-287, 2012. $IF = 1.86$

[R2] F. Rameau, D. Sidibé, C. Demonceaux, D. Fofi, "Visual Tracking with Omnidirectional Cameras : An Efficient Approach", Electronic Letters, vol. 47(21), pp. 1183-1184, 2011. $IF = 0.93$

[R1] D. Sidibé, P.Montesinos, S. Janaqi, "Matching Local Invariant Features with Contextual Information : An Experimental Evaluation", Electronic Letters on Computer Vision and Image Analysis, vol. 7(1) : 26-39, 2008.

Conférences Internationales avec comité de lecture

[C44] D. Sidibé, M. Rastgoo, F. Mériaudeau, "On spatio-temporal saliency detection in videos using multilinear PCA", ICPR 2016, Mexico

[C43] J. Massich, M. Rastgoo, G. Lemaître, C. Y. Chaung, T.Y. Wong, D. Sidibé, F. Mériaudeau, "Classifying DME vs Normal SD-OCT Volumes : A Review", ICPR 2016, Mexico.

[C42] K. Alsaih, G. Lemaitre, J. Massich, M. Rastgoo, D. Sidibe, T. Y. Wong, E. Lamoureux, D. Milea, C. Leung, and F. Meriaudeau, "Classification of SD-OCT volumes with multi-pyramids, LBP and HoG descriptors : Application to DME detection", IEEE Engineering in Medicine and Biology Society (EMBC) 2016. Orlando, USA.

[C41] S. Sankar, D. Sidibé, C. Y. Cheung, T. Y. Wong, E. Lamoureux, D. Milea, F. Meriaudeau, "Classification of SD-OCT volumes for DME detection : an anomaly detection approach", SPIE Medical Imaging 2016, San Diego, USA.

[C40] M. Rastgo, G. Lemaître, O. Morel, J. Massich, F. Marzani, R. Garcia, D. Sidibé, "Classification of melanoma lesions using sparse coded features and random forests", SPIE Medical Imaging 2016, San Diego, USA.

[C39] F. Rameau, D. Sidibé, C. Demonceaux, D. Fofi, "Structure from motion using a hybrid stereo-vision system", in IEEE URIA (Ubiquitous Robots and Ambient Intelligence), Goyank city, Korea, 2015.

[C38] M. M. Hittawe, S. M. Muddamsetty, D. Sidibé, F. Mériaudeau, "Multiple Features Extraction for Timber Defects Detection and Classification with SVM", in IEEE ICIP 2015, Quebec, Canada.

[C37] I. Sadek, D. Sidibé, F. Mériaudeau, "Automatic discrimination of color retinal images using the bag of words approach ", in SPIE Medical Imaging, 2015. USA. **Winner of Robert F. Wagner Best Student Paper Award.**

[C36] J. Diaz Barros, F. Garcia, D. Sidibé, “Real-Time Human Pose Estimation from Body-Scanned Point Clouds”, in VISAPP 2015, Berlin, Germany.

[C35] M. M. Hittawe, D. Sidibé, F. Meriaudeau, “Bag of words representation and SVM classifier for timber knots detection on color images”, in MVA 2015, Japan.

[C34] M. M. Hittawe, D. Sidibé, F. Meriaudeau, “A machine vision based approach for timber knots detection”, in QCAV 2015, France.

[C33] G. Lemaître, J. Massich, R. Marti, J. Freixenet, J. C. Vilanova, P. M. Walker, D. Sidibé, F. Mériaudeau, “A boosting approach for prostate cancer detection using multi-parametric MRI”, in QCAV 2015, France.

[C32] S. Muddamsetty, D. Sidibé, A. Trémeau, F. Mériaudeau, “Spatio-Temporal Saliency Detection in Dynamic Scenes using Local Binary Patterns”, in ICPR 2014. Stockholm, Sweden.

[C31] D. Venkatrayappa, D. Sidibé, F. Meriaudeau, P. Montesinos, “Adaptive Feature Selection for Object Tracking with Particle Filter”, in ICIAR 2014, Vilamoura, Algarve, Portugal.

[C30] Y. Salih, A. S. Malik, D. Sidibé, M. T ; Simsim, N. Saad, F. Mériaudeau, “Compressed VFH Descriptor for 3D Object Classification”, in 3DTV-Con 2014, Budapest, Hungary, 2014.

[C29] M. Nawaf, A. Trémeau, MD Abul Hasnat, D. Sidibé, “Color and Flow based Superpixels for 3D Geometry Respecting Meshing”, in WACV 214. Colorado, USA.

[C28] F. Rameau, C. Demonceaux, D. Sidibé, D. Fofi, “Control of a PTZ Camera in a Hybrid Vision Sysytem”, in VISAPP 2014. Lisbon, Portugal.

[C27] Y. Salih, A. Malik, N. Walter, D. Sidibé, N. Saad, F. Mériaudeau, “Noise Robustness Analysis of Point Cloud Descriptors”, in ACIVS 2013. Poznan, Ploand.

[C26] F. Meriaudeau, D. Sidibé, K. Adal, S. Ali, L. Giancardo, T. Karnowski, E. chaum, “Computer Aided Design for Diabetic Retinopathy”, in QCAV 2013. Fukuoka, Japan.

[C25] S. Ali, K. Adal, D. Sidibé, T. Karnoswki, E. Chaum, F. Meriaudeau, “Exudate Segmentation on Retinal Atlas Space”, in ISPA 2013. Trieste, Italy.

[C24] S. Muddamsetty, D. Sidibé, A. Trémeau, F. Mériaudeau, ”A Performance Evaluation of Fusion Techniques for Spatio-Temporal Saliency Detection in Dynamic Scenes”, in ICIP 2013. Melbourne, Australia.

[C23] K. Adal, S. Ali, D. Sidibé, T. Karnowski, F. Mériaudeau, “Automated Detection of Microaneurysm Using Robust Blob Descriptors”, in SPIE Medical Imaging, Orlando, Florida - USA, 9-14 February 2013.

[C22] S. Ali, K. Adal, D. Sidibé, T. Karnowski, F. Mériaudeau, “Steerable Transform for Atlas-Based Retinal Lesion Segmentation”, in SPIE Medical Imaging, Orlando, Florida - USA, 9-14 February 2013.

[C21] F. Rameau, A. Habed, C. Demonceaux, D. Sidibé, D. Fofi, “Self-calibration of PTZ camera using new LMI constraints”, ACCV 2012, Daejon, South Korea, November 2012.

[C20] A. Srikantha, D. Sidibé, F. Meriaudeau, ”An SVD-Based Approach for Ghost Detection and Removal in High Dynamic Range Images”, ICPR 2012 - Tsukuba, Japan, November 2012.

[C19] J. Mitra, Z. Kato, S. Ghose, D. Sidibé, R. Martí, X. Llado, A. Oliver, J. C. Vilanova, F. Meriaudeau, ”Spectral Clustering to Model Deformations for Fast Multimodal Prostate Registration”, ICPR 2012 - Tsukuba, Japan, November 2012.

[C18] S. Ghose, J. Mitra, A. Oliver, R. Marti, X. Llado, J. Freixenet, J. C. Vilanova, D. Sidibé, F. Meriaudeau, ”Graph Cut Energy Minimization in a Probabilistic Learning Framework for 3D Prostate Segmentation in MRI”, ICPR 2012 - Tsukuba, Japan, November 2012.

[C17] S. Ghose, J. Mitra, A. Oliver, R. Marti, X. Llado, J. Freixenet, J. C. Vilanova, J. Comet, D. Sidibé, F. Meriaudeau, ”A Mumford-Shah Functional based Variational Model with Contour, Shape, and Probability Prior information for Prostate Segmentation”, ICPR 2012 - Tsukuba, Japan, November 2012.

[C16] S. Ghose, J. Mitra, A. Oliver, R. Marti, X. Llado, J. Freixenet, J. C. Vilanova, D. Sidibé, F. Meriaudeau, “A Couple Schema of Probabilistic Atlas and Statistical Shape and Appearance Model for 3D Prostate Segmentation in MR Images”, IEEE ICIP 2012 – Orlando, Florida – USA, October 2012.

[C15] J. Mitra, S. Ghose, D. Sidibé, A. Oliver, R. Marti, X. Llado, J. C. Vilanova, J. Comet, F. Meriaudeau, “Weighted Likelihood Function of Multiple Statistical Parameters to Retrieve 2D TRUS-MRI Slice Correspondences for Prostate Biopsy”, IEEE ICIP 2012 - Orlando, Florida – USA, October 2012.

[C14] J. Mitra, S. Ghose, D. Sidibé, R. Marti, A. Oliver, X. Llado, J. C. Vilanova, J. Comet, F. Meriaudeau, ”Joint Probability of Shape and Image Similarities to Retrieve 2D TRUS-MR Slice Correspondence for Prostate Biopsy”, IEEE EMBC 2012 - San Diego, California, USA, August 2012.

[C13] S. Ghose, J. Mitra, A. Oliver, R. Marti, X. Llado, J. Freixenet, J. C. Vilanova, J. Comet, D. Sidibé, F. Meriaudeau, “Spectral Clustering of Shape and Probability Prior Models for Automatic Prostate Segmentation in Ultrasound Images”, IEEE EMBC 2012 - San Diego, California, USA, August 2012.

[C12] S. Ghose, J. Mitra, A. Oliver, R. Marti, X. Llado, J. Freixenet, J. C. Vilanova, J. Comet, D. Sidibé, F. Meriaudeau, ”A Supervised Learning Framework for Automatic Prostate Segmentation in Trans Rectal Ultrasound Images”, ACIVS 2012 – Brno, Czech Republic, September 2012.

[C11] V. Véga, D. Sidibé and Y. Fougerolle, “Road Signs Detection and Shape Reconstruction using Gielis Curves”, VISAPP 2012, Roma, Italy, 24-26 February 2012.

[C10] B. Khanal, S. Ali and D. Sidibé, “Robust Road Signs Segmentation in Color Images”, VISAPP 2012, Roma, Italy, 24-26 February 2012.

[C9] J. Mitra, A. Srikantha, D. Sidibé, R. Marti, A. Oliver, X. Llado, S. Ghose, J. C. Vilanova, J. Batle, F. Meriaudeau, “A shape-based statistical method to retrieve 2D TRUS-MR slice correspondence for prostate biopsy”, SPIE Medical Imaging 2012 - San Diego, California, USA, 5-9 February 2012.

[C8] B. Khanal, D. Sidibé, “Efficient Skin Detection under Illumination Changes and Shadows”, ICIRA 2011 – Aachen, Germany, December 2011.

[C7] D. Sidibé, D. Fofi, F. Mériaudeau, “Using Visual Saliency for Object Tracking with Particle Filters”, EUSIPCO 2010 - Aalborg, Denmark, 23-27 August 2010.

[C6] D. Sidibé, P. Montesinos, T. Tremeau, "Robust Facial Features Tracking using Geometric Constraints and Relaxation", IEEE MMSP 2009 - Rio de Janeiro, Brazil, October 5 – 7, 2009.

[C5] D. Sidibé, W. Puech and O. Strauss "Ghost Detection and Removal in High Dynamic Range Images", EUSIPCO 2009 - Glasgow, Scotland, 24 – 28 August, 2009.

[C4] D. Sidibé, O. Strauss and W. Puech "Automatic Background Generation from a Sequence of Images Based on Robust Mode Estimation", SPIE, IS&T Electronic Imaging, Digital Photography V. San Jose, California, USA, 18 – 22 January 2009.

[C3] D. Sidibé, P. Montesinos, S. Janaqi, "Fast and Robust Image Matching Using Contextual Information and Relaxation", VISAPP 07 - Barcelona, Spain, March 2007.

[C2] D. Sidibé, P. Montesinos, S. Janaqi, "Matching Local Invariant Features : How Can Contextual Information Help ?", SIPMCS 07 - Maribor, Slovenia, June 2007.

[C1] D. Sidibé, P. Montesinos, S. Janaqi, "A Simple and Efficient Eye Detection Method in Colour Images", IVCNZ 2006 - Great Barrier Island, New Zealand, November 2006.

Workshops Internationaux

[W6] G. Lemaître, J. Massich, M. Rastgoo, S. Sankar, F. Mériaudeau, D. Sidibé, "Classification of SD-OCT volumes using LBP : application to DME detection", Ophthalmic Images Analysis Workshop (OMIA), in conjunction with MICCAI 2015.

[W5] A. Q. Delgado, Y. Benezeth, D. Sidibé, "Automatic spatial and temporal organization of long range video sequences from low level motion features", Scene Uderstanding Workshop (SUN 2014), in conjunction with CVPR 2014.

[W4] Y. Benezeth, D. Sidibé, J.B. Thomas, "Background subtraction with multispectral videos sequences", OMNIVIS 2014.

[W3] A. P. Twinanda, M. Meilland, D. Sidibé, A. I. Comport, "On Keyframe Positioning for Pose Graphs Applied to Visual SLAM", IROS Workshop on Navigation, 2013. Tokyo, Japan

[W2] S. Ghose, J. Mitra, A. Oliver, R. Martí, X. Lladó, J. Freixenet, J.C. Vilanova, D. Sidibé, and F. Meriaudeau. "A Stochastic Approach to Prostate Segmentation in MRI", MICCAI Grand Challenge : Prostate MR Image Segmentation - Nice, France, October 2012.

[W1] F. Rameau, D. Sidibé, C. Demonceaux, D. Fofi, "Tracking a Moving Object with a Catadioptric Sensor using Particle Filter", OMNIVIS 2011 – Barcelona, Spain, November 2011.

Conférences nationales

[N4] F. Rameau, C. Demonceaux, D. Sidibé, D. Fofi, "Etude d'un système de vision hybride", ORASIS 2013, Cluny, France, juin 2013.

[N3] O. Strauss, D. Sidibé, W. Puech, "Détection robuste de mouvement par histogrammes quasi-continus", LFA 2012 – Compiègne, France, 15-16 novembre 2012.

[N2] F. Rameau, D. Sidibé, C. Demonceaux, D. Fofi, "Une approche performante de suivi visuel pour les caméras catadioptriques", RFIA 2012 – Lyon, France, janvier 2012.

[N1] D. Sidibé, P. Montesinos, S. Janaqi, "Mise en correspondance robuste d'invariants locaux par relaxation", ORASIS 2007 – Obernai, France, juin 2007.

II

DÉTECTION ET RECONNAISSANCE D'OBJETS

3

SAILLANCE VISUELLE DANS DES SCÈNES DYNAMIQUES

La détection de zones ou régions d'intérêt (ROI) dans une image (ou une séquence d'images) est une étape indispensable pour l'analyse et l'interprétation de scènes. Une approche intéressante consiste à détecter les régions visuellement *saillantes*, i.e. des régions qui se distinguent par certaines caractéristiques de leur voisinage. Si la détection de régions saillantes dans les images est un problème largement étudié dans la littérature, l'extension au domaine temporel a fait l'objet de peu d'études. Dans ce chapitre, nous nous intéressons donc à la détection de régions saillantes dans des scènes dynamiques complexes, i.e. avec des arrières plan dynamiques, et nous montrons l'intérêt d'une approche spatio-temporelle incluant la texture et la couleur.

3.1/ INTRODUCTION

L'attention visuelle spatiale est un mécanisme d'attention sélective qui permet de sélectionner et de traiter les parties les plus pertinentes du flot de données visuelles que nous percevons à chaque instant. Elle est en particulier déployée lorsqu'un observateur dirige son attention vers un ou plusieurs endroits précis du champ visuel. Ce mécanisme permet de focaliser l'attention de l'observateur vers une zone précise du champ visuelle, et si un stimulus apparait à cet endroit alors l'information de ce stimulus est traitée en priorité.

L'attention visuelle a fait l'objet de nombreuses études dans les domaines de la psychologie et des neurosciences [21, 42, 125, 65, 166]. Nous n'aborderons pas ces aspects biologiques/cognitifs de ce chapitre, et nous nous contenterons ici de mentioner quelques travaux dans le domaine de la vision par ordinateur.

Il est communément admis que l'attention visuelle repose sur deux grand types de processus [51] :

1. **Bottom-up :** Ce sont des processus basés sur la sélection automatique d'éléments de la scène qui sont ensuite associés du bas vers le haut, i.e. chaque niveau du processus associe des éléments séléctionnés au niveau inférieur. La sélection est basée sur la théorie d'intégration de primitives (Feature Integration Theory) de Treisman et Gelade, selon laquelle la vision précoce consiste à extraire de l'image des primitives (ou des attributs) sur des plans distincts (feature maps), en concordance avec la spécialisation des différentes aires du cortex visuel [163].

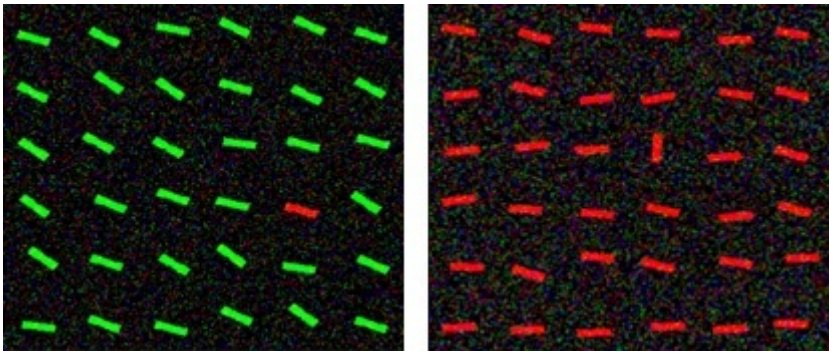


FIGURE 3.1 – Illustration de l'effet *pop-out*.

2. **Top-down :** Ce sont des processus contrôlés ou dirigés par des facteurs externes tels que les intentions ou la réalisation d'une tâche. Ils font donc intervenir la volonté du sujet.

De nombreux chercheurs se sont intéressés à l'étude des processus attentionnels qu'ils soient « bottom-up » ou « top-down » et aux liens existants entre les deux. Les études ont montré que les mécanismes « bottom-up » sont plus rapides et qu'ils précèdent les influences « top-down » plus longues à mettre en œuvre et qui durent dans le temps [176, 107].

Dans le domaine de la vision par ordinateur, la détection de régions d'intérêt permet de se focaliser sur quelques régions particulières de l'image. Ce qui a comme intérêt, d'une part, de réduire la taille des données à traiter, et, d'autre part, de limiter le temps de calcul particulièrement limité pour une application de robotique par exemple.

3.1.1/ CARTE DE SAILLANCE DANS LES SCÈNES DYNAMIQUES

La plupart des modèles proposés dans la littérature ont pour but de produire, à partir d'une image d'entrée I, une carte de saillance S qui fait ressortir les régions d'intérêt (ou régions saillantes). Une région saillante est caractérisée par ses attributs de couleur, de forme ou de texture, en contraste avec les régions voisines. C'est l'effet *pop-out* illustré par la figure 3.1, où sur l'image de gauche le trait rouge se distingue par sa couleur, et sur l'image de droite le trait vertical se distingue par son orientation.

Dans la pratique, on calcule une mesure de saillance pour chaque région centrée sur un pixel $\mathbf{x} = (x, y)$ de l'image. Deux approches sont possibles :

1. **Center-surround :** On extrait une fenêtre W_C centrée sur $\mathbf{x}$ qui caractérise la région à évaluer, ainsi qu'une fenêtre W_S centrée sur $\mathbf{x}$ qui représente l'arrière-plan. On extrait ensuite des attributs (couleur, texture, gradient) de chaque fenêtre et on calcule une mesure de dissimilarité qui indique la saillance du pixel $\mathbf{x}$. Cette approche est illustrée par la figure 3.2(a).
2. $k-$**nearest neighbours :** On extrait une fenêtre W_C centrée sur $\mathbf{x}$ qui caractérise la région à évaluer, et on la compare à k régions W_k voisines. En général, on considère des régions dans un rayon fixé. La somme des dissimilarités de W_C avec les W_k indique la saillance du pixel $\mathbf{x}$. Cette approche est illustrée par la figure 3.2(b).

Les deux approaches sont illustrées par les images de la figure 3.2. La première approche est plus rapide, mais la seconde est plus robuste comme le montre les auteurs

(a) Approche center-surround (b) Approche $k-$nearest neighbours

FIGURE 3.2 – Calcul de cartes de saillance.

dans [66].

Depuis les travaux de Itti, Koch et Neibur qui ont proposé en 1998 le premier modèle computationnel de détection de régions saillantes dans une image [94], il y a eu un très grand nombre de travaux et de modèles proposés. Un état de l'art complet sur ce sujet a été récemment réalisé par Borji et Itti [26]. Néanmoins, la plupart des méthodes sont limitées aux scènes statiques et peu de travaux concernent les scènes dynamiques. Nos travaux portant sur l'analyse des scènes dynamiques, nous ne nous intéresserons qu'aux méthodes proposées pour le calcul de la saillance dans des scènes dynamiques. Nous invitons le lecteur intéressé par le calcul de la saillance dans les images, à se référer à [26, 25]. Nous décrivons ici, brièvement, les principales idées mises en œuvre dans les différentes méthodes.

Une grande majorité des méthodes proposées dans le cas de scènes dynamiques est basée sur la fusion d'une carte statique calculée pour chaque image de la séquence, et d'une carte temporelle calculée à partir d'images successives. La carte statique M_S est généralement basée sur des attributs de couleur tandis que la carte temporelle ou dynamique M_D est elle basée sur les attributs de mouvement. Ces deux cartes sont combinées en une carte spatio-temporelle M_F.

On peut classer les différentes méthodes dans deux grandes catégories :

1. **Motion-based :** Les méthodes de cette catégorie estiment le déplacement de chaque pixel ou de chaque région, en utilisant le flot optique par exemple, pour en déduire la saillance du pixel ou de la région.
2. **Spatio-temporal center-surround :** Les méthodes de cette catégorie étendent le concept de *center-surround* au domaine temporel en utilisant des volumes spatio-temporels.

Parmi les méthodes de la première catégorie, citons les travaux de Marat *et al.* [105] qui proposent une méthode dans laquelle le mouvement de chaque pixel est estimé par le flot optique après compensation du mouvement de l'arrière plan entre deux vues successives. Le Meur *et al.* [87] utilisent également l'information du flot optique pour détecter la saillance visuelle dans des vidéos. Une approche similaire est utilisée par Tong *et al.* [182] qui utilisent en plus du déplacement, l'orientation et la phase du mouvement. Mancas *et al.* [104] utilisent aussi le flot optique, mais celui-ci est discritisé selon 4 directions (Nord, Sud, Est, Ouest) et 5 vitesses (très lent, lent, moyen, rapide, très rapide) pour créer 9 nouveaux attributs. Guo et Zhang [67] proposent une méthode dans laquelle l'attribut temporel est obtenu par simple différence d'images successives. Enfin, Zhou *et al.* [186] montrent qu'il est possible de détecter des objets en mouvement par rapport à

un fond dynamique en analysant le déphasage dans le domaine fréquentiel.

Dans la seconde catégorie, Chang *et al.* [31] construisent un volume spatio-temporel en tenant compte de l'image à l'instant t ainsi que des N images précédentes. Pour chaque pixel $\mathbf{x}$, le volume spatio-temporel est divisé en sous-volumes centrés sur $\mathbf{x}$ et la saillance du pixel $\mathbf{x}$ est calculée par une approche Bayesienne. Seo et Milanfar [147] calculent la saillance de chaque sous-volume en formulant le problème comme un problème de classification binaire. Les attributs utilisés pour la classification sont les sorties de filtres locaux orientés. La même idée est employée par Mahadevan et Vasconcelos [102] qui utilisent les textures dynamiques comme attributs.

3.1.2/ APPLICATIONS

La détection de régions saillantes est utile dans de nombreuses applications, dans des domaines aussi variés que la robotique ou l'imagerie médicale. Nous citons brièvement dans cette section quelques exemples d'applications.

— **Segmentation d'objets :** sans doute l'une des principales applications de la saillance visuelle. En effet, la carte de saillance indique les régions de l'image pouvant correspondre à des objets d'intérêt. En seuillant ces cartes, on obtient les objets de la scène. Par exemple, Achanta [3] propose une méthode dans laquelle l'image est dans un premier temps sur-segmentée en utilisant un algorithme de clustering, $k-$means, puis les segments dont la valeur moyenne de saillance est supérieur à un seuil sont retenus. D'autres approches de segmentation sont proposées dans [34, 5, 121, 127].

— **Re-dimensionnement d'images :** pour afficher les images sur des supports différents (smartphones, téléviseurs, tablettes, etc.), il faut pouvoir en modifier la taille sans altérer le contenu, i.e. en préservant les éléments de la scène les plus importants. Puisque la carte de saillance indique les régions visuellement saillantes de la scène, elle a été utilisée par de nombreux auteurs pour re-dimmensionner de manière *intelligente* les images [4, 97]. La principale idée de ces méthodes est basée sur l'algorithme *seam carving* de Avidan et Shamir [16], qui consiste à modifier (retirer ou ajouter des lignes et des colonnes) les régions de l'image qui contiennent le moins d'information.

— **Suivi d'objets :** les méthodes de suivi d'objets sont basées sur une représentation de l'apparence des objets en utilisant différents attributs tels que la couleur ou la texture. Dans [148, 101], les auteurs montrent que la prise en compte de la saillance dans le modèle d'apparence offre une plus grande robustesse face aux changement d'illumination et les occultations.

— **Localisation de robots :** pour se localiser, un robot mobile a besoin de détecter des points de repère dans son environnement. Dans [30, 52], les auteurs utilisent un système d'attention visuelle pour détecter les points de repère stables et faciles à re-détecter. Ceux-ci sont suivis dans plusieurs images d'une séquence, et leur position spatiale est estimée.

— **Détection de lésions rétiniennes :** la carte de saillance peut être employée

pour détecter les régions pouvant correspondre à des lésions dans une image rétinienne. C'est l'approche utilisée par Ujjwal *et al.* [165] qui segmentent la carte de saillance pour extraire les lésions potentielles. Chaque région potentielle est caractérisée par la moyenne et l'écart type de la saillance ainsi que la texture. Enfin, une étape de classification permet de détecter les lésions.

— **Evaluation de la qualité des images :** l'attention visuelle a été utilisée par de nombreux auteurs pour une évaluation « objective » de la qualité des images et des vidéos. Par exemple, les auteurs dans [118] définissent des métriques de comparaison qui tiennent compte du fait qu'un artéfact apparaissant dans une zone de forte saillance est plus gênant qu'un artéfact dans une zone de faible saillance.

3.1.3/ CONTRIBUTIONS

Nous apportons les contributions suivantes à la détection de la saillance visuelle dans les scènes dynamiques :

1. **Evaluation des méthodes de fusion :** La plupart des méthodes de la littérature sont basées sur la fusion d'une carte statique et d'une carte temporelle obtenues séparemment. Dans la section 3.2, nous réalisons la première évaluation (à notre caonnaissance) de différentes méthodes de fusion pour l'obtention de cartes de saillance spatio-temprelle.
2. **Saillance basée sur un opérateur de texture dynamique :** Nous proposons dans la section 3.3.1 une méthode de détection de saillance basée sur la fusion d'une carte statique (obtenue en utilisant la couleur) et d'une carte dynamique (obtenue en utilisant un descripteur de texture dynamique). Notre approche combine donc la couleur et la texture et donne des résultats satisfaisants par rapports aux méthodes de l'état de l'art.
3. **Saillance basée ACP multidimensionnelle :** Pour tenir compte des fortes corrélations spatio-temporelles qui existent entre les images successives d'une séquence vidéo, nous proposons dans la section 3.3.2 une méthode directe basée sur une ACP multidimensionnelle. Cette approche assez simple, ne necéssite pas de fusion.

3.2/ EVALUATION DES MÉTHODES DE FUSION DE CARTES STATIQUE ET TEMPORELLE

Comme indiqué dans la section 3.1.1, la plupart des méthodes de la littérature pour la détection de la saillance dans les scènes dynamique sont basées sur la fusion d'une carte statique et d'une carte temporelle obtenues séparemment. La figure 3.3 illustre le principe de la fusion pour l'estimation d'une carte de saillance spatio-temporelle. L'étape de fusion est importante car elle permet de pondérer l'importance de chacune des composantes statique et temporelle. Dans cette section, nous évaluons différentes méthodes de fusion proposées dans la littéraure.

Pour cette évaluation, nous utilisons la méthode de Goferman *et al.* [66], basée sur l'information de contexte pour calculer la carte statique, car cette méthode a montré de très

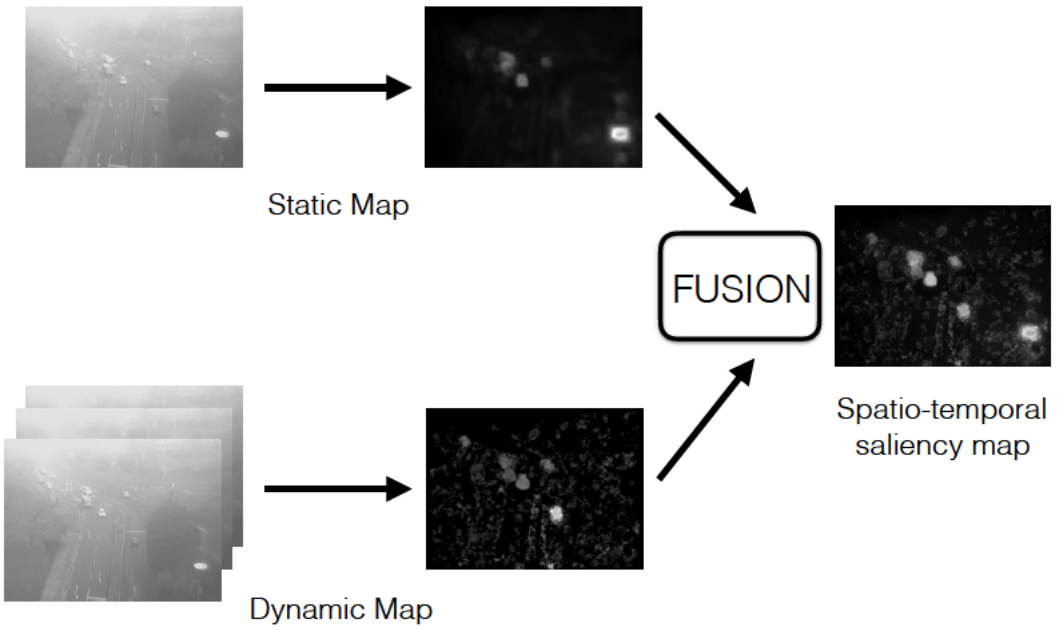


FIGURE 3.3 – Principe de la fusion pour l'obtention d'une carte de saillance spatio-temporelle.

bon résultats dans une récente évaluation [27]. Pour le calcul de la carte dynamique, nous utilisons une méthode basée sur l'estimation du flot optique [44]. Dans la suite de ce chapitre, on notera M_S la carte statique et M_D la carte dynamique.

Les différentes méthodes de fusion sont les suivantes :

— **Mean fusion (Mean) [94] :** Une simple moyenne des deux cartes :

$$M_F = (M_S + M_D)/2. \tag{3.1}$$

— **Max fusion (Max) [105] :** Une stratégie du type *winner takes all* (WTA), dans laquelle on retient la valeur maximale de la saillance dans les deux cartes :

$$M_F = \max(M_S, M_D). \tag{3.2}$$

— **Multiplication fusion (AND) [105] :** Une mulitiplication pixel par pixel correspondant à un *ET* logique :

$$M_F = M_S \times M_D. \tag{3.3}$$

— **Maximum skewness fusion (MSF) [105] :** Cette stratégie de fusion tient compte des caractéristiques de chacune des deux cartes. En particulier, la carte statique est modulée par sa valeur maximale α, tandis que la carte dynamique est modulée par son coefficient d'asymétrie (qui est le moment centré reduit d'ordre trois) β.

Enfin, le produit de ces deux termes γ renforcent les régions qui sont saillantes dans les deux cartes simultanément :

$$M_F = \alpha M_S + \beta M_D + \gamma(M_S \times M_D), \tag{3.4}$$

où $\alpha = \max(M_S)$, $\beta =$ asymétrie(M_D) et $\gamma = \alpha\beta$.

— **Binary thresholded fusion (BTF) [97] :** Un masque binaire M_B est généré à partir de la carte statique, en utilisant la valeur moyenne de M_S comme seuil. Ce masque est employé pour exclure les régions inconsistantes dans le domaine spatio-temporel, et pour accroître la robustesse de la carte finale lorsque les paramètres du mouvement ne sont pas correctement estimés :

$$M_F = \max(M_S, M_D \cap M_B). \tag{3.5}$$

— **Motion priority fusion (MPF) [120] :** Cette technique est basée sur l'hypothèse que le mouvement d'un objet mobile attire plus l'attention d'un observateur même lorsque le fond fixe est plus attrayant [120]. La perception des objets mobiles croît de manière non-linéaire avec le contraste du mouvement :

$$M_F = (1 - \alpha)M_S + \alpha M_D, \tag{3.6}$$

avec $\alpha = \lambda e^{1-\lambda}$ et $\lambda = \max(M_D) - \text{mean}(M_D)$.

— **Dynamic weight fusion (DWF) [178] :** Les poids des cartes statique et dynamique sont donnés par le rapport entre leurs moyennes pour chaque image de la séquence :

$$M_F = \alpha M_D + (1 - \alpha)M_S, \tag{3.7}$$

où $\alpha = \frac{\overline{M_D}}{\overline{M_D} + \overline{M_S}}$.

— **Information theory fusion (IFT) [68] :** Cette technique est basée sur la théorie de l'information et est formulée comme suit :

$$M_F = \alpha_S I(M_S) M_S + \alpha_D I(M_D) M_D, \tag{3.8}$$

où les poids α_S et α_D sont donnés par $\alpha_i = \max(M_i) I(M_i)$, et $I(M_i)$ est l'importance de la carte de saillance M_i.
Pour obtenir l'importance d'une carte de saillance M, on calcule la probabilité $p(M)$ en prenant le rapport entre le nombre de pixels dont la valeur est supérieur à un seuil τ, et le nombre total de pixels :

$$p(M) = \frac{\#(M(i,j)) > \tau}{\#(M(i,j))}.$$

L'importance de la carte est donnée par $I(M) = -\log(p(M))$.

— **Scale invariant fusion (SIF) [81] :** Cette approche multi-échelle combine les cartes à différentes échelles, et les différentes cartes ainsi obtenues sont combinées de manière linéaire :

$$M_F = \sum_{l=1}^{3} w_l M_F^l, \tag{3.9}$$

(a) Image extraites des séquences

(b) Segmentation manuelle

FIGURE 3.4 – Exemples d'images de la base de données SVCL.

Séquence	Mean	Max	AND	MSF	BTF	DWF	MPF	ITF	SIF
Birds	0.9713	0.9794	0.9023	0.9563	0.9852	0.9669	0.7639	0.9097	0.9245
Boats	0.9891	0.9745	0.9867	0.9881	0.9695	0.9827	0.9808	0.9889	0.9829
Cyclists	0.9628	0.9497	0.8862	0.9418	0.9533	0.9602	0.8394	0.9248	0.9498
Chopper	0.9784	0.9847	0.6891	0.6956	0.9852	0.9850	0.6791	0.9628	0.9711
Freeway	0.7128	0.6633	0.7023	0.7614	0.5087	0.5456	0.7581	0.6218	0.7452
Peds	0.9608	0.9435	0.8984	0.9380	0.9441	0.9512	0.8852	0.9400	0.9558
Jump	0.9395	0.9314	0.8949	0.9212	0.9459	0.9479	0.8535	0.8804	0.9197
Ocean	0.8273	0.7465	0.8108	0.8126	0.7535	0.7810	0.8032	0.8063	0.8412
Surfers	0.9453	0.9782	0.7993	0.9208	0.9844	0.9545	0.6251	0.9334	0.8757
Skiing	0.9678	0.9784	0.5195	0.6491	0.9807	0.9796	0.4905	0.9394	0.9365
Landing	0.9701	0.9524	0.9718	0.9703	0.9521	0.9579	0.9047	0.9353	0.9720
Traffic	0.9645	0.9566	0.8860	0.9540	0.8736	0.9615	0.9477	0.9640	0.9593
AUC moy	**0.9325**	0.9199	0.8289	0.8758	0.9030	0.9145	**0.7943**	0.9006	0.9200

TABLE 3.1 – Evaluation des méthodes de fusion : Mean (Mean fusion), Max (Max fusion), AND (Multiplication fusion), MSF (Maximum skewness fusion), BTF (Binary thresholded fusion), DWF (Dynamic weight fusion), MPF (Motion priority fusion), ITF (Information theory fusion), SIF (Scale invariant fusion).

où M_F^l est le résultat de la fusion à l'échelle l, et les coefficients de la combinaison linéaires sont donnés par $w_1 = 0.1$, $w_2 = 0.3$ et $w_3 = 0.6$ pour 3 échelles.

Nous évaluons ces différentes méthodes de fusion avec 12 séquences extraites de la base publique de vidéos de l'université de Californie à San Diego, SVCL [102], qui contient des séquences d'images capturées dans différentes conditions telles que la pluie, la neige, le brouillard, avec des voitures et des piétons, etc. Pour chaque séquence, une segmentation manuelle des objets d'intérêt, les objets saillants, est disponible pour un certain nombre d'images. Il est donc possible d'évaluer les différentes méthodes en segmentant les cartes de saillances obtenues et en comparant le résultat de la segmentation avec la vérité terrain. Nous utilisons ici comme mesure de performance, l'aire sous la courbe ROC, AUC (Area Under ROC Curve). La figure 3.4 montre quelques exemples d'images de la base de données et les segmentations manuelles correspondantes.

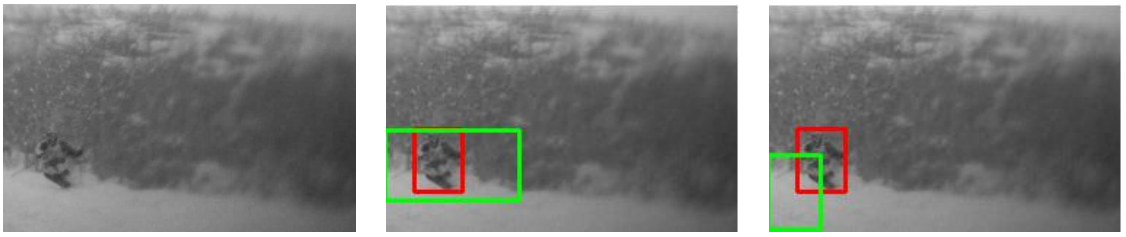


FIGURE 3.5 – Exemples de détection d'objet saillants avec la séquence *Skiing*. De gauche à droite : Image originale ; détection avec les méthodes de fusion *BTF* et avec *MPF*. Le rectangle rouge indique la vérité terrain, et le rectangle vert le résultat de la détection.

Les résultats obtenus par différentes méthodes de fusion sont rassemblés dans le tableau 3.1. On constate que les meilleurs résultats sont obtenus par les méthodes *Mean* [94], *SIF* [81], *Max* [105] et *DWF* [178] par ordre décroissant. En particulier, la méthode *Mean* obtient une AUC moyenne de 0.9325 pour l'ensemble des 12 séquences. Ces méthodes de fusion combinent intelligemment les deux cartes (statique et temporelle) : elles accordent plus d'importance à la valeur de la carte statique si elle est plus élevée, et vice versa. Par contre, les méthodes *MP* [120] and the *AND* [105] obtiennent les performances les plus faibles. La méthode *MP* obtient une AUC moyenne de 0.7943 pour les 12 séquences, i.e. 17% de moins que la méthode *Mean*. Cela peut s'expliquer par le fait que cette méthode de fusion accorde plus d'importance à l'information de mouvement (*motion priority*). Par conséquent, lorsque le contraste du mouvement, i.e. la différence du mouvement relatif, n'est pas correctement estimé, la carte spatio-temporelle obtenue n'est pas précise. Cela peut-être observé avec la séquence *Skiing* par exemple, pour laquelle la méthode *MP* obtient une AUC de 0.4905, alors que la méthode *BTF* obtient une valeur de 0.9807. La figure 3.5 montre les résultats de segmentation obtenus avec ces deux méthodes pour une image de la séquence *Skiing*. Comme on peut le voir, pour cette séquence avec un faible contraste de l'information de mouvement, la méthode *MP* produit un résultat incorrect car la carte dynamique est mal estimée.

Lorsqu'on analyse les résultats pour chacune des séquences, on constate que les meilleurs résultats sont obtenus avec la séquence *Boats*, tandis que les résultats les moins bons sont obtenus avec la séquence *Freeway*. La première est une séquence avec un bon contraste de couleur et de mouvement, donc les deux cartes statique et temporelle sont correctement estimées et toutes les méthodes de fusion donnent de bons résultats. La seconde a un contraste de couleur limité, et les méthodes *BTF* et *DWF* qui donnent une grande importance à la carte statique obtiennent des résultats moyens.

En résumé, ces résultats montrent que les méthodes de fusion qui s'appuient sur les caractéristiques de la scène étudiée (*Mean*, *Max*, *SIF* et *DWF*) fournissent des résultas plus stables que celles basées sur de forts a priori telle que la méthode *MPF*, dont les résultats dépendent de la validité de ces a priori pour la scène étudiée.

3.3/ PROPOSITION DE MÉTHODES DE SAILLANCE VISUELLE DANS LES SCÈNES DYNAMIQUES

Dans cette section, nous proposons deux approches pour le calcul de cartes de saillance dans les scènes dynamiques. La première méthode est basée sur la fusion de la texture

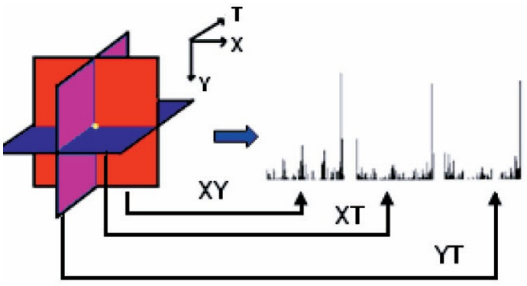


FIGURE 3.6 – Calcul du descripteur LBP-TOP (Image reproduite d'après [123]).

et de la couleur, section 3.3.1, et la seconde méthode est une approche directe exploitant la nature spatio-temporelle de la vidéo, section 3.3.2.

3.3.1/ COMBINAISON DE LA TEXTURE ET DE LA COULEUR

Dans la section 3.2, nous avons comparé différentes méthodes de fusion en utilisant une méthode basée sur l'estimation du flot optique pour le calcul de la carte temporelle. Toutefois, l'estimation du flot optique échoue dans des scènes avec un fond complexe comprenant des vagues d'eau, de la neige ou des arbres. Ici, nous proposons de représenter ces fonds complexes par une texture dynamique en utilisant l'opérateur LBP (Local Binary Patterns) étendu au domaine temporel.

3.3.1.1/ SAILLANCE DYNAMIQUE AVEC L'OPÉRATEUR LBP-TOP

L'opérateur LBP-TOP (pour Local Binary Patterns in Three Orthogonal Planes) est une extension de l'opérateur LBP au domaine temporel [185]. Cette extension est basée sur le calcul de la co-occurrence des descripteurs LBP dans trois plans orthogonaux tels que les plans XY, XT et YT. Les plans XT et YT donnent une information temporelle alors que le plan XY fournit une information spatiale. Un descripteur LBP est extrait dans chaque plan et le descripteur final est une concaténation des trois descripteurs, comme illlustré par la figure 3.6.

Pour chaque volume spatio-temporel $X \times Y \times T$, voir figure 3.7, nous calculons un histogramme de texture dynamique de la manière suivante :

$$H_{i,j} = \sum_{x,y,t} I\{f_j(x,y,t) = i\}, \quad i = 0,..,n_j - 1; \quad j = 0,1,2, \tag{3.10}$$

où n_j est le nombre de labels différents produit par l'opérateur LBP dans le j–ème plan ($j = 0$ pour XY, $j = 1$ pour XT et $j = 2$ pour YT). $f_j(x,y,t)$ est le code LBP du pixel central (x,y,t) dans le j–ème plan, et $I\{A\}$ est la fonction indicatrice de A : $I\{A\} = \begin{cases} 1 & \text{si A est vrai} \\ 0 & \text{si A est faux} \end{cases}$.

Une fois obtenus les histogrammes pour chaque volume, nous calculons la saillance du pixel central $\mathbf{x} = (x_c, y_c)$ en utilisant une approche *center-surround* (voir figure 3.2(a)). Plus précisément, si $\mathbf{h}_c$ et $\mathbf{h}_s$ désignent respectivement les histogrammes de la région

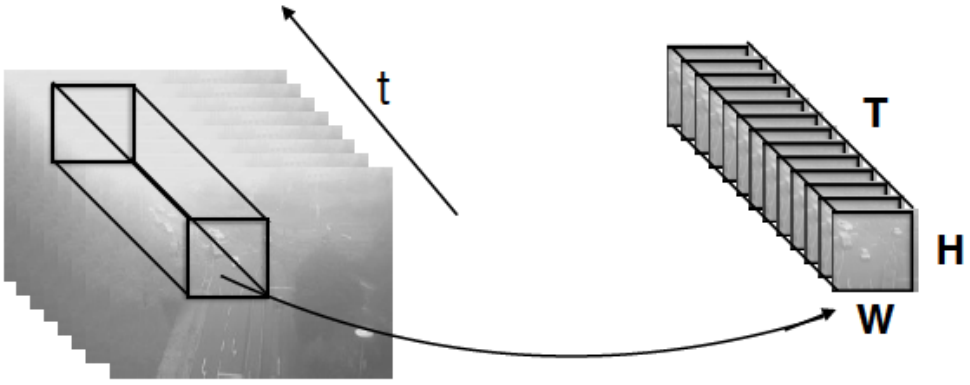


FIGURE 3.7 – Extraction de volumes spatio-temporels.

centrale et de la région avoisinante, alors la saillance du pixel central est donnée par :

$$S(\mathbf{x}) = \chi(\mathbf{h}_c, \mathbf{h}_s) = \sum_{i=1}^{B} \frac{(\mathbf{h}_c(i) - \mathbf{h}_s(i))^2}{(\mathbf{h}_s(i) + \mathbf{h}_c)/2}, \tag{3.11}$$

où B le nombre de classes de l'histogramme et χ est la distance *khi-2* entre les histogrammes.

Nous avons constaté qu'il est plus intéressant de calculer une carte de saillance pour chacun des plans XY, XT et YT, puis de les combiner en une carte spatio-temporelle. Cette fusion est effectuée en deux étapes :

1. **Carte dynamique :** Les deux cartes contenant une information temporelle, S_{XT} et S_{YT}, sont combinées en une carte dynamique M_D en utilisant la méthode de fusion *DWF* :
$$M_D = \alpha_D S_{YT} + (1 - \alpha_D) S_{XT}, \tag{3.12}$$
où $\alpha_D = \frac{mean(S_{YT})}{mean(S_{XT}) + mean(S_{YT})}$.
2. **Carte spatio-temporelle :** La carte dynamique ainsi obtenue est fusionnée avec la carte statique S_{XY} en utilisant la même méthode de fusion.

3.3.1.2/ FUSION DE LA TEXTURE ET DE LA COULEUR

Dans la méthode précédente, la carte statique est calculée en utilisant l'opérateur LBP dans le plan XY. Seule la texture est prise en compte, ce qui conduit à une carte statique incorrecte dans certains cas. Pour améliorer la performance de notre méthode, nous avons décidé de remplacer la carte statique par une carte de saillance basée sur la couleur. Nous avons opté pour la méthode de Goferman *et al.* [66], basée sur une information de contexte, que nous décrivons brièvement (c'est la même méthode employée dans la section 3.2 pour la comparaison des approches de fusion).

Dans un premier temps, une mesure de saillance locale est calculée pour chaque pixel de l'image. Celle-ci est basée sur la mesure de dissimilarité définie par :

$$d(p_i, q_k) = \frac{d_{color}(p_i, q_k)}{1 + c.d_{position}(p_i, q_k)}, \tag{3.13}$$

où $d_{color}(p_i, q_k)$ est la distance euclidiene, calculée dans l'espace CIELAB, entre une fenêtre p_i centré sur le pixel et les fenêtres $\{q_k\}_{k=1}^{K}$ qui sont les K fenêtres les plus similaires à p_i, $d_{position}(p_i, q_k)$ est la distance euclidienne entre les positions des fenêtres, et c une constante fixé à 3 dans nos expériences.

Une mesure de la saillance de chaque pixel, à une échelle r, est donnée par :

$$S_i^r = 1 - e^{\frac{1}{K}\sum_{k=1}^{K} d(p_i^r, q_k^r)}. \tag{3.14}$$

Cette mesure est calculée à différentes échelles, et la saillance du pixel est la valeur moyenne sur l'ensemble des échelles. Enfin, une information de contexte est incluse en utilisant le fait que les régions proches d'un centre d'attention doivent être explorées plus fréquemment que les régions éloignés de tout centre d'attention. Un pixel est considéré comme un centre d'attention, à une échelle r, si sa mesure de saillance est supérieure à un seuil $(S_i^r > 0.8)$. Finalement, la carte de saillance est définie par :

$$\hat{S}_i = \frac{1}{R}\sum_r S_i^r(1 - d_{foci}^r(i)), \tag{3.15}$$

où R est le nombre d'échelles, et $d_{foci}(i)$ est la distance euclidienne entre le pixel i et le centre d'attention le plus proche à l'échelle r.

Enfin, la carte de saillance statique M_S, obtenue par la méthode de Goferman *et al.* [66], est fusionée avec la carte dynamique M_D, obetnue par fusion des cartes S_{XT} et S_{YT} obtenues avec l'opérateur LBP-TOP, de la manière suivante :

$$M_F = \alpha M_D + (1 - \alpha)M_S, \tag{3.16}$$

avec $\alpha = \frac{mean(M_D)}{mean(M_D)+mean(M_S)}$.

3.3.1.3/ RÉSULTATS

Nous évaluons les méthodes proposées LBP-COLOR (combinaison de la couleur et de la texture) et LBP-TOP (texture uniquement), en utilisant les séquences de la base de données SVCL [102], et en utilisant comme mesure de performance, l'aire sous la courbe ROC, AUC (Area Under ROC Curve). Nous comparons également ces deux méthodes avec trois autres méthodes de l'état de l'art : une méthode basée sur le calcul du flot optique [112], la méthode basée sur l'auto-similarité ou *self-resemblance* (SR) [147] et la méthode basée sur la divergence de phase ou *phased discrepancy* (PD) [186].

Le tableau 3.2 montre les résultats obtenus par les différentes méthodes. Comme on peut le constater, la méthode proposée combinant la couleur et la texture, LBP-COLOR, obtient la meilleure performance pour l'ensemble des séquences avec une AUC moyenne de 0.914. On note aussi que la méthode basée uniquement sur la texture, LBP-TOP, obtient des résultats décevants (AUC = 0.7453). Ce qui confirme l'importance de la couleur pour calculer la carte statique, puisqu'on obtient un gain de l'ordre de 22% en combinant la couleur et la texture. La seconde meilleure performance est obtenue par la méthode combinant le flot optique et la couleur, avec une AUC moyenne de 0.9079. Toutefois, notre approche LBP-COLOR est supérieure grâce à l'utilisation de l'opérateur LBP-TOP.

Comme dans la section 3.2, les meilleurs résultats sont obtenus avec la séquence *Boats* tandis que les résultats les moins bons sont obtenus avec la séquence *Freeway*. Pour

Séquence	LBP-COLOR	LBP-TOP	OF [112]	SR [147]	PD [186]	AUC moy.
Birds	0.9586	0.7680	0.9664	0.9379	0.8221	0.8906
Boats	0.9794	0.8358	0.9827	0.9227	0.9765	**0.9394**
Bottle	0.9953	0.9413	0.8787	0.9961	0.8285	0.9279
Cyclists	0.9317	0.6737	0.9602	0.8682	0.9551	0.8777
Chopper	0.9717	0.9427	0.9850	0.7447	0.6470	0.8582
Freeway	0.7775	0.8684	0.5456	0.7760	0.7318	**0.7398**
Peds	0.9552	0.7376	0.9512	0.8603	0.8548	0.8718
Ocean	0.9271	0.8513	0.7810	0.8016	0.8235	0.8369
Surfers	0.9674	0.7489	0.9545	0.9455	0.9352	0.9103
Skiing	0.8389	0.3787	0.9796	0.8872	0.9367	0.8042
Jump	0.8957	0.6960	0.9481	0.8321	0.6616	0.8067
Traffic	0.7693	0.6088	0.9615	0.5491	0.8720	0.7521
AUC moy.	**0.9140**	**0.7453**	**0.9079**	**0.8434**	**0.8371**	

TABLE 3.2 – Evaluation de différentes méthodes de détection de saillance spatio-temporelle. LBP-COLOR (notre méthode combinant la couleur et la texture), LBP-TOP (la texture uniquement), OF (basée sur le flot optique), SR (basée sur l'auto-similarité) et PD (basée sur la divergence de phase).

cette dernière séquence, la couleur n'est pas très distinctive et c'est donc la carte dynamique qui est importante dans le processus de fusion. On remarque que c'est la méthode basée sur la texture, LPB-TOP, qui obtient le meilleur résultat pour cette séquence avec une AUC moyenne de 0.860, alors que la méthode basée sur le flot optique donne une AUC moyenne de 0.545. Cet exemple illustre l'avantage d'utiliser le descripteur LBP pour représenter les textures dynamiques. Les figures 3.8 et 3.9 montrent courbes ROC pour ces deux séquences.

3.3.2/ UNE APPROCHE DIRECTE PAR ACP MULTIDIMENSIONNELLE

La méthode proposée dans la section 3.3.1 est basée sur la fusion d'une carte statique avec une carte temporelle, chacune calculée de manière indépendente. Cette méthode, comme toute les méthodes basées sur la fusion, ignore les très fortes corrélations spatio-temporelles qui existent entre les images successives d'une vidéo. Dans cette section, nous proposons une approche qui considère une séquence d'images comme un volume spatio-temporel tridimensionnel. Notre méthode est basée sur la méthode de Margolin *et al.* [106] qui utilise une analyse en composantes principales (ACP) pour calculer une carte de saillance dans des images fixes. Nous étendons cette idée au domaine temporel pour calculer une carte de saillance spatio-temporelle par une approche multidimensionnelle.

3.3.2.1/ SAILLANCE STATIQUE BASÉE ACP

Nous décrivons très brièvement la méthode basée ACP [106] qui sert de base à notre approche. L'idée principale de la méthode est la comparaison de chaque fenêtre locale extraite de l'image avec toutes les autres fenêtres et pas uniquement avec ses k fenêtres

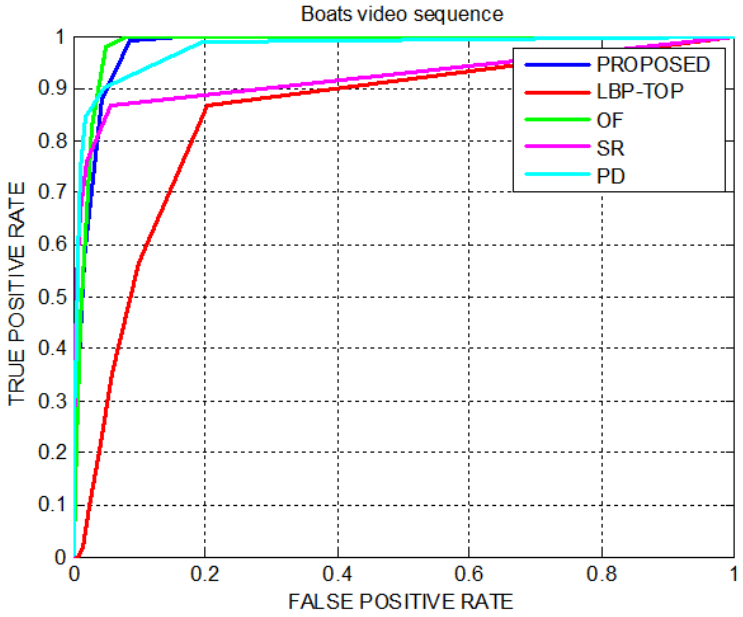


FIGURE 3.8 – Courbes ROC pour la séquence *Boats*.

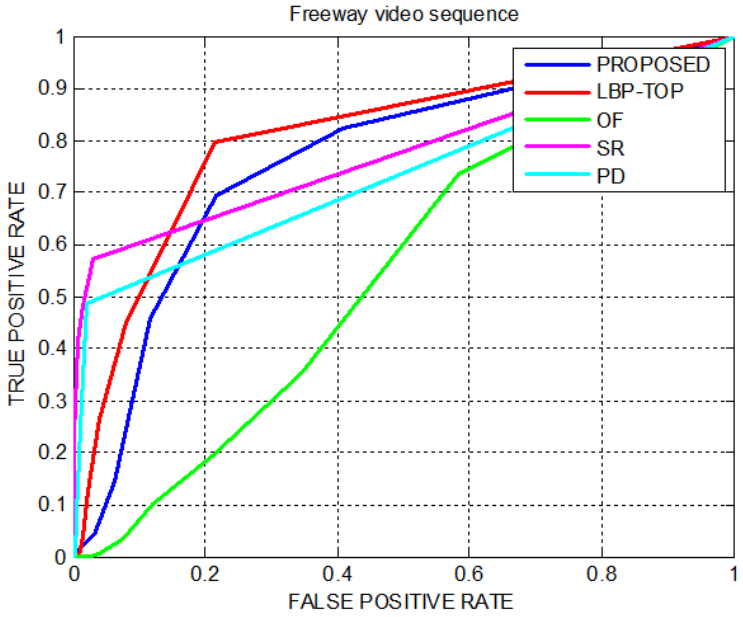


FIGURE 3.9 – Courbes ROC pour la séquence *Freeway*.

les plus similaires comme cela est fait dans de très nombreuses approches. Néanmoins, pour réduire la complexité élevée d'une comparaison avec toutes les fenêtres de l'image, une ACP est utilisée pour représenter les fenêtres et calculer leur saillance.

Plus précisément, étant donnée une image I, toutes les fenêtres $p_{\mathbf{x}}$ de taille $W \times H$ centrées sur les pixels $\mathbf{x} = (x, y)$ sont extraites. Chaque fenêtre est représentée par un vecteur de dimension $d = W \times H$, et l'ensemble de ces vecteurs forme une matrice $\mathbf{X}_I = [p_1, \ldots, p_M]$, où M est le nombre de fenêtres. Dans l'étape suivante de l'al-

gorithme, on applique une ACP à la matrice $\mathbf{X}_I$ et on représente chaque fenêtre par les coordonnées de sa projection sur les axes principaux $\widetilde{p}^k$, $k = 1, \ldots, K$: $p_{\mathbf{x}} = \sum_{k=1}^{K} \alpha_{\mathbf{x}}^k \widetilde{p}^k$. Finalement, la saillance de la fenêtre $p_{\mathbf{x}}$ est définie par :

$$P(p_{\mathbf{x}}) = \sum_{k=1}^{K} |\alpha_{\mathbf{x}}^k|. \tag{3.17}$$

Dans [106], cette mesure de saillance est complétée par une information couleur et est calculée à trois échelles différentes.

3.3.2.2/ SAILLANCE SPATIO-TEMPORELLE BASÉE ACP

Pour étendre cette idée au domaine temporel, plusieurs stratégies sont possibles, et nous proposons les trois suivantes :

ACP selon l'axe temporel L'idée la plus simple est de calculer une carte statique et une carte temporelle en utilisant la méthode basée ACP [106] , puis de fusionner les deux cartes. La carte statique M_S est calculée comme décrite à la section 3.3.2.1 en considérant des fenêtre spatiales centrées en chaque pixel de l'image. Pour la carte temporelle, nous considérons un ensemble de N images, et considérons pour chaque pixel l'ensemble de ses intensités dans les N images : $p(\mathbf{x}) = \{I_1(x, y), \ldots, I_N(x, y)\}$ comme illustré par la figure 3.10(a). L'ensemble des vecteurs $p(\mathbf{x}) \in \mathbb{R}^N$ forment une matrice X, et la carte temporelle M_D est calculée comme dans la section 3.3.2.1 en appliquant une ACP à X. Finalement, la carte spatio-temporelle est obtenue par fusion des cartes M_S et M_D : $M_F = \alpha M_D + (1 - \alpha) M_S$, avec $\alpha = \overline{M_D}/(\overline{M_D} + \overline{M_S})$.

ACP sur des sous-volumes 3D Une seconde approche consiste à extraire des petits volumes spatio-temporels 3D centrés sur chaque pixel, en considérant N images successives. Ces volumes sont ensuite transformés en vecteurs $\mathbf{x} \in \mathbb{R}^{NWH}$ comme illustré par la figure 3.10(b). Chaque vecteur contient à la fois les informations spatiales et temporelles, et la carte de saillance spatio-temporelle est obtenue par la méthode basée ACP comme décrite dans la section 3.3.2.1.

ACP multidimensionnelle Chaque volume spatio-temporel 3D est représenté sous forme de tenseur et non sous forme de vecteur, car la représentation vectorielle élimine la corrélation entre les axes spatiaux et temporels. A l'inverse, la représentation tensorielle préserve la structure 3D et la nature spatio-temporelle de la vidéo. Pour chaque pixel $\mathbf{x} = (x, y)$, on extrait un volume 3D qui est représenté par un tenseur d'ordre 3, $\mathcal{X} \in \mathbb{R}^{W \times H \times N}$, comme le montre la figure 3.10(c).

Nous utilisons une ACP multidimensionnelle pour représenter chaque tenseur. L'idée principale de l'ACP multidimensionnelle, ou MPCA (multilinear PCA) [95], est de transformer le tenseur $\mathcal{X}$ en un tenseur de taille réduite $\mathcal{Y} \in \mathbb{R}^{W' \times H' \times N'}$ en utilisant 3 matrices de projection :

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \times_3 \mathbf{U}^{(3)^T}, \tag{3.18}$$

avec $\mathbf{U}^{(1)} \in \mathbb{R}^{W \times W'}$ la matrice de projection selon le premier mode du tenseur, et de même pour $\mathbf{U}^{(2)}$ et $\mathbf{U}^{(3)}$. $\times_n$ est l'opérateur de projection selon le n-ième mode.

L'ACP multidimensionnelle utilise donc 3 matrices de projection, une pour chaque mode du tenseur. De plus, comme $W' < W$, $H' < H$ et $N' < N$, les dimensions du tenseur sont réduites de $W \times H \times N$ à $W' \times H' \times N'$. Les 3 matrices de projection sont obtenues par un processus itératif basé sur des projections alternées, chaque itération faisant appel à 3 décompositions spectrales. Pour plus de détails sur le calcul des matrices de projection, le lecteur est renvoyé à [95, 96].

Finallement, une mesure de saillance pour chaque pixel est calculée de manière similaire à la section 3.3.2.1 en sommant les coordonnées selon les 3 modes :

$$P(\mathbf{x}) = \sum_x \sum_y \sum_t |\mathcal{Y}(x, y, t)|. \tag{3.19}$$

3.3.2.3/ RÉSULTATS

Cette approche directe est évaluée en utilisant les séquences de la base de données SVCL [102], et en utilisant comme mesure de performance, l'aire sous la courbe ROC, AUC (Area Under ROC Curve).

Paramètres Les deux paramètres les plus importants à fixer dans cette méthode sont, d'une part, la taille de la fenêtre spatiale $W \times H$ et, d'autre part, la taille de la fenêtre temporelle N. Dans nos expériences, nous employons des fenêtres carrées de taille $W \times W$ est faisons varier W dans l'ensemble $\{5, 7, 9, 11, 13, 15, 17, 19, 21\}$.

La tableau 3.3 montre l'évolution de la valeur moyenne de l'AUC calculée avec toutes les séquences, en faisant varier W. Comme on peut le voir, la valeur de l'AUC augmente avec W jusqu'à se stabiliser autour de $W = 15$. Nous utilisons donc la valeur $W = 15$ dans la suite des expériences.

W	5	7	9	11	13	15	17	19	21
AUC moyenne	0.73	0.76	0.77	0.77	0.78	0.79	0.79	0.79	0.79

TABLE 3.3 – Variation de l'AUC moyenne avec la taille de la fenêtre spatiale.

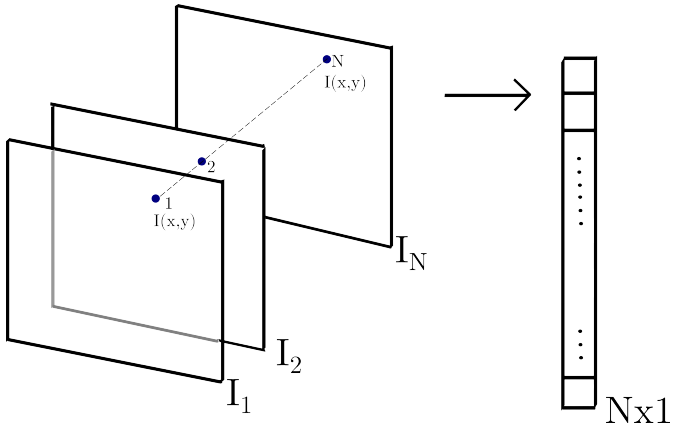
En fixant $W = 15$, nous faisons varier $N \in \{5, 7, 9, 11, 13\}$. Les résultats présentés dans le tableau 3.4 montrent que la valeur $N = 7$ est un bon compromis entre la précision des résultats et le temps de calcul. En effet, des valeurs plus importantes de N n'augmentent pas la valeur moyenne de l'AUC. Nous utiliserons donc la valeur $N = 7$ dans la suite des expériences.

N	5	7	9	11	13
AUC moyenne	0.81	0.84	0.84	0.83	0.83

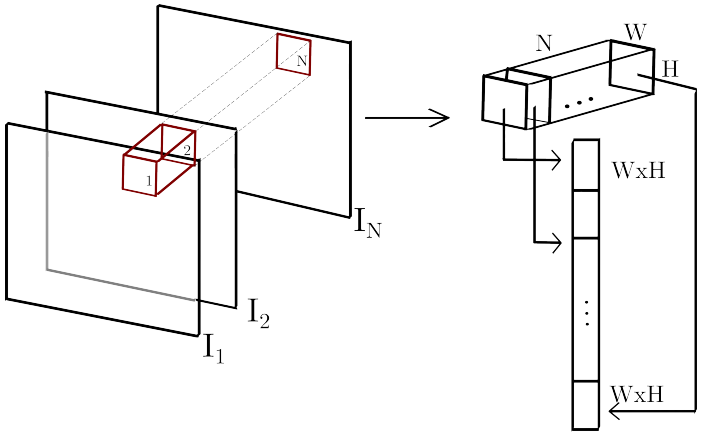
TABLE 3.4 – Variation de l'AUC moyenne avec la taille de la fenêtre temporelle.

Représentations Nous comparons les 3 méthodes de représentations décrites dans la section 3.3.2.2 :

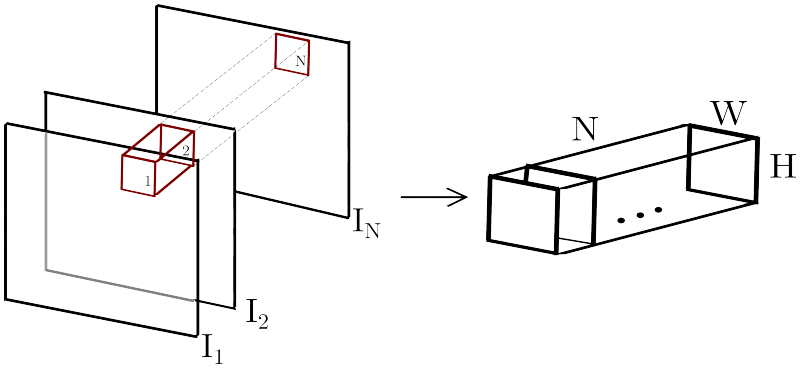
- **Fusion :** fusion des cartes statique et temporelle calculées en utilisant une ACP.
- **Vectorisation :** vectorisation des fenêtres spatio-temporelles et calcul de la carte de saillance en utilisant une ACP.



(a) ACP selon l'axe temporel



(b) Vectorisation du volume spatio-temporel



(c) Représentation tensorielle

FIGURE 3.10 – Différentes approches de représentation pour l'ACP appliquée à des séquences d'images.

— **MPCA :** représentation des fenêtres spatio-temporelles sous forme de tenseurs, et calcul de la carte de saillance en utilisant une ACP multidimensionnelle.

Les résultats présentés dans le tableau 3.5 montrent clairement que l'extraction de fenêtres spatio-temporelles 3D donne de biens meilleurs résultats. En particulier, les

Méthode	PCA [106]	Fusion	Vectorisation	MPCA
AUC moyenne	0.7930	0.7440	0.8436	0.9041

TABLE 3.5 – Comparaison des différentes approches de représentation.

Séquence	MPCA	LBP-COLOR [113]	OF [112]	SR [146]	PD [186]
Birds	0.9757	0.9586	0.9664	0.9379	0.8221
Boats	0.9059	0.9794	0.9827	0.9227	0.9765
Bottle	0.9936	0.9953	0.8787	0.9961	0.8285
Cyclists	0.9790	0.9317	0.9602	0.8682	0.9551
Chopper	0.9843	0.9717	0.9850	0.7447	0.6470
Freeway	0.8042	0.7775	0.5456	0.7760	0.7318
Peds	0.9405	0.9552	0.9512	0.8603	0.8548
Ocean	0.9037	0.9271	0.7810	0.8016	0.8235
Surfers	0.8448	0.9674	0.9545	0.9455	0.9352
Skiing	0.7857	0.8389	0.9796	0.8872	0.9367
Jump	0.9368	0.8957	0.9481	0.8321	0.6616
Traffic	0.7946	0.7693	0.9615	0.5491	0.8720
AUC moyenne	**0.9041**	**0.9140**	**0.9079**	**0.8434**	**0.8371**

TABLE 3.6 – Comparaison de différentes méthodes de détection de saillance spatio-temporelle.

approches *MPCA* et *Vectorisation* obtiennent une AUC moyenne de 0.9041 et 0.844, alors que l'approche *Fusion* donne une AUC moyenne de 0.7440, ce qui est moins élevé que l'application de la méthode de base à chaque image de la séquence de manière indépendente. Cela peut s'expliquer par le fait que l'extraction des intensités selon l'axe temporel, figure 3.10(a), est très locale car seule la valeur du pixel central est considérée. A l'inverse, les volumes spatio-temporels 3D incluent à la fois un contexte spatial et un contexte temporel, ce qui améliore les résultats. De plus, la représentation tensorielle, en préservant les corrélations spatio-temporelles, donne les meilleurs résultats.

Comparaison Enfin, nous comparons l'approche par ACP multidimensionnelle (*MPCA*) avec la méthode combinant la couleur et la texture (LBP-COLOR), la méthode basée sur le flot optique (OF) [112], la méthode basée sur l'auto-similarité ou *self-resemblance* (SR) [147] et la méthode basée sur la divergence de phase (PD) [186].

Les résultats rassemblés dans le tableau 3.6 montrent que l'approche directe par ACP multidimensionnelle donne des résultats très satisfaisants, avec une AUC moyenne de 0.9041. Elle obtient des résultats comparables à la méthode basée sur le flot optique (*OF*) et des résultats légèrement inférieurs à ceux obtenues par la méthode combinant la couleur et la texture (*LBP-COLOR*). Toutefois, il faut souligner que l'approche *MPCA* ici présentée n'utilise que les valeurs d'intensité des pixels alors que la méthode LBP-COLOR utilise des descripteurs LBP. On peut donc supposer qu'elle donnerait de meilleurs résultats si dans chaque fenêtre spatio-temporelle étaient extraits des descripteurs LBP-TOP, suivi de l'application d'une ACP multidimensionnelle.

3.4/ CONCLUSION ET DISCUSSION

Dans ce chapitre, nous avons abordé le problème de la détection de régions saillantes dans une séquence d'images. Ce problème a été moins abordé dans la littérature que celui de la détection de régions saillantes dans une image fixe.

Nous avons, dans un premier temps, évalué différentes méthodes de fusion de cartes de saillance. Cette étape de fusion est importante car, une grande majorité des méthodes proposées dans la littérature est basée sur la fusion d'une carte statique et d'une carte temporelle obtenues séparemment. Notre évaluation, la première à notre connaissance, montre que les méthodes de fusion qui s'appuient sur les caractéristiques de la scène étudiée (*Mean*, *Max*, *SIF* et *DWF*) fournissent des résultas plus stables que celles basées sur des a priori forts telle que la méthode *MPF*, dont les résultats dépendent de la validité de ces a priori pour chaque scène étudiée.

Nous avons ensuite proposé une méthode de détection de saillance basée sur une combinaison de la couleur et la texture. Les résultats obtenus montrent que l'utilisation de texture dynamique (avec des LBP-TOP) permet d'estimer correctement une carte de saillance dynamique lorsque le mouvement relatif de l'objet par rapport au fond de la scène est assez faible ou peu perceptible. L'approche proposée donne des résultats supérieurs par rapport aux méthodes de l'état de l'art.

Enfin, pour tenir compte des fortes corrélations spatio-temporelles qui existent entre les images successives d'une séquence vidéo, nous avons proposé une méthode directe basée sur une ACP multidimensionnelle. Cette approche assez simple, ne necéssite pas d'étape de fusion et donne des résultats satisfaisants. Toutefois, elle est assez coûteuse en temps de calcul. La méthode la plus intéressante reste donc celle combinant une carte statique (obtenue avec des attributs couleur) avec une carte dynamique (en décrivant le mouvement avec une texture dynamique).

Ce travail, réalisé dans le cadre de la thèse de Satya Muddamsetty (2011-2014) [112, 113], se poursuivi actuellement avec l'analyse de scènes en utilisant des caméras de profondeur de type Kinect. Dans ce travail, nous étudions la meilleure manière de prendre en compte l'information de profondeur disponible avec ce type de capteur pour obtenir des cartes de saillance spatio-temporelle.

4

DÉTECTION ET SUIVI D'OBJETS AVEC DES CAMÉRAS ATYPIQUES

La détection et le suivi d'objets mobiles dans une scène dynamique sont des problèmes fondamentaux en vision par ordinateur. Dans ce chapitre, nous nous intéressons à l'utilisation de caméras dites atypiques, par opposition à une caméra perspective RGB classique, pour l'analyse de scènes. En particulier, nous abordons le problème du suivi d'objets mobiles avec une caméra catadioptrique qui offre un grand champ de vue, ainsi que la détection et la reconnaissance d'objets avec une caméra de profondeur (de type Kinect) qui permet de capturer la scène en 3D.

4.1/ INTRODUCTION

Par caméras atypiques, nous entendons des caméras qui se distinguent des caméras RGB classiques soit par la géométrie, soit par le spectre d'acquisition, soit par le type de données produites. Ces caméras ont été dévelopées et sont employées dans différentes applications parce qu'elles offrent :

- **un champ de vision plus important :** c'est le cas de caméras dites omnidirectionnelles qui peuvent capturer une scène dans son intégralité, vision panoramique jusqu'à 360^o. Il en existe plusieurs configurations parmi lesquelles les caméras catadioptriques qui consistent en l'association d'un miroir et d'une caméra perspective. Ces caméras sont très utiles en robotique pour la localisation et la navigation autonome. Les images des figures 4.1(a) et (b) montrent des exemples de caméras catadioptriques et une image acquise avec ce type de caméras. Pour un aperçu de la vision panoramique, nous invitons le lecteur à se référer à [20].
- **un spectre plus large :** c'est le cas des caméras thermiques (ou infrarouge) qui permettent de capter les différents rayonnements infrarouge émis par un corps, ou des caméras polarimétriques permetant de mesurer l'état de polarisation de la lumière provenant de la scène. Les caméras polarimétriques ont notamment été utilisées pour la reconstruction 3D de surfaces réfléchissantes [111] ou dans des applications médicales [119]. Les figures 4.1(c) et (d) montrent une caméra infrarouge et une image acquise avec cette caméra.
- **une information 3D :** c'est le cas des caméras de profondeur, dont la Kinect de Microsoft, qui permettent d'obtenir une information sur la profondeur de chaque point de la scène acquise. Les figures 4.1(e) et (f) montrent une Kinect et une image acquise avec cette dernière. Ce type de caméra est de plus en plus utilisé

(a) Caméras catadioptriques.

(b) Image obtenue avec une caméra catadioptrique

(c) Caméra infrarouge.

(d) Image obtenue avec une caméra infrarouge

(e) Caméra de profondeur.

(f) Image obtenue avec une caméra de profondeur

FIGURE 4.1 – Différents types de caméras atypiques

dans les applications de navigation, visualisation et inspection [69].

4.1.1/ CONTRIBUTIONS

1. **Suivi avec une caméra catadioptrique :** Nous proposons une méthode d'adaptation des méthodes existantes de suivi en tenant compte de la géométrie particulière des capteurs et des images catadioptriques. Cette adaptation, basée sur le modèle sphérique pour représenter les images, permet une meilleure prise en compte de la résolution non-unifiorme des images, et offre une plus grande robustesse aux variations d'illumination de la scène.
2. **Reconnaissance d'objet avec une caméra de profondeur :** Nous proposons une méthodologie pour réduire la taille de différents descripteurs de nuages de points

3D acquis avec une Kinect, pour réduire la complexité sans sacrifier la performance des algorithmes de reconnaissance. Nous proposons ensuite un descripteur combinant des propriétés géométriques et de texture extraites du nuage de point pour définir un descripteur global, dont la taille est réduite par ACP. Le nouveau descripteur est plus performant en terme de robustesse au bruit (bruit Gaussien) et de reconnaissance d'objets et de catégorie d'objets.

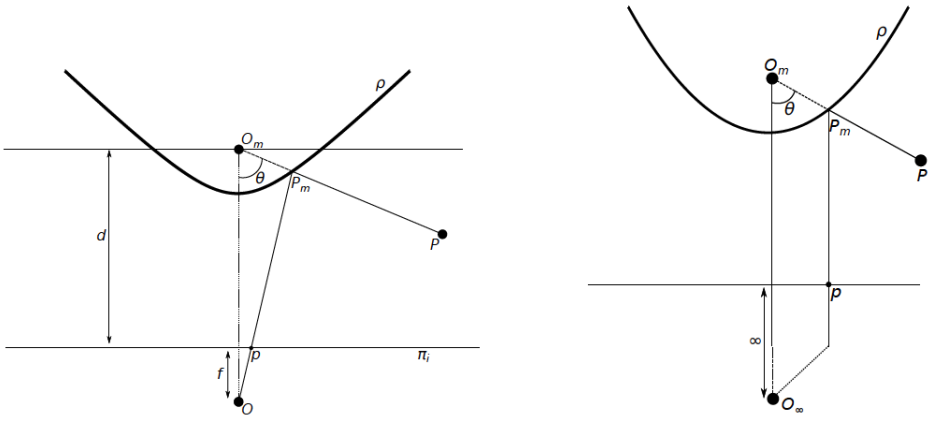
4.2/ SUIVI AVEC UNE CAMÉRA CATADIOPTRIQUE

Les caméras omnidirectionnelles permettent d'obtenir des images avec un grand angle de vue, mais elles ont souvent une résolution limitée et non-uniforme, et entraînent une forte distorsion géométrique de l'image. Les algorithmes classiques de traitement des images (filtrage, segmentation ,etc.) doivent donc être adaptés à ce type de caméras [39, 40]. Dans cette section, nous nous intéressons au suivi d'objets mobiles en utilisant une caméra catadioptrique. Ce travail a été réalisé dans le cadre de la thèse de François Rameau, dont l'objectif était la mise au point d'un système de vision hybride composé d'une caméra omnidirectionnelle, pour la vision panoramique, et d'une caméra active PTZ, pour la vision détaillée de zones d'intérêt.
Nous commençons par décire brièvement le fonctionnement d'une caméra catadioptrique, section 4.2.1, puis nous montrons comment adapter les méthodes classiques de suivi à ce type de capteurs, section 4.2.2 et 4.2.3.

4.2.1/ REPRÉSENTATION DES CAMÉRAS CATADIOPTRIQUES

Il existe différents dispositifs d'acquisition d'images omnidirectionnelles tels que les caméras rotatives (obtention d'images panoramiques à partir d'une caméra en mouvement), les caméras *fisheye* (utilisation d'une lentille spécifique) ou les systèmes polydioptriques (ensemble de caméras). Une caméra catadioptrique est constituée de l'association d'une caméra et d'un miroir permetant d'élargir le champ de vue. Afin d'obtenir un système de vision à point de vue unique, i.e. un sytème dans lequel les rayons lumineux convergent vers un centre de projection unique, Nayar et Baker montrent que seules 4 configurations utilisant différents types de miroir sont possibles [17]. Parmi celles-ci, la configuration utilisant un miroir hyperboloïde avec une caméra perspective (caméra *hypercatadioptrique*), et celle utilisant un miroir parabolique avec une caméra orthographique (caméra *paracatadioptrique*) sont les plus utilisées dans la pratique car elles offrent à la fois une centre de projection unique et une vision panoramique. La figure 4.2 montre les principes de formation d'images omnidirectionnelles avec ces deux types de caméras.

Des modèles de projection adaptés à chaque configuration ont été développés, cependant il est possible de représenter tous les types de caméras à point de vue unique à l'aide d'un modèle unifié de projection, appelé *modèle sphérique*, [57]. Celui-ci se traduit par une double projection via une sphère unité comme illustré par la figure 4.3. Dans un premier temps, un point de la scène 3D, $\mathbf{P} = [X, Y, Z]^T$, est projété sur la sphère unité en



(a) Caméra hypercatadioptrique. (b) Caméra paracatadioptrique

FIGURE 4.2 – Deux configuarations de caméras catadioptriques.

un point $\mathbf{P}_S$:

$$P_S = \frac{\mathbf{P}}{||\mathbf{P}||} = \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix}. \tag{4.1}$$

Ensuite, ce point $\mathbf{P}_S$ est projeté sur le plan image en un point $\mathbf{p}_i = [x, y]^T$ à partir d'un point O_C situé à une distance l du centre de la sphère :

$$x = \frac{X_S}{l + Z_S} \qquad \text{et} \qquad y = \frac{f_s Y_S}{l + Z_S}. \tag{4.2}$$

La projection globale se caractérise donc par l'équation suivante

$$\mathbf{p} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{f_s X}{l\sqrt{X^2+Y^2+Z^2}+Z} \\ \frac{f_s Y}{l\sqrt{X^2+Y^2+Z^2}+Z} \end{bmatrix}. \tag{4.3}$$

Le paramètre l définit le type de caméra représenté par le modèle sphérique. En effet, on montre que $l < 1$ correspond à une caméra hypercatadioptrique, tandis que $l = 1$ correspond à une caméra paracatadioptrique. De plus, il est aisé de vérifier que pour une caméra perspective conventionnelle, $l = 0$.

Le modèle sphérique permet donc de représenter toutes les caméras centrales, i.e. à point de vue unique.

4.2.2/ ADAPTATION DES MÉTHODES DE SUIVI

Le suivi d'objets mobiles dans une séquence d'images est nécessaire pour des applications telles que la vidéo surveillance, la robotique mobile, la réalité augmentée ou encore la compression vidéo [100]. La littérature comporte de nombreuses méthodes de suivi que nous pouvons, grossièrement, classer dans deux principaux groupes : les méthodes déterministes et les méthodes stochastiques. Les méthodes de suivi appartenant au premier groupe, par exemple l'algorithme KLT (Kanade-Lucas-Tomasi) [98] ou la méthode

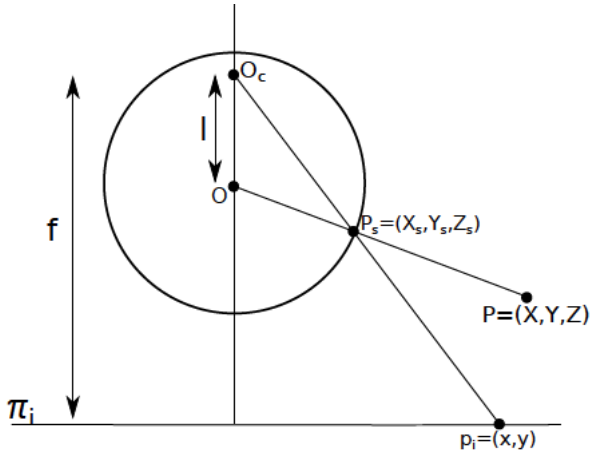
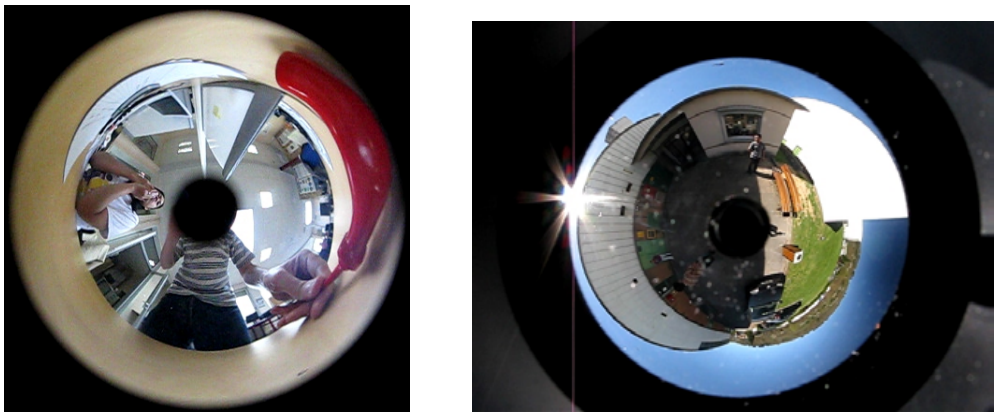


FIGURE 4.3 – Modèle sphérique de projection.



(a) Distorsion de l'image catadioptrique (b) Image catadioptrique avec éblouissement

FIGURE 4.4 – Difficultés engendrées par l'utilisation d'un miroir.

mean-shift [35], cherchent de manière itérative à maximiser une mesure de similarité locale entre un modèle de la cible et une région de l'image courante. Les méthodes stochastiques utilisent quant à elles une représentation d'état de l'objet en mouvement afin de modéliser sa dynamique. Parmi ces techniques nous pouvons notamment citer le filtre de Kalman et les filtres particulaires [73, 180].

Ces différentes méthodes sont couramment employées pour des séquences d'images perspectives, mais nous ne pouvons pas directement les appliquer à des vidéos acquises à l'aide d'une caméra catadioptrique. La principale raison est liée à la distorsion géométrique induite par l'utilisation d'un miroir. Comme nous pouvons le constater sur les images de la figure 4.4, le miroir implique une forte sensibilité à l'illumination de la scène, ainsi qu'une non-uniformité de la résolution de l'image qui rendent le suivi plus difficile. Ici, nous proposons une adaptation des algorithmes de suivi visuel, à savoir la méthode *mean-shift* et le filtrage particulaire, à la géométrie des capteurs catadioptriques.

Nous commençons par décrire, de manière succincte, le principe général du suivi de cibles et nous présenterons par la suite les adaptations proposées.

Principe du suivi visuel de cibles D'une manière générale, le suivi visuel de cibles

consiste à détecter dans chaque image d'une séquence, la position d'un ou plusieurs objets d'intérêt. Pour ce faire, l'objet est représenté par un modèle qui décrit soit sa forme, soit son apparence, soit sa position, etc. Etant donné ce modèle, que nous notons $\mathbf{q}$, et une position initiale $\mathbf{x}$ de l'objet, l'objectif est de déterminer la position $\mathbf{y}$ de l'objet dans l'image courante. Ce problème est équivalent à la recherche de la position $\mathbf{y}$ de l'image, dont l'apparence est représentée par $\mathbf{p}$, telle que, une certaine mesure de similarité $sim(\mathbf{p}, \mathbf{q})$ soit maximale.

Les différentes méthodes de suivi se distinguent par la représentation choisie et par l'approche utilisée pour résoudre ce problème d'optimisation.

— **Algorithme de suivi *mean-shift* :** Cet algorithme, proposé par Comaniciu *et al.* [35], représente l'objet à suivre par un histogramme couleur, $\mathbf{q}$, et repose sur la maximisation itérative de la similarité entre ce modèle $\mathbf{q}$ et différentes régions de l'image représentées par un histogramme $\mathbf{p}$. La mesure de similarité est définie par le coefficient de Bhattacharyya :

$$\beta(p, q) = \sum_{u=1}^{m} \sqrt{p(u)q(u)}, \tag{4.4}$$

L'algorithme procède de manière itérative comme suit. A partir d'une position initiale $\mathbf{x}$, qui est souvent la position détectée de l'objet dans l'image précédente, on calcule une position $\mathbf{y}$ comme étant le centre de gravité de toutes les fenêtres autour de $\mathbf{x}$ (les poids étant définis par les coefficients de similarité) :

$$\mathbf{y} = \frac{\sum_{i=1}^{n} w_i k(\|(\mathbf{x} - \mathbf{x}_i)/h\|^2)\mathbf{x}_i}{\sum_{i=1}^{n} w_i k(\|(\mathbf{x} - \mathbf{x}_i)/h\|^2)}. \tag{4.5}$$

Dans l'équation précédente, $k(.)$ est un noyau de lissage dont la largeur est h, et w_i est le poids de la fenêtre $\mathbf{x}_i$.
Intuitivement, la cible se déplace de $\mathbf{x}$ à $\mathbf{y}$ à chaque itération, et la position finale (définie par un critère d'arrêt) est la localisation de l'objet dans l'image courante.

— **Algorithme de suivi basée sur le filtrage particulaire :** Le filtrage particulaire, proposé pour le suivi d'objets par Isard et Blake [73], repose sur un mécanisme de prédiction de la nouvelle position de la cible à l'aide d'informations provenants des précédents états. Supposons que le système dynamique soit décrit par la représentation d'état suivante :

$$\begin{cases} x_t = f(x_{t-1}) + v_{t-1} \\ z_t = h(x_t) + n_t \end{cases}, \tag{4.6}$$

où f et h sont respectivement la fonction de transition de l'état et la fonction de mesure ou d'observation, et v_{t-1} et n_t sont les bruits du système et de mesure.
D'une manière générale, nous souhaitons estimer l'état de la cible à un instant t, $\mathbf{x}_t$, à partir des observations du système dans le temps, $Z_t = \{z_0, z_1, \ldots, z_t\}$. En d'autre termes, il faut estimer la probabilité conditionnelle $p(\mathbf{x}_t \mid Z_t)$.
Le filtre particulaire, comme toute méthode d'estimation Bayésienne, se décompose en deux étapes : une étape de prédiction et une étape de correction. L'étape de prédiction consiste à estimer la probabilité *a priori* :

$$p(\mathbf{x}_t \mid Z_{t-1}) = \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} \mid Z_{t-1}) d\mathbf{x}_{t-1}. \tag{4.7}$$

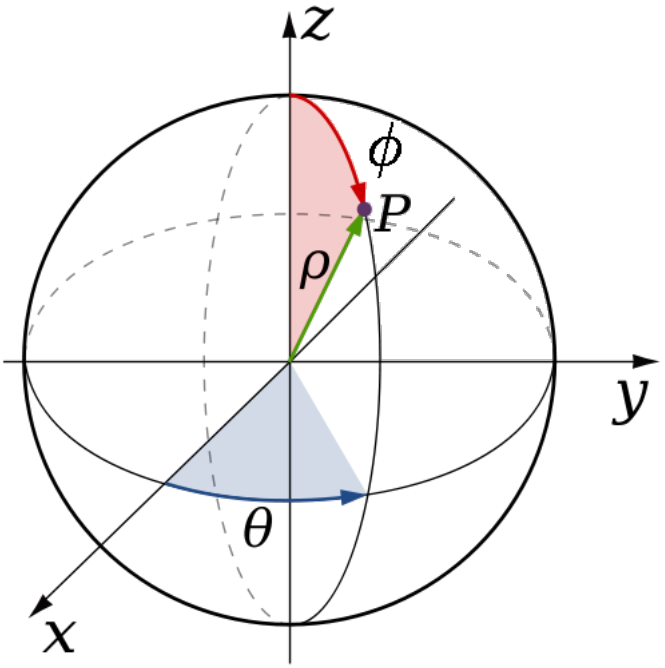


FIGURE 4.5 – Système de coordonnées sphérique.

L'étape de correction tient compte de l'observation à l'instant t pour calculer la probabilité *a posteriori* :

$$p(\mathbf{x}_t \mid Z_t) = \frac{p(z_t \mid \mathbf{x}_t)p(\mathbf{x}_t \mid Z_{t-1})}{p(z_t \mid Z_{t-1})}. \tag{4.8}$$

Adaptation du voisinage et représentation multi-parties Dans les méthodes de suivi, les objets sont décrits par des caractéristiques de couleur, de forme ou de texture, qui sont calculées dans une region d'intérêt (une fenêtre) définissant l'objet. Dans une image perspective une région d'intérêt est généralement définie par deux paramètres (largeur et hauteur) dans laquelle le voisinage est uniforme. On obtient alors un rectangle centré sur le pixel d'intérêt. Bien entendu, cet échantillonnage classique n'est pas adapté aux images catadioptriques car il considère la résolution comme étant uniforme sur l'ensemble de l'image et ne tient pas compte de la distorsion due à l'utilisation d'un miroir.
Comme évoqué à la section 4.2.1, nous pouvons utiliser le modèle sphérique pour représenter les images acquises avec une caméra catadioptrique. La projection sur la sphère nous permet de créer un voisinage adapté à la distorsion et à la non-uniformité de la résolution de l'image de la manière suivante.

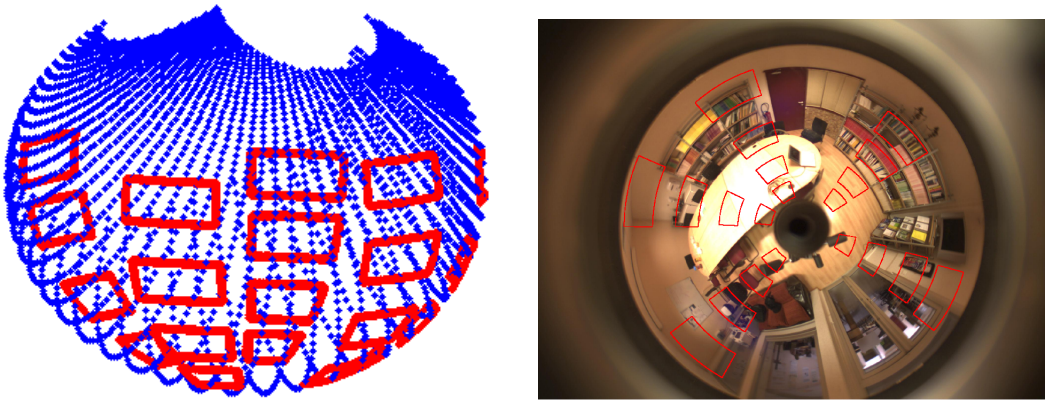
Chaque point est représenté par ses coordonnées sphériques définies comme suit :

$$\mathbf{x} = (\cos(\phi)\sin(\theta), \sin(\phi)\cos(\theta), \cos(\phi)), \tag{4.9}$$

où $\mathbf{x}$ est un point sur la sphère S^2, ϕ est la latitude (comprise entre 0 et π) et θ la longitude (de 0 à 2π). Par conséquent la localisation d'un point peut être définie à l'aide de 2 paramètres (θ, ϕ) comme nous pouvons le constater sur la figure 4.5.

Nous déterminons le voisinage à partir d'un point central sur la sphère $\mathbf{X}_{sph}(\theta,\phi)$ autour duquel nous définissons une plage de variation de θ et ϕ, respectivement $\delta\theta$ et $\delta\phi$:

$$N_S(\mathbf{X}_{sph}) = \{\mathbf{X}_S = (\theta', \phi')\epsilon S^2 | \left|\theta' - \theta\right| \leq \delta\theta \; et \left|\phi' - \phi\right| \leq \delta\phi\}.$$



(a) voisinages sur la sphère (b) voisinages projetés sur le plan image

FIGURE 4.6 – Voisinage avec des valeurs fixes $\delta\theta$=±0.2 et $\delta\phi$=±0.1.

De cette manière nous obtenons les coordonnées sur la sphère que nous reprojetons ensuite sur le plan image afin d'obtenir les coordonnées des pixels concernés. Comme nous pouvons le voir sur la figure 4.6, des fenêtres de dimension similaire sur la sphère auront des dimensions variables en fonction de leur position spatiale sur le plan image.

Ensuite, chaque fenêtre définie sur le plan image est représentée par un histogramme couleur. Comme les images catadioptriques sont très sensibles aux changements d'illumination, nous adoptons un espace couleur robuste à ces changements [169]. Cet espace repose sur la normalisation de chaque canal couleur de manière indépendante et est défini par :

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \frac{R-\mu_R}{\sigma_R} \\ \frac{G-\mu_G}{\sigma_G} \\ \frac{B-\mu_B}{\sigma_B} \end{pmatrix}, \tag{4.10}$$

σ et μ étant respectivement l'écart type et la moyenne du canal considéré.

Enfin, afin de gérer au mieux les changements d'échelle de la cible très rapides avec une caméra catadioptrique, nous représentons la cible par un histogramme multi-parties [99]. Au lieu de calculer un seul histogramme, cette approche consiste à en calculer 7 selon l'ajustement ilustré par la figure 4.7.

L'histogramme numéro 1 (fig.4.7(a)) correspond à l'histogramme de la région d'intérêt entière. Dans un deuxième temps la fenêtre est divisée en quatre parties, et un histogramme est calculé dans chacune d'entre elles. Cette organisation permet d'obtenir une information sur la rotation de la cible. La dernière étape est la division de la fenêtre en deux parties (intérieure et extérieure) qui permet une meilleure adaptation à l'échelle.

Le calcul de la similarité entre deux histogrammes multi-parties $\mathbf{q}$ et $\mathbf{p}$ est donné par :

$$\rho(\mathbf{q},\mathbf{p}) = \frac{\sum_{i=1}^{N} \beta[\mathbf{q}^i, \mathbf{p}^i]}{N}, \tag{4.11}$$

où, N est le nombre total d'histogrammes (7 dans notre cas), $\mathbf{p}^i$ et $\mathbf{q}^i$ sont, respectivement, les histogrammes de la i-ième sous-partition de $\mathbf{p}$ et de $\mathbf{q}$, et $\beta[,]$ est la mesure de similarité définie par l'équation 4.4.

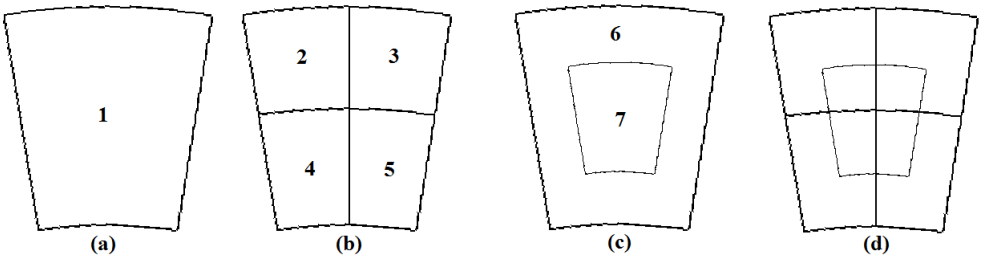


FIGURE 4.7 – Représentation multi-parties (a)région d'intérêt complète (b) division en 4 parties (c) division sensible aux changements d'échelle (d)représentation finale

	Localisation	Type de caméra	Changement d'illumination	Occultation	Cible	Images/sec	Nombre d'images
Séquence 1	Extérieure	Non centrale	Non	Non	Personne	30	780
Séquence 2	Intérieure	Non centrale	surface réfléchissante	Non	Objet	30	649
Séquence 3	Extérieure	Non centrale	Éblouissement	Partiel	Personne	30	602
Séquence 4	Intérieure	Centrale	Non	Total	Objet	15	481

TABLE 4.1 – Particularités des séquences utilisées

4.2.3/ RÉSULTATS

Pour évaluer les méthodes de suivi développées, nous avons fait l'acquisition de plusieurs séquences d'images avec une caméra catadioptrique et une caméra fisheye. Notons qu'une caméra fisheye n'est pas une caméra centrale, mais qu'elle peut être néanmoins représentée par le modèle sphérique [37, 181].

Ces séquences ont été acquises dans différentes conditions afin de couvrir des situations différentes : scène intérieure ou extérieure, objet ou humain en mouvement, avec ou sans occultation, caméra mobile ou statique. La résolution des images est de 640x480 pixels et la durée des séquences varie entre 500 et 800 images. Le tableau 4.1 rassemble l'ensemble des caractéristiques des séquences utilisées.

Nous avons, pour chacune des séquences, manuellement segmenté les objets pour avoir une vérité terrain. Nous comparons également les méthodes adaptées avec les méthodes conventionnelles pour mettre en évidence l'intérêt de prendre en compte la géométrie du capteur.

Critères d'évaluation

Il existe dans la littérature de nombreuses méthodes statistiques permettant l'évaluation des algorithmes de suivi. Trois critères représentatifs de la précision des différentes méthodes sont utilisés : la superposition spatiale, la superposition temporelle, et la distances des centres.

— **La superposition spatiale :** correspond au recouvrement spatial entre les deux boites englobantes (la fenêtre obtenue par l'algorithme de suivi S_t et la fenêtre issue de la vérité terrain G_t) :

$$A(S_t, G_t) = \frac{\text{Surface}(G_t \cap S_t)}{\text{Surface}(G_t \cup S_t)}. \tag{4.12}$$

— **La superposition temporelle :** correspond au nombre d'images de la séquence

Séquence	Méthode	superposition spatiale	distance des centres	superposition temporelle
Sequence 1	Méthode conventionnelle	45%	5.7	100%
	Notre méthode	71%	2.6	100%
Sequence 2	Méthode conventionnelle	30%	5.5	44%
	Notre méthode	60%	3.8	99.7%
Sequence 3	Méthode conventionnelle	33%	7.7	88%
	Notre méthode	65%	4.8	99.5%
Sequence 4	Méthode conventionnelle	45%	6	96.8%
	Notre méthode	66%	4.1	97.3%

TABLE 4.2 – Résultats du suvi avec un filtre particulaire conventionnel et avec la méthode adaptée.

pour lesquelles la superposition spatiale est supérieure à un seuil fixé :

$$\tau = \left(\frac{100}{N}\right) \sum_{i=1}^{N} TO(S_t, G_t), \tag{4.13}$$

avec N le nombre d'image de la séquence et $TO(S_t, G_t)$ le nombre d'images avec une superposition spatiale supérieure au seuil fixé T_{Ov} :

$$TO(S_t, G_t) = \begin{cases} 1 \; if \; A(S_t, G_t) > T_{Ov} \\ 0 \; if \; A(S_t, G_t) < T_{Ov} \end{cases} .$$

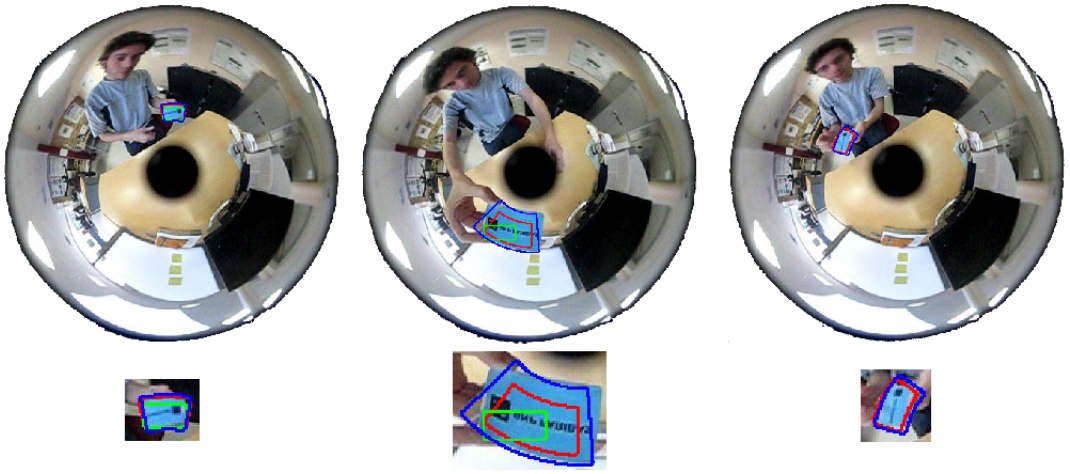
— **La distance entre les centres :** est la distance Euclidienne calculée entre le centre de la fenêtre de la vérité terrain et le centre de la fenêtre obtenue par la méthode de suivi.

Résultats Les deux tableaux 4.2 et 4.3 montrent les résultats obtenus avec, respectivement, un filtre particulaire (FP) et la méthode mean-shift (MS). On peut constater que la prise en compte de la géométrie du capteur catadioptrique permet d'améliorer la précision du suivi. On note aussi que le suivi basé sur le FP est plus fiable que le suivi basé MS. La figure 4.8 montre des résultats qualitatifs pour les séquences 2 et 3. La séquence 2 présente d'importants changements d'échelle de même qu'une variation de l'apparence de l'objet au cours du temps à cause des réflexions de la surface. Pour cette séquence, les méthodes non adaptées sont incapables de suivre la cible dans l'ensemble de la séquence. La séquence 3 fait intervenir de nombreux phénomènes typiques des capteurs catadioptriques, i.e. un fort changement de luminosité (éblouissement), une occultation et des changements d'échelle rapide avec une caméra mobile. Pour cette séquence, les algorithmes adaptés fournissent des résultats très convaincants en tous points comparativement aux méthodes conventionnelles.

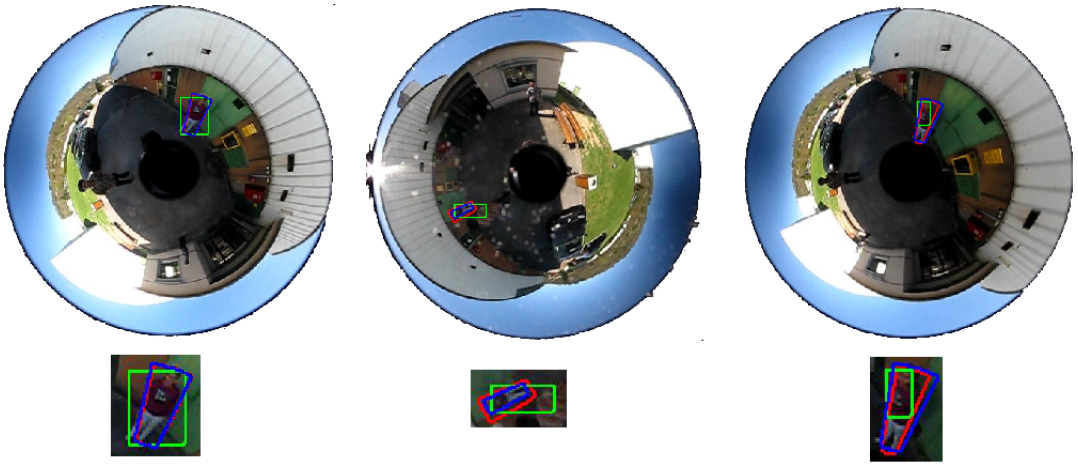
De manière quantitative et qualitative, les méthodes adaptées permettent une meilleure gestion des changements d'échelles (pour toutes les séquences), des occultations (séquence 3 et 4) et des forts changements d'illumination (séquence 2).

4.2.4/ CONCLUSION

Dans cette section, nous avons montré comment les méthodes conventionelles de suivi d'objets peuvent être adaptées à la géométrie particulière des images catadioptriques. Cette adaptation, basée sur le modèle sphérique, permet une meilleure prise en compte de la résolution non-unifiorme des images, et offre une plus grande robustesse aux variations d'illumination de la scène.



(a) séquence 2



(b) séquence 3

FIGURE 4.8 – Résultats obtenus avec un filtre particulaire conventionnel (fenêtre verte), d'un filtre particulaire adapté (fenêtre rouge) et de la vérité de terrain (fenêtre bleue).

Séquence	Méthode	superposition spatiale	distance des centres	superposition temporelle
Sequence 1	Méthode conventionnelle	32.4%	6.38	63.21%
	Notre méthode	64.5%	6.2	96.35%
Sequence 2	Méthode conventionnelle	42.5%	16.2	89%
	Notre méthode	47.79%	8.3	90%
Sequence 3	Méthode conventionnelle	40.55%	7.8	96%
	Notre méthode	61.8%	7.5	97.84%
Sequence 4	Méthode conventionnelle	34.72%	10.52	52.81%
	Notre méthode	67.67%	5.02	97.71%

TABLE 4.3 – Résultats du suivi avec l'algorithme *Mean-Shift* conventionnel et avec la méthode adaptée.

4.3/ DÉTECTION ET RECONNAISSANCE D'OBJETS 3D

Dans cette section, nous nous intéressons à l'analyse de scènes 3D et, plus particulièrement, à la reconnaissance d'objets dans une scène acquise avec une caméra

de profondeur de type Kinect. Le faible coût de la Kinect et son poids peu encombrant en font l'un des capteurs 3D les utilisés dans diverses applications de surveillance, de navigation ou d'analyse du mouvement [69]. Notons que nous employons, dans ce manuscrit, le terme de capteur 3D pour désigner tout capteur ou système de capteurs permettant de mesurer une profondeur et exprimant le résultat sous la forme d'une image de profondeur. Il existe différents types de capteurs 3D aussi bien actifs (comme la Kinect) que passifs (comme les paires stéréoscopiques).

Dans un premier temps, section 4.3.1, nous décrivons brièvement le fonctionnement du capteur. Puis, dans la section 4.3.2, nous proposons un descripteur de nuage de points 3D combinant des propriétés géométriques et de texture pour la reconnaissance d'objets.

4.3.1/ FONCTIONNEMENT DE LA KINECT

La Kinect, de Microsoft, initialement développé pour les jeux vidéos (console de jeux Xbox) est aujourd'hui l'un des capteurs de profondeur le plus utilisé par les chercheurs en vision par ordinateur et en robotique. Cela est principalement dû à son faible coût qui en fait l'un des capteurs 3D les plus abordables (comparé au paires stéréo et aux caméras à temps de vol), et au fait que la Kinect permet l'acquisition synchronisée d'une image RGB et d'une image de profondeur.

La Kinect, comme les autres caméras du même type, par exemple la Xtion Pro d'Asus, repose sur la projection d'un motif de lumière structurée (dans le domaine proche infrarouge) sur la scène observée. Une caméra infrarouge capture une image du motif projeté et il est possible d'estimer la profondeur de chaque point du motif par un simple calcul de disparité entre le motif et son image :

$$d = \frac{B \times f}{z}, \tag{4.14}$$

avec d la disparité, B la baseline (la distance entre le centre de la source et le centre du capteur), f la focale de la caméra et z la profondeur.

La figure 4.9 illustre le principe de la mesure de la profondeur par la Kinect.

Notons que ce capteur est également composé d'une caméra couleur RGB normale qui capture une image couleur de la scène. L'image de la figure 4.10 montre une caméra Kinect et ses différents constituants ainsi qu'une image couleur et l'image de profondeur correspondante.

Dans le cas de la Kinect, la lumière choisie se place dans le spectre du proche infrarouge la rendant invisible à l'oeil nu et à la caméra couleur RGB. Il est donc, a priori, possible d'utiliser ce capteur dans un environnement entièrement sombre. Cependant, même si la caméra 3D n'est pas sensible à la lumière visible, elle n'est pas utilisable de jour, en extérieur, du fait de la présence des longueurs d'onde du proche infrarouge dans la lumière émise par le Soleil. Ces rayons pertubent le calcul des profondeurs.

Plusieurs auteurs ont analysé les performances de la Kinect pour la mesure de profondeur. Ces différents travaux montrent que la Kinect est très adaptée pour la cartographie intérieure lorsque les distances de mesure sont faibles, i.e. inférieures à 3.5 mètres [152, 80]. A cette distance, la Kinect offre de meilleures performances qu'une caméra à temps de vol et est comparable à un télémètre laser [159]. D'autre part, Khoshelham *et al.* [80] montrent que l'erreur sur la mesure de la profondeur croît de manière

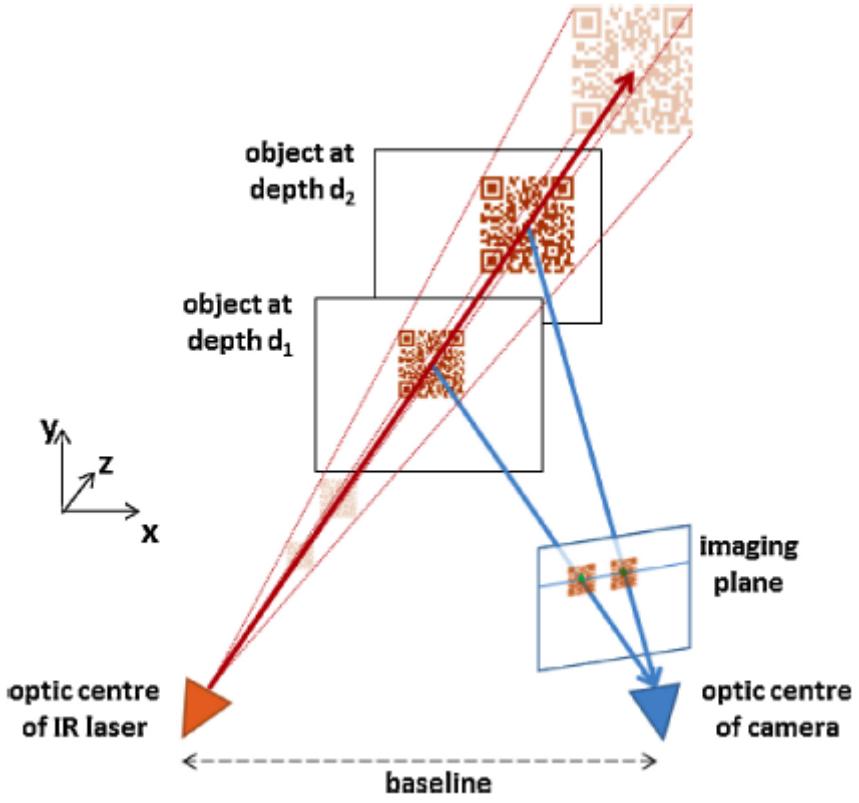


FIGURE 4.9 – Principe de la mesure de la profondeur par la Kinect. Image extraite de [69].

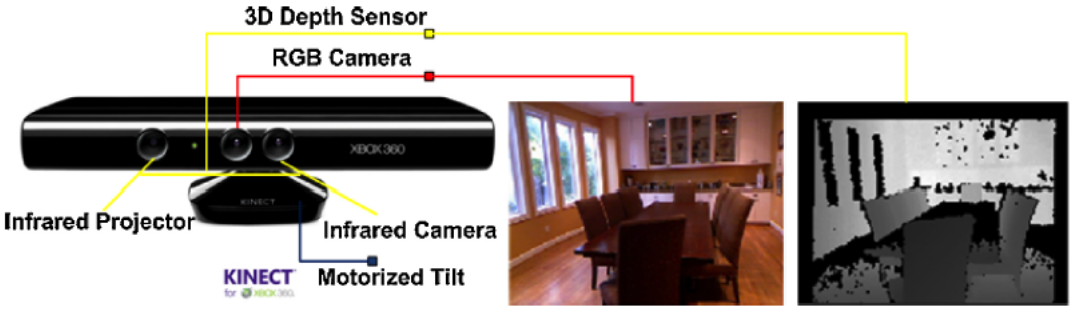


FIGURE 4.10 – Configuration de la Kinect. Image extraite de [69].

quadratique avec la distance de mesure, de quelques millimètres à environ 4 cm.

4.3.2/ RECONNAISSANCE D'OBJETS 3D

La reconnaissance d'objets respose, généralement, sur des descripteurs qui représentent l'objet de manière unique et facilitent son identification dans la scène. Parmi les nombreux descripteurs de surface proposés dans la littérature, ceux basés sur l'orientation des points 3D sont les plus couramment utilisés. Néanmoins, ils sont très peu robustes au bruit et aux variations de point de vue. Nous proposons, d'une part, une méthode qui permet de rendre ces descripteurs plus robuste en réduisant leur taille, et,

d'autre part, un descripteur qui combine des propriétés géométriques et de texture tout en assurant une bonne robustesse.

4.3.2.1/ UN ÉTAT DE L'ART DES DESCRIPTEURS 3D

De nombreux descripteurs 3D ont été proposés dans la littérature ces dernières années. Certains sont des extensions de descripteurs 2D, c'est le cas de 3D-SIFT ou 3D Shape Context qui sont des extensions des descripteurs 2D du même nom, tandis que d'autres ont été spécifiquement développés pour caractériser des nuages de points 3D [10, 137].

Les différents descripteurs peuvent être regroupés dans deux grandes catégories : les descripteurs locaux et les descripteurs globaux.

- **Descripteurs locaux :** Ils décrivent le voisinage de chaque point de la surface ou d'un ensemble de points d'intérêt définis sur la surface. Parmi les principaux descripteurs locaux, nous pouvons citer les descripteurs 3DSC (3D Shape Context) [53] et USC (Unique Shape Context) [160]. Le premier est une extension directe du descripteur 2D tandis que le second introduit un repère de référence unique pour le calcul du descripteur de forme. Des extensions telles que SHOT (Signature of Histogram of Orientation) [161] ont été proposées qui encodent également une différence d'orientation. Les descripteurs locaux sont généralement employés pour le recalage de surfaces.
- **Descripteurs globaux :** Ils décrivent une partie de la surface ou la surface entière à l'aide d'attributs géométriques ou colorimétriques. Le nuage de points est d'abord segmenté en régions (ou segments) et un descripteur est calculé pour chaque région. Parmi les principaux descripteurs globaux, soulignons ceux qui encodent les différences d'orientations comme les descripteurs PFH (Point Features Histogram) [135] et VFH(Viewpoint Feature Histogram) [136]. Le premier calcule une normale en chaque point 3D (la normale au plan tangent à la surface en ce point) et calcule la différence d'orientations entre un point 3D et tous les autres points de la surface. Enfin, ces différences sont encodées sous la forme d'un histogramme qui représente la surface. Le second descripteur est une extension du premier qui tient compte de l'orientation globale du nuage de point en calculant les différences d'orientations uniquement par rapport au centroïde du nuage. Ce qui réduit la complexité du calcul. Ces deux descripteurs sont représentatifs des descripteurs globaux et de nombreuses variantes ont été developpées telles que PFHRGB (PFH with RGB values) [13], CVFH (Clustered Viewpoint Feature Histogram) [12], OUR-CVFH (Oriented, Unique and Repeatable CVFH) [11], PPF (Point Pair Features) [135].

Pour une description plus détaillée des différents descripteurs de nuage de points, nous invitons le lecteur intéressé à consulter [162, 10, 140].

Plusieurs travaux ont analysé les performances de différents descripteurs 3D pour la reconnaissance d'objets [10] ou la classification de scènes urbaines [19]. Il ressort de ces études que les descripteurs qui encodent les différences d'orientations, tels que SHOT et PHF, et ceux qui encondent la distribution de points 3D, comme 3DSC et USC, sont les plus performants. Néanmoins, tous ces descripteurs représentent l'information sous forme d'histogrammes de grande taille. Par exemple, un descripteur SHOT est représenté par un vecteur de dimension 352, et PFH par un vecteur de dimension 125. Cela limite

leur application pour le traitement temp-réel, par exemple la reconnaissance d'objets avec un robot mobile. D'autre part, ces vecteurs de grande dimension sont souvent éparses, i.e. toutes les cellules de l'histogramme ne sont pas occupées. Et comme l'ont montré Salih *et al.* [142] cette représentation est peu robuste au bruit. On note une dégradation de la performance de l'ordre de 15% pour SHOT et l'ordre de 30% pour PFH, lorsqu'on ajoute un bruit Gaussien d'écart type égal à 5% à la position des points du nuage. Cette dégradation importante pouvant s'expliquer par le fait que ces descripteurs sont basés sur le calcul des normales et que cette estimation des normales est très sensible au bruit.

Dans la section suivante, nous proposons un descripteur, qui est une extension des descripteurs PHF et VFH, qui repose sur une analyse en composantes principales (ACP) pour réduire la dimension du descripteur et sur l'ajout de la couleur pour une plus grande robustesse.

4.3.2.2/ PROPOSITION D'UN DESCRIPTEUR 3D ROBUSTE ET DE TAILLE RÉDUITE

Nous proposons une méthodologie pour réduire la taille des descripteurs représentés sous forme d'histogramme, tout en maintenant une grande robustesse au bruit. Cette méthodologie, qui sera ici appliquée au descripteur PFH mais qui peut s'appliquer à d'autres descripteurs, est basée sur une ACP pour extraire les directions principales du nuage de point qui sont ensuite utilisées pour décrire le nuage. Enfin, nous combinons les propriétés géométriques et de texture extraites du nuage de point pour définir un descripteur global qui est comparé aux autres descripteurs présentés dans la littérature.

Réduction de la taille des descripteurs L'ACP a déja été employée pour réduire la taille du descripteur SIFT [78] et nous utilisons ici une approche similaire pour les descripteurs 3D. Dans [72], les auteurs utilisent également l'ACP pour compresser le descripteur « Spin Image », mais applique une ACP à chaque nuage de point de manière indépendante.

Dans notre approche, nous créons un modèle générique à partir d'un grand nombre de nuages de points représentant divers objets dans différentes positions et orientations. Pour ce faire, nous avons sélectionné un ensemble de $M = 5000$ nuages de points 3D de la base de données publique (RGB-D Dataset) [83]. Cette base, mise en place par des chercheurs de l'université de Whashington, est constituée d'environ 250 000 images RGB-D représentant 300 objets sous différents angles ou points de vue. Elle a à ce jour la base de données RGB-D la plus complète pour évaluer les algorithmes de reconnaissance d'objets. La figure 4.11 montre quelques exemples d'images de cette base de données.

L'approche proposée consiste à extraire de chaque nuage de points un descripteur $\mathbf{d}_i$, par exemple PFH, et à créer une matrice de données $\mathbf{X} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_M]$, avec $\mathbf{d}_i \in \mathcal{R}^d$ et d la dimension du descripteur. On applique une ACP en décomposant la matrice $\mathbf{X}$ en $\mathbf{X} = U\Sigma V^T$ (décomposition en valeurs singulières). Les colonnes de U sont les axes principaux sur lesquels on projete chaque descripteur : $\hat{\mathbf{d}}_i = U_k^T \mathbf{d}_i$, avec U_k la matrice composée des k premières colonnes de U correspondantes aux valeurs singulières les plus grandes (éléments diagonaux de Σ). On réduit donc la dimension du descripteur de d à k, avec $k < d$: $\mathbf{d}_i \in \mathcal{R}^d$ et $\hat{\mathbf{d}}_i \in \mathcal{R}^k$. Cette valeur k est l'un des paramètres de notre méthode et son influence sera étudiée dans la suite.

Pour appliquer cette méthodologie, nous devons nous assurer que tous les nuages de

FIGURE 4.11 – Quelques images RGB-D de la base de données RGB-D Datset [83].

points possèdent le même nombre de points 3D, ce qui n'est pas le cas. Il faut donc ré-échantillonner les nuages de points dont les tailles varient de quelques milliers à plusieurs dizaines de milliers de points. Supposons que l'on souhaite ré-échantillonner tous les nuages de points à une taille fixe de N points. Nous avons adopté une méthode d'échantillonnage assez simple qui consiste à d'abord calculer le centroïde du nuage, puis à ordonner les points en fonction de leur distance par rapport au centroïde. Enfin, nous sélectionnons, de manière uniforme, un ensemble de N points dans l'ensemble de points ordonnés. Dans nos expériences, nous avons fixé $N = 1000$ car cette valeur permet une bonne identification des objets à partir des nuages échantillonés.

Combinaison de la texture et de la géométrie La plupart des descripteurs présentés ci-dessus ne tiennent pas compte de la couleur, bien que la Kinect produise également une image couleur RGB de scène capturée. Quelques auteurs ont donc étendu ces descripteurs y en ajoutant une information couleur. C'est le cas de SHOT-COLOR [143] ou de PFHRGB (PFH with RGB values) [136] qui sont des extensions directes de SHOT et PFH, en ajoutant à l'histogramme des différences d'orientations, un histogramme RGB. Ce qui a pour effet d'augmenter la taille du descripteur ; SHOT-COLOR a une dimension de 1344, et PFH une dimension de 250 (soit deux fois la dimension de PFH).

Dans notre approche, nous incluons la texture de la surface en caractérisant chaque point du nuage par la valeur de la teinte (H) et la différence d'intensité ΔV par rapport au centre. Nous calculons un histogramme pour chacune de ces deux caractéristiques. Un nuage de point est donc décrit par 6 histogrammes :

- 3 histogrammes représentants les distributions des orientations de surfaces. Chaque histogramme encode les différences d'orientation entre les points du nuage et le centre du nuage, pour chacun des angles de roulis, de tangage et de lacet.
- 1 histogramme représentant la distribution des points 3D, i.e. les distances géométriques des points par rapport au centre du nuage.
- 2 histogrammes représentants les distributions de texture, i.e. teinte et intensité de chaque point du nuage.

Pour chacune de ces 6 caractéristiques, nous appliquons la méthodologie expliquée ci-dessus pour réduire les dimensions, et le descripteur final est obtenu comme la concaténation des 6 descripteurs projetés sur les axes principaux.

Données et critères d'évaluation Nous utilisons les données la base RGB-D Dataset pour notre évaluation. Nous considérons comme modèles d'objets, 300 nuages de points (un par objet) qui représentent le point de vue de référence, i.e. la vue de face de l'objet. Un nuage test donné est compararé à l'ensemble des 300 modèles d'objets, et le nuage le plus similaire est retourné. La reconnaissance de l'objet est considérée comme correcte, si le nuage retourné correspondant bien à l'objet. Notons que ces données d'évaluation n'ont pas été utilisées pour calculer les directions principales avec l'ACP.

Nous analysons les performances des descripteurs en réalisant plusieurs expériences :

- **Robustesse au bruit :** A chaque nuage de point représentant un modèle d'objet, nous ajoutons un bruit Gaussien, d'écart type variable. Les nuages bruités sont ensuite utilisés comme nuages de test pour évaluer la robustesse des descripteurs en calculant le nombre d'objets correctement identifiés.
- **Robustesse au changement de point de vue :** La base RGB-D Dataset comprend plusieurs images de chaque objet selon différents angles de vue. Nous évaluons la performance des différents descripteurs pour des angles de vues croissants.
- **Reconnaissance d'objets et de catégories d'objets :** Nous considérons 10 catégories d'objets, chacune contenant 5 objets différents. Les catégories définissant des groupes d'objets de la vie courante, par exemple des bols, des balles, des boîtes, etc. Pour chaque objet, 20 nuages de points 3D, i.e. 20 angles de vue différents, sont utilisés comme modèles et 10 autres nuages (i.e. 10 autres points de vue) sont utilisés comme données de test. Chacun des 500 nuages de test est comparé à l'ensemble des modèles, et on évalue la performance des descripteurs à deux niveaux : la reconnaissance de la catégorie de l'objet, et l'identification exacte de chaque objet.

Nous comparons plusieurs descripteurs :

- **PCA-PFH :** Le descripteur proposé basée sur une compression du descripteur PFH par ACP.
- **PCA-GTFH :** Le descripteur proposé combinant la texture et la géométrie (GTFH = Geometric and Texture Feature Histogram).
- **Divers :** Les descripteurs proposés dans la littérature tels que PFH, VFH, SHOT, PFHRGB, SHOTCOLOR, 3DSC et USC.

Résultats Nous commençons par comparer le descripteur obtenu en appliquant une ACP à PFH, noté PCA-PFH, avec le descripteur original, PFH, en présence de bruit. Nous faisons également varier la taille du descripteur PCA-PFH en faisant varier le nombre de composantes principales sur lesquelles la projection est effectuée. Comme le montre la figure 4.12, le descripteur PHF et la version compressée PCA-PFH sont très sensibles au bruit, puisqu'on obtient moins de 50% de détections correctes avec un niveau de bruit supérieur à 2.5%. Pour des niveaux de bruit plus faibles, on note que les résultats obtenus avec PCA-PFH sont au moins aussi bons que ceux obtenus avec PFH, et cela même lorsque la taille du descripteur est réduite à 75 au lieu de 125 pour PFH. Le meilleur compromis performance-taille est obtenu avec un descripteur de taille 90. C'est donc cette taille qui sera utilisée par la suite.

La figure 4.13 montre les résultats de l'évaluation de la robustesse au changement de point de vue. Comme on peut le constater, avec un descripteur PCA-PFH de taille 90 on obtient de meilleurs résultats qu'avec le descripteur PFH. En particulier, lorsque la différence d'angle de vue entre les objets est importante. Pour une différence de 65^o, on

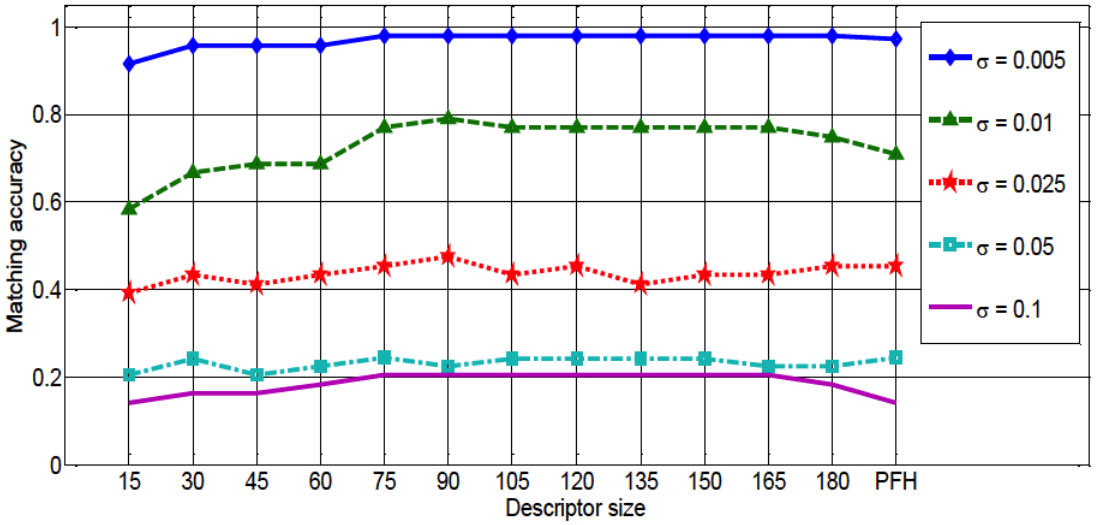


FIGURE 4.12 – Comparaison de PCA-PFH et PFH en présence de bruit.

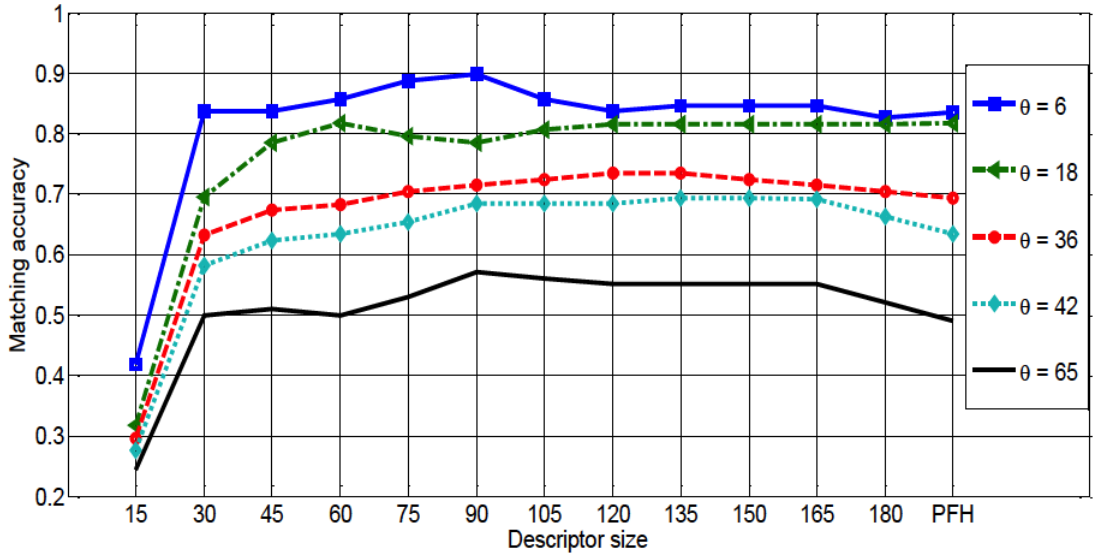


FIGURE 4.13 – Comparaison de PCA-PFH et PFH pour des points de vue variables.

obtient une performance de 57% avec PCA-PFH tandis qu'elle est de 49% avec PFH.

Ces deux premières expériences montrent que l'approche de réduction de la taille des descripteurs par ACP est pertinente et permet même d'accroître la robustesse au bruit et aux variations de point de vue. Nous appliquons donc cette méthodologie en ajoutant aux caractéristiques géométriques, des caractéristiques couleur pour décrire les nuages de points 3D. Le descripteur obtenu, PCA-GTFH, est comparé avec divers autres descripteurs proposés dans la littérature. La figure 4.14 montre les résultats de l'évaluation de la robustesse de ces descripteurs au bruit. On note que pour des faibles niveaux de bruit, mois de 2%, tous les descripteurs, à l'exception de USC, SHOT et SHOT-COLOR, donnent de très bons résultats, supérieurs à 80%. La mauvaise performance de USC (Unique Shape Context) et de ses extensions SHOT et SHOT-COLOR, est due au fait que ce descripteur calcule des différences d'orientations par rapport à un repère de référence

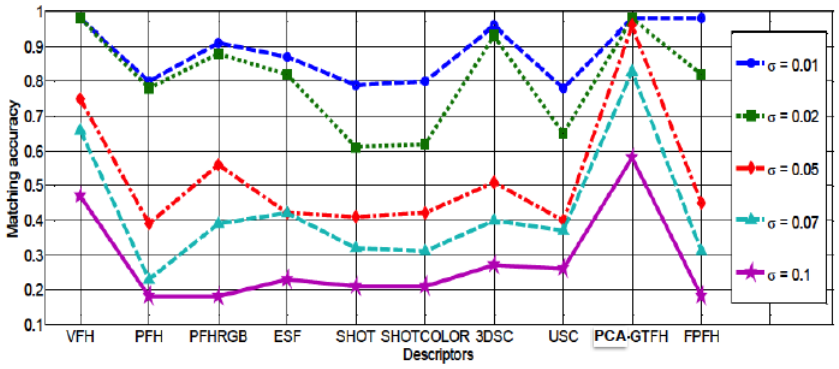


FIGURE 4.14 – Comparaison de PCA-GTFH avec différents autres descripteurs en présence de bruit.

unique. Or ce repère est sensible au bruit, ce qui dégrade le descripteur obtenu. Les autres descripteurs, par exemple PFH, calculent toutes les différences d'orientation entre toutes les paires de points. Ils sont donc plus robustes au bruit. On note également que le descripteur proposé, PCA-GTFH, est plus robuste que les autres lorsque le niveau de bruit augmente. Pour un bruit d'écart type $\sigma = 0.1$, c'est l'unique descripteur qui obtient une performance supérieure à 50%.

Dans l'expérience précédente (voir figure 4.12) la performance de PCA-PFH pour ce niveau de bruit était inférieure à 20%. Ce qui confirme que la combinaison de la texture et de la géométrie permet d'améliorer les résultats. Néanmoins, comme le montre aussi la figure 4.14, le simple ajout de la couleur, comme c'est le cas avec PFHRGB et SHOTCOLOR, n'améliore pas sensiblement les performances. En effet, nous constatons que la performance de PFHRGB est comparable à celle de PFH, et celle de SHOT-COLOR est quasi-identique à celle de SHOT. C'est donc bien l'approche de représentation par ACP qui permet une plus grande robustesse au bruit.

Dans la dernière expérience, nous évaluons la capacité des différents descripteurs à identifier correctement un objet et la catégorie à laquelle il appartient. Le tableau 4.4 rassemble les résultats obtenus. Bien entendu, la reconnaissance des catégories est plus aisée et les scores obtenus sont donc plus élevés. On notera que le descripteur proposé, PCA-GTFH, obtient le meilleur score avec un taux de reconnaissance correcte de 97.5%, bien supérieur à ceux des autres descripteurs. Pour la reconnaissance des objets, PCA-GTFH obtient un bon score de 78.75% et n'est dépassé que par le descripteur SHOT-COLOR qui obtient un score de 80.30%.

4.3.3/ CONCLUSION

Dans cette section, nous avons dans un premier temps montré l'intérêt de la réduction de la taille des descripteurs géométriques de nuages de points en utilisant une ACP. Cette réduction permet, non seulement, un gain en temps de calcul et en espace mémoire, mais améliore la robustesse du descripteur au bruit et aux variations de points de vue de l'objet. Il est à souligner que cette méthodologie peut être appliquée à différents descripteurs représentés sous la forme d'histogrammes. Nous avons ensuite proposé un descripteur basé sur la combinaison d'attributs géométriques et de texture, et les expériences avec

Descripteurs	Résultats (%)		Temps CPU
	Objets	Catégories	millisecondes
PCA-GTFH	78.75	**97.50**	125.00
PFHRGB [13]	73.75	91.25	2856.25
PFH [135]	60.61	72.73	1775.00
ESF [174]	41.25	65.00	150.00
VFH [136]	27.50	46.25	112.50
FPFH [136]	60.61	69.70	815.25
SHOTCOLOR [143]	**80.30**	81.82	121.00
SHOT [161]	66.67	77.27	106.25
3DSC [53]	46.97	59.09	143.75
USC [160]	65.15	81.82	156.25

TABLE 4.4 – Comparaison de PCA-GTFH avec différents autres descripteurs pour la reconnaissance d'objets et de catégories d'objets.

une base de données publique ont montré la bonne performance de notre descripteur par comparaison avec différents descripteurs proposés dans la littérature.

4.4/ CONLUSIONS ET DISCUSSION

Dans ce chapitre, nous avons abordé le problème de l'analyse de scènes avec des caméras atypiques, i.e. autres que des caméras RGB perspectives classiques. Nous nous sommes, en particulier, intéressés au suivi d'objets mobiles avec une caméra catadioptrique et à la reconnaissance d'objets 3D acquis avec une caméra de profondeur de type Kinect.

Dans une première partie, nous avons proposé une méthode d'adaptation des méthodes existantes de suivi, qui permet de tenir compte de la géométrie particulière des capteurs et des images catadioptriques. Les résultats obtenus montrent que les méthodes déterministes (mean-shift par exemple), tout comme les approches probabilistes (filtre particulaire) peuvent être adaptées pour un suivi robuste avec une caméra catadioptrique, et même une caméra fisheye.

Dans une seconde partie, nous avons montré qu'il est possible de réduire la taille de différents descripteurs de nuages de points 3D acquis avec une Kinect, pour réduire la complexité sans sacrifier la performance des algorithmes de reconnaissance. Cette réduction est basée sur une ACP qui extrait les directions principales du nuage de points qui sont ensuite utilisées pour décrire le nuage. Enfin, nous combinons les propriétés géométriques et de texture extraites du nuage de point pour définir un descripteur global, dont la taille est réduite par ACP. Le nouveau descripteur ainsi obtenu est plus performant en terme de robustesse au bruit (bruit Gaussien) et de reconnaissance d'objets et de catégorie d'objets.

Ces travaux ont été réalisés dans le cadre de deux thèses de doctorat. La thèse de François Rameau, co-financée par la DGA et la région Bourgogne, avait pour objectif la mise au point d'un système de vision hybride composé d'une caméra omnidirectionnelle, pour la vision panoramique, et d'une caméra active PTZ, pour la vision détaillée de zones d'intérêt. Le suivi avec la caméra omnidirectionnelle permet d'orienter la caméra PTZ sur l'objet pour en avoir un vue détaillée [131, 130, 129, 128]. La thèse de Yasir Salih, en co-

tutelle avec l'UTP en Malaisie, avait pour but l'analyse de scènes 3D à l'aide d'un robot mobile équipé de Kinect. L'accent a donc été mis sur la réduction de la complexité des calculs [142, 141].

Ce travail sur les images de profondeur se poursuit dans le cadre de l'ANR PLATINUM (2015-2019), projet dans le cadre duquel je co-encadre une thèse de doctorat à partir de la rentrée 2016, portant sur la localisation d'un agent (un robot mobile ou une personne) dans un environnement urbain à l'aide de données hétérogènes (nuage de points 3D, image de profondeur, image couleur, information sémantique). Dans ce projet, nous nous intéresserons à l'utilisation conjointe de ces différents types d'informations pour en extraire des caractéristiques permettant l'identification des objets et des points de repère dans l'environnement.

III

Le dépistage de la rétinopathie diabétique

5

ANALYSE D'IMAGES DE FOND D'ŒIL

La rétinopathie diabétique (RD) est une complication oculaire du diabète qui se manifeste par l'apparition de lésions sur la rétine du patient diabétique. Si elle n'est pas traitée très tôt, la RD conduit à de graves déficiences visuelles dont la cécité. Elle est d'ailleurs la principale cause de cécité chez les personnes de moins de 65 ans en France. Dans ce chapitre, nous présentons différentes méthodes pour la détection de lésions rétiniennes et le diagnostic de la RD en utilisant des images de fond d'œil. En tenant compte de la difficulté d'obtention d'un grand nombre d'exemples manuellement annotés dans le domaine médical, nous proposons des approches nécissitant peu ou pas d'exemples annotés. Enfin, nous explorons aussi les méthodes d'extraction automatique de caractéristiques dans les images pour la classification.

5.1/ INTRODUCTION

La rétinopathie diabétique (RD), une conséquence du diabète, en particulier de l'hyperglycémie chronique, est l'une des principales causes de cécité dans le monde avec la cataracte, le glaucome et la dégénérescence maculaire liée à l'âge (DMLA). Si la cataracte peut être traitée efficacement par un traitement chirurgical, ce n'est pas le cas des autres pathologies. Néanmoins, une détection précoce permet de réduire, voire d'éviter le risque de cécité dans la plupart des cas [79]. Dans ce travail, nous nous intéressons particulièrement à la RD qui se manifeste par l'apparition de lésions rétiniennes dues à l'endommagement des capillaires rétiniens (petits vaisseaux de la rétine).

Avec la progression du diabète dans nos sociétes, la prévalence de la RD est amenée à croître de manière importante dans les prochaines années. On estime en effet que la RD affectera environ 300 millions de personnes dans le monde à l'horizon 2025 [48, 173]. L'une des difficultés du dépistage de la RD est liée à la nature silencieuse de la maladie qui ne devient symptomatique qu'au stage des complications. Un suivi régulier des patients diabétiques est donc un enjeu important pour limiter la perte de vue. Or, dans une recommandation de décembre 2010, la Haute Autorité de Santé (HAS) note une insuffisance du suivi ophtalmologique des patients diabétiques, avec moins de 50% d'entre eux consultant un ophtalmologiste chaque année. La HAS conclut que la pratique actuelle du dépistage de la RD ne permet pas d'atteindre les objectifs fixés et elle recommande le dépistage par lecture différée de photographies du fond d'œil [70]. Ce dépistage repose sur des systèmes d'aide à la décision (CAD pour *Computer Aided Diagnosis*) qui offrent les bénéfices suivants :

— **Dépistage d'une population importante :** Les photographies peuvent être

(a) Vision normale (b) Vision affectée par la RD

FIGURE 5.1 – Effet de la rétinopathie diabétique sur la vision.

réalisées par des personnes formées, non nécéssairement des médécins, et de manière itinérante pour accéder aux populations éloignées des centres de soins.

— **Réduction de temps et de coût :** Un premier dépistage automatique permet d'orienter uniquement les personnes à risque vers les ophtalmologistes.

5.1.1/ MOYENS DE DÉPISTAGE DE LA RD

D'une manière générale, la RD endommage les capillaires rétiniens ce qui entraîne un dysfonctionnement des photo-récepteurs de la rétine et une détérioration de la vue comme le montre l'exemple de la figure 5.1.

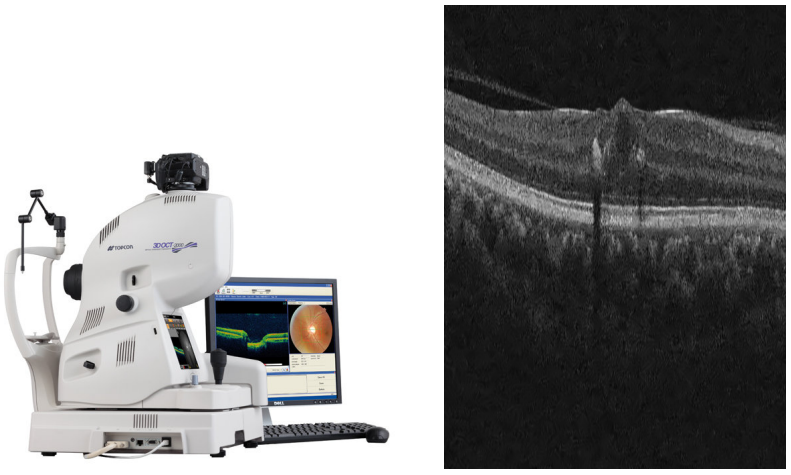
Le dépistage de la RD repose sur des systèmes d'imagerie qui permettent d'observer l'intérieur de l'œil de manière non invasive. Les deux principaux outils de dépistage sont la photographie du fond d'œil et la tomographie par cohérence optique.

La photographie du fond d'œil
Elle constitue l'examen de référence à la fois pour le dépistage et la surveillance de la RD. Elle permet de visualiser différents signes cliniques de la RD et de quantifer l'évolution de la maladie. Une photographie du fond d'œil est obtenue à l'aide d'une caméra spécifique, appelée caméra de fond d'œil (ou fundus camera en anglais), qui permet de capturer la surface intérieure de l'œil incluant la rétine, la papille optique et la macula. La caméra de fond d'œil est composé d'un système microscopique, d'une source lumineuse (un flash) et d'une caméra [139]. La figure 5.2 montre une caméra de fond d'œil ainsi que deux images rétiniennes acquises avec cette caméra.

La tomographie par cohérence optique
La tomographie par cohérence optique (Optical Coherence Tomography [OCT]) est une technique d'imagerie du fond d'œil qui permet d'obtenir in vivo des images en coupe optique de la rétine avec une résolution de l'ordre de 5 μm. Le principe de l'OCT est analogue à celui de l'échographie, sauf qu'elle utilise la lumière et non le son, et le principe de l'interféromètre de Michelson. L'image reconstruite dépend donc de l'absorption et de

(a) Caméra (b) Image de l'œil droit (c) Image de l'œil gauche

FIGURE 5.2 – Caméra de fond d'œil et exemple d'images rétiniennes.



(a) Caméra OCT (b) Image OCT reconstruite

FIGURE 5.3 – Caméra OCT et exemple d'image OCT obtenue.

la réflexion de cette lumière par les tissus biologiques. L'OCT est aujourd'hui la seule technique permettant de visualiser les différentes couches constitutives de la rétine *in situ*. La figure 5.3 montre un exemple d'image OCT de l'œil.

Chacun de ces deux outils offre des avantages pour le dépistage de la RD. La photographie du fond d'œil est simple à utiliser, largement disponible à un faible coût, et permet de visualiser les lésions apparaissant à la surface de la rétine. L'OCT a un coût plus élévé mais permet d'obtenir des informations non visibles sur la surface de la rétine. Elle permet en particulier de mesurer l'épaisseur de la rétine et est indispensable pour le traitement et le suivi des œdèmes maculaires. Dans ce chapitre, nous aborderons la détection de la RD à l'aide d'images du fond d'œil uniquement. L'analyse des images OCT sera abordée dans le chapitre 6.

5.1.2/ DÉTECTION DE LÉSIONS RÉTINIENNES

La photographie du fond d'œil permet d'obtenir des images couleur de la retine du patient, sur laquelle il est possible de voir différentes lésions caractéristiques de la RD. En effet, la rétine est irriguée par des vaisseaux sanguins qui lui apportent les éléments nutritifs

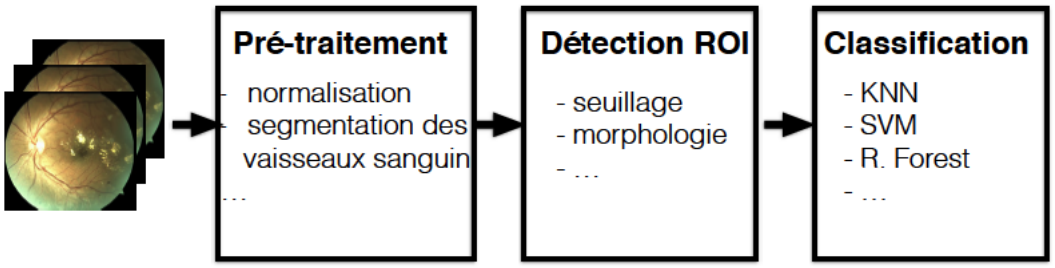


FIGURE 5.4 – Procédure de détection de lésions dans les images de fond d'œil.

nécessaires à son fonctionnement. Ceux-ci donnent des capillaires très fins qui irriguent les cellules nerveuses de la rétine. Lorsque la glycémie reste élevée, le diabète provoque des déformations et des lésions des capillaires rétiniens [79]. La déformation de la paroi des capillaires entraîne des **microanévrismes** qui sont considérés comme les premiers signes visibles de la RD. La perméabilité de la paroi, entraîne le passage de liquide qui provoque le gonflement de la rétine, i.e. un **œdème maculaire**. Elle peut aussi entraîner le passage des lipides du sang qui se déposent sur la rétine formant des **exsudats**. Parfois, l'occlusion des capillaires entraîne l'arrêt de la circulation sanguine dans une zone de la rétine, qui en réaction, stimule la prolifération de vaisseaux anormaux, les **néovaisseaux**, qui peuvent saigner et provoquer une **hémorragie intravitréenne**.

On distingue deux types de RD :

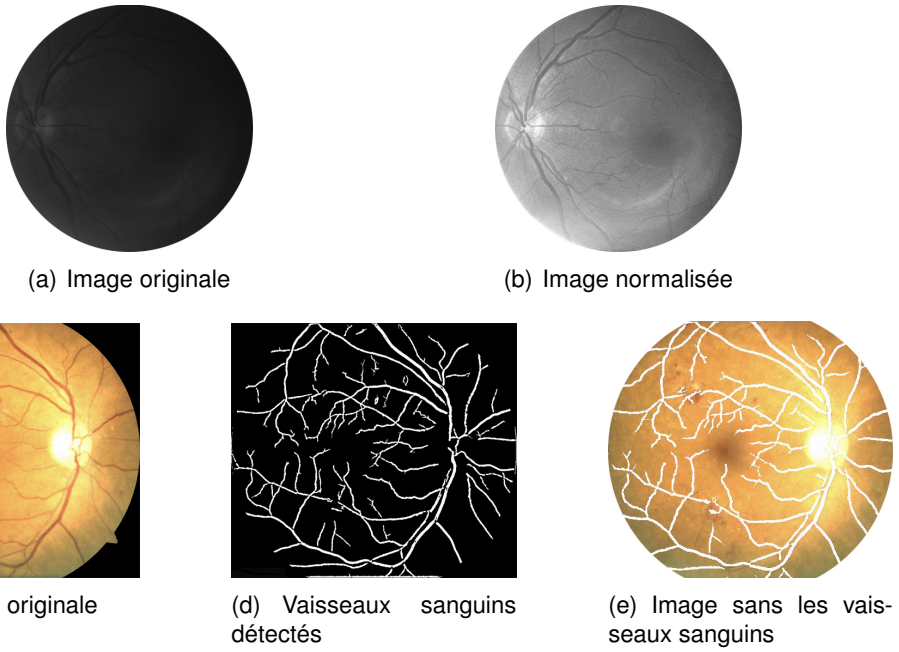
— **La RD non proliférante :** qui se manifeste par la présence de microanévrismes, d'œdème maculaire ou des exsudats.

— **La RD proliférante :** lorsqu'il y a déjà des néovaisseaux ou une hémorragie intravitréenne.

Le dépistage de la RD est donc basée sur la détection dans les photographies du fond d'œil de différentes lésions telles que les microanévrismes, les exsudats, les œdèmes maculaires, les hémorragies, etc. [122].

De nombreuses méthodes ont été proposées dans la littérature pour la détection de lésions rétiniennes. Certaines méthodes sont spécifiques à un type de lésion, tandis que d'autres peuvent détecter plusieurs lésions dans la même image. Une grande majorité de ces méthodes adopte une méthodologie en trois étapes [2, 164, 168] : i) le pré-traitement des images, ii) la détection de lésions potentielles (ou régions d'intérêt), et iii) la classification de ces régions. La figure 5.4 montre la procédure généralement adoptée pour la détection de lésions dans les images de fond d'œil.

1. **Pré-traitement :** Cette étape a plusieurs objectifs parmi lesquels la compensation de la grande variabilité de contraste et d'illumination des images de fond d'œil. En effet, en imagerie de fond d'œil, il est difficile d'obtenir une illumination uniforme de la rétine [2]. Des méthodes d'égalisation d'histogrammes ou de soustraction d'arrière plan sont employées pour réduire les effets d'illumination et améliorer le contraste des images [38, 49]. Une autre méthode basée sur une décomposition en valeurs singulières (SVD) est proposée par Adal *et al.* [6]. Le second objectif du pré-taitement est l'élimination des structures vasculaires de la rétine et la papille optique pour réduire le nombre de faux positifs lors de la détection de lésions. La détection des vaisseaux sanguins est en soi un problème important pour l'estimation de la qualité des images de fond d'œil [60] ou pour le diagnos-



(a) Image originale (b) Image normalisée

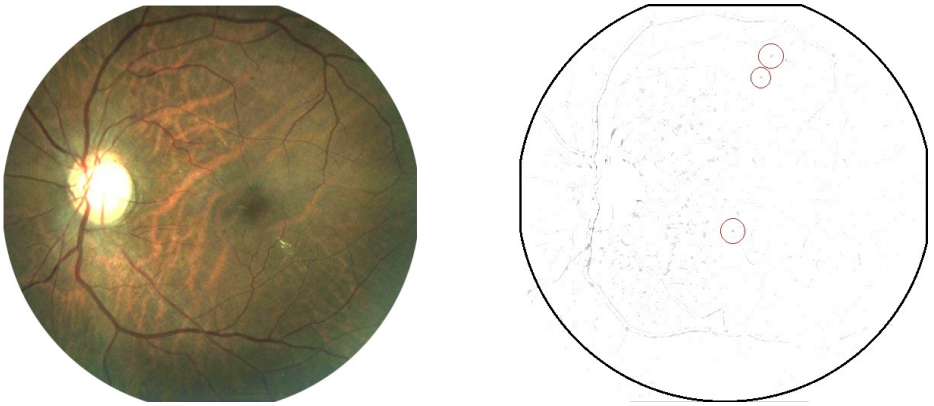
(c) Image originale (d) Vaisseaux sanguins détectés (e) Image sans les vaisseaux sanguins

FIGURE 5.5 – Pré-traitement des images de fond d'œil ; (a) et (b) amélioration du contraste ; (c)-(e) élimination des vaisseaux sanguins.

tic de certaines pathalogies (il est établit que la structure et le calibre des vaisseaux sont caractéristiques de certaines maladies cardiovasculaires [108]). Il existe donc un grand nombre d'algorithmes pour la segmentation des vaisseaux sanguins de la rétine [183, 158, 132, 153, 50]. La figure 5.5 montre des exemples de pré-traitements appliqués à des images de fond d'œil. Soulignons également que le canal vert de l'image couleur est généralement employé pour la détection de lésions, car il offre un meilleur contraste que les deux autres canaux.

2. **Détection de ROI :** La seconde étape consiste à détecter des régions d'intérêt (ROI) dans l'image, celles-ci pouvant correspondre aux lésions recherchées. Chaque lésion étant caractérisée par une certaine apparance, plusieurs approches sont utilisées. La plus simple est basée sur un seuillage de l'image suivi d'une application de filtres morphologiques pour obtenir les ROIs [155, 117]. En général, un seuillage adaptatif est employé [184], et certains auteurs utilisent la transformée de Hough pour détecter des régions circulaires (pour la détection de microanévrisme notamment) [1]. Enfin, les méthodes de segmentation telles que la croissance de région (*region growing*) ou le clustering sont aussi utilisées pour détecter les ROI [154, 115]. Un exemple de détection de ROIs par seuillage est présenté à la figure 5.6.

3. **Classification :** La dernière étape consiste à sélectionner parmi les ROIs détectées, celles qui correspondent à des lésions rétiniennes en utilisant une méthode de classification supervisée. Pour ce faire, on dispose d'un ensemble d'apprentissage comprenant des exemples de lésions manuellement segmentées (exemples positifs), ainsi que des régions ne correspondant pas à des lésions (exemples négatifs). Différents attributs, de forme, de couleur ou de texture, sont extraits de cet ensemble et une fonction de prédiction est estimée par minimisation d'une fonction d'erreur définie sur cet ensemble. Cette fonction de prédiction associe à chaque ROI



(a) Image originale (b) Détection de ROI

FIGURE 5.6 – Détection de régions d'intérêt ; (a) image originale ; (b) Détection de ROIs par seuillage : les vrais lésions sont indiquées par les cercles.

une classe ou étiquette {lésion, pas lésion}. Les méthodes de classification les plus couramment employées sont les plus proches voisins (K-NN) [117], les machines à support de vecteurs (SVM) [61, 168], le classifieur de Bayes (Naive Bayes) [7], les réseaux de neurones (Neural networks) [110] et les forêts aléatoires (Random Forest) [7].

5.1.3/ CONTRIBUTIONS

Nous apportons les contributions importantes suivantes à l'analyse d'images de fond d'œil :

1. **Classification semi-supervisée :** Les méthodes de classification supervisée nécessitent un nombre important de données annotées pour l'apprentissage. Or l'annotation manuelle d'images médicales est une tâche fastidieuse et sujette à erreur, même pour les spécialistes. Nous proposons donc une approche de classification qui ne nécessite qu'un nombre limité d'exemples annotés et nous l'appliquons avec succès à la détection de microanévrismes dans la section 5.2.

2. **Détection basée atlas :** La plupart des méthodes de détection d'exsudats nécessitent des étapes de pré-traitement pour éliminer les structures rétiniennes telles que les vaisseaux sanguins et la papille optique. Nous proposons, dans le section 5.3, une méthode de segmentation des exsudats basée sur un atlas, qui ne nécessite pas ces étapes de pré-traitement et qui permet une détection rapide et robuste des lésions.

3. **Discrimination automatique :** Les lésions, par exemple les exsudats et les druses, peuvent avoir une apparence très similaire mais caractériser des pathologies différentes. Cette ambiguïté est l'une des causes de faux positifs obtenus avec les méthodes de détection de lésions. Dans la section 5.4, nous proposons une méthode de discrimination d'images de fond d'œil en fonction du type de lésions présentes dans l'image en utilisant une représentation parcimonieuse des images.

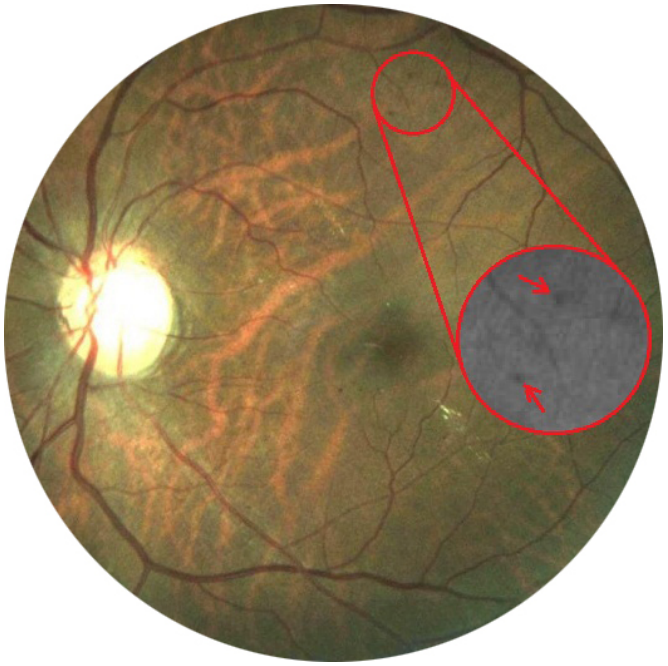


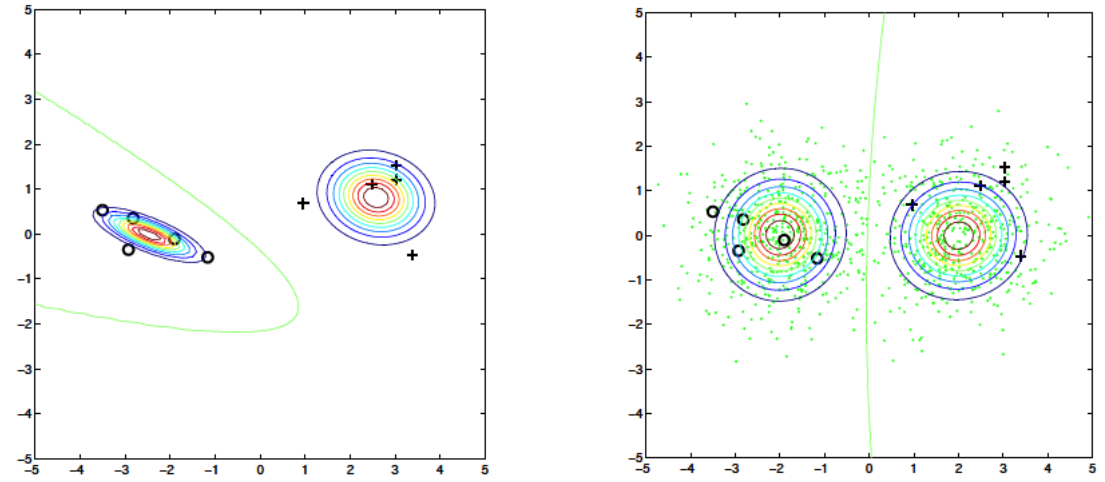
FIGURE 5.7 – Exemples de microanévrismes dans une image de fond d'œil. Les MAs sont indiqués par flêches rouges.

5.2/ UNE MÉTHODE SEMI-SUPERVISÉE POUR LA DÉTECTION DE MICROANÉVRISMES

La méthodologie générale décrite à la section 5.1.2, i.e. pré-traitement, détection de ROIs et classification, repose sur un apprentissage supervisé qui nécessite un nombre important d'exemples de lésions manuellement segmentées. En effet, plus on a d'exemples d'apprentissage, mieux l'algorithme d'apprentissage se comporte (meilleure généralisation). Néanmoins, si l'obtention d'un grand nombre d'images ne représente aucun problème, l'annotation manuelle des images est une tâche fastidieuse et sujette à erreur, même pour les spécialistes. Ceci est particulièrement vrai dans le cadre de l'analyse des images de fond d'œil, mais est vrai dans le domaine médical en général.

On dispose donc, très souvent, d'un grand nombre d'images non annotées qu'il faut utiliser pour apprendre un classifieur de manière automatique. Une première idée serait l'utilisation d'algorithmes de classification non supervisée, par exemple des algorithmes de type clustering, mais cette idée n'est pas pertinente dans le cadre de notre application. En effet, comme le montre la figure 5.7, les microanévrismes sont très difficiles à distinguer, même à l'œil, et une méthode entièrement non supervisée entraîne de nombreux faux positifs. L'approche que nous proposons est donc une méthode semi-supervisée qui permet de tenir compte de l'expertise des spécialistes, mais de manière minimale en annotant quelques exemples. Nous avons donc un ensemble d'apprentissage qui contient à la fois des exemples labellisés en nombre réduit, et des exemples non labellisés en nombre plus important.

Nous commençons par décrire brièvement le principe de l'apprentissage semi-supervisé dans la section 5.2.1, avant de présenter notre méthode de détection de microanévrismes (MAs) dans la section 5.2.2.



(a) Apprentissage avec peu de données labellisées

(b) Apprentissage avec peu de données labellisées et beaucoup de données non labellisées

FIGURE 5.8 – Illustration de l'apprentissage semi-supervisé ; (a) Frontière de décision obtenue avec peu de données labellisées ; (b) Frontière de décision en tenant compte des données non labellisées. Image reproduite d'après [187].

5.2.1/ APPRENTISSAGE SEMI-SUPERVISÉ

Nous présentons brièvement ici le principe de l'apprentissage semi-supervisé et les principaux algorithmes. Pour un état de l'art plus détaillé, nous renvoyons le lecteur à [32, 187].

Dans un problème d'apprentissage supervisé, nous disposons d'un ensemble d'apprentissage $\mathcal{X} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, où chaque $\mathbf{x}_i \in \mathbb{R}^d$ est un vecteur de caractéristiques, et chaque $y_i \in \mathcal{Y} = \{1, \ldots, C\}$ est la classe, ou le label, de l'exemple $\mathbf{x}_i$. L'objectif est d'estimer une fonction de décision $f : \mathcal{X} \rightarrow \mathcal{Y}$, qui minimise une fonction de coût sur l'ensemble d'apprentissage $\mathcal{X}$.

En apprentissage semi-supervisé, nous disposons d'un ensemble d'apprentissage qui contient à la fois des exemples labellisés $(\mathbf{x}_i, y_i) \in \mathcal{L}$, et des exemples non labellisés $\mathbf{x}_j \in \mathcal{U}$:

$$\mathcal{X} = \{\mathcal{L}, \mathcal{U}\}, \text{ avec } \mathcal{L} = \{\mathcal{X}_l, \mathcal{Y}_l\} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}, \ \mathcal{U} = \{\mathcal{X}_u\} = \{\mathbf{x}_j\}_{j=1}^{u}, \ l + u = N, \text{ et } u \gg l.$$

L'objectif est d'utiliser ces deux types de données pour estimer une fonction de décision f, avec l'hypothèse que les données non labellisées $\mathcal{U}$ peuvent aider à l'obtention d'une meilleure fonction f comme le montre l'exemple de la figure 5.8.

Toutefois, les données non labellisées n'aident pas toujours à une meilleure prédiction, et cela est possible seulement si la densité de probabilité $p(\mathbf{x})$ des données non labellisées contient une information utile pour l'estimation de la densité conditionnelle $p(y \mid \mathbf{x})$. Les algorithmes de classification semi-supervisée nécessitent que certaines conditions soient satisfaites [32] :

— **Smoothness :** Deux points proches dans une région de forte densité, ont des labels proches.
— **Cluster assumption :** Des points appartenant à un même cluster, ont une grande

probabilité d'appartenir à la même classe.
— **Density boundary :** La fonction de décision doit se trouver dans une région de faible densité.
— **Manifold assumption :** Dans l'espace de dimension d, les exemples d'apprentissage appartiennent à un sous-espace de dimension plus faible.

L'apprentissage semi-supervisé peut être soit *transductif*, soit *inductif* [187]. Dans le premier cas, on s'intéresse uniquement à la prédiction des labels des données non labellisées $\mathbf{x}_j \in \mathcal{U}$ de l'ensemble d'apprentissage. Dans le second cas, on s'intéresse à la prédiction de labels pour des données non disponibles au moment de l'apprentissage, i.e. à la bonne généralisation de l'algorithme.

Notre objectif étant la détection automatique de lésions dans des images de fond d'œil, nous nous intéresserons uniquement aux algorithmes inductifs tels que l'auto-apprentissage (self-training), le co-apprentissage (co-training) et les modèles de mélanges (mixture models) [32, 187].

5.2.1.1/ SELF-TRAINING

L'auto-apprentissage (self-training) consiste à entraîner un classifieur f avec les données labellisées $\mathcal{L}$. Le classifieur est ensuite utilisé pour prédire les labels des données incomplètes $\mathcal{U}$. Les données de $\mathcal{U}$ dont les labels sont prédits avec une forte probabilité sont ajoutées à $\mathcal{L}$, et le classifieur est ré-appris avec $\mathcal{L}$. La procédure est répétée jusqu'à satisfaire un critère d'arrêt.

L'auto-apprentissage est sans doute l'algorithme d'apprentissage semi-supervisé le plus simple à mettre en œuvre. On peut utiliser n'importe quel classifieur supervisé pour étiquetter les donnés labelisées $\mathcal{L}$ du moment que celui-ci fournit une probabilité pour la classe estimée.

La principale limitation de cette méthode est le risque de renforcement des erreurs de prédictions. Toutefois, ce risque peut être limité en employant un classifieur supervisé robuste, i.e. avec une confiance de prédiction élevée [187].

5.2.1.2/ CO-TRAINING

L'idée du co-apprentissage (co-training) repose sur l'hypthèse que s'il existe deux projections indépendantes d'un même espace de données, deux classifieurs entraînés selon ces deux projections doivent produire les mêmes labels pour les mêmes données. Ils peuvent donc mutuellement s'entraîner [23]. L'algorithme procède de la manière suivante :

- L'ensemble des attributs est divisé en 2 ensembles indépendants, i.e. $\mathbf{x}_i = [\mathbf{x}_i^{(1)}\ \mathbf{x}_i^{(2)}]^T$ et
$$\mathcal{L} = \{\mathcal{L}^{(1)}, \mathcal{L}^{(2)}\} = \{(\mathbf{x}_i^{(1)}, y_i),\ (\mathbf{x}_i^{(2)}, y_i)\}_{i=1}^{l}.$$
- Deux classifieurs sont entraînés séparément en utilisant respectivement $\mathcal{L}^{(1)}$ et $\mathcal{L}^{(2)}$
- Ces classifieurs sont utilisés pour étiqueter les données non labellisées $\mathcal{U}^{(1)}$ et $\mathcal{U}^{(2)}$

- Les données de $\mathcal{U}^{(1)}$ étiquetées avec une bonne confiance par le classifieur (1) sont ajoutées à $\mathcal{L}^{(2)}$ et les données de $\mathcal{U}^{(2)}$ étiquetées avec une bonne confiance par le classifieur (2) sont ajoutées à $\mathcal{L}^{(1)}$. Les classifieurs sont ré-entraînés avec ces nouvelles données.
- Lorsque l'apprentissage est terminé, les deux classifieurs sont combinés.

Le co-apprentissage est semblable à l'auto-apprentissage à la différence que les deux classifieurs s'entraînent mutuellement à partir de deux représentations différentes des données. Dans le cas des images, nous pouvons par exemple décrire une image par ses attributs couleur et par sa texture. Dans ce cas, nous représentons l'image de la manière suivante : $\mathbf{x}_i = [\mathbf{x}_i^{(1)} = \text{couleur} \mid \mathbf{x}_i^{(2)} = \text{texture}]^T$. Le classifieur (1) est entraîné avec les attributs de couleur et renforce le classifieur (2) qui est entraîné avec les attributs de texture, et vice-versa.

Le co-apprentissage repose sur les deux hypothèses suivantes :

1. **Redondance :** Chacun des deux sous-ensembles d'attributs peut être utilisé pour entraîner un classifieur.
2. **Indépendance conditionnelle :** Etant donné les labels, les deux sous ensembles d'attributs sont indépendants :

$$P(\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \mid y) = P(\mathbf{x}^{(1)} \mid y)\, P(\mathbf{x}^{(2)} \mid y).$$

La seconde hypothèse peut paraître forte, mais dans la pratique le co-apprentissage donne de bons résultats même lorsque cette hypothèse n'est pas satisfaite [23]. C'est aussi le cas d'algorithmes de classification supervisée tels que *Naive Bayes* qui sont basés sur la même hypothèse d'indépendance conditionnelle. Enfin, le co-apprentissage est l'exemple le plus représentatif d'une famille plus large de méthodes appelée *multi-view learning* [187].

5.2.1.3/ MIXTURE MODELS

Les algorithmes de mélanges de modèles (mixture models) sont des méthodes génératives qui estiment la densité de probabilité jointe des données labélisées et non labelisées :

$$P(\mathcal{X}_l, \mathcal{Y}_l, \mathcal{X}_u \mid \theta).$$

En supposant connue la forme de la densité de probabilité $P(\mathcal{X}, \mathcal{Y} \mid \theta)$, par exemple une Gaussienne, l'objectif est donc d'estimer les paramètres θ qui maximisent la densité jointe

$$\arg\max_{\theta} P(\mathcal{X}_l, \mathcal{Y}_l, \mathcal{X}_u \mid \theta) = \sum_{\mathcal{Y}_u} P(\mathcal{X}_l, \mathcal{Y}_l, \mathcal{X}_u, \mathcal{Y}_u \mid \theta),$$

à l'aide d'une méthode itérative telle que l'algorithme EM (Expectation-Maximization) [187].

Une procédure simple est la suivante :

— Répéter jusqu'à convergence
  - Entraîner un modèle $\hat{\theta}$ avec les données labellisées $\mathcal{L}$.
  - Utiliser ce modèle pour prédire les labels des données non étiquetées $\mathcal{U}$. Par exemple, dans le cas d'un problème de classification binaire $y \in \{+1, -1\}$,

$$\mathbb{E}[y_j] = (+1) \times P(y_j = +1 \mid \hat{\theta}, \mathbf{x}_j) + (-1) \times P(y_j = -1 \mid \hat{\theta}, \mathbf{x}_j)$$

- Ré-entraîner le modèle avec l'ensemble des données $\mathcal{L}$ et $\mathcal{U}$ avec les labels prédits.
- Utiliser le nouveau modèle $\hat{\theta}$ pour corriger les labels $\mathbb{E}[y_j]$ de $\mathcal{U}$.

Cette méthode dépend fortement de l'hypothèse sur la forme de la densité de probabilité $P(\mathcal{X}, \mathcal{Y} \mid \theta)$. Si on suppose un mélange de Gaussiennes, alors $\theta = \{w_i, \mu_i, \Sigma_i\}_{i=1}^{K}$, où w_i est le poids associé à chaque Gaussienne de moyenne μ_i et de matrice de covariance Σ_i.

5.2.2/ DÉTECTION DE MICROANÉVRISMES

Les microanévrismes (MAs) sont les premiers signes visibles de la RD et apparaissent sous forme de points rouges de petite taille [122]. La figure 5.7 montre une image de fond d'œil avec des MAs manuellement annotés. Comme on peut le voir sur cette figure, il très difficile, même à l'œil de distinguer les MAs des autres structures présentes sur la rétine. La détection automatique des MAs est donc une tâche difficile et la plupart des méthodes génère un nombre élevé de faux positifs. Nous décrivons dans cette section notre approche de détection de MAs dans les images de fond d'œil. Elle suit la procédure générale décrite plus haut, à la section 5.1.2, et comporte les 3 étapes : i) pré-traitement, ii) détection de ROIs et iii) classification.

Pour la première étape, nous employons la méthode d'amélioration de contraste décrite dans [6] basée sur une SVD. Pour la seconde étape, nous proposons une approche inspirée de la détection de points d'intérêt en vision par ordinateur et basée sur une analyse multi-échelle de l'image [93, 18]. Enfin, pour la troisième étape, nous proposons une approche semi-supervisée pour pallier la difficulté d'obtention d'un grand nombre d'exemples manuellement annotés.

5.2.2.1/ DÉTECTION DE ROIS

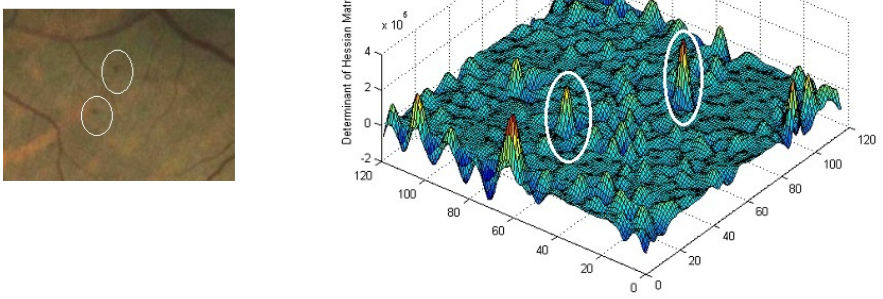
La détection de ROIs est une étape cruciale de toute méthode de détection de lésions dans les images de fond d'œil. En effet, il est souhaitable de réduire le nombre de régions détectées, pour réduire les faux positifs, tout en gardant toutes les régions qui correspondent à des lésions, pour réduire le nombre de faux négatifs.

Nous exploitons la Hessienne calculée en chaque pixel de l'image pour détecter les régions noires et circulaires. La Hessienne en un pixel $\mathbf{x} = (x, y)$ est définie par :

$$H(\mathbf{x}; \sigma) = \begin{bmatrix} I_{xx}(\mathbf{x}; \sigma) & I_{xy}(\mathbf{x}; \sigma) \\ I_{xy}(\mathbf{x}; \sigma) & I_{yy}(\mathbf{x}; \sigma) \end{bmatrix}, \tag{5.1}$$

où $I_{xx}(\mathbf{x}; \sigma) = I(\mathbf{x}) * G_{xx}(\sigma)$ est la dérivée seconde de l'image calculée selon l'axe x avec un filtre Gaussian de paramètre σ.

En analysant les valeurs propres de H, λ_1 et λ_2, il est possible de détecter les régions correspondant à des MAs potentiels. En particulier, nous constatons que les MAs correspondent généralement aux régions où le déterminant de H est élevé comme le montre l'exemple de la figure 5.9. De plus, comme les MAs sont de forme circulaire, il faut $\lambda_1 > 0$, $\lambda_2 > 0$ et $\lambda_1 \approx \lambda_2$. L'algorithme complet de détection des ROIs est décrit dans le tableau 5.1, et la figure 5.10 montre un exemple de détection de ROIs avec cet algorithme.



(a) Deux exemples de microanévrismes

(b) Déterminant de la matrice Hessienne, pour $\sigma = 2$

FIGURE 5.9 – Réponse de l'opérateur Hessienne. Les microanévrismes et leurs réponses sont indiqués par les ellipses blanches.

- Etant donné une image I
  - Appliquer les méthodes de pré-traitement
  - Calculer la Hessienne H, Eq. (5.1)
  - Calculer λ_1 et λ_2, et calculer leur rapport et le déterminant de H

    $A_r = \frac{|\lambda_1|}{|\lambda_2|}, \qquad |H| = \lambda_1.\lambda_2$
- Déterminer la carte des candidtas I_{ROIs} comme suit :

  $I_{ROIs} = |H| > Th_1 \bigcap A_r < Th_2 \bigcap \lambda_1 > 0 \bigcap \lambda_2 > 0$

TABLE 5.1 – Algorithme de détection de ROIs pour la détection de microanévrismes : Les valeurs suivantes sont utilisés pour les seuils ; $Th_1 = 3 \times 10^4$ et $Th_2 = 2$.

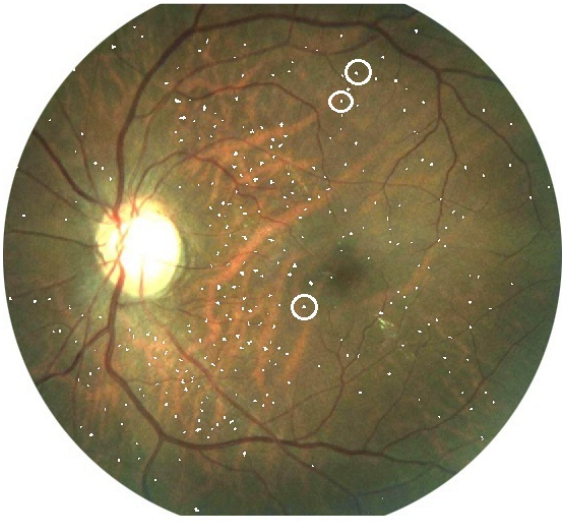
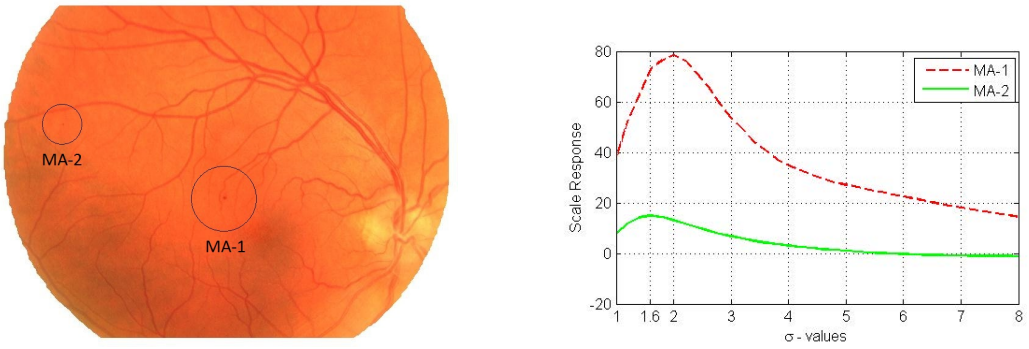


FIGURE 5.10 – Exemples de régions d'intérêt détectées en utilisant l'algorithme décrit dans le tableau 5.1. Les vrais MAs sont indiqués par des cercles blancs.

Le paramètre le plus important de notre méthode de détection de ROIs est le paramètre d'échelle σ, i.e. la taille de la Gaussienne utilisée pour le calcul des dérivées. On peut



(a) Deux exemples de microanévrismes

(b) Réponse de la Hessienne dans l'espace échelle

FIGURE 5.11 – Estimation de l'échelle locale des microanévrismes ; la réponse maximale de l'opérateur Hessienne, Eq. 5.2, indique l'échelle locale du MA.

utiliser une valeur fixe, mais cela ne tient pas compte, d'une part, de la variabilité de la taille des MAs (leur taille varie de 20 à 100 microns), et, d'autre part, de la variabilité de la résolution des images de fond d'œil acquises avec des appareils différents. Il faut donc estimer le meilleur paramètre σ pour chaque ROIs. Nous utilisons la méthode de détermination de l'échelle proposée par Lindeberg [90] et largement utilisée dans le domaine de la vision par ordinateur notamment pour la détection de points d'intérêt [93, 18]. Cette approche consiste à trouver le maximum local de la Hessienne dans l'espace échelle :

$$(\mathbf{x}, \hat{t}) = \arg\max_{t}(|H(\mathbf{x}, t|), \tag{5.2}$$

avec $|H(\mathbf{x}, t|$ le déterminant de H et $t = \sigma$.

Notons que la variable d'optimisation est t, car la position $\mathbf{x}$ est fixée. La valeur optimale $\hat{t}$ définie l'échelle local de la ROI au pixel $\mathbf{x} = (x, y)$.

La figure 5.11 illustre cette procédure d'estimation de l'échelle locale pour deux ROIs correspondant à des MAs.

5.2.2.2/ DÉTECTION DE MAS

Une fois les régions d'intérêt (i.e. les lésions potentielles) détectées, l'étape finale de la méthode consiste à distinguer les régions correspondant réellement à des microanévrismes. Pour ce faire, nous adoptons une méthode de classification semi-supervisée qui permet d'apprendre de manière efficace un classifieur avec peu de données annotées.

Données d'apprentissage Pour l'apprentissage nous disposons d'une dizaine d'images manuellement annotées par un ophtalmologiste (Dr. Chaum de l'université du Tennesse aux USA) et d'environ 300 images non annotées. Les images annotées contiennent au total 80 MAs, tandis que le nombres de MAs dans les images non annotées est inconnu. Ces images sont pré-traitées et les ROIs sont détectées selon la méthode décrite à la section 5.2.2.1.

Les ROIs extraites de la dizaine d'images annotées représentent les données d'apprentissage labellisées, 80 exemples positifs et plusieurs centaines d'exemples négatifs. Les

ROIs extraites des 300 images non annotées représentent les données d'apprentissage non labellisées, plusieurs milliers.

Caractéristiques utilisées Nous représentons chaque ROI par un vecteur de caractéristiques $\mathbf{x}$ correspondant à divers attributs. Le premier ensemble d'attributs est calculé à partir de la réponse de l'opérateur Hessienne. En particulier, nous utilisons la moyenne et le max des valeurs propres et du déterminant de H :

$$\mathbf{x}_{Hessian} = \{\bar{\lambda_1}, \lambda_1^{\max}, \bar{\lambda_2}, \lambda_2^{\max}, |\bar{H}|, |H|^{\max}\}.$$

Ce premier ensemble d'attribut décrit la forme de la ROI.

Le deuxième ensemble de caractéristiques décrit la texture de la ROI en employant le descripteur SURF qui calcule la distribution locale des orientations du gradient dans la région [18] : $\mathbf{x}_{SURF}$.

Enfin, le dernier ensemble de caractéristiques est obtenu en appliquant une transformée de Radon à la ROI [64, 63]. Pour assurer une invariance à la luminosité, chaque ROI est normalisée de manière à avoir une moyenne nulle et un écart-type égal à un. La transformée de Radon est calculée avec des angles de projection $\theta \in [0, 180)$, avec un pas de 5^o. Les attributs retenus sont :

$$\mathbf{x}_{Radon} = \{\max(R_\mu(ROI), \sigma_{R_\mu(ROI)}, \max(R_\sigma(ROI))\},$$

où $R_\mu(ROI)$ et $R_\sigma(ROI)$ sont respectivement la moyenne et l' écart-type des transformées de Radon R calculées selon les différentes orientations θ.

L'ensemble final de caractéristiques pour chaque ROI est donc $\mathbf{x} = [\mathbf{x}_{Hessian}\, \mathbf{x}_{SURF}\, \mathbf{x}_{Radon}]^T$.

Méthodes d'apprentissage Nous comparons deux approches d'apprentissage semi-supervisé : l'auto-apprentissage et le co-apprentissage. Comme mentionné à la section 5.2.1, chacune de ces deux approches utilisent un ou des classifieurs supervisés comme modèles de base. Nous avons utilisé 4 classifieurs communément employés dans la littérature : les k plus proches voisins (kNN), le classifieur de Bayes (Naïve Bayes), les forêts aléatoires (random forest, RF) et les machines à vecteurs de support (SVM).

Données de test La performance de la méthode de détection de MAs proposée est évaluée avec deux bases de données. La première est une base publique développée par l'Université de l'Iowa dans le cadre du challenge *Retinopathy Online Challenge* (ROC) [116]. Elle comporte 50 images d'apprentissage dans lesquelles les MAs sont manuellement annotés, et 50 images de test dont l'annotation n'est pas disponible. Il faut soumettre ses résultats de détection sur ces images aux organisateurs qui évaluent la performance de la méthode. Dans nos expériences, nous utilisons l'ensemble des 100 images de la base ROC comme images de test pour évaluer notre algorithme de détection de MAs.

Une autre base de donnée fournie par le département d'ophtalmologie de l'université de Tennesse (UTHSC), avec lequel nous travaillons depuis plusieurs années, permet d'évaluer la capacité de notre approche à détecter des patient atteints de la RD. Cette base comporte 50 images dont 37 de patients identifiés comme malades et 13 de patients sains. Le but ici n'est pas de détecter précisément les MAs dans les images, car leurs

Approche	Méthodologie	Sensibilité	FPs/Image
Spencer et al. [155]	Top-hat transform	12%	20.3
Abdelazeem [1]	Circular Hough-transform	28%	505.85
Walter et al. [172]	Diameter closing	36%	154.42
Zhang et al. [184]	Multiple-Gaussian mask	33%	328.3
Lazar et al. [85]	Croos-section profile	48%	73.94
Notre méthode	**Hessian operator**	**44.64%**	**35.20**

TABLE 5.2 – Comparaison des méthodes de détection de ROIs avec la base de données ROC. Notons que les valeurs données dans ce tableau, à l'exception de celle obtenue par notre méthode, sont extraites de [84].

nombre et positions ne sont pas connus, mais de détecter si un patient est atteint de la RD ou pas sur la base des MAs détectés.

Résultats Dans une première expérience, nous comparons notre méthode de détection de ROIs avec diverses méthodes de la littérature sur la base de donnée d'apprentissage du challenge ROC. Les résultats rassemblés dans le tableau 5.2 montrent la pertinence de notre approche basée sur une détection de régions circulaires en analysant la réponse de l'opérateur Hessienne. En effet, notre méthode de détection de ROIs réduit le nombre de faux positifs moyens par image d'environ 52% par rapport à la meilleure approche proposée par Lazar et Hajdu [86, 85], tout en maintenant une sensibilité comparable. La réduction du nombre de faux positifs à l'étape d'extraction des ROIs est importante pour facilier l'étape de classification des MAs.

Ensuite, nous avons utilisé l'approche d'auto-apprentissage pour entraîner chacun des 4 classieurs, kNN, SVM, RF et Naïve Bayes, avec toutes les données d'apprentissage. Chaque classifieur ainsi entraîné est testé avec les données du challenge ROC dont la vérité terrain est connue. Les résultats obtenus sont décrits par une courbe FROC (Free-response ROC curve) qui représente la fraction de lésions correctement détectées dans les images en fonction du nombre moyen de faux positifs (FP) détectés par image. Cette représentation est couramment utilisée pour évaluer les algorithmes de détection de lésions [109]. Comme le montre la figure 5.12, les classifieurs SVM et kNN sont les plus performants.

Dans un deuxième temps, les deux meilleurs classifieurs, kNN et SVM, sont utilisés dans une approche de co-apprentissage. Les résultats de la figure 5.13 montre que le co-apprentissage améliore légèrement les résultats de détection, notamment pour des taux moyens de faux positifs inférieurs à 1 FP/image.

Pour la suite des expériences nous utilisons la méthode de co-apprentissage et adoptons le classieur SVM entraîné avec l'ensemble des données d'apprentissage. Nous avons participé au challenge ROC, http ://webeye.ophth.uiowa.edu/ROC/, en soumettant nos résultats de détection obtenus sur l'ensemble test. Les organisateurs du challenge évaluent chaque méthode en calculant un score (competition performance score, CPM) qui est calculé comme la sensibilité moyenne à différents taux de FPs ou points d'opération (operating points, OP). Deux ensembles de points d'opérations sont utilisés pour évaluer les algorithmes à des très faibles taux de FPs, ainsi qu'à des taux de FPs moyens : $OP_1 = \{1/8, 1/4, 1/2, 1, 2, 4, 8\}$ et $OP_1 = \{2, 4, 8, 12, 16, 20\}$. Le tableau 5.3 montre les résultats obtenus par notre méthode de détection de MAs et la comparai-

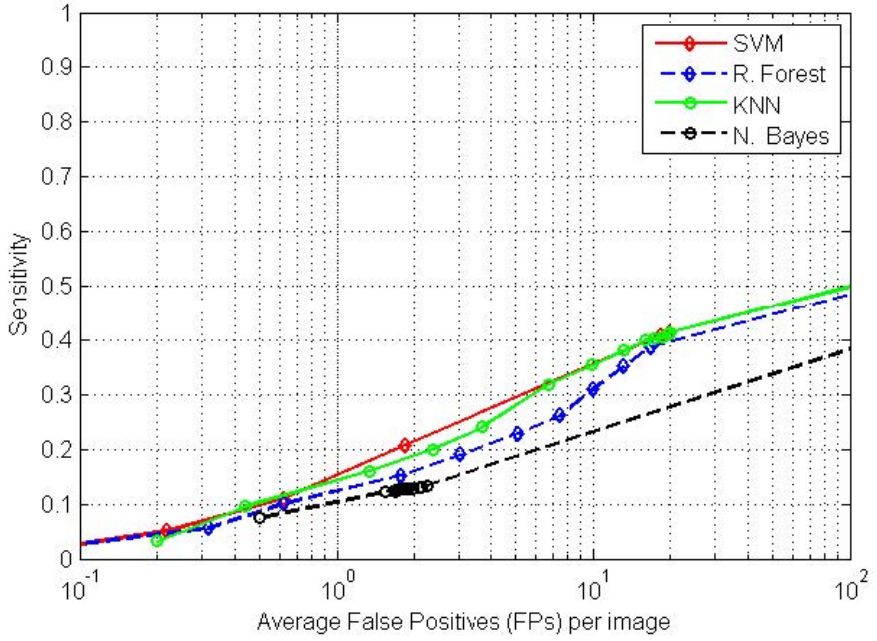


FIGURE 5.12 – Résultats obtenus avec la méthode d'auto-apprentissage. Notons que l'axe des abscisses est dans une échelle logarithmique.

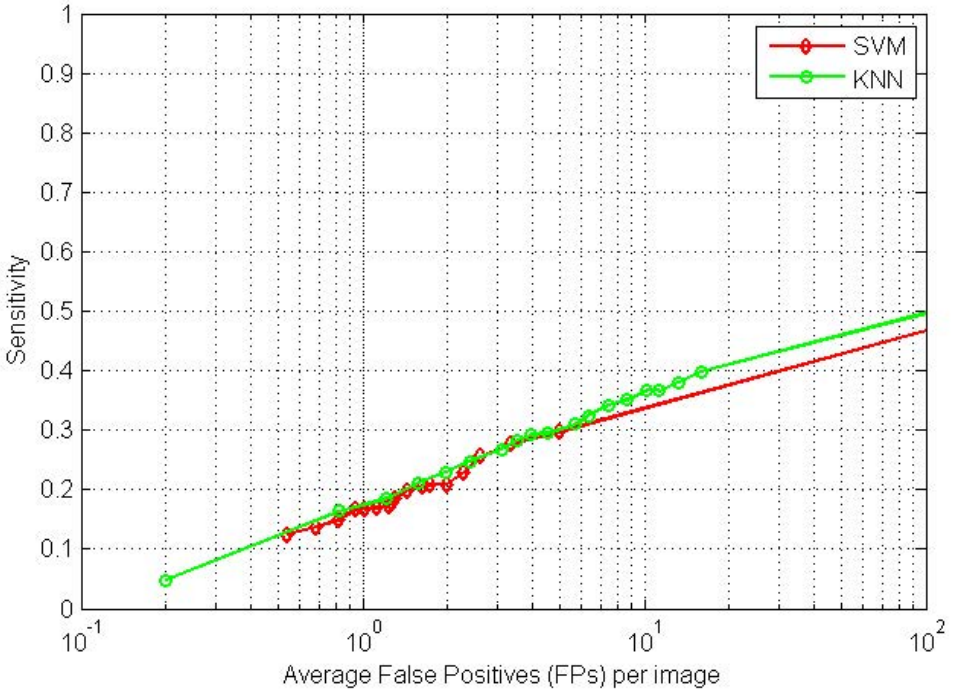


FIGURE 5.13 – Résultats obtenus avec la méthode de co-apprentissage. Notons que l'axe des abscisses est dans une échelle logarithmique.

son avec différentes approches proposées dans la littérature. Comme on peut le voir, notre approche obtient des résultats comparables aux meilleurs approches bien que ne nécessitant qu'une dizaine d'image manuellement annotées pour l'apprentissage. Ces

Nom du groupe	CMP at OP_1	CMP at OP_2
Waikato	0.206	0.335
IRIA-Group	0.264	0.503
Fujita Lab	0.310	0.468
GIB Valladolid	0.322	0.514
OKmedical II	0.369	0.502
ISMV	0.375	0.469
LaTIM	0.381	0.565
Niemeijer	0.395	0.558
DRSCREEN	0.434	0.666
Notre méthode	**0.364**	**0.538**

TABLE 5.3 – Comparaison de différentes méthodes de détection de MAs avec la base de données ROC. Les scores sont calculé à deux points d'opération différents : $OP_1 = \{1/8, 1/4, 1/2, 1, 2, 4, 8\}$, $OP_1 = \{2, 4, 8, 12, 16, 20\}$.

résultats montrent l'intérêt d'une approche semi-supervisée pour la détection de lésion dans les images de fond d'œil.

La seconde expérience a pour but d'évaluer l'efficacité de notre méthode de détection de MAs dans une application de dépistage à grande échelle, où les patients peuvent rapidement avoir un premier diagnostic à distance par exemple. Le but ici n'est pas de détecter précisément les MAs dans les images, mais de détecter si un patient est atteint de la RD ou pas sur la base des MAs détectés. Les patients ainsi détectés peuvent ensuite être orientés vers un spécialiste pour un examen plus approfondi. Nous utilisons dans cette expérience la base de données fournie par UTHSC et les résultats de la figure 5.14 montrent la performance de notre méthode par apprentissage semi-supervisé. Soulignons que le même système de détection entraîné pour les données du challenge ROC est ici utilisé. La méthode de co-appprentissage avec un classifieur SVM atteint une sensibilité de plus 80% pour une spécificité de 92%.

5.2.3/ CONCLUSION

Dans cette section, nous avons montré l'intérêt d'une approche semi-supervisée pour la détection de microanévrismes dans des images de fond d'œil. Les résultats obtenus avec deux bases de données différentes montrent que notre méthode, entraînée avec seulement une dizaine d'images annotées, obtient des résultats comparables avec les meilleurs approches de la littérature qui nécessitent des ensembles d'apprentissage très importants. D'autre part, notre méthode de détection de ROIs basée sur une analyse des structures locales de l'image dans un espace échelle, permet de réduire de manière significative le nombre de faux positifs, ce qui améliore les résultats de la classification.

5.3/ UNE MÉTHODE DE DÉTECTION D'EXSUDATS BASÉE ATLAS

Dans cette section, nous nous intéressons à la détection des œdèmes maculaires, i.e. une augmentation d'épaisseur de la macula par accumulation de liquide. Dans les images de fond d'œil, les œdèmes maculaires sont caractérisés par la présence d'exsudats qui

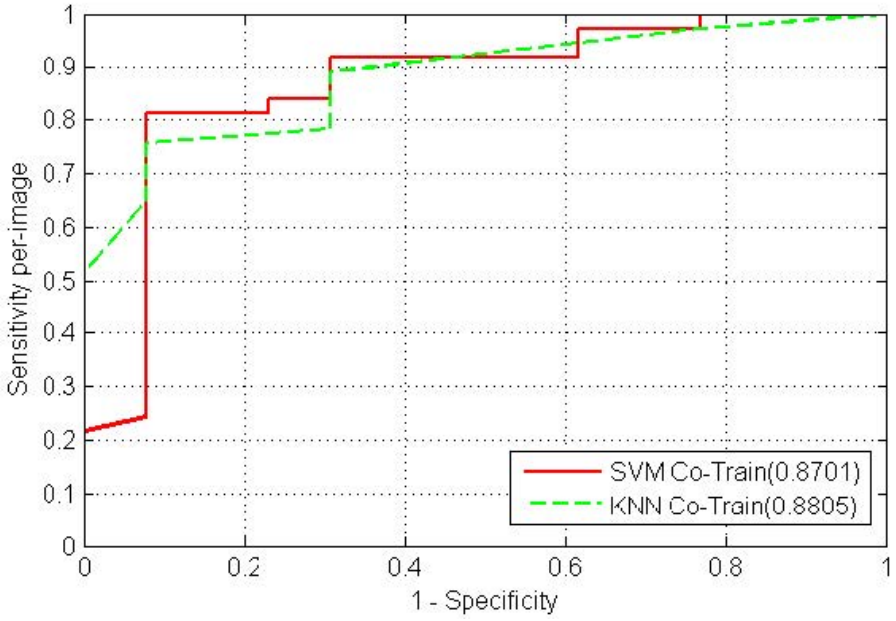
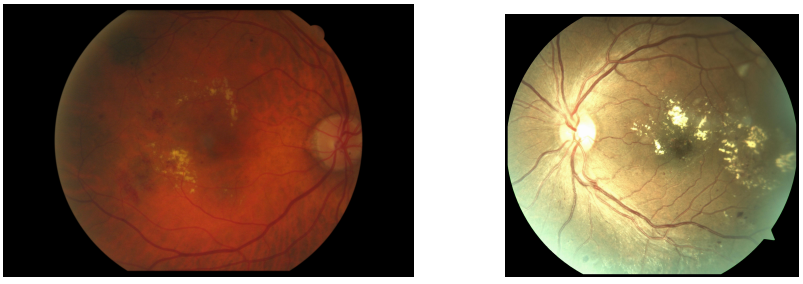


FIGURE 5.14 – Résultats de détection de la RD avec la base de données de UTHSC.



(a) (b)

FIGURE 5.15 – Images de fond d'œil présentant des exsudats (dépôts jaunâtres).

sont des accumulations de lipoprotéines dans l'épaisseur de la rétine, et qui apparaissent sous forme de dépôts jaunâtres. La figure 5.15 montre des exemples d'images de fond d'œil présentant des exsudats.

Les méthodes de détection d'exsudats proposées dans la littérature [156, 154, 115, 55, 62, 61], adoptent généralement la méthodologie classique décrite à la section 5.1.2, i.e. pré-traitement, détection de ROIs et classification. L'étape de pré-traitement est ici particulièrement importante, notamment la détection de la papille optique qui a une apparence similaire à celle des lésions que l'on souhaite détecter. La papille optique est généralement détectée et éliminée manuellement [62, 74], ce qui est laborieux, ou en utilisant des méthodes de segmentation telles que les contours actifs [92], ce qui est sujet à erreur et coûteux en temps de calcul.

Nous proposons une méthode de segmentation des exsudats dans les images de fond d'œil basée sur un atlas, et qui ne nécessite pas les étapes de pré-traitement telles que la segmentation des vaisseaux sanguins et la détection de la papille optique. En effet, l'atlas nous permet d'avoir une image de référence qui contient les différentes struc-

tures de la rétine (papille optique, macula, principaux vaisseaux sanguins). En recalant une image test avec cette image de référence, ces structures rétiniennes sont facilement supprimées. De plus, une simple différence entre l'image de référence et l'image test recalée permet de faire ressortir les possibles lésions, et une étape finale de post-traitement permet de détecter les lésions d'intérêt.

Nous commençons par présenter la méthode de création de l'atlas dans la section 5.3.1, puis nous présentons la méthode de détection des exsudats dans la section 5.3.2.

5.3.1/ CRÉATION D'UN ATLAS

L'utilisation d'un atlas consiste à apparier une image de référence (l'atlas) et une image à traiter en utilisant une méthode de recalage. On superpose ainsi les informations contenues dans l'image de référence et l'image à segmenter, en particulier les structures anatomiques de la rétine telles que la papille optique, la macula et les principaux vaisseaux sanguins.

Données utilisées Nous créons deux atlas, un pour chaque œil, en utilisant un ensemble de 400 images de bonne qualité extraites de la base de données TRIAD (Tele-medical Retinal Image Analysis and Diagnosis) [89]. Cette base de données comporte environ 5200 images de fond d'œil collectées entre février 2009 et août 2011 dans plusieurs cliniques de la région du sud-est des USA. Chacune de ces images comporte un indice de qualité qui mesure la bonne visibilité des structure rétiniennes dans l'image [60]. Cet ensemble de 400 images constitue l'ensemble d'apprentissage utilisé pour créer notre atlas. Il contient uniquement des images de patients sains.

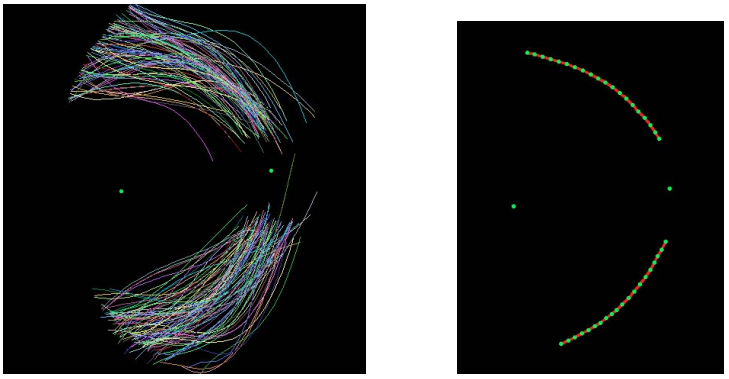
Système de coordonnées de référence Pour obtenir l'image de référence (l'atlas), il faut recaler toutes les images d'apprentissage dans un référentiel commun. Nous utilisons comme référentiel le système de coordonnées défini par les positions du centre de la papille optique, du centre de la macula, et les deux vaisseaux sanguins principaux.

Plus précisémment, nous commençons par détecter la position de la papille optique et de la macula, et par détecter les vaisseaux sanguins dans toutes les images d'apprentissage. Ensuite, en considérant comme points de référence les positions moyennes de la papille optique (p_{oc}) et de la macula (p_{mc}) dans ces images, nous recalons tous les vaisseaux sanguins en estimant une transformation rigide (rotation d'angle θ et translation de vecteur $T = [t_x, t_y]^T$) de la manière suivante :

$$\arg\min_{\theta,T} \left(\sum_{i=1}^{N} \|p_{oc} - \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} (p_{oc}^i - T)\|^2 + \sum_{i=1}^{N} \|p_{mc} - \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} (p_{mc}^i - T)\|^2 \right), \tag{5.3}$$

avec p_{oc}^i et p_{mc}^i les positions de la papille optique et de la macula dans l'image i, et N le nombre d'image.

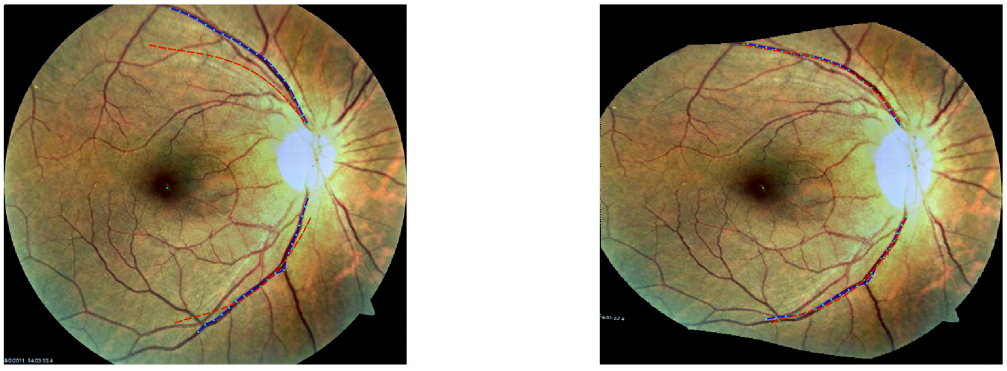
Le résultat de l'application de cette transformation rigide est illustré par l'image de la figure 5.16(a). Pour chaque vaisseaux ainsi recalé, nous sélectionnons $M = 20$ points uniformément répartis le long du vaisseau, et appliquons une ACP à l'ensemble de ces points. Les axes principaux obtenus par ACP définissent les deux vaisseaux principaux utilisés dans le sytème de coordonnées de référence. La figure 5.16(b) montre le système de référence qui sera utilisé par la suite pour recaler les images.



(a) Tous les vaisseaus sanguins détectés dans les 400 images

(b) Système de coordonnées de référence

FIGURE 5.16 – Système de coordonnées de référence de l'atlas.



(a) Image à recaler

(b) Image recalée avec TPS

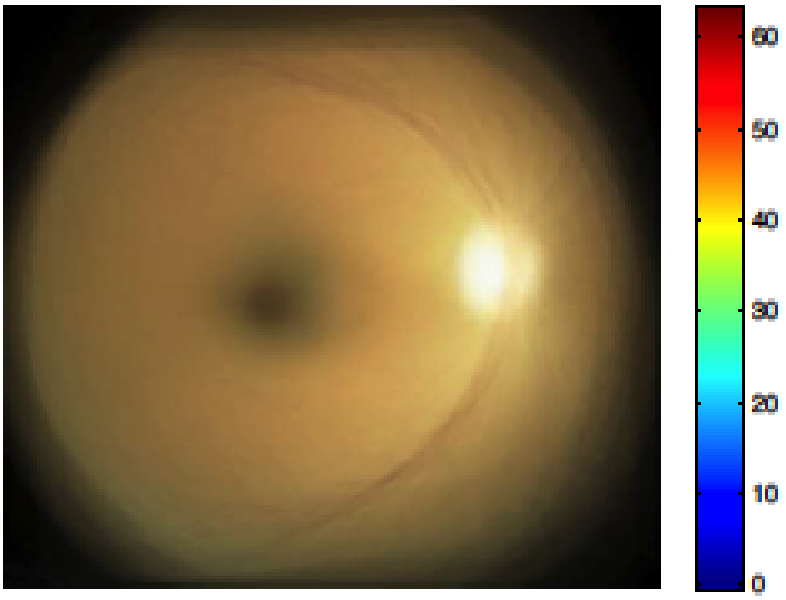
FIGURE 5.17 – Illustration du recalage des images ; (a) la courbe rouge correspond au référentiel commun et la courbe bleue aux vaisseaux détectés dans l'image ; (b) après recalage, les vaisseaux détectés dans l'image sont alignés avec les axes de référence.

Recalage des images et obtention de l'atlas Une fois le référentiel commun obtenu, chaque image d'apprentissage est recalée de manière fine en utilisant la méthode des TPS (thin-plate splines) [24] comme illustrée par la figure 5.17. Enfin, nous obtenons notre image de référence (atlas) en prenant la moyenne de toutes les images recalées. La figure 5.18 montre l'atlas obtenu pour l'œil droit. Comme on peut le voir, celui-ci contient les principales structures de la rétine (figure 5.18(b)) ainsi que la distribution chromatique moyenne de la population étudiée (figure 5.18(a)).

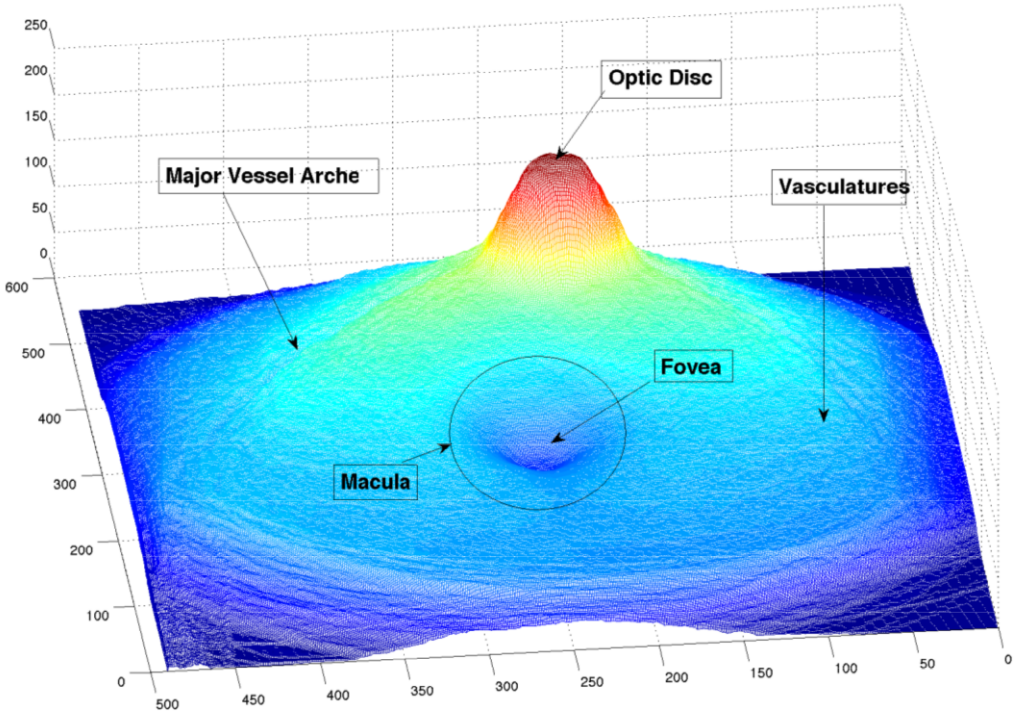
5.3.2/ DÉTECTION D'EXSUDATS

L'atlas créé avec des images de patients sains est utilisé pour détecter des lésions dans des images tests de la manière suivante :

— L'image test est d'abord recalée dans le référentiel commun (figure 5.17(b)) et nous calculons la différence entre l'image transformée et l'image de référence. Cette opération permet d'éliminer les structures anatomiques (papille optique, ma-



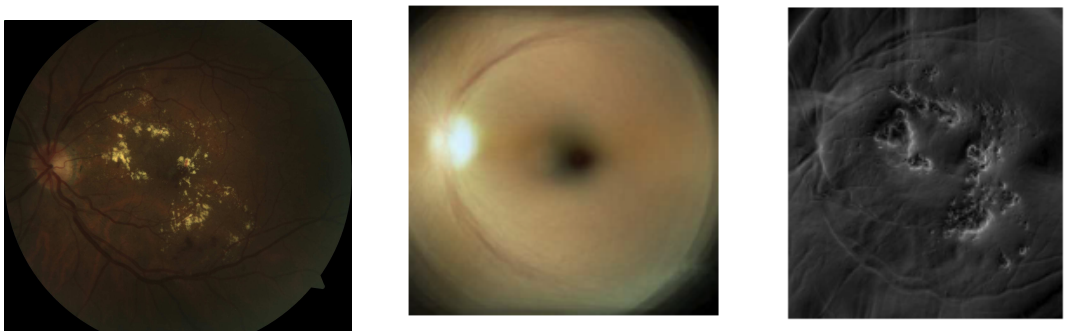
(a) Distribution chromatique moyenne



(b) Différentes structures anatomiques

FIGURE 5.18 – Image de référence obtenue avec indication des structures anatomiques.

cula, vaisseaux sanguins) et fait ressortir les possibles lésions comme l'illustre la figure 5.19.



(a) Image test (b) Image de référence (c) Image de différence

FIGURE 5.19 – Détection de lésions par recalage et soustraction avec l'image de référence.

— Enfin, les régions correspondant aux exsudats sont détectées en analysant la réponse d'un opérateur de contour pour tenir compte du fait que les exsudats présentent des contours prononcés.

Données test Pour valider notre méthode de détection d'exsudats, nous employons la base de données publique HEI-MED [61] utilisée par plusieurs auteurs dans la littérature. Elle comporte 169 images de fond d'œil de patients sains et de patients avec des œdèmes maculaires, et les images comportant des exsudats sont manuellement annotées.

Méthodes de détections L'image de différence obtenue ci-dessus peut être simplement seuillée pour obtenir les exsudats. Cependant, la différence d'une image test avec l'image de référence ne fait pas uniquement ressortir les exsudats mais toutes les structures qui diffèrent par la forme ou par l'apparence du model de référence. Aussi, pour réduire le nombre de faux positifs, nous employons deux opérateurs qui accentuent les contours prononcés des exsudtats pour les distinguer d'autre régions. Nous utilisons les opérateur de Riesz [167] et de Kirsch [82].

Nous comparons également notre approche avec les méthodes proposées par Sanchez et al. [144], Sopharak et al. [154] et Giancardo et al. [62] qui sont toutes des méthodes basées sur un seuillage et des règles de décision (aucun classifieur n'est utilisé). Les deux dernières méthodes utilisent également l'opérateur de Kirsch pour distinguer les exsudats.

Résultats Les résultats obtenus sont décrits par des courbes FROC (Free-response ROC curve) et les aires sous les courbes (AUC). Notons que cette évaluation est faite au niveaux des régions et non des pixels, i.e. une détection est considérée comme correcte (vrai positif) si une partie de la région détectée a une intersection non vide avec une région de l'image de la vérité terrain.

Nous commençons par évaluer les différentes méthodes de post-traitement (différents opérateurs de contour). Les résultats, donnés dans le tableau 5.4 montrent que l'emploi de chaque opérateur de contour améliore la détection des exsudats mais cette amélioration est très faible. Par contre, l'usage simultané des deux opérateurs offre une amélioration notable des résultats. Nous passons d'une valeur d'AUC de 0.7612 pour un

Méthode	AUC
Seuil sans post-traitement	0.7612
Kirsch + seuillage	0.7832
Riesz + seuillage	0.7866
Kirsch & Riesz + seuillage	0.8258

TABLE 5.4 – Comparaison de différentes méthodes de post-traitement pour la détection d'exsudats. La valeur AUC indique l'aire sous la courbe FROC.

Méthode	AUC
Sopharak [154]	0.58
Sanchez [144]	0.80
Giancardo [62]	0.83
Notre méthode	**0.83**

TABLE 5.5 – Comparaison de différentes méthodes de détection d'exsudats. La valeur AUC indique l'aire sous la courbe FROC.

seuillage direct de l'image de différence, à une valeur de 0.8258 pour un seuillage après application des deux opérateurs. L'intérêt de combiner les deux opérateurs est dû au fait que l'opérateur de Riesz permet de détecter les contours dans toutes les directions mais a une large bande passante qui le rend sensible au bruit généré par les pixels voisins, tandis que l'opérateur de Kirsch a une bande passante limitée mais ne détecte les contours que dans un nombre limité de directions. La combinaison permet donc d'accentuer les contours dans toutes les directions tout en limitant le nombre de fausses détections. La figure 5.20 montre des exemples de détection d'exsudats dans différentes images de fond d'œil.

Enfin, nous comparons notre méthode basée atlas à trois méthodes de la littérature et les résultats sont présentés sur la figure 5.21 et rassemblés dans le tableau 5.5. Comme on peut le voir, la méthode proposée est comparable aux approches de la littéraure. Elle est légèrement supérieure à la méthode de Sanchez [144] et a des résultats similaires à ceux obtenus par la méthode de Giancardo [62]. Toutefois, notre méthode de détection ne nécessite pas les étapes de pré-traitement telles que la détection et l'élimination des vaisseaux sanguins ou de la papille optique. Notons également que ces résultats de comparaison sont à prendre avec quelques précautions. En effet, notre méthode détecte les exsudats dans le référentiel de l'atlas et il faut ensuite reprojeter les régions détectées dans l'image de fond d'œil pour comparaison avec la vérité terrain. Cette étape nécessite des interpolations qui peuvent introduire des erreurs dans le calcul des FPs et donc la génération de la courbe FROC.

5.3.3/ CONCLUSION

Dans cette section, nous avons présenté une méthode de détection d'exsudats dans les images de fond d'œil basée sur un atlas. Cette méthode facilite la détection des lésions en recalant l'image test avec une image de référence et en calculant simplement la distance entre ces deux images. Les résultats obtenus avec la base de données publique HEI-MED montrent que cette approche obtient des résultats au moins comparables à ceux des meilleures méthodes de la littérature. Notons enfin que cette méthodologie

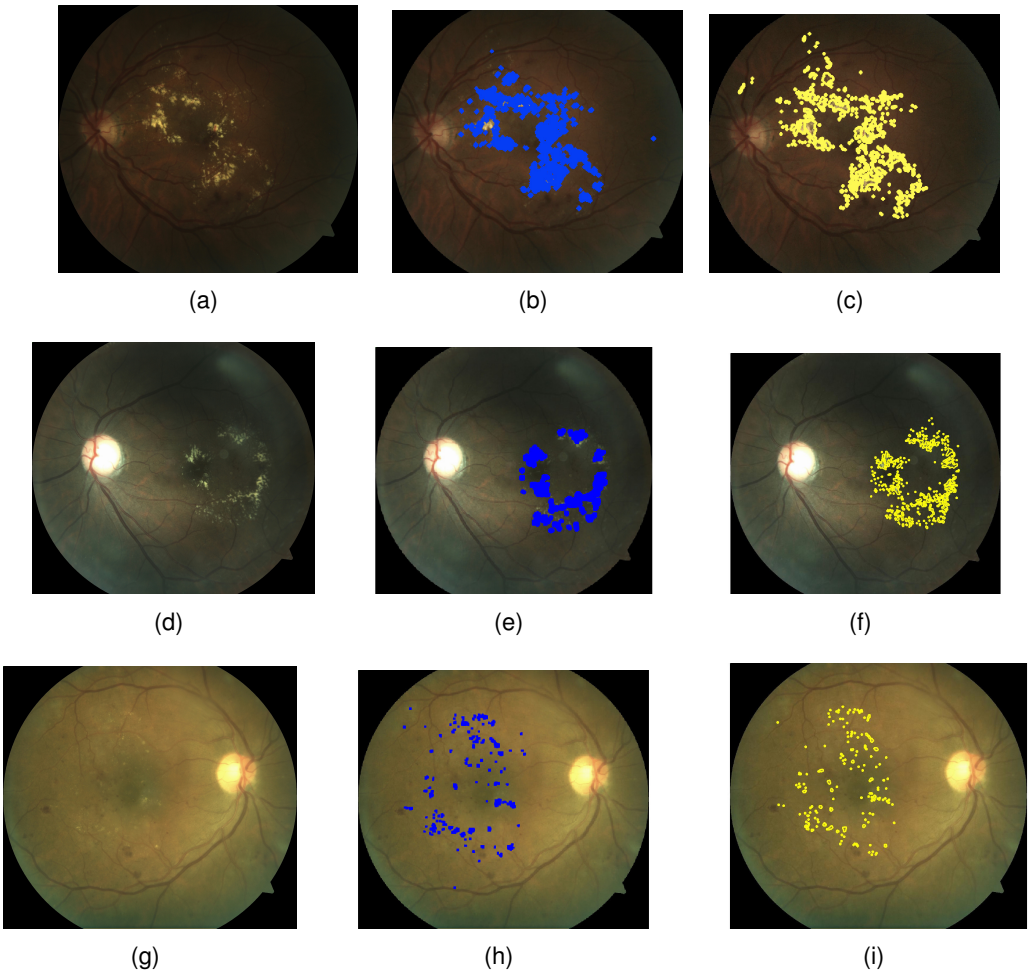


FIGURE 5.20 – Exemples de détection d'exsudats par la méthode proposée. Sur chaque ligne, on a de gauche à droite, l'image test (a, d, g), le résultat de la détection (b, e, h), et la vérité terrain (c, f, i).

peut être appliquée à la détection d'autres types de lésions rétiniennes, et que l'atlas peut également servir comme un détecteur de ROIs suivi de l'emploi d'un classifieur pour la décision finale. Néanmoins, cele nécessite d'avoir un nombre d'images d'apprentissage important pour entraîner le classifieur.

5.4/ DISCRIMINATION D'IMAGES DE FOND D'ŒIL

Dans les deux sections précédentes, nous nous sommes intéressés à la détection de lésions particulières, les MAs dans la section 5.2 et les exsudats dans la section 5.3, dans les images de fond d'œil. Cependant, une image peut contenir plusieurs lésions différentes et des lésions différentes peuvent avoir une apparence très similaire. C'est notamment le cas pour les exsudats et les druses qui sont très semblables, de petites régions jaunâtres, bien que signes de pathologies assez différentes. Les exsudats sont caractéristiques des œdèmes maculaires tandis que les druses sont caractéristiques de la dégénérescence maculaire liée à l'âge. La discrimination de ces deux types de lésions

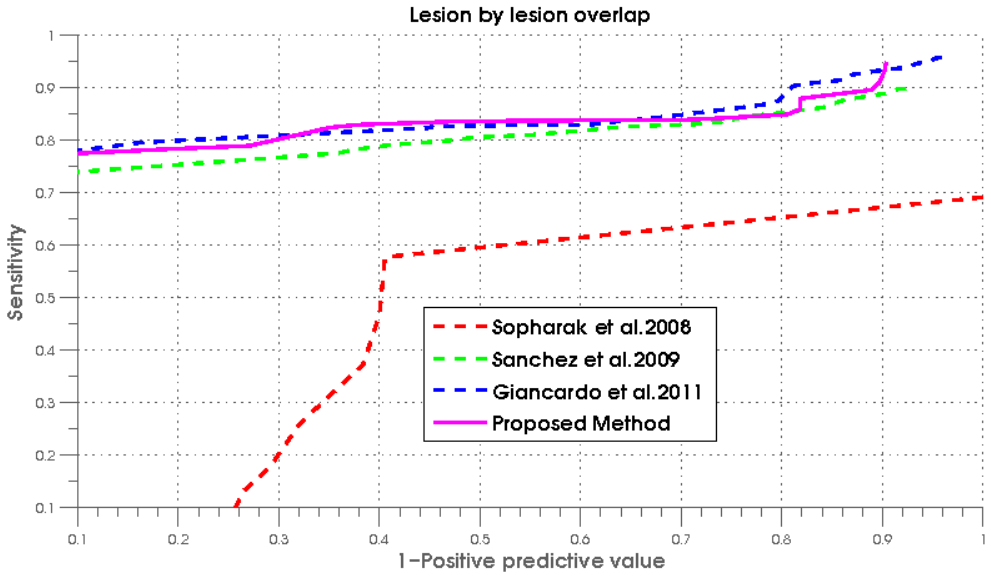


FIGURE 5.21 – Comparaison de différentes méthodes de détection d'exsudats.

est donc importante pour les systèmes CAD.

Dans cette section, nous proposons une méthode de discrimination d'images de fond d'œil en fonction du type de lésions présentes dans l'image. En particulier nous nous intéressons à la discrimination d'images contenant des exsudats ou des druses. Notre méthode repose sur l'extraction automatique de caractéristiques dans les images pour la classification, et ne nécessite pas d'étapes de pré-traitement telles que la détection de vaisseaux sanguins ou de la papille optique. Dans la section 5.4.1 suivante, nous décrivons la méthode d'extraction de caractéristques basée sur une représentation parcimonieuse des images, et nous présentons les résultats obtenus par notre approche dans la section 5.4.2.

5.4.1/ EXTRACTION AUTOMATIQUE DE CARACTÉRISTIQUES DISCRIMINANTES

Plusieurs méthodes ont été proposées pour la discrimination automatique d'images de fond d'œil et nous proposons un état de l'art dans [149]. Ces méthodes sont basées sur le schéma général présenté à la section 5.1.2, et nécessitent donc des méthodes de pré-traitement comme la détection de vaisseaux sanguins [138] ou la localisation de la papille optique [170]. Ensuite, différents descripteurs tels que les LBP (local binary patterns), HOG (histogram of oriented gradients) ou SIFT (scale invariant feature transform) sont extraits pour caractériser les images d'apprentissage, et un dictionnaire visuel est créé pour représenter les images. L'étape finale est une classification en utilisant un SVM [138, 124] ou un kNN [170, 133].

Nous proposons une approche qui ne nécessite aucun pré-traitement des images et qui extrait de manière automatique les caractéristiques discriminantes des images en utilisant une représentation éparse [177, 43].

Représentation parcimonieuse L'objet de la représentation parcimonieuse est de décrire, le plus correctement possible, un signal (ou une image interprétée comme un

signal 2D) comme une combinaison linéaire d'un nombre limité d'éléments, aussi appelés atomes, d'un dictionnaire [177, 43]. Plus précisément, soit un dictionnaire $\mathbf{D} \in \mathbb{R}^{n \times K}$ dont chaque atome est un vecteur $\mathbf{d}_j \in \mathbb{R}^n$, $j = 1, \ldots, K$, et un signal représenté par le vecteur $\mathbf{x} \in \mathbb{R}^n$. On souhaite trouver le vecteur $\mathbf{y} \in \mathbb{R}^K$ tel que $\mathbf{x} \approx \mathbf{D}\mathbf{y}$ sous la contrainte que $\mathbf{y}$ soit un vecteur éparse, i.e. avec un grand nombre d'éléments égaux à zéro. On doit donc résoudre le problème d'optimisation suivant :

$$\min_{\mathbf{y}} \|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2 \text{ avec } \|\mathbf{y}\|_0 \leq \lambda, \tag{5.4}$$

où $\|\mathbf{y}\|_0$ est la pseudo norme l_0 défini comme le nombre d'éléments non nuls du vecteur $\mathbf{y}$, et λ est un niveau de parcimonie fixé.

Les éléments non nuls de $\mathbf{y}$ sont les coefficients de la représentation du signal $\mathbf{x}$ sur la base définie par le dictionnaire $\mathbf{D}$. Contrairement à des représentations telles que l'ACP qui représentent le signal comme une combinaison linéaire de tous les atomes du dictionnaire, la représentation parcimonieuse restreint le nombre d'atomes utilisés offrant une plus grande flexibilité et permettant la prise en compte d'informations a priori [43]. Néanmoins, la résolution exacte du problème d'optimisation ci-dessus, équation (5.4), est un problème NP difficile [43] et des solutions approchées sont obtenues soit par utilisation de méthodes gloutonnes telles que la « poursuite adaptative » (*matching pursuit* (MP) et *orthogonal matching pursuit* (OMP)) [103], soit par des méthodes de ''relaxation'' qui remplacent la norme l_0 par une norme l_1 afin de résoudre un problème de programmation linéaire [33]. Cette dernière famille de méthodes est désignée par le terme « poursuite de base » (*basis pursuit*).

Apprentissage du dictionnaire et description des images Le choix du dictionnaire $\mathbf{D}$ est crucial pour obtenir une bonne représentation des images. Nous pouvons utiliser un dictionnaire existant, par exemple un dictionnaire appris sur une grande base de données telle que ImageNet [134], mais il est plus intéressant d'apprendre un dictionnaire spécifique pour le problème à résoudre. Nous employons l'algorithme K-SVD [8] qui est une méthode non supervisée d'apprentissage de dictionnaire.

Etant donné un ensemble d'apprentissage $\{I_1, \ldots, I_N\}$, nous commençons par diviser chaque image en patchs (régions) de taille 8×8 et représentons chaque patch par un vecteur de caractéristiques $\mathbf{x}_i \in \mathbb{R}^n$ comme illustré par la figure 5.22. Notons que toutes les images sont re-dimmensionnées pour avoir une taille fixe de 512×512, et le nombre de patchs extraits dans chaque image est donc égal à $p = 4096$. L'ensemble de tous les patchs extraits de toutes les images d'apprentissage forme la matrice des données d'apprentissage $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M]$, avec $M = N \times p$ le nombre total de patchs d'apprentissage.

L'objectif est donc de trouver un dictionnaire $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K]$ de taille K donné, et une matrice de coefficients de représentation $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_M]$ telle que $\mathbf{X} \approx \mathbf{D}\mathbf{Y}$:

$$\underbrace{\begin{bmatrix} \\ \mathbf{X} \\ \\ \end{bmatrix}_{n \times M}}_{\text{données d'apprentissage}} = \underbrace{\begin{bmatrix} \\ \mathbf{D} \\ \\ \end{bmatrix}_{n \times K}}_{\text{dictionnaire}} \underbrace{\begin{bmatrix} \\ \mathbf{Y} \\ \\ \end{bmatrix}_{K \times M}}_{\text{matrice de coefficients}} \tag{5.5}$$

Le problème peut donc être écrit sous la forme d'une optimisation sous contraintes :

$$\min_{\mathbf{D},\mathbf{Y}} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_2 \text{ avec } \forall i \, \|\mathbf{y}_i\|_1 \leq \lambda. \tag{5.6}$$

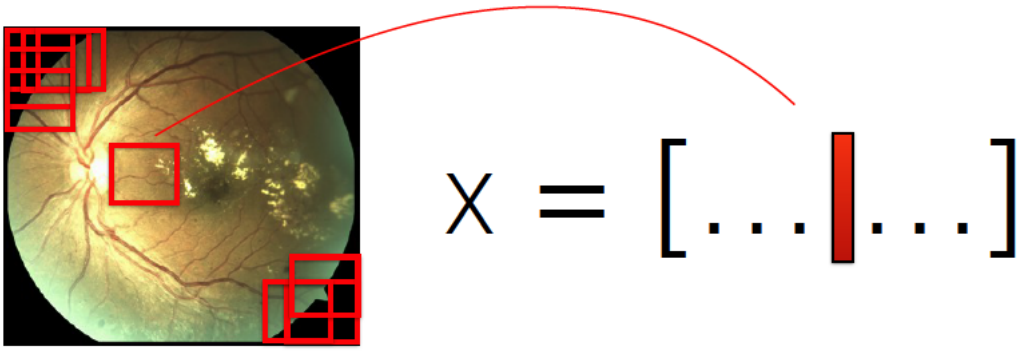


FIGURE 5.22 – Extraction de patchs dans les images de fond d'œil.

L'algorithme K-SVD résoud ce problème d'optimisation de manière itérative en alternant les étapes de calcul de la matrice de représentation $\mathbf{Y}$ et de calcul du dictionnaire $\mathbf{D}$. Etant donné le dictionnaire $\mathbf{D}$, la matrice $\mathbf{Y}$ est obtenue en résolvant, pour chaque patch $\mathbf{x}_i$, le problème défini par l'équation (5.4). Etant donnée la matrice $\mathbf{Y}$, le dictionnaire est mis à jour de manière séquentielle, i.e. un atome à la fois, en résolvant un système linéaire à l'aide d'une décomposition en valeur sigulières (SVD).

Une fois le dictionnaire obtenu, il est employé pour représenter chaque image I. Chacun des patch $\mathbf{x}_i$ de l'image I est représenté par un vecteur éparse de coefficient $\mathbf{y}_i$ solution du problème d'optimisation (équation 5.4). Nous obtenons donc une matrice de représentation $\mathbf{Y}_I \in \mathbb{R}^{K \times p}$ pour l'image I, p étant le nombre de patchs extraits de l'image. A partir de cette matrice $\mathbf{Y}_I$, nous calculons un descripteur global de l'image $\mathbf{f} \in \mathbb{R}^K$ de la manière suivante :

$$\begin{bmatrix} \\ \mathbf{Y}_I \\ \\ \end{bmatrix}_{K \times p} \Longrightarrow \mathbf{f} = \begin{bmatrix} \vdots \\ \mathbf{f_j} \\ \vdots \end{bmatrix}_{K \times 1} \quad \forall j,\ \mathbf{f_j} = g(\mathbf{Y}_I(j,:)). \tag{5.7}$$

Cette dernière étape est appelée *pooling*, et plusieurs fonctions de pooling g peut être employées :

— **Moyenne**

$$\mathbf{f_j} = \frac{1}{p} \sum_{l=1}^{p} \mathbf{Y}_I(j,l)$$

— **Max**

$$\mathbf{f_j} = \max_l |\mathbf{Y}_I(j,l)|$$

— **Abs**

$$\mathbf{f_j} = \frac{1}{p} \sum_{l=1}^{p} |\mathbf{Y}_I(j,l)|$$

Notons que la fonction de pooling *Moyenne* revient à calculer un histogramme, i.e. la fréquence d'occurence de chaque atome du dictionnaire dans l'image. La fonction de pooling *Max* est généralement employée car elle produit de meilleurs résultats de classification [179].

La figure 5.23 montre la procédure globale de discrimination d'images de fond d'œil.

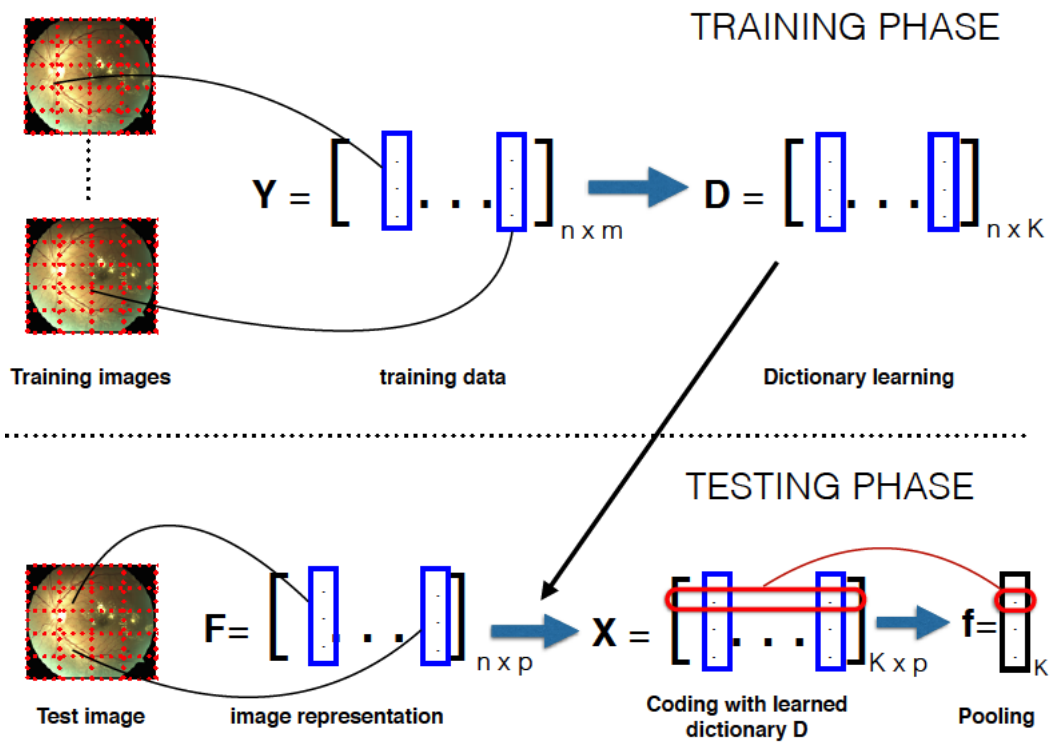


FIGURE 5.23 – Procédure d'apprentissage et de discrimination d'images de fond d'œil.

5.4.2/ DISCRIMINATION D'IMAGES CONTENANT DES DRUSES ET DES EXSUDATS

Nous évaluons les performances de l'approche développée pour la discrimination d'images de fond d'œil contenant des exsudats ou des druses (patients atteints respectivement de RD et de DMLA) et d'images ne contenant aucune lésion (patients sains). Nous comparons également notre approche avec des méthodes basées sur l'approche de représentation par *sacs de mots* (bag of features) [138, 170].

Données et procédure d'évaluation Nous avons constitué un ensemble de 828 images comprenant 452 images saines, 85 images avec des druses et 291 images avec des exsudats. Ces images sont extraites de différentes bases de données publiques acquises dans conditions différentes avec des appareils différents. L'utilisation d'images de diverses sources permet de tester la robustesse de la méthode de classification aux conditions d'acquisition. La figure 5.24 montre des exemples d'images formant la base de données et le tableau 5.6 indique la provenance et la répartion des images.

Pour l'évaluation, nous utilisons l'approche par validation croisée, k-fold cross-validation avec $k = 10$. L'ensemble des images est divisé en 10 parties, en tenant compte de la porportion d'images de chacune des catégories. Nous sélectionnons une des parties qui constitue l'ensemble de validation tandis que les 9 parties restantes constituent un ensemble d'apprentissage dont les images sont utilisées pour la création d'un dictionnaire. Les images sont représentées en utilisant ce dictionnaire et un classifieur SVM est entraîné avec les descripteurs obtenus. Le classifieur ainsi appris est évalué avec l'ensemble de validation. Ce procédé est répété 10 fois et nous calculons la performance moyenne (précision, sensibilité et spécificité moyennes). Cette approche permet, d'une

Dataset	Saines	Druses	Exsudats
ORNL [1]	36	61	20
HEI-MED [2]	20	-	26
STARE [3]	-	24	-
HRF [4]	15	-	-
DRIDB [5]	10	-	27
DRIVE [6]	20	-	-
MESSIDOR [7]	351	-	218
Total	**452**	**85**	**291**

TABLE 5.6 – Provenance des images et répartition dans les 3 catégories.

(a) Image saine (b) Image avec des druses (c) Image avec des exsudats

FIGURE 5.24 – Quelques exemples d'images de la base de données utilisée pour la discrimination d'images de fond d'œil.

part, d'assurer que chaque image est utilisée au moins une fois pour l'apprentissage et pour la validation, et, d'autre part, de mesurer la variablité des résultats.

D'autre part, puisque nous avons un problème de classifcation à 3 classes, respectivement *saines*, *druses* et *exsudats*, nous adoptons une approche *un contre tous* pour l'évaluation, i.e. que nous utilisons les images d'une classe comme exemples positifs et toutes les images des autres classes comme exemples négatifs.

Caractéristiques utilisées Pour décrire les images, nous utilisons les descripteurs suivants, extraits des patchs sélectionnés dans l'image (voir figure 5.22) :

— **COULEUR :** Pour chaque patch de taille 8×8, nous calculons un histogramme de taille 8 pour chacun des plans r, g et b de l'espace RGB normalisé, ainsi que pour les plans h, s, v, C_r et C_b des espaces HSV et YC$_r$C$_b$. Le descripteur final est la concaténation des 8 histogrammes obtenus, et donc de dimension égale à 64.

— **SIFT :** Pour chaque patch, nous calculons un descripteur SIFT en prenant comme point d'intérêt, le centre du patch. Chaque descripteur est de dimension 128.

— **LBP :** Pour chaque patch, nous calculons un descripteur LBP en prenant comme

1. private dataset kindly provided by T. Karnowski from ORNL.
2. see (http://vibot.u-bourgogne.fr/luca/heimed.php)
3. see (http://www.ces.clemson.edu/~ahoover/stare/)
4. see (http://www5.cs.fau.de/research/data/fundus-images/)
5. see (http://www.fer.unizg.hr/ipg/resources/image_database)
6. see (http://www.isi.uu.nl/Research/Databases/DRIVE/)
7. see (http://messidor.crihan.fr)

		Attributs			
		COULEUR	LBP	HOG	SIFT
Saines	Prec	92.60 (±5.42)	88.30 (±3.77)	92.30 (±3.13)	97.50 (±2.84)
	Sens	80.60 (±14.61)	75.10 (±16.35)	80.90(±17.88)	96.50 (±5.76)
	Spec	96.20 (±4.02)	92.10 (±4.63)	95.00 (±2.00)	97.7 (±3.50)
Druses	Prec	98.00 (±2.36)	96.90 (±2.38)	95.90 (±2.56)	99.80 (±0.63)
	Sens	95.60 (±10.63)	91.60 (±9.26)	94.30 (±6.34)	99.10 (±2.85)
	Spec	98.20 (±1.99)	98.20 (±2.10)	97.00 (±2.58)	100 (±0)
Exsudats	Prec	93.80 (±3.91)	88.30 (±2.87)	93.50 (±2.27)	97.70 (±2.54)
	Sens	95.40 (±2.59)	90.30 (±5.44)	93.50 (±3.63)	97.40 (±3.75)
	Spec	91.70 (±9.04)	86.40 (±7.66)	93.40 (±7.50)	98.20 (±3.05)

TABLE 5.7 – Comparaison des différents attributs. Pour chaque classe, nous indiquons la précision (Prec), la sensibilité (Sens) et la spécificité (Spec) moyennes, ainsi que les écarts-types calculés par validation croisée.

point d'intérêt, le centre du patch. Chaque descripteur est de dimension 58.

— **HOG :** Pour chaque patch, nous calculons un descripteur HOG en prenant comme point d'intérêt, le centre du patch. Chaque descripteur est de dimension 31.

Le descripteur couleur est un bon indicateur de la présence de lésions telles que les druses et les exsudats, tandis que le descripteur LBP capture la texture de la région et SIFT et HOG décrivent la forme de la région.

Résultats Dans toutes les expériences, nous fixons le niveau de parcimonie (le paramètre λ de l'équation (5.4)) égal à 3, car cette valeur nous donne les meilleurs résultats. Notons que le but des expériences de cette section, est l'identification de la catégorie de chaque image, i.e. associer à chaque image test une étiquette choisie dans l'ensemble {*Saine*, *Druse*, *Exsudat*}.

Nous commençons par comparer les différents descripteurs présentés ci-dessus, COULEUR, SIFT, HOG et LBP. Pour cela, nous fixons la taille du dictionnaire à $K = 100$ (les observations sont identiques pour des tailles variables du dictionnaire). Les résultats rassemblés dans le tableau 5.7 montrent que l'apprentissage avec le descripteur SIFT donne les meilleurs résultats. Par exemple, pour la classe *Saines* nous obtenons une sensibilité de 96.5% avec SIFT alors que le second meilleur descripteur donne une sensibilité de 80.9%. On note aussi que le descripteur SIFT est celui qui donne les résultats les plus stables (faibles écarts-types). Nous avons également tester différentes combinaisons de ces descripteurs, mais n'avons pas constaté d'amélioration notable des performances. Nous utiliserons donc le descripteur SIFT dans la suite des expériences.

Un paramètre important de la méthode est la taille K du dictionnaire utilisé. Nous faisons varier cette valeur de 10 à 1000 pour l'identification d'images appartenant à chacune des 3 classes. Les résultats rassemblés dans le tableau 5.9 montrent que les résultats augmentent avec la taille du dictionnaire et on obtient des résultats presque parfaits pour une taille $K = 300$ avec une sensibilité et une spécifité entre 99% et 100%.

Enfin nous comparons les résultats obtenues par notre approche, basée sur une représentation éparse, et la méthode de classification basée sur les sacs de mots (Bag of Words) utilisée par Grinsven *et al.* [170] et par Sadek et al. [138]. Les résultats du tableau 5.8 montrent que les deux approches obtiennent une précision moyenne compa-

		Taille du dictionnaire			
		50	100	500	1000
Méthode proposée	Prec	93.70 (±3.71)	97.50 (±2.84)	99.40 (±0.97)	99.80 (±0.63)
	Sens	92.40 (±5.33)	96.50 (±5.76)	98.50 (±3.17)	100 (±0)
	Spec	96.60 (±3.17)	97.70 (±3.50)	99.70 (±0.95)	99.70 (±0.95)
Bag-of-Words	Prec	93.70 (±2.58)	95.30 (±2.06)	97.20 (±2.04)	97.70 (±2.06)
	Sens	90.20 (±8.11)	87.30 (±12.59)	92.50 (±6.57)	92.20 (±12.04)
	Spec	94.60 (±3.50)	96.60 (±3.50)	98.20 (±1.55)	98.80 (±1.55)

TABLE 5.8 – Comparaison entre l'approche par sac de mots (Bag of Words) et l'approche par représentation parcimonieuse (Sparse coding) pour la classe *Saine*.

rable, respectivement de 97.5% et 95.3% pour un dictionairte de taille 100. Toutefois, il y a une différence significative lorsque l'on considère la sensibilité. Par exemple, pour un dictionnaire de 100 atomes, notre méthode donne une sensibilité moyenne de 96.5% tandis que l'approche par sac de mots donne 87.5%. Soulignons également que ces deux méthodes basées sur les sacs de mots nécessitent des étapes de pré-traitement (suppression des vaisseaux sanguins, etc), ce qui n'est pas le cas avec l'approche proposée.

5.4.3/ CONCLUSION

Dans cette section, nous avons présenté une méthode de discrimination d'images de fond d'œil basée sur une représentation éparse des images. Cette méthode présente l'avantage de ne nécessiter aucune étape de pré-traitement telle que la suppression des vaisseaux sanguins ou de la papille optique. Les résultats obtenus montrent que notre méthode est capable de correctement distinguer des images saines, de celles contenant des druses et celles contenant des exsudats, avec une spécificité et une sensibilité proches de 100%.

5.5/ CONCLUSIONS ET DISCUSSION

Dans ce chapitre, nous avons abordé la problématique de l'analyse d'images de fond d'œil pour le dépistage de la rétinopathie diabétique (RD). Ce problème est très largement traité dans la littérature, mais quasiment toutes les méthodes sont basées sur un apprentissage supervisé et nécessitent donc un travail d'annotation manuelle important. Par rapport à ces méthodes, les approches que nous avons proposées nécessitent peu ou pas d'images manuellement annotées au niveau des lésions à détecter, ce qui constituent un avantage important. Plus précisément :

— Nous avons proposé une méthode performante de détection de microanévrismes basée sur un apprentissage semi-supervisé. Ce type d'apprentissage utilise un nombre limité d'exemples annotés (dans notre cas une dizaine) et un grand nombre d'exemples non annotés (dans notre cas plusieurs centaines) pour entraîner un classifieur. Nous avons montré qu'une telle approche est performante pour la détection de lésions dans les images de fond d'œil, à condition que les caractéristiques utilisées pour l'apprentissage soient bien définies. Les résultats obtenus avec une base de données publique montrent que notre méthode d'extrac-

tion de caractéristiques basée sur une analyse des structures locales de l'image dans un espace échelle, permet de réduire de manière significative le nombre de faux positifs et obtient des résultats comparables avec les meilleurs approches de la littérature qui nécessitent un gros effort d'annotation pour l'apprentissage.
— Nous avons proposé une méthode de détection d'exsudats basée sur un atlas facile à mettre en œuvre car elle ne nécessite pas d'étapes de pré-traitement telles que la segmentation des vaisseaux sanguins et la détection de la papille optique. De plus, la détection des lésions se réduit à une simple différence d'images suivie de quelques opérations de post-traitement. Cela ne nécessite pas d'annotation manuelle des lésions pour l'apprentissage.
— Pour faciliter la détection de lésions, nous proposons une méthode qui permet de discriminer les images de fond d'œil en fonction des lésions qu'elles contiennent. Notre méthode, basée sur une représentation éparse des images, ne nécessite aucune étape de pré-traitement et donne des résultats de discrimination presque parfaits, une spécificité et une sensibilité proches de 100%.

Ces différents points peuvent être intégrés dans un système complet d'analyse d'images de fond d'œil. Nous pouvons commencer par savoir si l'image comporte ou non des lésions, en utilisant la méthode de discrimination automatique. Puis, si elle comporte des lésions, utiliser un atlas construit avec des images saines pour détecter les lésions potentielles (régions d'intérêt). Enfin, en fonction de la décision de la première étape, i.e. du type de lésions présentes, procéder à l'extraction de caractéristiques adéquates pour obtenir une détection finale. Cette détection finale pouvant être réalisée avec un classifieur entraîné sur un nombre réduit d'exemples annotés en utilisant une approche semi-supervisée.

Ce travail a été, pour une grande part, réalisé dans le cadre d'un projet de collaboration internationale avec Oak Ridge National Laboratory (ORNL) aux USA. Il a donné lieu à plusieurs publications (3 revues [14, 7, 149], 5 conférences internationales). Il se poursuit actuellement avec l'emploi d'une modalité d'imagerie complémentaire aux images de fond d'œil, la tomographie par cohérence optique. L'analyse de ce type d'image est abordée dans le chapitre 6 suivant.

		taille du dictionnaire (K)							
		10	20	30	50	100	200	300	500	1000
Saines	Prec	83.80 (±5.29)	88.20 (±4.24)	90.30 (±4.27)	93.70 (±3.71)	97.50 (±2.84)	99.40 (±0.97)	99.80 (±0.63)	99.60 (±0.84)	99.80 (±0.63)
	Sens	56.50 (±9.83)	77.20 (±10.32)	79.20 (±17.81)	82.40 (±15.33)	96.50 (±5.76)	98.50 (±3.17)	100 (±0)	100 (±0)	100 (±0)
	Spec	90.70 (±4.06)	91.00 (±4.27)	92.90 (±4.63)	96.60 (±3.17)	97.70 (±3.50)	99.70 (±0.95)	99.70 (±0.95)	99.50 (±1.08)	99.70 (±0.95)
Druses	Prec	91.90 (±3.38)	96.30 (±3.33)	96.60 (±2.41)	98.20 (±1.93)	99.80 (±0.63)	99.80 (±0.63)	100 (±0)	100 (±0)	100 (±0)
	Sens	81.20 (±12.59)	87.90 (±15.44)	87.90 (±5.67)	92.40 (±9.57)	99.10 (±2.85)	99.20 (±2.53)	100 (±0)	100 (±0)	100 (±0)
	Spec	95.00 (±2.79)	98.60 (±1.90)	98.60 (±2.37)	99.40 (±1.90)	100 (±0)	100 (±0)	100 (±0)	100 (±0)	100 (±0)
Exsudats	Prec	88.30 (±6.73)	91.50 (±3.66)	93.30 (±3.68)	95.70 (±3.06)	97.00 (±2.54)	99.60 (±0.84)	99.80 (±0.63)	99.80 (±0.63)	99.80 (±0.63)
	Sens	91.80 (±9.83)	92.10 (±10.32)	94.10 (±17.81)	97.70 (±15.33)	97.40 (±5.76)	100 (±0)	100 (±0)	100 (±0)	100 (±0)
	Spec	83.4 (±11.37)	90.90 (±7.06)	91.50 (±9.18)	92.40 (±6.72)	98.20 (±3.05)	99.00 (±2.11)	100 (±0)	100 (±0)	100 (±0)

TABLE 5.9 – Variation des résultats de la classification avec la taille du dictionnaire. Pour chaque classe, nous indiquons la précision (Prec), la sensibilité (Sens) et la spécificité (Spec) moyennes, ainsi que les écarts-types calculés par validation croisée.

6

CLASSIFICATION D'IMAGES OCT

La tomographie par cohérence optique (OCT) est de plus en plus utilisée pour le dépistage des pathologies de l'œil car elle permet d'obtenir des images en coupe de la rétine et la visualisation de lésions internes non visibles sur la surface de la rétine. Dans ce chapitre, nous nous intéressons à l'analyse et à la classification automatique de données OCT pour le dépistage des œdèmes maculaires diabétiques (OMD) associés à la rétinopathie diabétique (RD).

6.1/ INTRODUCTION

L'œdème maculaire diabétique (OMD) est une complication de la RD caractérisée par un glonflement de la rétine (œdème maculaire) ainsi que des dépôts de lipides et de lipoprotéines sur la rétine (exsudats secs). L'OMD est considéré comme une des principales causes de perte de vision chez les diabétiques.

Si l'analyse d'images de fond d'œil peut permettre la détection des exsudats, comme nous l'avons évoqué à la section 5.3, elle ne permet pas la détection du principal signe qui est la présence d'œdème maculaire. L'OCT, par contre, permet une mesure précise de l'épaisseur de la rétine ainsi que la détection de signes non visibles sur la surface de la rétine tels que les kystes rétiniens. Elle est aujourd'hui devenu un standard en ophtalmologie.

6.1.1/ INTRODUCTION À L'IMAGERIE OCT

La tomographie par cohérence optique ou OCT pour l'acronyme anglais de *Optical Coherence Tomography*, est une technique d'imagerie *in vivo* développée au début des années 1990 par des chercheurs du MIT [71, 54]. L'OCT est analogue à l'imagerie ultrasonore (échographie) mais elle est basée sur l'utilisation d'ondes lumineuses pouvant traverser les tissus biologiques. Le principe physique de l'OCT respose sur l'utilisation d'un interféromètre de Michelson permettant de produire des franges d'interférence à partir desquelles sont déduites les informations nécessaires à la formation des images [45, 46, 58].

Fonctionnement Le principe de fonctionnement d'un dispositif d'OCT est illustré par la figure 6.1. Un interféromètre de Michelson est éclairé par une source de lumière qui est divisée en deux parties : un faisceau de référence envoyé sur le miroir de référence et un

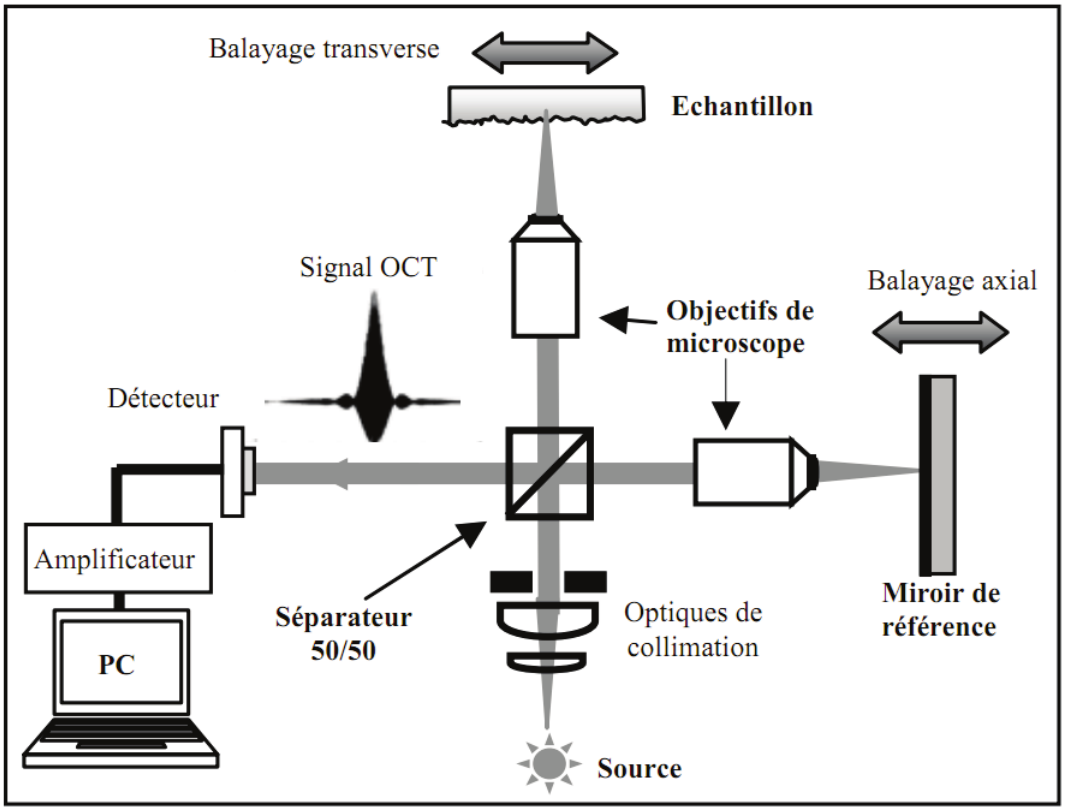


FIGURE 6.1 – Principe de l'imagerie OCT basée sur l'interféromètre de Michelson. Image reproduite d'après [58].

faisceau envoyé sur l'échantillon étudié. A la sortie de l'interféromètre, des interférences se produisent si la différence de marche (différence de longueur optique entre les deux faisceaux) est inférieure à la longueur de cohérence de la lumière.
Un signal OCT est généré en faisant varier la position axiale du miroir de référence et la mesure en profondeur est obtenue grâce à la modification du trajet optique du faisceau de référence [28, 58]. En effet, lorsque l'égalité des trajets optiques dans les 2 bras de l'interféromètre correspond à une profondeur dans l'échantillon où se trouve une structure réfléchissante (ou rétro-diffusante), des interférences se produisent. En enregistrant l'amplitude des interférences au cours du déplacement du miroir de référence, on peut accéder à la distribution des structures internes de l'échantillon en fonction de leur profondeur. Ce « sondage » de la profondeur est réalisé à différents endroits dans l'échantillon en balayant le faisceau lumineux. On obtient ainsi une image tomographique orientée perpendiculairement à la surface de l'échantillon [41].

L'intensité détectée en un point P quelconque de l'échantillon est donné par [28] :

$$I(P) = I_{ref}(P) + I_{ech}(P) + 2\sqrt{I_{ref}(P)I_{ech}(P)}\, e^{-(\frac{\pi\Delta\nu\tau}{2\sqrt{ln2}})^2} \cos\left(\frac{2\pi}{\lambda}\tau\right), \tag{6.1}$$

où $I_{ref}(P)$ et $I_{ech}(P)$ sont respectivement les intensités issues de la surface de référence et de l'échantillon, τ est le retard temporel introduit par le déplacement Δl de la surface de référence ($\tau = 2\frac{\Delta l}{c}$, avec c la vitesse de la lumière), et λ est la longueur d'onde de la source lumineuse utilisée, de largeur à mi-hauteur $\Delta\lambda$ et $\Delta\nu = c\frac{\Delta\lambda}{\lambda^2}$.

La résolution axiale de l'OCT est déterminée par la longueur de cohérence de la lumière

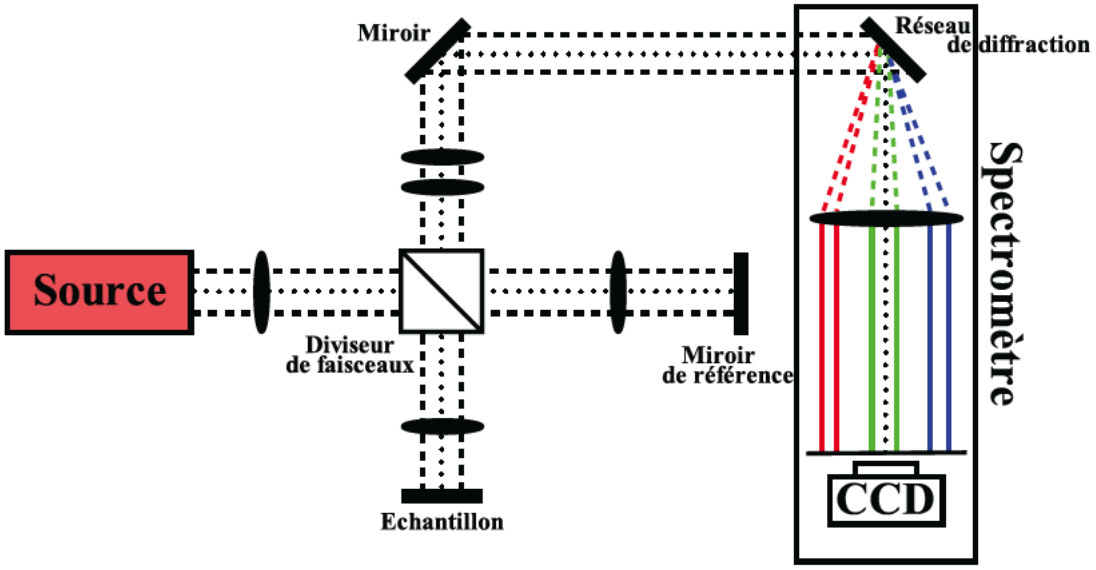


FIGURE 6.2 – Principe de l'OCT spectrale basée sur l'utilisation d'un spectromètre. Image reproduite d'après [58].

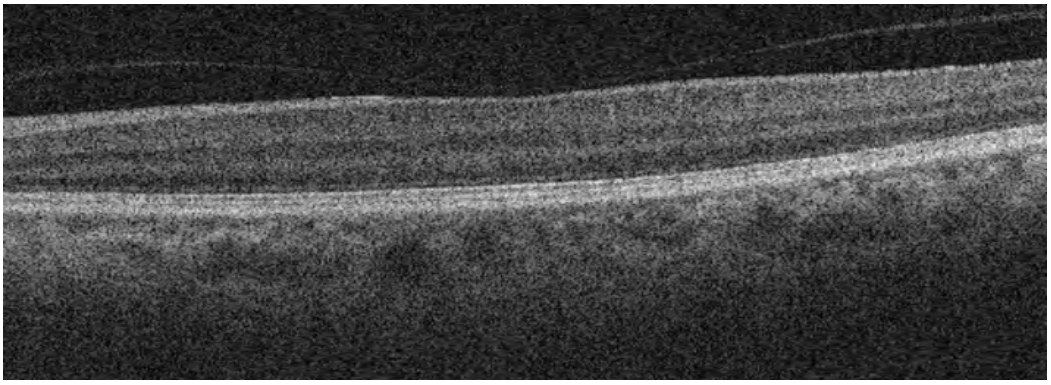
détectée. L'équation suivante donne la résolution axiale Δz en fonction de la largeur spectrale $\Delta\lambda$, de le longueur d'onde λ de la lumière dont le spectre est supposé Gaussien, et de l'indice de réfraction n du milieu [46] :

$$\Delta z = \frac{2ln2}{n\pi}\left(\frac{\lambda^2}{\Delta\lambda}\right). \tag{6.2}$$

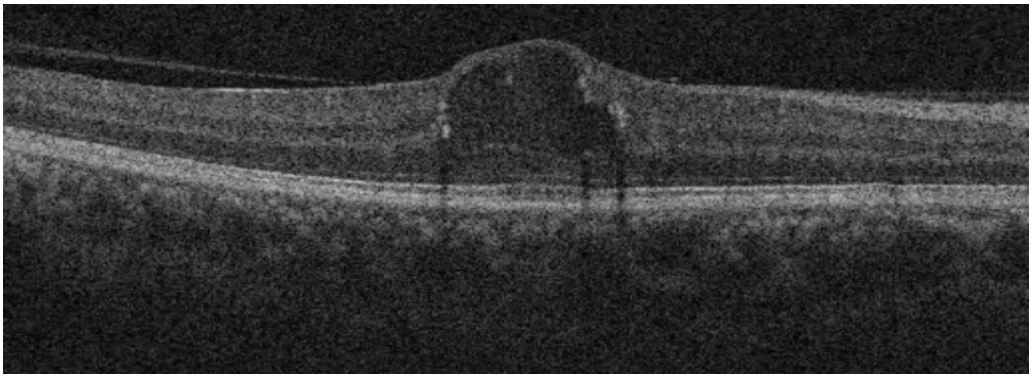
Cette résolution correspond à la distance minimale que l'on peut détecter entre deux couches réflectives selon la direction de propagation du faisceau lumineux. L'OCT requiert donc une source lumineuse de large spectre pour accroitre cette résolution axiale. Les principales sources de lumière utilisées sont les diodes super-luminescentes (SSL) et les lasers femtosecondes.

Domaine temporel vs. Domaine spectral Dans le domaine temporel, il faut faire varier la position du miroir de référence pour créer des franges d'interférences, on parle alors d'OCT temporelle (time-domain OCT, TD-OCT). Une autre solution consiste à analyser le spectre du signal d'interférence pour en déduire la profondeur des strucrures réfléchissantes de l'échantillon [47]. On parle dans ce cas d'OCT spectrale (spectral-domain OCT, SD-OCT). L'OCT spectrale repose sur l'utilisation d'un spectromètre qui réalise une décomposition spectrale de l'intensité enregistrée. La transformée de Fourier de cette intensité spectrale donne accès au profil de reflectivité et les pics de la transformée permettent de localiser les interfaces de l'échantillon [58]. La figure 6.2 montre le schéma d'un dispositif d'OCT spectral.

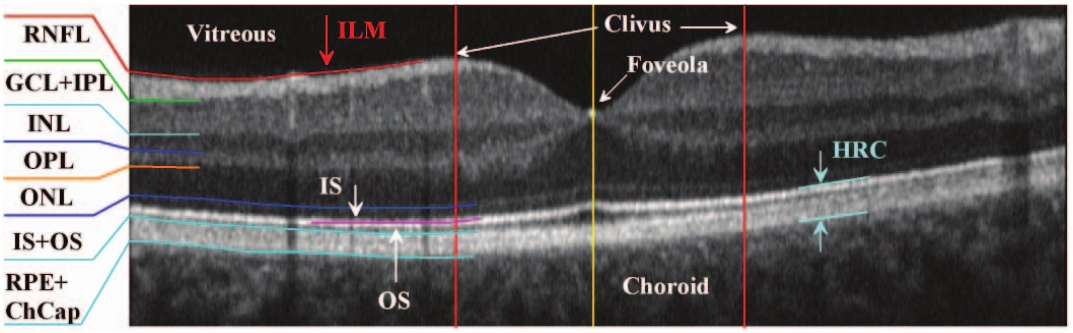
L'OCT spectrale offre plusieurs avantages par rapport à l'OCT temporelle. D'abord, puisqu'il n'est plus nécessaire de réaliser un balayage du miroir de référence, l'acquisition est plus rapide d'un facteur de 50 à 100 [175]. De plus, la sensibilité de détection ne dépend pas de la largeur spectrale de la source lumineuse.



(a) Exemple d'image OCT d'un œil sain.



(b) Exemple d'image OCT d'un œil atteint d'OMD.



(c) Différentes structures de l'œil détectées dans les images OCT [59].

FIGURE 6.3 – Exemples d'image OCT et structure de l'œil.

Utilisation en ophtalmologie Bien qu'utilisée dans de nombreux domaines (dermatologie, gastro-entérologie ou dentaire), la principale application de l'OCT depuis son développement reste le domaine de d'ophtalmologie. Elle est en effet la seule technique permettant de visualiser les différentes couches constitutives de la rétine *in vivo*. L'OCT est aujourd'hui devenu un standard en ophtalmologie.

La figure 6.3 montre des exemples d'image OCT ainsi que les couches rétiniennes visibles sur une image OCT. Comme on peut le voir sur la figure 6.3(c), les différentes couches sont identifiables sous forme de bandes longitudinales claires et sombres en

fonction de la réflectivité des différents tissus. On note que la couche correspondant à l'épithélium pigmentaire (retinal pigment epithelium, RPE) et celle correspondant au réseau de fibres optiques (retinal nerve fiber layer, RNFL) possède des réflectivités élevées et apparaissent donc plus claires que les autres couches. D'autre part, la réflectivité étant plus faible pour les kystes, les œdèmes ou l'atrophie des tissus, ceux-ci sont également clairement identifiables comme sur l'exemple de la figure 6.3(b).

6.1.2/ ETAT DE L'ART ET POSITIONNEMENT DU TRAVAIL

Dans cette section, nous décrivons brièvement différents travaux portant sur l'anaylse d'images OCT pour le dépistage de pathologies liées à l'œil et nous situons notre travail par rapport à ces travaux. Notons qu'une part importante des travaux de la littérature est consacrée à la segmentation des couches rétiniennes. Néanmoins, notre objectif étant ici la classification des images OCT, nous ne détaillerons pas cette étape et renvoyons le lecteur intéressé aux références données ci-dessous.

6.1.2.1/ PRÉ-TRAITEMENTS

Toutes les méthodes d'analyse d'images OCT commencent par une étape de pré-traitement visant à améliorer la qualité des images. En effet comme le montre les exemples de la figure 6.3, les images OCT sont affectées par un bruit de type speckle (ou chatoiement en français) qui est dû à la rétrodiffusion des ondes dans des milieux de taille et densité variables [145]. Différentes techniques de débruitage peuvent être utilisées pour les images OCT. Les méthodes classiques telles que le filtrage passe-bas, moyenneur, médian ou Gaussien ne sont pas adaptées car elles provoquent un lissage des contours. Garvin *et al.* [56] introduisent des filtres de diffusion anisotropiques qui rehaussent les contours et facilitent leur préservations lors du lissage. Les résultats obtenus sont de meilleure qualité. Les méthodes basées sur un seuillage des coefficients de la décomposition en ondelettes du signal ont également été employées [77, 126]. Enfin, Coupe *et al.* [36] ont montré que la méthode des moyennes non locales (non local means, NLM), introduite par Buades *et al.* [29], est une méthode de débruitage assez performante en présence de bruit de type speckle. Nous employons donc cette méthode qui préserve les structures fines de l'image ainsi que les zones homogènes dans nos travaux.

Les images OCT sont généralement organisées en volumes 3D comme présentés sur la figure 6.4(a). Dans l'exemple présenté, les images sont acquises en réalisant un balayage axial selon l'axe y, et chaque image, appelée B-scan, correspond à une image en coupe de la rétine dans le plan $x - z$. Le volume OCT est donc une série de B-scans, dont le nombre varie en fonction de l'appareil de mesure utilisé et de la résolution fixée par l'utilisateur. Du fait du positionnement de l'œil, des angles d'inclinaison et de la courbure de la rétine, les B-scans obtenus pendant une même acquisition peuvent ne pas être parfaitement alignés. Un pré-traitement couramment employé consiste à aligner les différents B-scans formant un même volume 3D. Par exemple, les auteurs dans [91] proposent une approche qui consiste déplier l'image de telle sorte que la couche correspondant à l'épithélium pigmentaire (RPE) soit plus ou moins horizontal. Cette méthode est illustrée par les images des figures 6.4(b) et (c). Notons toutefois que cette étape d'alignement n'est pas nécessaire avec les appareils plus récents qui réalisent un suivi des

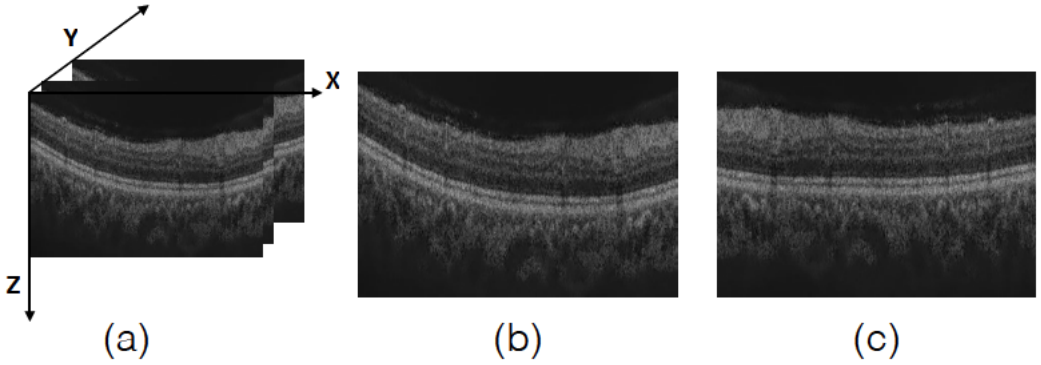


FIGURE 6.4 – Alignement des scans OCT à l'intérieur d'un volume. (a) Volume 3D formé d'une série de B-scans ; (b) B-scan avant alignement ; (c) B-scan après alignement.

mouvements de l'œil.

6.1.2.2/ SEGMENTATION DES COUCHES RÉTINIENNES

La segmentation des différentes couches rétiniennes dans les images OCT a fait l'objet d'un grand nombre de travaux dans la littéraure, car elle permet de mesurer l'épaisseur rétinienne totale ou l'épaisseur de la RFNL qui sont des mesures utilisées pour le dépistage de pathalogies. D'ailleurs, la plupart des appreils d'acquisition sont livrés avec des logiciels « maison » qui donnent une estimation de ces mesures.

Différentes approches sont employées pour segmenter les couches rétiniennes, parmi lesquelles les modèles déformables (contours actifs), les méthodes par coupe de graphes (graph cuts) ou des méthodes basées sur la classification des pixels de l'image. Comme mentionné dans l'introduction de cette section, nous ne détaillerons pas plus les méthodes de segmentation des couches rétiniennes, et nous invitons le lecteur intéressé à consulter les travaux décrits dans [59, 75, 76].

6.1.2.3/ CLASSIFICATION

Contrairement à la segmentation des couches rétiniennes, peu de travaux ont été consacrés à la classification automatique de données OCT (sans segmentation préalable des couches) pour distinguer des patients sains de ceux atteints de pathalogies telles que l'œdème maculaire diabétique (OMD). Or, cette classification est très importante pour plusieurs raisons :

- l'analyse manuelle des volumes OCT, 3D, pour rechercher des signes spécifiques de l'OMD est fastidieuse, coûteuse en temps et en énergie.
- elle permet d'éliminer les données de patients sains, les plus nombreux, et de consacrer le temps à l'analyse approfondie de cas réellement importants, cas pathologiques.

Quasiment toutes les méthodes proposées dans la littérature pour la classification d'images OCT sont des méthodes supervisées [91, 171, 157, 9, 15]. Elles nécessitent donc un ensemble d'apprentissage manuellement annoté qui comporte des exemples positifs (cas pathologiques) et des exemples négatifs (cas normaux ou sains). On peut

décomposer ces approches en quatres principales étapes : i) *pré-traitement* pour réduire le bruit speckle ; ii) *extraction de caractéristques* telles que la texture ou la forme ; et iii) *répresentation* locale ou globale des images ; iv) *classification* des images.

Par exemple, Srinivasan *et al.* [157] extraient un descripteur HOG de chaque B-scan d'un volume OCT et entraînent un classifieur SVM. Chaque B-scan (image OCT) est classé individuellement et la catégorie du volume OCT est décidé par un vote majoritaire. Liu *et al.* [91] proposent une approche similaire mais utilisent le descripteur LBP pour décrire chaque B-scan. Ces deux méthodes classifient chaque B-scan individuellement, tandis que d'autres méthodes traitent les données OCT comme des volumes 3D. C'est le cas de l'approche proposée par Venhuizen *et al.* [171] qui commence par extraire des points d'intérêt dans chaque B-scan, qui sont caractérisés par le vecteur d'intensité calculé dans un voisinage 9×9. La dimension de ce vecteur de caractéristiques est réduit à 9 par une ACP et un dictionaire visuel est créé en utilisant l'ensemble d'apprentissage. Enfin, chaque volume OCT est représenté par un histogramme obtenu par une méthode de type sac de mots (bag of words) et les forêts aléatoires (random forests) sont utilisées pour la classification. Albarrak *et al.* [9] extraient des sous-volumes 3D qui sont décrits par des LBP et le volume OCT est représenté par un vecteur qui est la concaténation des descripteurs de chaque sous-volume.

6.1.2.4/ CONTRIBUTIONS

Nous apportons les contributions suivantes à la classification d'images OCT :

1. **Analyse des méthodes de représentation :** Dans un premier temps, nous proposons une méthode basée sur des descripteurs locaux, semblable aux méthodes de la littérature. Toutefois, nous analysons en détail les différentes approches de représentation des images OCT (locale, globale, patchs 2D et 3D) pour en sélectionner les meilleures. Ce travail est décrit dans la section 6.2, et les résultats montrent une amélioration par rapport aux méthodes de l'état de l'art.
2. **Classification non supervisée :** Dans un second temps, nous proposons une méthode qui ne nécessite pas un ensemble d'apprentissage avec des exemples positifs et négatifs. Nous adoptons une approche basée sur la « détection d'anomalie », en modélisant l'apparence des images OCT saines par un modèle de mélanges de Gaussiennes (GMM) et en détectant les images OCT anormales (pathologiques) par rapport à ce modèle. Cette méthode, décrite dans la section 6.3, donne de très bons résultats de classification.
3. **Identification de B-scans pathologiques :** Notre méthode de classification non supervisée, section 6.3, permet d'identifier de manière automatique les B-scans pathologiques à l'intérieur d'un volume OCT. Ceci est un avantage par rapport aux autres approches existantes, car nous pouvons appliquer des algorithmes spécifiques de détections (de kystes ou d'exsudats) aux images sélectionées.

6.2/ CLASSIFICATION BASÉE SUR DES DESCRIPTEURS LOCAUX

Dans cette section, nous nous intéressons à la classfication de volumes OCT en deux catégories : *anormale* pour les patients atteints de pathologies telles que l'OMD, et *normale* pour les patients sains. Nous adoptons une approche de classification supervisée

basée sur l'extraction de caractéristiques dans les images OCT. Pour chaque patient, nous avons un volume OCT composé de 128 B-scans (ou images 2D) comme illustré par la figure 6.4(a). La question qui se pose est donc celle de la réprésentation de ce volume par un descripteur unique qui sera utilisé pour entraîner un classifieur.

Description et représentation Dans ce travail nous utilisons comme caractéristique, le decripteur de texture LBP dont la pertinence pour la images OCT a été prouvée [91, 88]. Celui-ci peut être extrait de manière *globale* ou *locale*.

— **Description globale :** Un descripteur LBP est calculé pour chaque B-scan du volume qui est donc décrit par un vecteur qui est la concaténation des différents descripteurs. On peut aussi extraire un descripteur LBP-TOP pour le volume OCT (le descripteur LBP-TOP est présenté dans la section 3.3.1.1 et illustré par la figure 3.6). Ces deux approches de représentation globale sont illustrées par les figures 6.5(a) et (c).

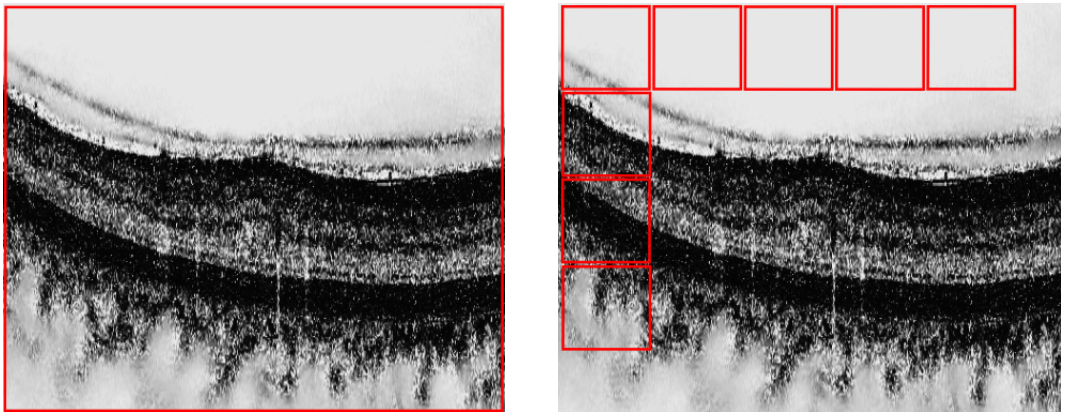
— **Description locale :** Les descripteurs LBP, respectivement LBP-TOP, sont calculés pour chaque patch de taille $m \times m$ extrait de l'image, respectivement pour chaque sous-volume de taille $m \times m \times m$ extrait du volume OCT. Ces deux approches de représentation locale sont illustrées par les figures 6.5(b) et (d). Dans nos expériences, la taille des patchs est fixée à $m = 7$.
Ensuite, une approche de type sac de mots (bag of words) est employée pour représenter les volumes [151]. L'ensemble des descripteurs obtenus est utilisé pour créer un dictionaire visuel par agglomération (k-means) et chaque volume OCT est représenté par un histogramme qui représente les fréquences d'apparition de chaque mot du dictionaire dans le volume.

Une fois les volumes décrits, i.e. représentés par un vecteur de caractéristiques, un classifieur est entraîné avec les données d'apprentissage. Nous avons évalué 4 classifieurs différents qui sont les plus proches voisins (k-NN), la régression logistique (LR), les forêts alétoires (RF) et les machines à support de vecteurs (SVM).

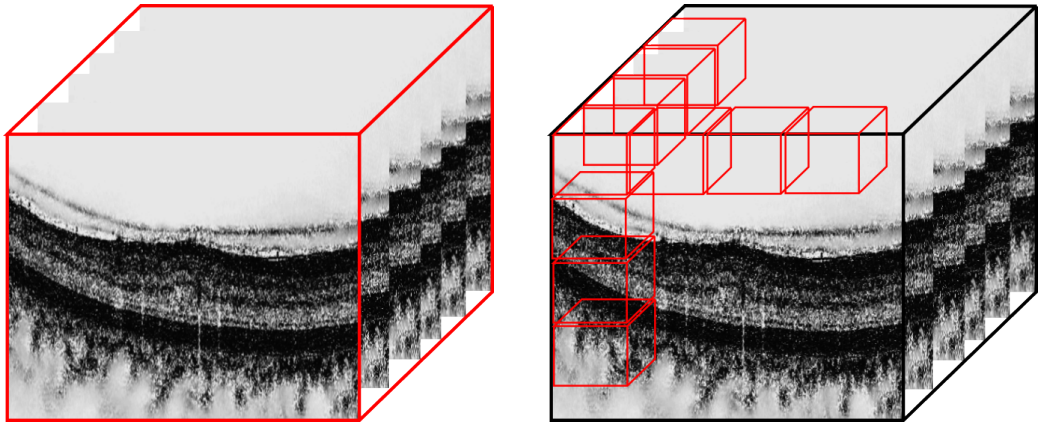
Données d'évaluation Nous disposons d'un ensemble de 32 volumes OCT (16 volumes sains et 16 volumes anormaux). Ces données acquises avec un appareil CIRRUS TM (Carl Zeiss Meditec, Inc. Dublin, CA) sont fournies par nos collègues du SERI (Singapore Eye Reserach Institute). Chaque volume OCT est composé de 128 B-scans (voir figure 6.4(a) pour illustration), chacun de taille 512×1024. Chaque volume a été analysé par un spécialiste et classé dans l'une des catégories en fonction de la présence ou non de signes pathologiques dans les B-scans.

Résultats Puisque nous avons peu de données, les résultats d'évaluation sont obtenus avec une stratégie de type « leave-one-out cross-validation » (LOOCV), i.e. que le classifieur est entraîné avec 30 volumes (15 volumes de chaque catégorie) et testé sur les 2 volumes restants, et cette procédure est répétée 16 fois.

Nous avons effectué plusieurs expériences en associant chaque réprésentation (globale/locale, 2D/3D) à chaque classifieur, pour déterminer la meilleure association. Le tableau 6.1 montre les résultats obtenus, avec pour chaque représentation, le classifieur qui donne les meilleurs résultats. Les résultats montrent que les représentation locales donnent des résultats de classification bien meilleurs que les représentation globales. En particulier, la représentation locale 2D (extraction de patch 7×7 dans les B-scans suivi d'une approche BoW), associée à un classifieur de type SVM, donne une sensibilité



(a) description globale (2D) (b) description locale (2D)



(c) description globale (3D) (d) description locale (3D)

FIGURE 6.5 – Description de volumes OCT. (a) description globale par B-scan ; (b) description locale par B-scan ; (c) description globale par volume ; (d) description locale par volume. **NOTE : les images sont présentées ici en couleur inverse pour une meilleure visualisation**.

Représentation	Classifieur	Sensibilité (%)	Spécificité (%)
Globale (2D)	RF	56.2	75
Globale (3D)	RF	81.2	81.2
Locale (2D)	SVM	81.2	93.7
Locale (3D)	SVM	75	100

TABLE 6.1 – Evaluation des différentes méthodes de représentation.

de 81.2% et une spécificité de 93.7%. La représentation globale 3D, avec un RF, donne des résultats satisfaisants avec une sensibilité et une spécificité de 81.2%, tandis que la représentation globale 2D donne une sensibilité de 56.2% et une spécificité de 75%.

Nous avons également comparé notre méthode de classification avec plusieurs approches de l'état de l'art. Les résultats rassemblés dans le tableau 6.2 montre que nous

Méthode	Classifieur	Sensibilité (%)	Spécificité (%)
Proposée (locale - 2D)	SVM	81.2	93.7
Srinivasan *et al.* [157]	SVM	68.8	93.8
Liu *et al.* [91]	SVM	68.8	93.8
Venhuizen *et al.* [171]	RF	61.5	58.8

TABLE 6.2 – Comparaison avec d'autres méthodes.

obtenons de meilleurs résultats, notamment en termes de sensibilité. En effet, notre méthode obtient une spécificité égale à celles obtenues par les approches de Srinivasan *et al.* [157] et de Liu *et al.* [91], i.e. 93.7%. Mais ces deux méthodes donnent une sensibilité de 68.8% quand nous obtenons 81.2% avec notre méthode, soit un gain d'environ 18%. Notons enfin que la méthode de Venhuizen *et al.* [171] basée elle aussi sur les sacs de mots, mais en utilisant comme caractéristique l'intensité, donne des résultats décevants. Ce qui confirme que l'opérateur de texture LBP est très adapté pour la classification d'image OCT. Notons également que nous avons ré-implémenté ces différentes méthodes pour cette évaluation.

Conclusion Nous pouvons conclure que la texture est un bon descripteur des images OCT comparée à l'intensité [171] ou au gradient [157], et qu'une représentation locale par extraction de patchs dans les images donne de bons résultats de classification.

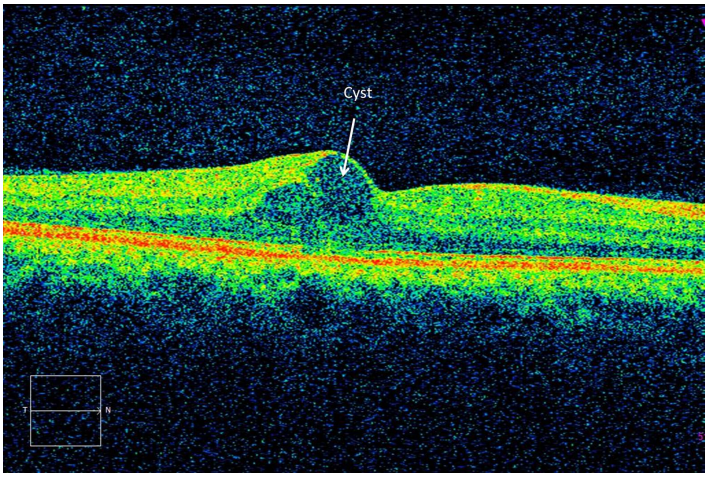
6.3/ UNE APPROCHE BASÉE « DÉTECTION D'ANOMALIE »

Dans cette section, nous abordons le problème de la classfication de volumes OCT sous un angle différent, à savoir la détection, dans un volume OCT, des images (B-scans) présentant des signes pathologiques. En effet, l'étape suivante du processus d'analyse consiste à segmenter dans les images les signes caractéristiques de la pathalogie tels que les kystes et évaluer leur taille (la figure 6.6 montre quelques signes de l'OMD). Pour ce faire, il faut identifier les B-scans qui présentent ces signes. Or, à l'intérieur d'un volume OCT (de 128 B-scans comme dans le cas des données que nous utilisons) tous les B-scans ne présentent pas de signes pathologiques même pour un patient atteint d'OMD.

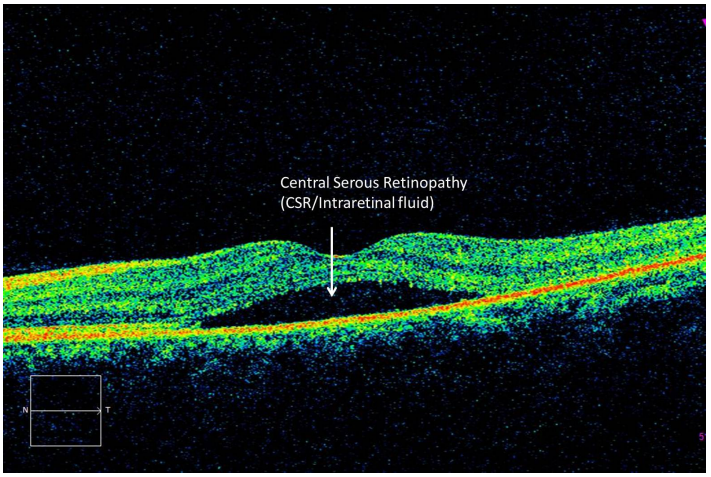
On pourait utiliser une approche de classification supervisée comme dans la section précédente, mais cela nécessite un effort important d'annotation pour constituer un ensemble d'apprentissage avec des B-scans sains et des B-scans pathologiques. Nous proposons donc une méthode qui ne nécessite pas un ensemble d'apprentissage avec des exemples positifs et négatifs. Nous adoptons une approche basée sur la « détection d'anomalie », en modélisant l'apparence des images OCT saines par un modèle de mélange de Gaussiennes (GMM) et en détectant les images OCT anormales par rapport à ce modèle.

La méthode se décompose en trois principales étapes :

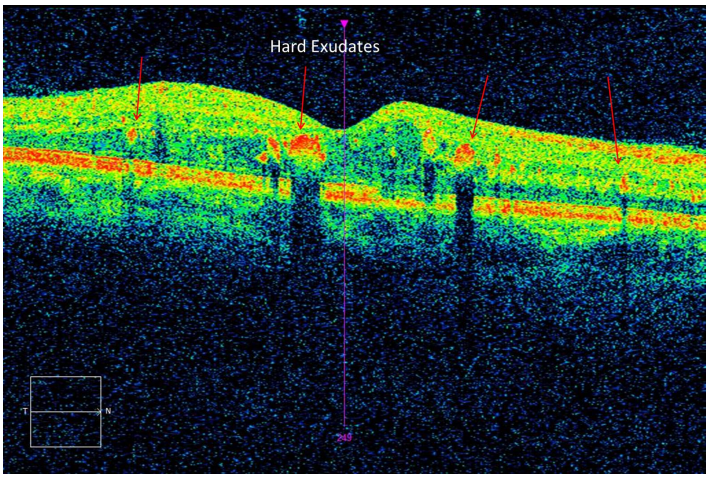
1. La création d'un modèle, GMM, qui représente l'apparence moyenne (distribution des intensités) des B-scans normaux.



(a) Kyste rétinien.



(b) Accumulation de fluide dans l'espace sous rétinien.



(c) Exsudats secs.

FIGURE 6.6 – Quelques exemples de signes caractéristiques de l'OMD dans les images OCT. NOTE : les images sont présentées en « fausses couleurs » uniquement pour visualisation.

2. La détection de B-scans anormaux en les condésirant comme des « anomalies » par rapport au modèle, i.e. les B-scans dont la distribution d'intensité dévie par rapport au modèle.
3. La classification de volumes OCT en fonction du nombre de B-scans anormaux détectés.

Création du modèle Nous représentons l'apparence globale des B-scans normaux par un modèle de mélange de Gaussiennes (GMM). Plus précisémment, nous disposons d'un ensemble de N volumes OCT qui sont tous de patients sains. Comme illustré par la figure 6.4(a), chaque volume est composé d'un nombre n_B de B-scans de taille $W \times H$. Nous représentons chaque B-scan par un vecteur $\mathbf{b}$ de dimension $d = WH$, et un volume OCT est donc représenté par un ensemble de vecteurs $V = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{n_B}\}$, où chaque $\mathbf{b}_i \in \mathcal{R}^d$.

En regroupant tous les B-scans des N volumes sains, nous obtenons une matrice de données $\mathbf{X} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_M]$, avec $M = Nn_B$ le nombre total de B-scans sains. Dans un premier temps, nous appliquons une ACP pour réduire la dimension des données car chaque B-scan est représenté par un vecteur $\mathbf{b}_i \in \mathcal{R}^d$, et $d = WH$ est assez grand ($d = 204800$ pour des B-scans de taille 400×512). L'ACP reduit la dimension à, typiquement, $p = 300$ ou $p = 500$ en retenant 95% de la variance totale des données.

Finalement, dans l'espace de dimension p, nous décrivons l'apparence globale des B-scans sains par un GMM défini comme combinaison linéaire de K Gaussiennes :

$$p(\mathbf{x} \mid \theta) = \sum_{i=1}^{K} w_i g_i(\mathbf{x} \mid \mu_i, \Sigma_i), \tag{6.3}$$

où $\mathbf{x} \in \mathcal{R}^p$, les w_i, $i = 1, \ldots, K$ sont les poids de la combination tels que $\sum_{i=1}^{K} w_i = 1$, et $g_i(\mathbf{x} \mid \mu_i, \Sigma_i)$ est la i-ème composante du mélange.

Chaque composante du mélange est une Gaussienne multivariée définie par :

$$g_i(\mathbf{x} \mid \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i)\}, \tag{6.4}$$

avec $\mu_i \in \mathcal{R}^p$ la moyenne, et Σ_i la matrice de covariance de taille $p \times p$.

Les paramètres du modèle $\theta = \{w_i, \mu_i, \Sigma_i, i = 1, \ldots, K\}$ sont estimés à l'aide d'un algorithme itératif tel que l'espérance-maximisation (expectation-maximization [EM]) [114]. Dans la pratique, un seul paramètre global définit le modèle. Il s'agit du nombre K de composantes. Une fois fixée la valeur de K, l'algorithme EM estime l'ensemble des paramètres θ. D'autre part, pour une meilleure initialisation, nous utilisons d'abord un algorithme de clustering (par exemple k-means) pour créer K clusters à partir des M images d'apprentissage. Et nous représentons chaque cluster par une Gaussienne multivariée pour initialiser l'algorithme EM.

Le nombre K de composantes du modèle est choisi par validation croisée comme nous l'expliquons ci-dessous, et la figure 6.7 montre un aperçu global de la méthode de création du modèle.

Identification des B-scans pathologiques Une fois le modèle créé a partir d'un ensemble de volumes normaux, nous détectons les B-scans anormaux dans un volume test de la manière suivante :

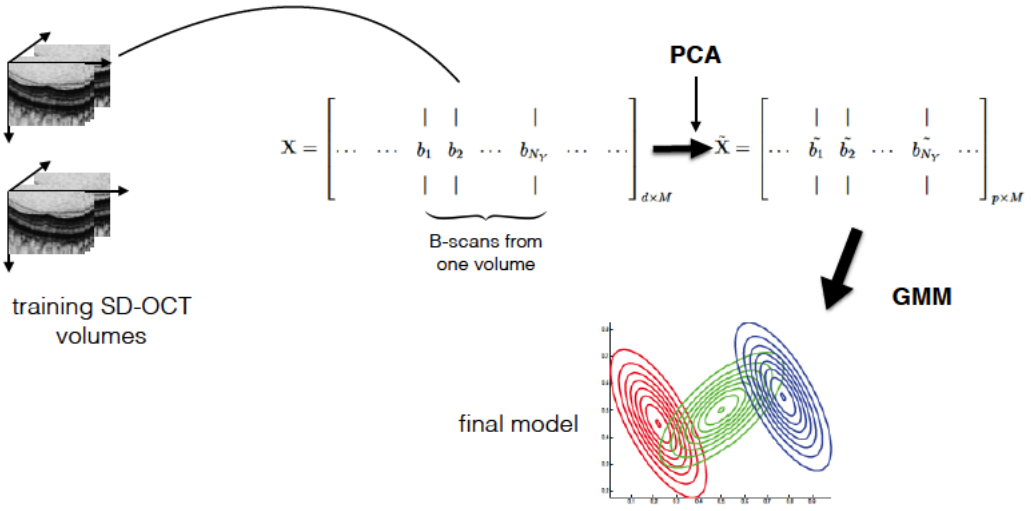


FIGURE 6.7 – Schéma général de la méthode de création du modèle GMM.

— Chaque B-scan $\mathbf{b} \in \mathcal{R}^d$ est projeté dans le sous-espace défini par les composantes principales calculées à l'étape précédente. On obtient un vecteur $\tilde{\mathbf{b}} \in \mathcal{R}^p$.
— Nous calculons la distance de Mahalanobis de $\tilde{\mathbf{b}}$ par rapport au modèle GMM :

$$\Delta_{GMM}(\tilde{\mathbf{b}}) = \arg\min_i \Delta_i(\tilde{\mathbf{b}}), \tag{6.5}$$

où $\Delta_i(\tilde{\mathbf{b}})$ est la distance de $\tilde{\mathbf{b}}$ par rapport à la i-ème composante du modèle et est définie par :

$$\Delta_i(\tilde{\mathbf{b}}) = (\tilde{\mathbf{b}} - \mu_i)^T \Sigma_i^{-1} (\tilde{\mathbf{b}} - \mu_i). \tag{6.6}$$

— Le B-scan est considéré comme anormal si la distance est supérieur à un seuil δ :

$$\mathbf{b} \quad \text{est anormal si} \quad \Delta_{GMM}(\tilde{\mathbf{b}}) > \delta. \tag{6.7}$$

Le choix du seuil δ est important pour limiter les fausses détections. Nous fixons ce seuil de manière automatique en généralisant la règle des « trois sigmas » (règle des 3σ) au cas multidimensionnel.

Rappelons que si $\mathbf{x}$ est une variable aléatoire 1D qui suit une distribution normale de moyenne μ est de variance σ^2, alors $P(|\mathbf{x} - \mu| \leq 3\sigma) = 0.997$. En d'autres termes, la quasi-totalité des valeurs se situe à une distance inférieure ou égale à 3σ de la moyenne, et les valeurs qui dévient de cette règle peuvent être considérées comme des « anomalies » (i.e. des valeurs aberrantes). Cette règle peut être étendue au cas de Gaussiennes multivariées en constatant que la distance de Mahalanobis suit une loi du χ^2 à p degrés de liberté [114, 22] : $\Delta_{GMM} \sim \chi^2_p$. On peut donc considérer comme « anomalies » les points de $\mathcal{R}^p$ dont la distance de Mahalanobis par rapport au modèle, se situe au delà de la 99ème centile de la loi du χ^2_p, soit $\delta = \chi^2_{p:0.99}$.

La figure 6.8 montre, de manière schématique, la procédure de détection des B-scans anormaux.

Classification des volumes OCT Nous adoptons une règle de classification simple : un volume OCT est considéré comme normal (ou anormal) en fonction du nombre de B-scans anormaux détectés dans le volume. Idéalement, un volume normal ne devrait

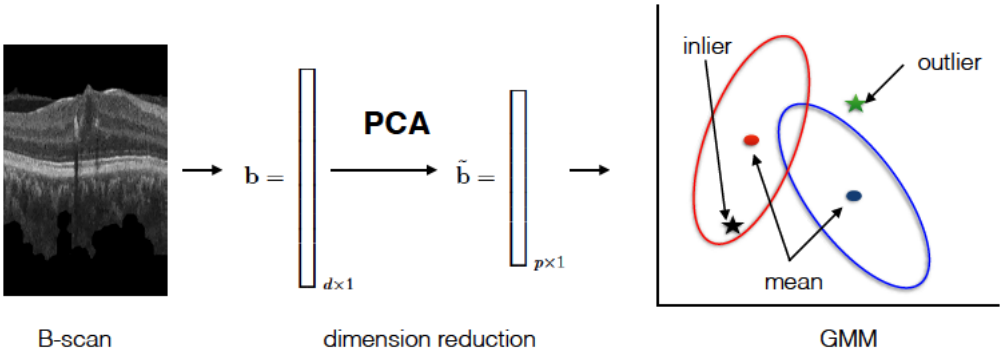


FIGURE 6.8 – Procédure de détection des B-scans anormaux. Illustration dans le cas $p = 2$ et $K = 2$ Gausiennes.

comporter aucun B-scan anormal, et, inversement, un volume anormal doit comporter plusieurs B-scans pathologiques. Dans la pratique, nous définissons un seuil empirique N_a sur le nombre de B-scans anormaux détectés.

De plus, pour accroître la robustesse de la méthode, nous constatons que les signes de la pathologie sont visibles dans plusieurs B-scans successifs. Plus précisément nous imposons que si un B-scan $\mathbf{b}_i \in V = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{n_B}\}$ est détecté comme anormal, alors au moins l'un de ses voisins $\mathbf{b}_{i-1}$ ou $\mathbf{b}_{i+1}$ doît également être détecté comme anormal. Cette règle simple, réduit le nombre de fausses détections.

Résultats Nous utilisons les mêmes données que dans la section 6.2 pour l'évaluation, i.e. un ensemble de 32 volumes OCT (16 volumes normaux et 16 volumes anormaux).

Les deux paramètres importants de la méthode proposée sont le nombre de composantes du modèle GMM (K) et le seuil (N_a) pour décider si un volume est normal ou non. Le premier paramètre est fixé pendant la phase de création du modèle par validation croisée, et le second est fixé de manière empirique pendant la phase de test de façon à optimiser le taux de bonne classification.

Pour l'apprentissage (i.e. la création du modèle GMM) nous sélectionnons de manière aléatoire 11 des 16 volumes normaux. Pour une valeur de K fixé, un ensemble de 8 volumes sur les 11 sélectionnés (ce qui correspond à 1024 B-scans) est utilisé pour créé le modèle et les 3 volumes restants (i.e. 384 B-scans) sont utilisés pour la validation. Ce processus de validation croisée est répété 10 fois (en choisissant à chaque fois 8 volumes pour la création du modèle et 3 pour la validation) et nous calculons le taux moyen de bonne classification. La figure 6.9 montre les résultats obtenus et on note que la valeur $K = 5$ donne les meilleurs résulats. Nous créons donc le modèle final avec $K = 5$ et en utilisant l'ensemble des 11 volumes d'entraînement.

Pour tester le modèle, nous utilisons les 5 volumes normaux (non employés pour créer le modèle GMM) et les 16 volumes anormaux. Pour chaque volume, nous détectons le nombre de B-scans anormaux et le volume est considéré comme anormal si ce nombre est supérieur au seuil N_a fixé. Les résultats du tableau 6.3 montre qu'avec un seuil bas des volumes normaux sont incorrectement classés comme anormaux, d'où une faible spécificité. A l'inverse, avec un seuil élevé, des volumes anormaux sont incorrectement classés comme normaux, et on a une faible sensibilité. Les meilleurs résultats sont ob-

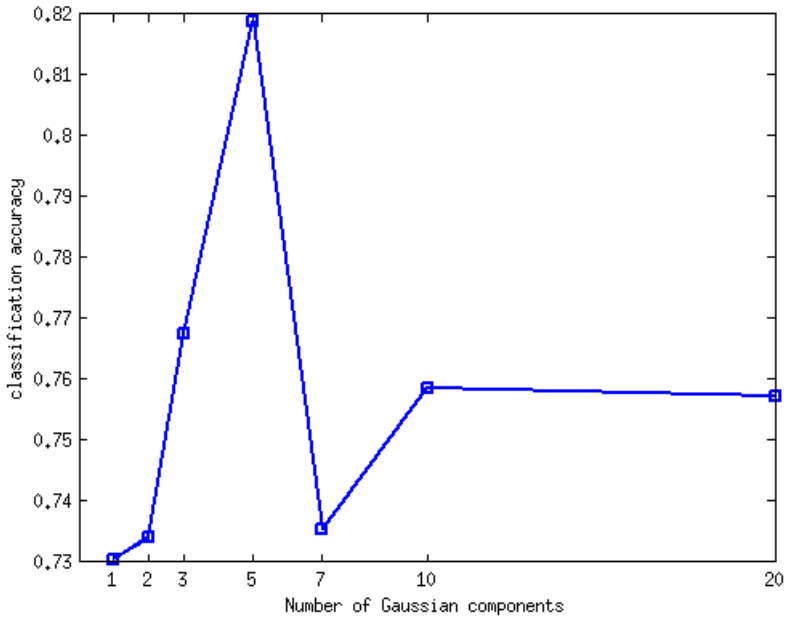


FIGURE 6.9 – Variation du taux de classification correct des B-scans en fonction du nombre K de composantes.

Seuil (N_a)	2	4	6	8	10
Sensibilité (%)	100	93.75	87.50	75.00	62.50
Spécificité (%)	60.00	80.00	80.00	100	100

TABLE 6.3 – Variation du taux de classification correct des volumes en fonction du seuil N_a.

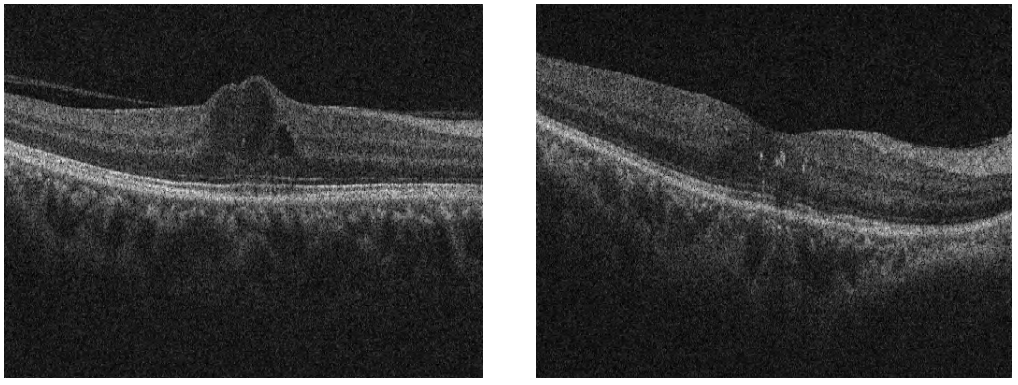
tenus avec un seuil $N_a = 4$ qui donne une sensibilité de 93.75% et une spécificité de 80%.

Enfin, nous comparons la méthode proposée avec différentes autres méthodes de la littérature, de même qu'avec la méthode décrite à la section 6.2 basée sur l'extraction de caractéristiques locales. Comme le montre les résultats rassemblés dans le tableau 6.4 notre approche basée « détection d'anomalie » obteint de très bons résultats comparés aux autres approches. Elle obtient la meilleure sensibilité avec une valeur de 93.75%, mais une spécificité inférieur par rapport à l'approche proposée dans la section 6.2. Néanmoins, un avantage considérable de la nouvelle méthode est qu'elle permet l'identification des B-scans anormaux à l'intérieur d'un volume OCT. Cela est important car on évite ainsi l'inspection de tous les B-scans du volume. De plus, une fois les B-scans anormaux détectés on peut s'intéresser à la segmentation de signes spécifiques tels que les kystes ou les exsudats. La figure 6.10 montre des exemples de B-scans anormaux identifiés dans un volume OCT.

Conclusions La méthode proposée permet d'identifier les B-scans présentant des signes pathologiques et, par la même, de distinguer les volumes sains de ceux anormaux. D'autre part, elle ne nécessite pas l'annotation manuelle des B-scans pour l'ap-

Méthode	Classifieur	Sensibilité (%)	Spécificité (%)
Proposée (GMM)	-	93.75	80
Proposée (section 6.2, locale - 2D)	SVM	81.2	93.7
Srinivasan *et al.* [157]	SVM	68.8	93.8
Liu *et al.* [91]	SVM	68.8	93.8
Venhuizen *et al.* [171]	RF	61.5	58.8
Lemaître et al. [88]	RF	87.5	75

TABLE 6.4 – Comparaison de différentes méthodes de classification de volumes OCT.



(a) Image OCT présentatnt un kyste. (b) Image OCT présentant des exsudats.

FIGURE 6.10 – Exemples de B-scans anormaux détectés à l'intérieur d'un volume OCT.

prentissage, ce qui permet une mise en œuvre rapide.

6.4/ CONCLUSIONS ET DISCUSSION

Dans ce chapitre, nous avons abordé le problème de la classification automatique d'images OCT pour la détection de patients atteints d'œdèmes maculaires diabétiques (OMD) associés à la rétinopathie diabétique (RD).

Nous avons, dans un premier temps, abordé ce problème sous l'angle classiquement adopté dans la littérature, à savoir par une méthode de classification supervisée. En analysant en détail différentes méthodes de représentation des volumes et images OCT, nous avons proposé une méthode basée sur une représentation par sac de mots et le descripteur LBP. Ce travail a également mis en évidence la supériorité d'un descripteur de texture, par rapport au gradient ou à l'intensité.

Dans un second temps, nous avons voulu identifier les B-scans (i.e. les images 2D formant un volume OCT) présentant des signes pathologiques de manière automatique. En effet, c'est en analysant individuellemnt chaque B-scan d'un volume, que le patient est déclaré sain ou non. On perçoit donc clairement l'intérêt d'une méthode automatique. Nous avons proposé une méthode qui ne nécessite pas un ensemble d'apprentissage annoté avec des exemples positifs et négatifs, ce qui est indisponible. Notre approche est basée sur la « détection d'anomalie », en modélisant l'apparence des images OCT saines par un modèle de mélanges de Gaussiennes (GMM) et en détectant les images

OCT anormales (pathalogiques) par comparaison avec ce modèle.

Les résultats obtenus montrent que nos deux approches sont très compétitives par rapport aux travaux de la littérature. Toutefois, les méthodes proposées peuvent être améliorées à plusieurs niveaux :

- Dans la méthode basée GMM, nous utilisons comme caractéristique l'intensité pour construire le modèle. Or, l'approche supervisée a montré que les descripteurs de texture (type LBP) sont plus performants. On peut facilement intégerer ces caractéristiques dans le modèle basé GMM pour améliorer les résultats.
- Nous avons employé une base de données OCT composé d'un nombre limité de volumes, 32 exactement. Cela est principalement dû à l'absence d'une base de donnée publique plus large correspondant à nos besoins. Il existe bien une base publique [157], mais elle contient uniquement 45 volumes (15 cas sains, 15 cas d'OMD et 15 cas de DMLA). De plus, les auteurs ne fournissent pas les données brutes, i.e. les volumes OCT non traités, mais uniquement les B-scans déjà pré-traités, et différents volumes ont un nombre différent de B-scans. Les choix de pré-traitement ne sont pas indiqués. Nous avons donc uniquement travaillé avec les 32 volumes fournis par nos collègues du SERI (Singapore Eye Rsearch Institute). Dans un futur proche, nous allons évaluer nos méthodes avec une base de données plus importante.

Ce travail a été, en grande partie, réalisé dans le cadre d'un projet international, le PHC Merlion en partenariat avec le SERI (2015-2016) [88, 150]. Cette collaboration internationale se poursuit et a été étendue à Honk Kong (Chineese University) depuis cette année. Nous constituons en ce moment, avec les collègues de Honk Kong, une base de données importante de plusieurs centaines de volumes OCT (soit plusieurs dizaines de milliers d'images OCT) que nous envisageons de rendre accessible à l'ensemble de la communauté. Cela permettra d'évaluer plus précisément les méthodes développées et de faciliter les comparaisons.

IV

CONCLUSION ET PERSPECTIVES

7

CONCLUSION GÉNÉRALE

Dans ce dernier chapitre, nous faisons un bilan de nos activités de recherche et proposons un projet de recherche pour les années à venir. Celui-ci a pour but de poursuivre et finaliser les travaux en cours, mais aussi d'ouvrir de nouvelles perspectives de travail.

7.1/ BILAN

Dans ce mémoire, nous avons présenté nos principales contributions scientifiques à l'analyse et à l'interprétation d'images de différents types : images RGB, images catadioptriques, images de profondeur et images OCT.

Dans chacun des deux domaines d'application considérés, à savoir l'analyse de scènes dynamiques et le dépistage de la rétinopathie diabétique, nous avons obtenu des résultats satisfaisants en proposant des méthodes rapides et nécessitant peu de supervision :

— **Saillance spatio-temporelle :** Nous avons proposé une méthode de détection de régions visuellement saillantes basée sur une combinaison de la couleur et la texture. En particulier, les résultats obtenus montrent que l'utilisation de texture dynamique (avec des LBP-TOP) permet d'estimer correctement une carte de saillance dynamique lorsque le mouvement relatif de l'objet par rapport au fond de la scène est assez faible ou peu perceptible. Nous avons également évalué différentes méthodes de fusion pour la combinaison des cartes de saillance statique et temporelle. Enfin, pour tenir compte des fortes corrélations spatio-temporelles qui existent entre les images successives d'une séquence vidéo, nous avons proposé une autre méthode directe basée sur une ACP multidimensionnelle. Cette approche assez simple, ne necéssite pas d'étape de fusion et donne de bons résultats. Toutefois, elle est assez coûteuse en temps de calcul.

— **Suivis d'objets dans les images catadioptriques :** En utilisant le modèle sphérique pour la représentation des images, nous avons proposé une méthode d'adaptation des méthodes existantes de suivi, qui permet de tenir compte de la géométrie particulière des capteurs et des images catadioptriques. Les résultats obtenus montrent que les méthodes déterministes (mean-shift par exemple), tout comme les approches probabilistes (filtre particulaire) peuvent être adaptées pour un suivi robuste avec une caméra catadioptrique, et même une caméra fisheye.

— **Descripteurs d'objets 3D :** Pour permettre l'emploi des capteurs de profondeur dans des sytèmes embarqués, nous avons montré qu'il est possible de réduire la

taille de différents descripteurs de nuages de points 3D acquis avec une Kinect, pour réduire la complexité sans sacrifier la performance des algorithmes de reconnaissance. De plus, en combinant les propriétés géométriques et de texture extraites du nuage de point, nous définissons un descripteur global plus performant en terme de robustesse au bruit (bruit Gaussien) et de reconnaissance d'objets et de catégorie d'objets.

— **Détection de lésions rétiniennes :** Le dépistage de la rétinopathie diabétique étant basé sur la détection de lésions dans les images de fond d'œil, nous avons proposé différentes méthodes pour la détection de microanévrismes et d'exsudats. Par rapport aux méthodes proposées dans la littérature, nos approches nécessitent peu ou pas d'images manuellement annotées au niveau des lésions à détecter, tout en obtenant de très bons résultats de détection.

— **Classification d'images OCT :** Pour la détection de patients atteints d'œdèmes maculaires diabétiques (OMD) associés à la rétinopathie diabétique (RD), nous avons proposé une méthode efficace de classification de volumes OCT. Notre approche permet, non seulement, de classer un volume entier, mais également de détecter les B-scans pathologiques à l'intérieur du volume 3D.

7.2/ PERSPECTIVES

Même si nous avons obtenu des résultats satisfaisants tout en limitant le degré de supervision nécessaire pour l'apprentissage, les méthodes de classification employées restent simples et nos approches peuvent être améliorées.

La tendance actuelle pour l'extraction de caractéristiques dans les images et les vidéos est l'utilisation de réseaux de neurones profonds (deep learning architectures). Ces réseaux, en particulier les CNNs (convolutional neural networks), sont capables d'extraire, de manière automatique, des représentations pertinentes des données pour obtenir des résultats de classifications impressionants dans des domaines d'applications très variés (médicale, robotique, fouille de données, etc). La raison principale du succès actuel de ces approches connues depuis des décennies, tient, d'une part, au fait que nous disposons à présent de très grandes bases de données annotées pour apprendre les nombreux paramètres de ces réseaux, et, d'autre part, au fait que nous disposons de moyens de calcul importants (GPU) pour l'apprentissage. Toutefois, ces approches nécessitent des adaptations pour chaque application. Par exemple, pour la classification de volumes OCT, il est nécessaire de définir des opérations de convolutions et d'agrégation 3D pour traiter le volume dans son intégralité. Nous travaillons actuellement dans ce sens dans le cadre d'une collaboration avec le LITIS à Rouen.

Un autre aspect important que nous souhaitons aborder, est l'extraction automatique de caractéristiques dans des données hétérogènes, i.e. des données de différentes natures ou de modalités différentes. Par exemple, un véhicule autonome est généralement équipé de différents capteurs qui fournissent des informations de natures différentes : des images RGB, des images de profondeurs, un nuage de points 3D, etc. L'utilisation conjointe de toutes ces informations constitue un enjeu important pour la localisation et la navigation autonome. Dans le domaine médical, en plus des données images (OCT par exemple), nous avons accès à diverses autres informations sur le patient telles que son âge, son sexe, ses antécédents, des données génétiques, etc. Toutes ces informations peuvent être combinées pour fournir une meilleure interprétation des images.

Dans les deux cas, plutôt que de traiter séparément chaque information, puis de les combiner a posteriori, ils nous semble plus intéressant d'adopter une approche qui traite simultanément toutes les informations disponibles pour en extraire les attributs les plus pertinents pour une application donnée. En effet, chacune des informations, ou chaque modalité, peut être considérée comme une vue différente (ou une projection différente) de l'espace des données. En utilisant conjointement toutes les vues disponibles, l'algorithme d'apprentissage exploite la redondance des données. Nous nous intéresserons donc aux algorithmes d'apprentissage multi-vues (multiview learning) en les combinant avec les réseaux profonds pour l'extraction de caractéristiques. Ce travail sera, en partie, réalisé dans le cadre de la thèse de Nathan Piasco (2016-2019) que je co-encadre dans le cadre de l'ANR PLATINUM. L'objectif de ce travail est la localisation d'un agent mobile (piéton, véhicule, etc.) dans une scène 3D dynamique représentée par des sphères de vues qui contiennent à la fois des informations couleur, 3D et sémantiques. Il faut donc comparer des données hétérogènes et estimer la pose 3D de l'agent.

A plus long terme, je pense que la combinaison des approches semi-supervisées avec les méthodes actives d'apprentissage (apprentissage par renforcement par exemple) est un sujet prometteur. En effet, l'utilisation d'un nombre croissant d'exemples annotés lors de l'apprentissage de méthodes partiellement supervisées accroît généralement leur performances. Il semble donc intéressant de pouvoir inclure quelques exemples annotés « bien choisis » au cours de la phase d'apprentissage. Les problèmes à résoudre pour une amélioration optimale des performances des algorithmes concernent :

— **la sélection des exemples :** quel critère utiliser pour sélectionner les « bons » exemples à inclure dans l'ensemble d'apprentisage ?

— **l'arrêt de l'apprentissage :** quel critère pour terminer l'apprentissage du model ?

Cette approche me semble particulièrement intéressante pour l'analyse d'images médicales, mais elle peut très bien s'appliquer à l'analyse de scènes dynamiques dans le domaine de la robotique.

BIBLIOGRAPHIE

[1] S. Abdelazeem. Micro-aneurysm detection using vessels removal and circular hough transform. In *Radio Science Conference, 2002. (NRSC 2002). Proceedings of the Nineteenth National*, pages 421 – 426, 2002.

[2] M.D. Abràmoff, M.K. Garvin, and M. Sonka. Retinal imaging and image analysis. *Biomedical Engineering, IEEE Reviews in*, 3 :169–208, 2010.

[3] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned Salient Region Detection. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pages 1597 – 1604, 2009.

[4] R. Achanta and S. Susstrunk. Saliency detection for content-aware image resizing. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1005–1008. IEEE, 2009.

[5] R. Achanta and S Susstrunk. Saliency detection using maximum symmetric surround. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2653–2656. IEEE, 2010.

[6] K. M. Adal, S. Ali, D. Sidibé, T. Karnowski, E. Chaum, and F. Meriaudeau. Automated detection of microaneurysms using robust blob descriptors. In *Proc. of SPIE Vol*, volume 8670, 2013.

[7] K. M. Adal, D. Sidibé, S. Ali, E. Chaum, T. Karnowski, and F. Meriaudeau. Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning. *Computer Methods and Programs in Biomedicine*, 114(1) :1–10, 2014.

[8] M. Aharon, M. Elad, and A. Bruckstein. K-svd : An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. on Signal Processing*, 45(11) :4311–4322, 2006.

[9] A. Albarrak, F. Coenen, and Y. Zheng. Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction. In *17th Annual Conference in Medical Image Understanding and Analysi*, pages 59–64, 2013.

[10] A. Aldoma, Z.C. Marton, F. Tombari, W. Wohlkinger, C. Potthast, B. Zeisl, R. B. Rusu, S. Gedikli, and M. Vincze. Point cloud library. *IEEE Robotics & Automation Magazine*, 1070(9932/12), 2012.

[11] A. Aldoma, F. Tombari, R. B. Rusu, and M. Vincze. Our-cvfh–oriented, unique and repeatable clustered viewpoint feature histogram for object recognition and 6dof pose estimation. In *Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium*, pages 113–122. Springer, 2012.

[12] A. Aldoma, M. Vincze, N. Blodow, D. Gossow, S. Gedikli, R. B. Rusu, and G. Bradski. Cad-model recognition and 6dof pose estimation using 3d cues. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 585–592. IEEE, 2011.

[13] L. A. Alexandre. 3d descriptors for object and category recognition : a comparative evaluation. In *Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Portugal*, volume 1, page 7. Citeseer, 2012.

[14] S. Ali, D. Sidibé, K.M. Adal, L. Giancardo, E. Chaum, T. Karnowski, and F. Meriaudeau. Statistical atlas based exudate segmentation. *Computerized Medical Imaging and Graphics*, 37(5-6) :358–368, 2013.

[15] N. Anantrasirichai, A. Achim, J.E. Morgan, I. Erchova, and L. Nicholson. Svm-based texture classification in optical coherence tomography. In *IEEE Symposium on Biomedical Imaging*, pages 1332–1335. IEEE, 2013.

[16] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics (TOG)*, 26(3) :10, 2007.

[17] S. Baker and S. K. Nayar. A theory of catadioptric image formation. In *Proceedings of the 6th International Conference on Computer Vision*, pages 35–42, 1998.

[18] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3) :346–359, 2008.

[19] J. Behley, V. Steinhage, and A. B. Cremers. Performance of histogram descriptors for the classification of 3d laser range data in urban environments. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 4391–4398. IEEE, 2012.

[20] R. Benosman and S. B. Kang. *Panoramic Vision : Sensors, Theory, and Applications*. Springer-Verlag, May 2001.

[21] N. P. Bichot. Attention, eye movements, and neurons : Linking physiology and behavior. In *Vision and attention*, pages 209–232. Springer, 2001.

[22] C. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[23] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100. ACM, 1998.

[24] F. L. Bookstein. Principal warps : thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6) :567–585, 1989.

[25] A. Borji, M.M. Cheng, H. Jiang, and J. Li. Salient object detection : A benchmark. *IEEE TIP*, 24(12) :5706–5722, 2015.

[26] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on PAMI,*, 35(1) :185–207, 2013.

[27] A. Borji, D.N. Sihite, and L. Itti. Salient object detection : A benchmark. In *ECCV (2)*, pages 414–429, 2012.

[28] M. Born and W. E. *Principles of Optics*, chapter Interference and Diffraction with Partially Coherent Light, pages 491–505. Pargamon Press, 1970.

[29] A. Buades, B. Coll, and J.M. Morel. A non-local algorithm for image denoising. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 60–65. IEEE, 2005.

[30] C. K. Chang, C. Siagian, and L. Itti. Mobile robot vision navigation & localization using gist and saliency. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 4147–4154. IEEE, 2010.

[31] L. Chang, P. C. Yuen, and G. Qiu. Object motion detection using information theoretic spatio-temporal saliency. *Pattern Recognition*, 42(11) :2897–2906, 2009.

[32] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-supervised learning*, volume 2. MIT press Cambridge, MA :, 2006.

[33] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuits. *SIAM Review*, 43(1) :129–159, 2001.

[34] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S.M. Hu. Global contrast based salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 409–416. IEEE, 2011.

[35] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 :564–575, 2003.

[36] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot. Nonlocal means-based speckle filtering for ultrasound images. *IEEE TIP*, pages 2221–2229, Oct 2009.

[37] J. Courbon, Y. Mezouar, L. Eckt, and P. Martinet. A generic fisheye camera model for robotic applications. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1683–1688. IEEE, 2007.

[38] M.J. Cree, E. Gamble, and DJ Cornforth. Colour normalisation to reduce inter-patient and intra-patient variability in microaneurysm detection in colour retinal images. In *Proceedings of WDIC*, pages 163–168, 2005.

[39] K. Daniilidis, A. Makadia, and T. Bulow. Image processing in catadioptric planes : spatiotemporal derivatives and optical flow computation. In *In IEEE Workshop on Omnidirectional Vision*, pages 3–10, 2002.

[40] C. Demonceaux and P. Vasseur. Omnidirectional image processing using geodesic metric. In *Proceedings of IEEE International Conference on Image processing*, ICIP'09, pages 221–224, 2009.

[41] A. Dubois. Tomographie par cohérence optique, 2011. Institut d'Optique, Université Paris Sud.

[42] J. Duncan and G. W. Humphreys. Visual search and stimulus similarity. *Psychological physics*, 57 :117–120, 1989.

[43] M. Elad. *Sparse and Redundant Representations : From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[44] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th SCIA*, pages 363–370, 2003.

[45] A. F. Fercher. Optical coherence tomography. *J. Biomed. Opt*, 1 :157–173, 1996.

[46] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser. Optical coherence tomography - principles and applications. *Repports on Progress in Physics*, 66(2) :239–303, 2003.

[47] A. F. Fercher, C. K. Hitzenberger, G. Kamp, and S ; Y. El-Zaiat. Measurement of intraocular distances by backscattering spectral interferometry. *Optics Communications*, 117(1) :43–48, 1995.

[48] D.S. Fong, L.P. Aiello, F.L. Ferris, and Klein P. Diabetic retinopathy. *Diabetes Care*, 27(10) :2540–2553, 2004.

[49] M. Foracchia, E. Grisan, and A. Ruggeri. Luminosity and contrast normalization in retinal images. *Medical Image Analysis*, 9(3) :179–190, 2005.

[50] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, and A. R. Rudnicka. Blood vessel segmentation methodologies in retinal images - a survey. *Computer Methods and Programs in Biomedicine*, 108(1) :407–433, 2012.

[51] S. Frintrop. *Computational Visual Attention*. Springer, 2011.

[52] S. Frintrop and P. Jensfelt. Attentional landmarks and active gaze control for visual slam. *Trans. Rob.*, 24(5) :1054–1065, October 2008.

[53] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *European conference on computer vision*, pages 224–237. Springer, 2004.

[54] F.J. Fujimoto, M.E. Brezinski, G.J. Tearney, S.A. Boppart, B.E. Bouma, M.R. hee, J.F. Southern, and E.A. Swanson. Optical biopsy and imaging using optical coherence tomography. *Nature Med.*, 1 :970–972, 1995.

[55] M. García, C. I. Sánchez, M. I. López, D. Ab''asolo, and R. Hornero. Neural network based detction of hard exudates in retinal images. *Computer Methods and Programs in Biomedicine*, 93(1) :9–19, 2009.

[56] M.K. Garvin, M.D. Abramoff, R. Kardon, S.R ; Russell, W. Xiaodong, and M. Sonka. Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-d graph search. *IEEE Trans. on Medical Imaging*, 27(10) :1495–1505, 2008.

[57] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 45 :223–243, 2002.

[58] I. Ghorbel. *Segmentation et quantification des couches rétiniennes dans des images de tomographie de cohérence optique, dans le cas de sujets sains et pathologiques*. PhD thesis, Telecom ParisTech, 2012.

[59] I. Ghorbel, F. Rossant, I. Bloch, S. Tick, and M. Paques. Automated segmentation of macular layers in oct images and quantitative evaluation of performances. *Pattern Recognition*, 44(8) :1590–1603, 2011.

[60] L. Giancardo, MD Abramoff, E. Chaum, TP Karnowski, F. Meriaudeau, and KW Tobin. Elliptical local vessel density : a fast and robust quality metric for retinal images. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 3534–3537. IEEE, 2008.

[61] L. Giancardo, F. Meriaudeau, T. Karnowski, Y. Li, S. Garg, K. Tobin Jr, and E. Chaum. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical Image Analysis*, 16(1) :216–226, 2012.

[62] L. Giancardo, F. Meriaudeau, T.P. Karnowski, Y. Li, K.W. Tobin, and E. Chaum. Automatic retina exudates segmentation without a manually labelled training set. In *Biomedical Imaging : From Nano to Macro, 2011 IEEE International Symposium on*, pages 1396–1400. IEEE, 2011.

[63] L. Giancardo, F. Meriaudeau, TP Karnowski, Y. Li, KW Tobin, and E. Chaum. Microaneurysm detection with radon transform-based classification on retina images. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 5939–5942. IEEE, 2011.

[64] L. Giancardo, F. Meriaudeau, T.P. Karnowski, K.W. Tobin, Y. Li, and E. Chaum. Microaneurysms detection with the radon cliff operator in retinal fundus images. In *Proc. of SPIE Vol*, volume 7623, pages 76230U–1, 2010.

[65] R. Goebel, L. Mucklil, and D. S. Kim. *The Human Nervous System (Second Edition)*, chapter Visual System, pages 1280–1305. Elsiveir, 2004.

[66] S. Goferman, L. Zelnik-manor, and A. Tal. Context-aware saliency detection. In *IEEE Conf. on CVPR*, 2010.

[67] C. L. Guo and L. M. Zhang. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Transactions on Image Processing,*, 19(1) :185 –198, 2010.

[68] B. Han and B. Zhou. High speed visual saliency computation on gpu. In *Proc of ICIP*, 2007.

[69] J. Han, L. Shao, D. Xu, and J. Shotton. Enhanced computer vision with microsoft kinect sensor : A review. *IEEE Trans. Cybernetics*, 43(5) :1318–1334, 2013.

[70] Haute Autorité de Santé. Dépistage de la rétinopathie diabétique par lecture différée de photographies du fond d'œil. Technical report, Rapport de la Haute Autorité de Santé, décembre 2010.

[71] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stison, M. R ; Hee, T. Flotte, K. Grogory, C. A. Puliafito, and J. G. Fujimoto. Optical coherence tomography. *Science*, 254 :1178–1181, 1991.

[72] H. Hwang, S. Hyung, S. Yoon, and K. Roh. Robust descriptors for 3d point clouds using geometric and photometric local feature. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4027–4033. IEEE, 2012.

[73] M. Isard and A. Blake. Condensation ?conditional density propagation for visual tracking. *International journal of computer vision*, 29(1) :5–28, 1998.

[74] H. F. Jaafar, A. K Nandi, and W. Al-Nuaimy. Automated detection of red lesions from digital colour fundus photographs. In *Conf Proc IEEE Eng Med Biol Soc*, pages 6232–5, 2011.

[75] R. Kafieh, H. Rabbani, M. D Abramoff, and M. Sonka. Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map. *Medical image analysis*, 17(8) :907–928, 2013.

[76] R. Kafieh, H. Rabbani, and S. Kermani. A review of algorithms for segmentation of optical coherence tomography from retina. *Journal of Medicals Signals and Sensors*, 3(1) :45–60, 2013.

[77] V. Kajić, B. Povazay, B. Hermann, B. Hofer, D. Marshall, P.L. Rosin, and W. Drexler. Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis. *Optics Express*, 18(14) :14730–14744, 2010.

[78] Y. Ke and R. Sukthankar. Pca-sift : A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–506. IEEE, 2004.

[79] P.J. Kertes and T.M. Johnson. *Evidence-based eye care*. Lippincott Williams & Wilkins, 2007.

[80] K. Khoshelham and S. O. Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2) :1437–1454, 2012.

[81] W. Kim, C. Jung, and C. Kim. Spatiotemporal saliency detection and its applications in static and dynamic scenes. *IEEE Trans. Circuits Syst. Video Techn*, 21(4) :446–456, 2011.

[82] R. A. Kirsch. Computer determination of the constituent structure of biological images. *Computers and Biomedical Research*, pages 315–328, 1970.

[83] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1817–1824. IEEE, 2011.

[84] I. Lazar and A. Hajdu. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Trans. Biomed. Eng.*, 59(6) :1720–1726, 2012.

[85] I. Lazar and A. Hajdu. Retinal microaneurysm detection through local rotating cross-section profile analysis. *IEEE Trans. Medical Imaging*, 32(2) :400–407, 2013.

[86] I. Lazar, A. Hajdu, and R.J. Quareshi. Retinal microaneurysm detection based on intensity profile analysis. In *8th International Conference on Applied Informatics*, 2010.

[87] O. Le Meur, P. Le Callet, and D. Barba. Predicting visual fixations on video based on low-level visual features. *Vision research*, 47(19) :2483–2498, 2007.

[88] G. Lemaître, M. Rastgoo, J. Massich, S. Sankar, F. Mériaudeau, and D. Sidibé. Classification of SD-OCT volumes with LBP : Application to dme detection. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI), Ophthalmic Medical Image Analysis Workshop (OMIA)*, 2015.

[89] Y. Li, T.P. Karnowski, K.W. Tobin, L. Giancardo, S. Morris, S.E. Sparrow, S. Garg, K. Fox, and E. Chaum. A health insurance portability and accountability act–compliant ocular telehealth network for the remote diagnosis and management of diabetic retinopathy. *Telemedicine and e-Health*, 2011.

[90] T. Lindeberg. *Scale-space theory in computer vision*. Springer, 1994.

[91] Y-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and J. M. Rehg. Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Medical image analysis*, 15(5) :748–759, 2011.

[92] M. I. Lopez, C. Sanchez, and R. Hornero. Retinal image analysis to detect and quantify lesions associated with diabetic retinopathy. In *Association for Research in Vision and Ophthalmology*, 2003.

[93] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2) :91–110, 2004.

[94] L. ltti, C. Koch, and E. Neibur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on PAMI*, 20 :1254–1259, 1998.

[95] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. MPCA : Multilinear principal component analysis of tensor objects. *IEEE Trans. on Neural Networks*, 19(1) :18–39, 2008.

[96] H. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. A survey of multilinear subspace learning for tensor data. *Pattern Recognition*, 44(7) :1540–1551, 2011.

[97] T. Lu, Z. Yuan, Y. Huang, D. Wu, and H. Yu. Video retargeting with nonlinear spatial-temporal saliency fusion. In *ICIP*, pages 1522–4880, 2010.

[98] B.D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[99] E. Maggio and A. Cavallaro. Multi-part target representation for color tracking. In *Proc. Int. Conf. Image Processing*, pages 729–732, 2005.

[100] E. Maggio and A. Cavallaro. *Video Tracking : Theory and Practice*. Wiley, February 2011.

[101] V. Mahadevan and N. Vasconcelos. Saliency-based discriminant tracking. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1007–1013. IEEE, 2009.

[102] V. Mahadevan and N. Vasconcelos. Spatiotemporal saliency in dynamic scenes. *IEEE Transactions on PAMI,*, 32(1) :171 –177, 2010.

[103] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionarie. *IEEE Trans. on Signal Processing*, 41(12) :3397–3415, 1993.

[104] M. Mancas, N. Riche, J. Leroy, and B. Gosselin. Abnormal motion selection in crowds using bottom-up saliency. In *ICIP*, pages 229–232, 2011.

[105] S. Marat, T. H. Phuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Guérin-Dugué. Modelling spatio-temporal saliency to predict gaze direction for short videos. *IJCV*, 82(3) :231–243, 2009.

[106] R. Margolin, T. Ayellet, and L. Zelnik-Manor. What makes a patch salient. In *IEEE CVPR*, pages 1139–1146, 2013.

[107] S. Marrat. *Modeles de saillance visuelle par fusion d'informations sur la luminancen le mouvement et les visages pour la prédiction de mouvements oculaires lors de l'exploration de vidéos*. PhD thesis, Université de Grenoble, 2009.

[108] B. R. McClintic, J. I. McClintic, J. D. Bisofnamo, and R. C. Block. The relationship between retinal microvascular abnormalities and coronary heart disease : a review. *Am. J. Med.*, 83(4) :374.e1–374.e7, 2010.

[109] C.E. Metz. Receiver operating characteristic analysis : a tool for the quantitative evaluation of observer performance and imaging systems. *Journal of the American College of Radiology*, 3(6) :413–422, 2006.

[110] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara, and H. Fujita. Automated microaneurysm detection method based on double ring filter in retinal fundus images. *SPIE Medical Imaging 2009 : Computer-Aided Diagnosis*, 7260(1) :72601N, 2009.

[111] O. Morel, C. Stolz, F. Meriaudeau, and P. Gorria. Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging. *Applied optics*, 45(17) :4062–4068, 2006.

[112] S. Muddamsetty, D. Sidibé, A. Trémeau, and F. Mériaudeau. A performance evaluation of fusion techniques for spatio-temporal saliency detection in dynamic scenes. In *ICIP*, 2013.

[113] S. Muddamsetty, D. Sidibé, A. Trémeau, and F. Mériaudeau. Spatio-temporal saliency detection in dynamic scenes using local binary patterns. In *ICPR*, pages 2353–2358, 2014.

[114] K.P. Murphy. *Machine learning : a probabilistic perspective*. MIT Press, Cambridge, MA., 2012.

[115] M. Neimeijer, B. van Ginneken, S. Russell, M. Suttorp-Schulten, and M. Abramoff. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Investigative Ophthalmology & Visual Science*, 48(5) :2260–2267, 2007.

[116] M. Niemeijer, B. Van Ginneken, M.J. Cree, A. Mizutani, G. Quellec, C.I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, et al. Retinopathy online challenge : Automatic detection of microaneurysms in digital color fundus photographs. *Medical Imaging, IEEE Transactions on*, 29(1) :185–195, 2010.

[117] M. Niemeijer, B. Van Ginneken, J. Staal, M.S.A. Suttorp-Schulten, and M.D. Abràmoff. Automatic detection of red lesions in digital color fundus photographs. *Medical Imaging, IEEE Transactions on*, 24(5) :584–592, 2005.

[118] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba. Considering temporal variations of spatial visual distortions in video quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 3(2) :253–265, 2009.

[119] T. Novikova, A. Pierangelo, A. De Martino, A. Benali, and P Validire. Polarimetric imaging for cancer diagnosis and staging. *Opt. Photon. News*, 23(10) :26–33, 2012.

[120] J. Peng and Q. Xiaolin. Keyframe-based video summary using visual attention clues. *IEEE on MultiMedia,*, 17(2) :64 –73, 2010.

[121] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters : Contrast based filtering for salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 733–740. IEEE, 2012.

[122] P.J. Phillips. Visible manifestations of diabetic retinopathy. *Medicine Today*, 5(5) :83, 2004.

[123] M. Pietikäinen, G. Zhao, A. Hadid, and T. Ahonen. *Computer Vision Using Local Binary Patterns*. Number 40 in Computational Imaging and Vision. Springer, 2011.

[124] R. Pires, H. Jelinek, J. Wainer, S. Goldenstein, E. Valle, and A. Rocha. Assessing the need for referral in automatic diabetic retinopathy detection. *IEEE Trans. on Biomedical Engineering*, 60 :3391–3398, 2013.

[125] M. I. Posner and S. E. Petersen. The attention system of the human brain. *Annaual Reviews of Neuroscience*, 13 :25–42, 1990.

[126] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, and M. Sonka. Three-dimensional analysis of retinal layer texture : identification of fluid-filled regions in sd-oct of the macula. *IEEE Trans. on Medical Imaging*, 29 :1321–1330, 2010.

[127] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä. Segmenting salient objects from images and videos. In *Computer Vision–ECCV 2010*, pages 366–379. Springer, 2010.

[128] F. Rameau, C. Demonceaux, D. Sidibé, and D. Fofi. Control of a ptz camera in a hybrid vision system. In *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, volume 3, pages 397–405. IEEE, 2014.

[129] F. Rameau, A. Habed, C. Demonceaux, D. Sidibé, and D. Fofi. Self-calibration of a ptz camera using new lmi constraints. In *Asian Conference on Computer Vision*, pages 297–308. Springer, 2012.

[130] F. Rameau, D. Sidibé, C. Demonceaux, and D. Fofi. Tracking moving objects with a catadioptric sensor using particle filter. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 328–334. IEEE, 2011.

[131] F. Rameau, D. Sidibé, C. Demonceaux, and D. Fofi. Visual tracking with omnidirectional cameras : an efficient approach. *Electronics letters*, 47(21) :1, 2011.

[132] E. Ricci and R. Perfetti. Retinal blood vessel segmentation using line operators and support vector classification. *Medical Imaging, IEEE Transactions on*, 26(10) :1357–1365, 2007.

[133] A. Rocha, T. carvalho, H.F. Jelinek, S. Goldenstein, and J. Wainer. Points of interest and visual dictionaries for automatic retinal lesion detection. *IEEE Trans. on Biomedical Engineering*, 59 :2244–2253, 2012.

[134] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3) :211–252, 2015.

[135] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz. Aligning point cloud views using persistent feature histograms. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3384–3391. IEEE, 2008.

[136] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu. Fast 3d recognition and pose using the viewpoint feature histogram. In *Intelligent Robots and Systems (IROS)*, pages 2155–2162. IEEE, 2010.

[137] R. B. Rusu and S. Cousins. 3d is here : Point cloud library (pcl). In *Robotics and Automation, IEEE International Conference on*, pages 1–4. IEEE, 2011.

[138] I. Sadek, D. Sidibé, and F. Meriaudeau. Automatic discrimination of color retinal images using the bag of words approach. In *SPIE Medical Imaging*, pages 94141J–8, 2015.

[139] PJ Saine. Fundus photography : What is a fundus camera. *Ophthalmic Photographers Society*, 2006.

[140] Y. Salih. *Development of Point Cloud Descriptor for Robust 3D recognition*. PhD thesis, Université De Bourgogne, 2015.

[141] Y. Salih, A. S. Malik, D. Sidibé, MT Simsim, N. Saad, and F. Meriaudeau. Compressed vfh descriptor for 3d object classification. In *2014 3DTV-Conference : The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4. IEEE, 2014.

[142] Y. Salih, A. S. Malik, N. Walter, D. Sidibé, N. Saad, and F. Meriaudeau. Noise robustness analysis of point cloud descriptors. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 68–79. Springer, 2013.

[143] S. Salti, F. Tombari, and L. Di Stefano. Shot : unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125 :251–264, 2014.

[144] C. I. Sánchez, M. García, A. Mayo, M. I. López, and R. Hornero. Retinal image analysis based on mixture models to detect hard exudates. *Medical Image Analysis*, 13(4) :650–658, 2009.

[145] J.M. Schmitt, S. Xiang, and K.M. Yung. Speckle in optical coherence tomography. *Journal of Biomedical Optics*, 4(1) :95–105, 1999.

[146] H. J. Seo and P. Milanfar. Nonparametric bottom-up saliency detection by self-resemblance. In *Computer Vision and Pattern Recognition Workshops, 2009*, pages 45 –52, 2009.

[147] H. J. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*, 9(12) :15, 2009.

[148] D. Sidibé, D. Fofi, and F. Mériaudeau. Using visual saliency for object tracking with particle filters. In *Proc of EUSIPCO*, 2010.

[149] D. Sidibé, I. Sadek, and F. Meriaudeau. Discrimination of retinal images containing bright lesions using sparse coded features and svm. *Computers in Biology and Medicine*, 62 :175–184, 2015.

[150] D. Sidibé, S. Sankar, G. Lemaître, M. Rastgoo, C. Y. Massich, J. Cheung, G. S.W. Tan, D. Milea, E. Lamoureux, T. Y. Wong, and F. Mériaudeau. An anomaly detection approach for the identification of dme patients using spectral domain optical coherence tomography images. *Computer Methods and Programs in Biomedicine*, 139 :109–117, 2017.

[151] J. Sivic and A. Zisserman. Video google : a text retrieval approach to object matching in videos. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognitiong*, pages 1470–1477, 2003.

[152] J. Smisek, M. Jancosek, and T. Pajdla. 3d with kinect. In *Consumer Depth Cameras for Computer Vision*, pages 3–25. Springer, 2013.

[153] J.V.B. Soares, J.J.G. Leandro, R.M. Cesar, H.F. Jelinek, and M.J. Cree. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *Medical Imaging, IEEE Transactions on*, 25(9) :1214–1222, 2006.

[154] A. Sopharak, B. Uyyanonvara, S. Barman, and T.H. Williamson. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Computerized Medical Imaging and Graphics*, 32(8) :720–727, 2008.

[155] T. Spencer, John A. Olson, Kenneth C. McHardy, Peter F. Sharp, and John V. Forrester. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput. Biomed. Res.*, 29(4) :284–302, August 1996.

[156] T. Spencer, R.P. Phillips, P.F. Sharp, and J.V. Forrester. Automated detection and quantification of microaneurysms in fluorescein angiograms. *Graefe's archive for clinical and experimental ophthalmology*, 230(1) :36–41, 1992.

[157] P.P. Srinivasan, L.A. Kim, P.S. Mettu, S.W. Cousins, G.M. Comer, J.A. Izatt, and S. Farsiu. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomedical Optical Express*, 5(10) :3568–3577, 2014.

[158] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken. Ridge based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4) :501–509, 2004.

[159] T. Stoyanov, A. Louloudi, H. Andreasson, and A. J. Lilienthal. Comparative evaluation of range sensor accuracy in indoor environments. In *5th European Conference on Mobile Robots*, pages 19–24, 2011.

[160] F. Tombari, S. Salti, and L. Di Stefano. Unique shape context for 3d data description. In *Proceedings of the ACM workshop on 3D object retrieval*, pages 57–62. ACM, 2010.

[161] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *European conference on computer vision*, pages 356–369. Springer, 2010.

[162] F. Tombari, S. Salti, and L. Di Stefano. Performance evaluation of 3d keypoint detectors. *International Journal of Computer Vision*, 102(1-3) :198–220, 2013.

[163] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1) :97–136, 1980.

[164] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J.P. Hubschman, B. al Diri, C.Y. Cheung, D. Wong, M. Abrámoff, G. Lim, D. Kumar, P. Burlina, N.M. Bressler, H. F. Jelinek, F. Meriaudeau, G. Quellec, T. MacGillivray, and B. Dhillon. Validation retinal fundus image analysis algorithms : issues and proposal. *Investigative Ophthalmology & Visual Science*, 54(5) :3546–3569, 2013.

[165] Ujjwal, Deepak K. S., Chakravarty A., and Sivaswamy J. Visual saliency based bright lesion detection and discrimination in retinal images. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1436–1439, 2013.

[166] S. K. Ungerleider and G. Leslie. Mechanisms of visual attention in the human cortex. *Annual review of neuroscience*, 23(1) :315–341, 2000.

[167] M. Unser and D. Van De Ville. Wavelet steerability and the higher-order Riesz transform. *IEEE Transactions on Image Processing*, 19(3) :636–652, 2010.

[168] M. Usman Akram, S. Khalid, A. Tariq, S. A. Khan, and F. Azam. Detection and classification of retinal lesions for grading of diabetic retinopathy. *Computers in Biology and Medicine*, 161-171 :45, 2014.

[169] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9) :1582–1596, 2010.

[170] M. van Grinsven, A. Chakravarty, J. Sivaswamy, T. Theelen, B. van Ginneken, and C. Sanchez. A bag of words approach for discriminating between retinal images containing exudates or drusen. In *IEEE International Symposium on Biomedical Imaging*, pages 1444–1447, 2013.

[171] F. G. Venhuizen, B. van Ginneken, B. Bloemen, M. JJP. van Grinsven, R. Philipsen, C. Hoyng, T. Theelen, and C. I. Sánchez. Automated age-related macular degeneration classification in oct using unsupervised feature learning. In *SPIE Medical Imaging*, pages 94141I–94141I. International Society for Optics and Photonics, 2015.

[172] T. Walter, Pascale Massin, Ali Erginay, Richard Ordonez, Clotilde Jeulin, and Jean-Claude Klein. Automatic detection of microaneurysms in color fundus images. *Medical Image Analysis*, 11(6) :555 – 566, 2007.

[173] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King. Global prevalence of diabetes estimates for the year 2000 and projections for 2030. *Diabetes Care*, 27(5) :1047–1053, 2004.

[174] W. Wohlkinger and M. Vincze. Ensemble of shape functions for 3d object classification. In *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*, pages 2987–2992. IEEE, 2011.

[175] M. Wojtkowski, L. Leitgeb, A ; Kowalczyk, T. Bajraszewski, and A. F. Fercher. In-vivo human retinal imaging by fourier domain optical coherence tomography. *J. Biomed. Opt.*, 7 :457–463, 2002.

[176] J. M. Wolfe, G. A. Alvarez, and T. S. Horowitz. Attention is fast but volition is slow. *Nature*, 2000.

[177] J. Wright, Y. Ma, J. Mairal, G. Sapiro, and A. Zisserman. Sparse representation for computer vision an pattern recognition. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognitiong*, pages 1031–1044, 2010.

[178] X. Xiao, C. Xu, and Y. Rui. Video based 3d reconstruction using spatio-temporal attention analysis. In *Multimedia and Expo (ICME)*, 2010.

[179] J. Yang, K. Yu, Y. Gong, and T Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognitiong*, pages 1794–1801, 2009.

[180] A. Yilmaz, O. Javed, and M. Shah. Object tracking : A survey. *Acm computing surveys (CSUR)*, 38(4) :13, 2006.

[181] X. Ying and Z. Hu. Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model. In *European Conference on Computer Vision*, pages 442–455. Springer, 2004.

[182] T. Yubing, F. A. Cheikh, F. F. Elahi Guraya, H . Konik, and A. Trémeau. A spatio-temporal saliency model for video surveillance. *Cognitive Compuation*, Volume 3, Issue 1 :pp 241–263, 2011.

[183] F. Zana and J.C. Klein. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *Image Processing, IEEE Transactions on*, 10(7) :1010–1019, 2001.

[184] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray. Detection of microaneurysms using multi-scale correlation coefficients. *Pattern Recognition*, 43(6) :2237–2248, 2010.

[185] G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on PAMI*, 29(6) :915–928, 2007.

[186] B. Zhou, X. Hou, and L. Zhang. A phase discrepancy analysis of object motion. In *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part III*, ACCV'10, pages 225–238. Springer-Verlag, 2011.

[187] X. Zhu. Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison, 2005.

TABLE DES FIGURES

LISTE DES TABLES

Document réalisé avec LaTeX et :
le style LaTeX pour Thèse de Doctorat créé par S. Galland — http://www.multiagent.fr/ThesisStyle
la collection de paquets `tex-upmethodology` — http://www.arakhne.org/tex-upmethodology/