



HAL
open science

De la couleur au mouvement. De l'apparence à l'action.

Michèle Gouiffès

► **To cite this version:**

Michèle Gouiffès. De la couleur au mouvement. De l'apparence à l'action.. Traitement des images [eess.IV]. Université Paris Sud, 2018. tel-03520874

HAL Id: tel-03520874

<https://hal.science/tel-03520874>

Submitted on 27 Jan 2022

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

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

HABILITATION À DIRIGER DES RECHERCHES
DE L'UNIVERSITÉ PARIS-SUD

École doctorale n°580

Traitement du Signal et des images, Télécommunications,
Automatique, Génie Informatique

par

MICHÈLE GOUIFFÈS

De la couleur au mouvement. De l'apparence à l'action.

le 26 janvier 2018

Composition du Jury :

M.	PATRICK BOUTHEMY	Directeur de recherche INRIA Rennes	(Rapporteur)
Mme.	ALICE CAPLIER	Professeure des Universités Institut Polytechnique de Grenoble	(Rapporteur)
M.	LUDOVIC MACAIRE	Professeur des Universités Université de Lille 1	(Rapporteur)
M.	YACINE BELLIK	Maître de Conférences HDR Université Paris Sud	(Examinateur)
M.	MOUNÎM A. EL YACOUBI	Professeur des Universités Telecom SudParis / Institut Mines-Telecom	(Examinateur)
Mme	LUCE MORIN	Professeure des Universités INSA de Rennes	(Examinatrice)
Mme	SAMIA BOUCHAFA-BRUNEAU	Professeure des Universités Université Evry Val d'Essonne	(Invitée)

Table des matières

1	Introduction	5
1.1	Applications : contraintes et inspirations	6
1.2	Problématiques	10
1.3	Organisation du mémoire	11
2	Couleur	13
2.1	Modèles photométriques	13
2.1.1	Modèle d'acquisition/projection	14
2.1.2	Modèle dichromatique	14
2.1.3	Modèles de variations photométriques	15
2.2	La couleur du pixel	17
2.3	Caractéristiques locales ou ponctuelles	17
2.3.1	Descripteurs couleur	18
2.3.2	Descripteurs binaires	20
2.4	Cartes topographiques	22
2.5	Fiabilité de la couleur	24
2.6	Effacer la couleur	26
2.7	Conclusion	27
3	Apparence	31
3.1	Le choix d'une représentation	31
3.1.1	Critères liés à l'objet	32
3.1.2	Critères externes à l'objet	33
3.1.3	Discussion	34
3.2	Descripteurs par distributions de primitives	35
3.2.1	Degré de connexité couleur	36
3.2.2	Sélection automatique d'attributs couleur	36
3.3	Descripteurs par covariance	39
3.3.1	Covariance et texture : le descripteur ELBCM	40
3.3.2	Suivi d'objets multiples	42
3.4	Coopération de descripteurs	45
3.4.1	Coopération Points-Covariance	46
3.4.2	Coopération Distribution - Covariance	48
3.5	Conclusion	48
4	Mouvement	49
4.1	Estimer le mouvement	49
4.1.1	Le flot optique	49
4.1.2	Les trajectoires	52
4.2	Analyser les scènes en mouvement	52
4.2.1	Extraire la structure de la scène	52

TABLE DES MATIÈRES

4.2.2	Détection d'anomalies	53
4.3	Décrire le mouvement	54
4.3.1	Histogrammes de mouvement en représentation polaire	56
4.3.2	Représentation par mélange de gaussiennes	57
4.4	Conclusion	58
5	Conclusion	59
5.1	Applications	59
5.2	Pluralité des descripteurs	61
5.3	Temporalité d'une action	61
5.4	Modéliser le contexte	63
5.5	Reconfigurabilité	63
A	Curriculum Vitae	79
A.1	Parcours et diplômes	79
A.2	Activités de recherche	80
A.2.1	Publications et communications	80
A.2.2	Participation significative à des projets de recherche	80
A.2.3	Encadrement	84
A.2.4	Participation à des comités	86
A.3	Responsabilités administratives	87
A.3.1	Responsabilités pédagogiques	87
A.3.2	Participation aux instances locales	87
A.3.3	Participation à des comités de recrutement	87
A.4	Activités d'enseignement	88
A.4.1	Postes occupés	88
A.4.2	Principaux enseignements dispensés depuis 2006	89
A.4.3	Participation à la vie collective de Polytech'Paris Sud	90
A.5	Diffusion de la culture scientifique et technique	91
B	Liste de publications	93

Chapitre 1

Introduction

LES recherches menées en vision par ordinateur visent à définir des modèles computationnels dotés de fonctionnalités similaires à celles du système visuel humain (SVH), en tentant d'imiter les tâches de la rétine et du cortex visuel.

Le système visuel humain est capable de constance chromatique [vK05], c'est-à-dire que l'œil s'adapte rapidement aux changements de source de lumière et interprète correctement les teintes malgré un éclairage ambiant fortement coloré. Le cortex visuel primaire se charge des traitements bas niveau à savoir la détection du contraste, des lignes orientées, tandis que l'aire visuelle secondaire fournit une carte d'organisation topographique du champ visuel. Ces informations sont ensuite traitées en parallèle ; d'une part dans la voie dorsale qui est dédiée à la détection du mouvement et à une première analyse du contexte ; d'autre part dans la voie ventrale, qui fournit une première interprétation en objets [Ben07].

La plupart des systèmes de vision artificielle se fondent sur une succession de traitements analogues. Les primitives visuelles locales extraites au niveau du pixel, sont combinées pour exhiber une cohérence spatiale, temporelle ou spatio-temporelle et produire des descriptions plus globales de l'image, plus facilement interprétables par le système. Alors que le SVH est fiable¹ et rapide, la définition d'un système de vision artificielle passe par des compromis entre fiabilité et rapidité.

Cette synthèse porte principalement sur les premiers étages du système de vision (le bas-niveau), en particulier sur la définition de primitives visuelles et de descripteurs d'objets, dans le but d'améliorer la fiabilité des traitements de plus haut niveau. Cependant, si une plus grande importance est attribuée à la fiabilité, nous garderons toujours à l'esprit l'objectif de rapidité.

Les travaux décrits dans cette synthèse ont été développés dans des contextes variés où l'objectif visé pose une contrainte qui guide le choix des approches tout en représentant un moteur pour la créativité. La section 1.1 est dédiée à ces applications, qui sous-tendront la suite du document. De ces applications se dégagent des problématiques communes qui sont décrites dans la section 1.2. Enfin, la section 1.3 détaille le plan de la suite du manuscrit.

Le lecteur pourra se référer à la figure 1.1 qui synthétise les thèses co-encadrées ainsi que les principales collaborations financées dans le cadre de projets scientifiques. Notons en particulier le changement de laboratoire qui a eu lieu en 2013 de l'IEF² vers le LIMSI³, suite à des réflexions commencées en 2010 autour des restructurations au sein de l'IEF. En 2011, nous avons bénéficié de co-financement IEF-LIMSI pour démarrer des projets de collaboration et tenter d'identifier des pistes de recherche en vue d'une possible intégration.

1. Tout en pouvant être trompé par des effets d'optique.

2. Institut d'Électronique Fondamentale

3. Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur

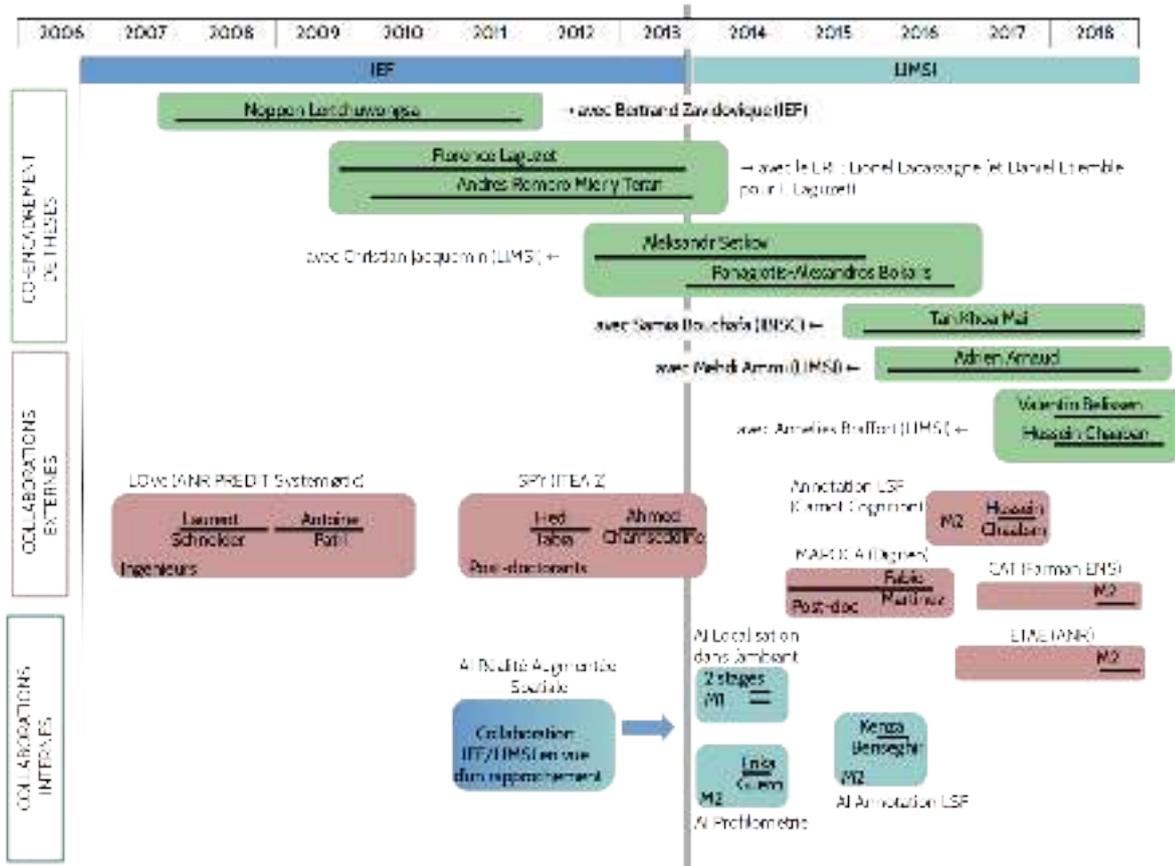


FIGURE 1.1 – Synthèse des thèses co-encadrées et des projets de collaboration scientifiques. Les segments correspondent aux périodes de contrats de stage, d’ingénieur, de thèse ou de post-doctorat. Les AI, pour Actions Initiatives, sont des projets internes au LIMSIS.

1.1 Applications : contraintes et inspirations

Mon expérience en recherche débute à l’IRSTEA⁴ (appelé à l’époque Cemagref) sur la vision industrielle pour la traçabilité de produits carnés et s’enrichit depuis 2006 à l’Université Paris-Sud autour d’applications variées : l’assistance à la conduite, la vidéo-surveillance (au laboratoire IEF), la réalité augmentée, l’interaction homme-machine (au LIMSIS).

Vision industrielle. Du fait des multiples problèmes d’hygiène alimentaire survenant dans la chaîne de production carnée, la traçabilité est devenue une priorité, permettant de rassurer un consommateur de plus en plus méfiant. Au niveau des abattoirs porcins, la solution retenue en 2005 était le tatouage des animaux et le marquage à l’encre des pièces par un numéro d’abattage unique normalisé par l’AFNOR⁵ en 1999. Cet identifiant permet ensuite d’accéder à l’information de provenance de l’animal.

Le numéro d’abattage est retranscrit par impression sur les huit pièces de découpe avant qu’elles soient entreposées dans l’abattoir dans l’attente de leur expédition. Avant leur sortie de l’abattoir, l’identifiant doit être lu. Cette tâche, auparavant réalisée par un opérateur, a fait l’objet d’études par l’IRSTEA, en partenariat avec l’OFIVAL⁶, pour la mise en œuvre d’un système robotisé constitué d’un bras muni d’une caméra couleur, permettant la détection puis la lecture automatique de

4. Institut de Recherche en Sciences et Technologies pour l’Environnement et l’Agriculture
 5. Association Française de NORmalisation
 6. Office national interprofessionnel des viandes, de l’élevage et de l’aviculture

cet identifiant. Dans un premier temps, le système localise le marquage puis se déplace automatiquement en face du marquage pour en faciliter la lecture. Mes travaux de thèse [Gou05] ont porté sur la détection du marquage à l'encre par segmentation couleur et sur le suivi de primitives visuelles ponctuelles extraites sur ce marquage de façon à guider le déplacement du système par asservissement visuel. La détection du marquage se heurtait à des difficultés telles que la variation de concentration d'encre utilisée, la variété des couleurs de couenne et la présence possible de défauts sur la pièce ayant des caractéristiques couleur et structurelles parfois proches du marquage à l'encre. En outre, les couleurs d'encre utilisées pouvaient varier d'un abattoir à l'autre. Pour répondre à cette problématique, nous avons défini des caractéristiques couleur invariantes à la transparence de l'encre et proposé une approche de segmentation fondée sur le degré de connexité couleur. En ce qui concerne le suivi de primitives, le caractère spéculaire du matériau considéré et le changement de point de vue au cours du déplacement induisent des variations photométriques susceptibles de nuire à cette tâche. Cela m'a orientée vers l'étude des modèles photométriques, le suivi de primitives de types points d'intérêt et l'étude de l'invariance couleur.

Assistance à la conduite. Face à un marché de l'automobile européen saturé et à une forte concurrence asiatique, la plupart des constructeurs et des équipementiers français ont accru leur effort en R&D et se sont rapprochés des laboratoires de recherche afin de mieux répondre à l'enjeu sociétal qu'est celui de la sécurité routière. Les véhicules récents sont ainsi équipés d'options évoluées de sécurité active comme la détection de fatigue ou la détection d'obstacles. C'est dans ce contexte que le projet LOVE (ANR Predit Systemtic Logiciel d'Observation des Vulnérables) a permis entre 2007 et 2010, la collaboration de Renault, Valeo et neuf partenaires académiques. Ce projet, qui s'inscrit en parfaite cohérence avec les axes de recherche menés par le département ACCIS de l'IEF auquel j'appartenais entre 2006 et 2013, visait à développer un système embarqué d'acquisition et de traitement multi-capteurs (stéréovision et LIDAR⁷). En termes de traitement d'images, il s'agissait de proposer des algorithmes temps-réel et fiables permettant la détection de piétons malgré de multiples difficultés : le mouvement de la caméra, la qualité dégradée des images, les variations d'illumination.

Dans ce contexte, nous avons travaillé sur la construction de la carte de profondeur par stéréovision dans la thèse de N. Lertchuwonga [LGZ11, Ler11, LGZ12], sur l'analyse de ces cartes par des méthodes cumulatives *UV-disparité* pour la détection d'obstacles (la Fig. 1.2 montre un exemple de détection), ainsi que sur des méthodes de suivi d'objets [GZ09, GLL10a, GPV10]. Depuis 2015, je renoue avec cette application par le co-encadrement de la thèse de Tan-Khoa Mai avec Samia Bouchafa (IBISC), pour laquelle nous nous intéressons au calcul et à l'analyse du mouvement ainsi qu'à la détection de structures 3D à partir d'une approche cumulative appelée *UV-velocity* [MGB17a, MGB17b].

Analyse automatique de vidéos et surveillance de sites. Depuis quelques années, les caméras de surveillance se multiplient dans l'espace urbain et les solutions clé-en-main progressent dans l'espace privé. Souvent simple moyen de dissuasion, la surveillance de certains sites est suffisamment critique pour affecter un opérateur en continu sur le poste de contrôle.

Parallèlement à cela, les solutions logicielles se développent progressivement, soit pour enregistrer les données et les annoter automatiquement de manière à faciliter une future recherche soit pour détecter automatiquement des alarmes.

L'objectif du projet Européen ITEA2⁸ SPY pour *Surveillance imProved sYstem*⁹ (2010-2013) était de concevoir, développer et tester un système intelligent multi-capteurs de surveillance adapté aux besoins des agents de la sécurité publique (pompiers et police), c'est-à-dire une solution intégrée mobile, facilement reconfigurable et prenant pleinement en compte les infrastructures et les

7. LIDAR pour *Light Detection and Ranging*

8. Information Technology for European Advancement

9. <http://www.itea2-spy.org/>

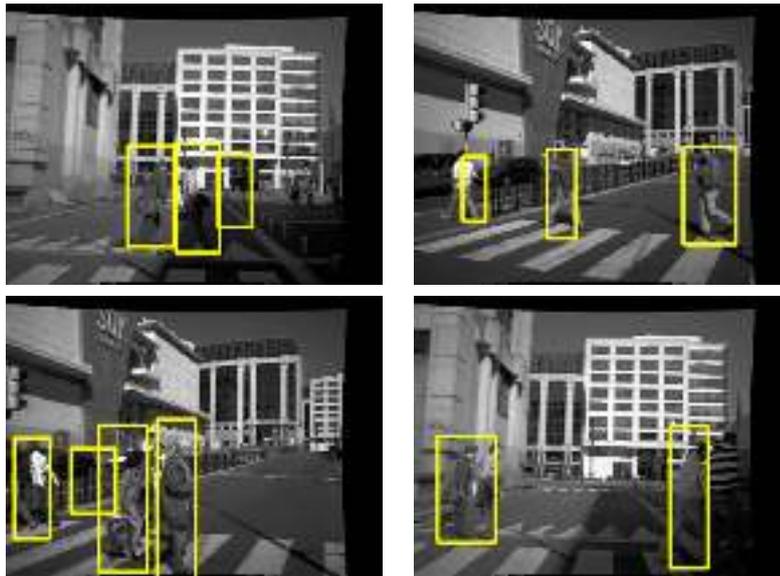


FIGURE 1.2 – Exemples de détections d’obstacles à partir d’un capteur stéréoscopique embarqué dans un véhicule. L’approche utilisée partait de l’espace cumulatif appelé *UV-disparité* permettant de résumer l’extraction des plans frontaux (ici des piétons) dans la scène à une simple détection de segments dans l’espace *UV-disparité*.

réseaux de surveillance fixes existants. Les scénarios considérés concernaient notamment la patrouille classique, la surveillance de sites, la sécurisation d’événements, la détection de départs de feux, l’escorte de véhicule. Ils prévoyaient notamment d’annoter les vidéos acquises avant leur sauvegarde, de façon à simplifier leur indexation et leur usage futur. Ce projet impliquait 14 partenaires dans 5 pays. À l’IEF, ce projet impliquait Lionel Lacassagne avec lequel je collabore depuis lors sur les aspects adéquation algorithme-architecture (AAA).

La multiplicité des conditions d’acquisitions (caméra embarquée sur un véhicule ou fixée au sol, conditions d’illumination variables) et les contraintes d’embarquabilité des algorithmes constituent ici les deux principaux défis. Nous avons proposé des algorithmes de suivi d’objets et de détection d’anomalies avec le souci de l’invariance et de la rapidité des traitements. Le projet SPY a permis de recruter des post-doctorants, parmi lesquels Hedi Tabia [TGL12b, TGL12a] et Ahmed Chamseddine [CBAGL16] pour ce qui concerne l’analyse des vidéos et Ali Zahraee [yTLZG13, RLGHZ14] pour l’étude de l’AAA de certains de ces algorithmes afin de les intégrer sur les architectures ciblées dans le projet. En parallèle de ce projet, nous avons eu la chance d’obtenir deux bourses de thèses, celles de Florence Laguzet et d’Andres Romero Mier y Terán, dont les travaux se sont axés vers des problématiques connexes à celles abordées dans le projet SPY, à savoir le suivi et la ré-identification d’objets [LGLE11, LRG⁺15].

Par ailleurs, le projet SPY nous a permis d’identifier des pistes de collaboration avec Antoine Manzanera (ENSTA Paritech), qui était également partenaire du projet. Le travail de caractérisation des anomalies s’est ainsi étendu à celle des activités humaines au cours du projet Digiteo MAPOCA¹⁰ avec le post-doctorat de Fabio Martinez y Carillo entre 2014 et 2016 [MMGB15, CaMGN16].

Réalité augmentée spatiale et systèmes caméra-projecteurs. La Réalité Augmentée Spatiale (RAS) consiste à adapter une projection vidéo aux propriétés géométriques et colorimétriques de la surface de projection. Les systèmes utilisés sont des projecteurs dits intelligents, dans le sens où ils sont équipés d’une ou plusieurs caméras et de capteurs proprioceptifs afin d’acquérir des informations sur l’environnement. Les applications visées sont particulièrement variées, allant du

10. Modélisation des Activités par Processus ponctuels et Covariance de caractéristiques.

travail collaboratif à la valorisation du patrimoine, en passant par le prototypage [MSP⁺11].

À travers le co-encadrement de deux thèses avec Christian Jacquemin au LIMSI (Aleksandr Setkov [Set15] et Panagiotis-Alexandros Bokaris [Bok16]), je me suis penchée sur la compensation géométrique et colorimétrique de la projection en vue d'améliorer le confort de visualisation quelles que soient les conditions de projection. Ces deux problématiques nécessitent la mise en correspondance entre l'image initiale et la prise de vue de l'image projetée. En plus des variations de couleurs dues aux spectres de réponse du capteur et au spectre d'émission du projecteur, les distorsions subies par l'image initiale sont complexes, en particulier si la surface de projection est non-planaire et de réflectance non-uniforme.

Nous avons notamment mené une étude sur les caractéristiques couleur invariante adaptées à ce contexte et sur leur appariement [SGJ13, SGJ16]. Une base de données de vidéo-projections (*DacImPro*) a été construite et rendue accessible à la communauté [SCG⁺15]. Ensuite, une nouvelle approche de compensation colorimétrique des vidéo-projections a été proposée. En utilisant un apprentissage *hors-ligne*, une seule image de la scène suffit pour estimer la réflectance en tout point de la surface de projection colorée [BGJC14]. Une fois connues, les couleurs des surfaces sont atténuées visuellement pour améliorer le rendu visuel de la projection. La figure 1.3 provenant de [Bok16] donne les principaux éléments de ce système.

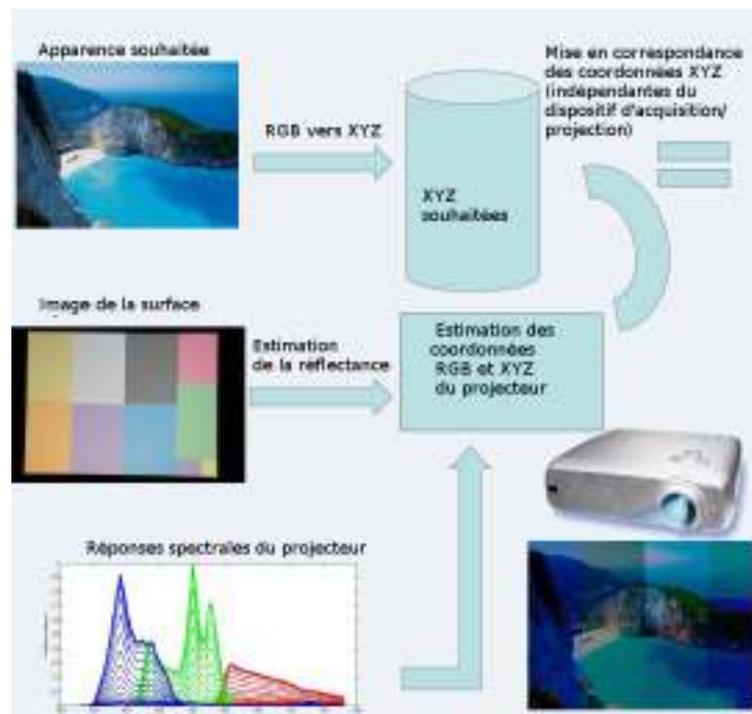


FIGURE 1.3 – Illustration du système de compensation couleur par vidéo-projection [Bok16]. En partant de l'image idéale (avant projection) et d'une image de la surface de projection, et en connaissant les spectres d'émission du projecteur, l'image est corrigée avant projection pour réduire l'influence de la surface.

Image et interaction. La miniaturisation accrue des composants électroniques et les progrès récents des développements de systèmes embarqués et des réseaux de communication sans fil ouvre la voie à de nouvelles possibilités d'interaction entre l'utilisateur et son environnement par l'émergence d'objets du quotidien communicants et organisés en réseaux. L'interaction devient ubiquitaire dans le sens où l'utilisateur peut interagir n'importe où dans cet environnement avec un système global en permanence attentif aux activités de l'utilisateur. Cela peut se faire par le biais de capteurs caméra, radio-fréquence ou biométriques. En 2014, j'ai découvert l'Interaction Ambiante via une

collaboration avec Yacine Bellik visant à fusionner des données de localisation des utilisateurs issus de capteurs Radio-Fréquences avec des données image afin d'améliorer la précision et la robustesse de la localisation des utilisateurs dans la salle d'expérimentation iRoom du LIMSI.

L'utilisateur peut interagir de façon naturelle, c'est-à-dire sans avoir recours à une interface classique (clavier, souris) mais directement par le biais de ses activités, de ses gestes ou la manipulation d'objets. Une application majeure concerne l'assistance à la personne et le maintien à domicile des personnes en perte d'autonomie. En termes de traitement d'images, cela nécessite de mettre en place des algorithmes de détection, de reconnaissance, de suivi d'objets et d'activités. Le projet Digiteo MAPOCA a permis de proposer des méthodes d'analyse d'activités humaines pouvant également s'appliquer à l'analyse de la communication visuo-gestuelle et à l'annotation de vidéos de langue des signes [MMGB15].

Enfin, je m'intéresse actuellement à la reconstruction 3D d'environnements intérieurs pour la rénovation à l'aide d'une tablette (thèse CIFRE en cours d'Adrien Arnaud en collaboration avec Mehdi Ammi) et à l'interaction tangible avec la maquette 3D obtenue [ACGA16, ACC⁺16].

Photographie et spectacle vivant. Observer, questionner, imaginer, expérimenter, exposer, étendre le champ des connaissances : la recherche scientifique et la recherche artistique ne se meuvent-elles pas d'une même dynamique ? Le dialogue art-science est fécond en soi puisqu'à la frontière entre objectivité et subjectivité, l'art et la science s'inspirent l'un-e de l'autre. J'ai pu en faire l'expérience à travers des projets art-science collaboratifs impliquant le plasticien Bertrand Planes (projets *HTRI* et *Pixel's Pixels*), la metteur en scène et cinéaste Véronique Caye (projet *Gardien du Temple*) ou encore en participant à des projets collaboratifs au *sas*¹¹ (science art société). *HTRI* (pour *High Time Range Imaging*) [MPJ13] visait à fusionner plusieurs photographies d'une même scène acquises à des instants très différents de la journée pour n'en former qu'une seule. Cette image juxtapose les couleurs *saillantes* tout en saisissant les événements éphémères, de forte *saillance temporelle*, tels que l'apparition d'objets ou les éclairages artificiels. *Gardien du Temple* s'appuie sur les travaux de thèse de Panagiotis-Alexandros Bokaris en créant une installation interactive autour du thème de l'effacement et de la révélation de présences, une métaphore de l'effacement de la mémoire et de la culture [BGC⁺17].

Ces travaux ont profité de la dynamique de la Diagonale Paris-Saclay¹² et de l'axe transversal VIDA¹³ (Virtualités, Interaction, Design et Art) du LIMSI dont le but est de faciliter et de promouvoir des recherches associant scientifiques, industriels, artistes et designers sur des projets Arts/Sciences.

1.2 Problématiques

Bien que différentes, les applications décrites précédemment soulèvent des problématiques communes. En effet, la plupart des approches sur lesquelles j'ai proposé des contributions reposent tout d'abord sur une extraction de primitives visuelles de bas niveau, puis sur une modélisation de l'apparence visuelle ou de la dynamique d'un objet pour aboutir à un appariement. Chaque étape réduit la quantité de données tout en enrichissant l'information vers une abstraction plus importante. Dans les tâches de reconnaissance, un modèle (ou un ensemble de modèles) de référence est préalablement construit à partir d'une image ou d'une base d'apprentissage. La réussite de l'appariement repose sur le choix d'une mesure de similarité, qui est contrainte par la nature des modèles et des attributs.

Les méthodes et descripteurs doivent tenir compte de la variabilité des données : variabilité des images à traiter, variabilité de la forme des objets présents dans la scène (piétons et véhicules

11. <http://www.e-sas.org/>

12. <http://www.ladiagonale-paris-saclay.fr/>

13. <https://vida.limsi.fr/doku.php>

dans un paysage urbain), variabilité d'apparence (changement d'échelle, rotations, déplacements non-rigides, occultations), variabilité des conditions d'acquisition (flou, changements d'illumination), variabilité de l'environnement (dans le cas d'une application mobile). Il s'agit de définir des modèles représentatifs et discriminants de l'objet d'intérêt qui soient par ailleurs stables face aux variabilités internes (inhérentes à l'objet ou à la classe d'objets) et externes.

Ainsi, la définition des descripteurs et de la mesure de similarité qui leur est associée peut être guidée par un ensemble d'attentes parfois difficilement conciliables : unicité, invariance, rapidité, compacité.

Unicité. La distinction des primitives visuelles (ou des objets) entre elles doit être reflétée par une dissimilarité entre leurs descripteurs : à primitive visuelle (ou objet) unique, descripteur unique ; à primitives similaires, descripteurs similaires.

Invariance de la description, d'un point de vue géométrique ou photométrique. La représentation doit être stable au cours du temps malgré les modifications des conditions externes à l'objet (changements d'environnement, conditions d'observation) et de sa condition interne (changements d'illumination, déformation géométrique).

Rapidité. Le temps d'exécution des traitements s'avère parfois prioritaire par rapport aux deux critères précédents. Cela peut amener à privilégier l'usage de descripteurs rapides au détriment de l'unicité et de l'invariance, ou encore à fournir un effort d'accélération des codes en passant par des optimisations, des parallélisations ou des instructions dédiées.

Compacité. Une bonne compacité permet de réduire les besoins en mémoire de stockage, permet d'accélérer l'accès aux données ou encore leur transmission sur un réseau.

Tout en gardant à l'esprit la satisfaction de ces quatre objectifs, nos réflexions ont été menées autour de trois axes : la couleur, l'apparence et le mouvement. Sur la base des descripteurs proposés, nous avons également cherché à proposer des méthodes adaptatives par rapport à un contexte (externe ou interne à l'objet) et à faire coopérer les descripteurs ou les attributs.

1.3 Organisation du mémoire

La suite de ce manuscrit s'organise ainsi autour de trois volets : couleur, apparence et mouvement.

- **Couleur.** Le chapitre 2 s'intéresse aux primitives visuelles définies au niveau du pixel ou d'un ensemble de pixels, avec comme principal objectif l'invariance vis-à-vis des variations d'illumination.
- **Apparence.** Les caractéristiques couleur et texture sont combinées pour construire une représentation d'apparence des objets, servant à l'appariement.
- **Mouvement.** Dans le chapitre 4, le mouvement est vu comme une primitive d'entrée qui est analysée pour détecter des anomalies dans les vidéos et reconnaître des actions.

Nous aurons compris que cette organisation correspond à trois niveaux d'abstraction croissante : la couleur est vue comme une primitive de bas niveau exploitée ensuite dans la définition de modèles d'apparence. Ces modèles de couleur ou d'apparence sont appariés temporellement pour définir un mouvement. Lui-même peut-être caractérisé et analysé. Il s'avère que cette structuration en trois volets est cohérente avec l'ordre chronologique dans lequel j'ai abordé ces sujets depuis le début de notre activité.

La conclusion du chapitre 5 propose un bilan des travaux réalisés et ouvre plusieurs perspectives de travail. Notons enfin que le lecteur trouvera une synthèse des activités de recherche et d'enseignement en annexe A ainsi qu'une liste des publications en annexe B.

Chapitre 2

Couleur

Issue de phénomènes physiques et psychologiques, instrument artistique et élément culturel, la couleur est un sujet de recherche pluridisciplinaire. C'est en partant des modèles photométriques qui seront décrits dans la section 2.1, que se déclinent des attributs couleur invariants vis-à-vis des changements d'illumination et dont l'usage permet de simplifier les traitements ultérieurs. L'extension des algorithmes à la couleur est une question en soi lorsqu'il s'agit de prendre en compte les spécificités mathématiques des espaces couleur, l'intercorrélation entre les trois canaux ou encore la propriétés du Système Visuel Humain (SVH). Cette synthèse s'articule autour de cinq axes de réflexion, allant de la couleur du pixel jusqu'à l'effacement de la couleur.

1) *La couleur du pixel.* Chaque pixel porte une information photométrique allant de la simple luminance à l'information multi-spectrale. Parmi les multiples modèles de représentation de la couleur, nous nous sommes intéressés à ceux qui présentent des propriétés d'invariance vis-à-vis des changements d'illumination [TFMB04, GGvdWG12] (voir partie 2.2).

2) *La couleur texturée.* Les points « d'intérêt » sont dénommés ainsi de par leur caractère *saillant*, et correspondent généralement à des coins ou à des points-clés caractéristiques, possédant au moins deux directions de fort gradient. L'utilisation de tels ensembles de pixels permet de définir des invariants géométriques ainsi que des invariants photométriques supplémentaires qui apportent une réelle plus-value à l'extension de descripteurs vers la couleur (voir Section 2.3).

3) *Les couleurs connexes,* c'est-à-dire la couleur d'un ensemble de pixels connexes et homogènes, leur contour, la couleur des ensembles et lignes de niveaux. La partie 2.4 aborde deux approches d'extraction d'ensembles et de lignes couleur définies en accord avec le contenu de l'image.

4) *Fiabilité de la couleur.* Les invariants couleur s'avèrent peu exploitables lorsque les couleurs sont peu saturées. La partie 2.5 évoque nos travaux sur la prise en compte d'une mesure de pertinence des invariants couleur pour en modérer l'usage.

5) *Effacer la couleur.* Les couleurs des objets dans la scène peuvent être virtuellement effacées par vidéo-projection (voir section 2.6).

Finalement, la partie 2.7 permettra de conclure et d'ouvrir quelques perspectives à ces travaux.

2.1 Modèles photométriques

Après avoir introduit la modélisation du système d'acquisition (ou d'acquisition/projection) telle qu'elle est couramment admise, nous abordons quelques simplifications de ce modèle, qui seront utiles dans la suite de cette synthèse.

2.1.1 Modèle d'acquisition/projection

Soit P un point de la scène se projetant au point p dans le plan image d'un capteur à K canaux spectraux. La réponse sur la composante k , notée $f_k(p)$, provient de l'intégration suivant les longueurs d'onde λ du produit de trois signaux : 1) le spectre de l'illuminant $\mathcal{E}(\lambda, p)$, 2) la réflectance de la surface $\mathcal{R}(\lambda, p)$ qui définit la proportion de lumière incidente réfléchi par la surface pour chaque longueur d'onde λ ; et enfin 3) les trois courbes de sensibilité spectrale du capteur $S_k(\lambda)$. Ainsi, la réponse en chaque canal est donnée par :

$$f_k(p) = \int_{\lambda} \mathcal{E}(\lambda, p) \mathcal{R}(\lambda, p) S_k(\lambda) d\lambda \quad (2.1)$$

Dans le cas d'un système caméra-projecteur, \mathcal{E} correspond au spectre d'émission du vidéoprojecteur \mathcal{L} auquel se combine le spectre d'illumination de la scène \mathcal{I} :

$$\mathcal{E}(\lambda) = \mathcal{L}(\lambda) + \mathcal{I}(\lambda) \quad (2.2)$$

Afin de définir des attributs couleur invariants vis-à-vis des changements d'illumination, il est nécessaire de modéliser les variations couleur dues aux conditions d'acquisition. Cependant, l'équation de l'acquisition (2.1) est difficile à exploiter directement à moins de poser quelques hypothèses simplificatrices telles que celles qui ont abouti au modèle dichromatique [Sha85], au modèle d'illumination affine, utilisé notamment par Finlayson *et. al.* [FHST05, vdSGS10, SGJ13] pour la définition d'invariants, ou encore au modèle de type *gamma* très largement utilisé pour modéliser les réponses des moniteurs [TFMB04]. Nous reviendrons sur ces modèles dans la section 2.3.1 qui s'intéressera à l'invariance couleur dans le cas de système caméra-projecteur. Ces modèles sont également largement exploités dans les travaux que nous avons menés sur la compensation couleur et qui sont abordés dans la Section 2.6.

2.1.2 Modèle dichromatique

Le modèle dichromatique proposé par Shafer [Sha85] stipule qu'un matériau diélectrique non-homogène, uniformément coloré et mat, réfléchit la lumière de deux façons, à savoir par réflexion de surface ou d'*interface* et par réflexion de *corps* ou intrinsèque, diffuse. Selon ce modèle, les composantes RGB c au pixel p s'écrivent :

$$c(p) = m_b(p) c_b(p) + m_s(p) c_s(p) \quad (2.3)$$

Dans cette équation, $c_b(p) = (c_b^R, c_b^G, c_b^B)^T$ est la couleur intrinsèque (*body color*) qui dépend notamment des propriétés physiques du matériau. Le triplet $c_s(p) = (c_s^R, c_s^G, c_s^B)^T$ est associé à la réflexion spéculaire et est donc reliée à la couleur de l'illuminant. Finalement, m_b et m_s sont deux fonctions relatives à la géométrie de la scène. L'équation (2.3) permet de modéliser la distribution des couleurs comme un ensemble de plans ou d'arêtes [KSK90, NFT97, VBvdWV11].

Dans le cas lambertien, la contribution spéculaire peut être négligée et le terme $m_s(p) c_s(p)$ s'annule dans (2.3). Dans ce cas, la plupart des images naturelles se concentrent autour d'un nombre fini de lignes droites dans le cube RGB, c'est-à-dire le long de chaque vecteur de couleur intrinsèque c_b . Par exemple, la figure 2.1(a) montre une image couleur¹ et la figure 2.1(b) la distribution RGB de ses couleurs. Cette image correspond à un exemple idéal puisque les objets sont très colorés et peu texturés.

1. issue de la base Kodak : <http://r0k.us/graphics/kodak/>

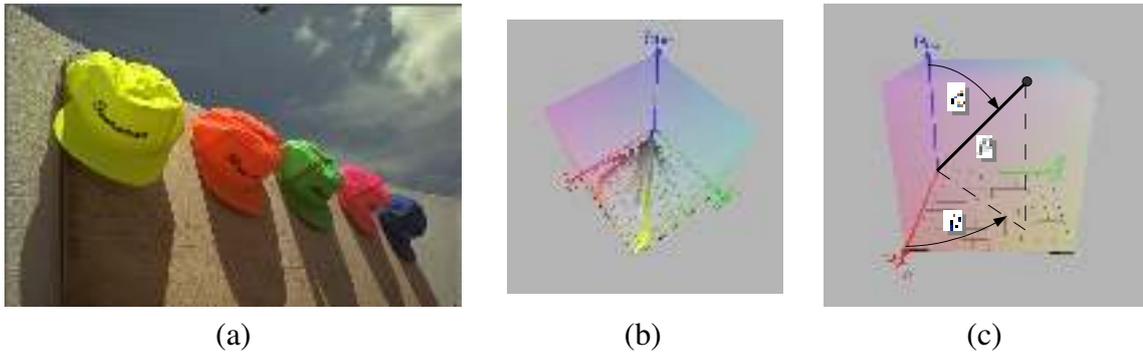


FIGURE 2.1 – (a) Exemple d'image couleur et (b) distribution des couleurs dans le cube RGB. Les vecteurs c_b sont bien visibles. (c) Représentation sphérique (ρ, θ, ϕ) des vecteurs de couleur intrinsèque dans le cube RGB.

2.1.3 Modèles de variations photométriques

Les variations photométriques intervenant entre deux images couleur $\mathbf{f} = \{f_k\}_{k \in RGB}$ et $\mathbf{g} = \{g_k\}_{k \in RGB}$, provenant de deux points de vue ou de deux instants différents, peuvent être approximées en chaque point par un modèle linéaire à 12 paramètres (homographie) [OMC03, MJ05] :

$$\mathbf{g}(t(p)) = \mathbf{A}\mathbf{f}(p) + \mathbf{B} \quad (2.4)$$

où \mathbf{A} est une matrice 3×3 , \mathbf{B} un vecteur 3×1 et t la transformation géométrique du point p entre les deux prises de vues. Le modèle affine est un cas particulier de ce modèle pour lequel les variations sont indépendantes sur chaque composante k , tel que : $g_k(t(p)) = a_k f_k(p) + b_k \forall k \in R, G, B$. Les courbes de sensibilité des caméras ou les courbes d'émission des projecteurs étant généralement de forme exponentielle [NPG03, DZR⁺08], le modèle de variation de type *gamma* est communément employé pour exprimer la transformation photométrique entre une image \mathbf{f} et sa projection \mathbf{g} .

Variations spatialement uniformes. En considérant une portion d'image \mathcal{W} dans \mathbf{f} et son homologue dans une deuxième image \mathbf{g} , les paramètres du modèle affine peuvent être estimés à partir des statistiques couleur (moyenne et écart-type) sur chaque canal k et peuvent être compensés en ligne, image par image [JFS01]. La méthode *High Time Range Imaging* [MPJ13] que nous avons développée lors d'une collaboration art-science² fusionne plusieurs images d'une même scène acquises à des instants très différents d'une journée, en collectant les couleurs et les objets visuellement *saillants*. La Figure 2.2 donne un exemple de résultat. La méthode commence par compenser les variations d'illumination globales apparaissant entre images successives \mathbf{f}_{t-1} et \mathbf{f}_t en estimant la transformation \mathbf{M} telle que $\mathbf{f}_t \simeq \mathbf{M}\mathbf{f}_{t-1}$ avec $\mathbf{f}(p) = [R, G, B, 1]^\top$ et \mathbf{M} une matrice 3×4 telle que $\mathbf{M} = [\mathbf{A}_{3 \times 3} \mathbf{B}_{3 \times 1}]$. L'expression reliant les deux images couleur peut alors s'écrire :

$$\mathbf{f}_t(p) = \mathbf{M}\mathbf{f}_{t-1}(p) + \mathbf{E}(p) \quad (2.5)$$

Le terme $\mathbf{E}(p)$ contient les erreurs de la normalisation globale qui correspondent aux changements locaux qui ne sont pas compensés par la transformation globale. Il s'agit soit de l'apparition éphémère d'un nouvel objet dans la scène, du déplacement des ombres, de réflexions ou de l'éclairage changeant. Ces images d'erreurs servent donc à extraire ces événements et à les cumuler dans l'image résultat. Quant au fond, il est calculé par une combinaison des images d'entrée, avec une pondération plus ou moins importante suivant leur saillance visuelle couleur calculée en utilisant conjointement la teinte et la saturation.

2. avec Christian Jacquemin du LIMSI et l'artiste plasticien Bertrand Planes : <http://www.bertrandplanes.com/>.



FIGURE 2.2 – Exemples de résultats de fusion HTRI (en bas) à partir de 16 photographies (en haut) d’une scène vue à différents moments d’une journée. Ces photographies ont été prises par l’artiste plasticien Bertrand Planes.

En guise d’évaluation de l’approche, nous avons exploité le critère de qualité appelé NIQE [MSB13] parce qu’il ne requiert aucune référence. Nous avons comparé nos résultats avec l’approche de fusion d’images proposée dans [DKMB09] et différentes approches de soustraction de fond. L’approche proposée obtient de meilleures valeurs de NIQE sur les séquences qui montrent de forts changements photométriques.

Variations spatialement non-uniformes. Dans [GCFMT12], nous avons proposé deux approximations linéaires des variations photométriques s’appliquant à des zones d’intérêt \mathcal{W} de taille limitée. Dans ce cas, les paramètres a_k et b_k du modèle affine (spatialement uniforme) sont remplacés par des fonctions approximées par série de Taylor de degré 1 ou 2. Nous avons étudié la validité des

modèles proposés, en fonction des caractéristiques de la surface et des conditions d'acquisition, en supposant que la surface est différentiable en termes de géométrie et de réflectance. Il s'agit d'une hypothèse raisonnable : 1) pour des surfaces de taille restreinte comme des points d'intérêt, 2) pour des applications industrielles où des surfaces naturelles et des objets manufacturés sont considérés. Finalement, nous avons introduit ces modèles dans une approche de suivi de type KLT (Kanade-Lucas-Tomasi [LK⁺81, TK91, BM04]). Comparé aux méthodes analogues, telles que [JFS01], il produit les meilleurs résultats en termes du nombre de points correctement suivis et de résidus de convergence.

2.2 La couleur du pixel

La plupart des invariants pixelliques couleur se fondent sur l'analyse du modèle dichromatique de Shafer décrit par l'équation (2.3), en posant des hypothèses quant aux variables impliquées dans ce modèle, sur les conditions d'acquisition (illuminant d'énergie uniforme ou non), sur le type de matériau considéré (matériau mat ou non) et sur sa couleur (uniforme ou non).

Dans le cas d'un illuminant blanc³, un invariant couleur peut être défini en chaque point indépendamment de ses voisins. Si le matériau est lambertien, le modèle dichromatique (2.3) se réduit au seul terme de réflexion diffuse. En chaque pixel, le rapport de deux composantes RGB est alors un invariant par rapport à l'orientation de surface, la direction d'observation, la direction de l'illumination et l'intensité d'éclairage. C'est, parmi d'autres [GS99, vdSGS10], le cas pour la norme L_1 des composantes RGB, parmi lesquelles la composante rouge r s'écrit :

$$r = \frac{R}{R + G + B} \text{ est équivalent à } \frac{c_b^R}{c_b^R + c_b^G + c_b^B}, \quad (2.6)$$

où la couleur c_b^R, c_b^G, c_b^B de (2.3) ne dépend que des caractéristiques du matériau de la sensibilité spectrale de la caméra.

Dans le cas non lambertien, l'hypothèse d'illuminant blanc suppose que le terme spéculaire est identique sur chaque composante k . Alors, la soustraction de deux composantes annule la composante spéculaire [vdSGS10] puis le rapport de différences supprime la contribution de la géométrie m_b . C'est le cas notamment de la composante de teinte.

Ensuite, la correspondance entre les composantes RGB et le modèle dichromatique peut être simplifiée en les convertissant dans un espace sphérique (ρ, θ, ϕ) comme nous l'avons proposé dans [GZ11]. La figure 2.1(c) illustre cette transformation, où ρ exprime la distance entre l'origine du cube $\mathbf{O} = (0, 0, 0)$ et la couleur \mathbf{c} . Ainsi dans le cas achromatique, ρ correspond à la luminance. Pour les pixels chromatiques, ρ correspond à la distance de \mathbf{c} sur le vecteur couleur correspondant, les valeurs (θ, ϕ) restant invariantes. Pour des matériaux lambertiens, ces deux angles ne dépendent que de la position de la couleur intrinsèque et sont donc invariants vis-à-vis des ombres et des changements d'intensité. À partir de cette représentation couleur, j'ai proposé une extension des cartes topographiques qui sera abordée dans la section 2.4.

2.3 Caractéristiques locales ou ponctuelles

En supposant une certaine uniformité locale (sur une fenêtre \mathcal{W}) des caractéristiques physiques dans la scène⁴, des opérations inter-pixel élémentaires permettent de définir de nouveaux invariants couleur ou binaires. La prise en compte de deux pixels voisins [Hea92, GS99] suffit pour définir des

3. On appelle illuminant blanc un illuminant dont le spectre est uniforme suivant les longueurs d'ondes et qui ne varie qu'en intensité.

4. Les caractéristiques physiques sont ici les caractéristiques de réflectance des surfaces, la forme, l'illuminant.

invariants couleur dans le cas complexe d'objets spéculaires vus sous un illuminant coloré. Cette partie s'intéresse plus généralement aux points d'intérêt invariants, tout d'abord ceux définis en couleur (section 2.3.1) puis en binaire (section 2.3.2).

2.3.1 Descripteurs couleur

Les descripteurs de points d'intérêt définis à partir de gradients sont par nature robustes aux changements uniformes de contraste : Harris [ST94], SIFT [Low04], SURF [BETV08]. Ils sont encore plus stables s'ils sont calculés à partir d'invariants couleur. Pour ce qui est de SIFT, les composantes opposées $O_1O_2O_3$ offrent le meilleur compromis entre invariance et séparabilité dans les applications de reconnaissance d'objets [vdSGS10].

La prise en compte d'un voisinage ouvre également la voie aux approches de constance chromatique telles que *retinex* [FCM00, KFC10] qui permettent de déterminer les réflectances réelles malgré un illuminant coloré. Reposant sur l'hypothèse que la luminosité d'un pixel dépend de sa propre réflectance et de celle de ses voisins, cette méthode estime localement la couleur de l'illuminant et la soustrait de chaque couleur dans la région d'étude. La constance chromatique peut également être obtenue par simple modification d'histogrammes, par exemple par égalisation (EH). Finlayson *et al* [FHST05] ont montré qu'appliquer l'EH marginalement aux images couleur permet de préserver l'ordre de rang des couleurs. L'EH s'avère invariant à l'ensemble des changements d'illumination pouvant être modélisés par une fonction monotone, ce qui est le cas de la plupart des modèles à commencer par ceux décrits dans la Section 2.1.

Alors que la plupart des invariants couleur ont été définis dans le cadre classique de la vision par ordinateur, les applications de réalité augmentée spatiale font émerger de nouveaux verrous. Une image f subit une série de déformations avant d'être rendue dans la scène, à savoir des modifications spectrales (liées aux caractéristiques du vidéo-projecteur, à la réflectance de la surface de projection, à l'éclairage ambiant) et géométriques (distorsions radiales du vidéo-projecteur, irrégularités de la surface de projection), avant d'être vue par une caméra pour former l'image captée g . Dans le cadre de la thèse d'Aleksandr Setkov [Set15, SG16, SGJ13, SGJ16], nous nous sommes intéressés à l'invariance couleur entre ces deux images f et g . L'objectif de la thèse était de compenser géométriquement la projection de façon non-intrusive, sans lumière structurée ni télémètre laser, mais par l'appariement entre f et g puis par l'estimation et l'inversion du modèle géométrique de déformation, comme l'illustre la figure 2.3.

Des points-clés de type SURF couleur sont exploités pour l'appariement, notre contribution se situant au niveau de la comparaison des variantes couleur invariantes. Nous avons montré, à la fois théoriquement et expérimentalement, que l'égalisation d'histogramme était particulièrement adaptée à notre cas d'étude. En effet, la forme du spectre $\mathcal{E}(\lambda, q)$ produit par le projecteur dépend de la technologie, qui peut être tri-LCD⁵ ou DLP⁶. Dans le cas des vidéoprojecteurs tri-LCD auxquels nous nous sommes intéressés, les lampes contiennent de la vapeur de mercure, dont le spectre d'émission est discret et possède généralement trois faisceaux monochromatiques dominants, un par canal couleur k . En supposant l'absence d'illumination de la scène, le terme \mathcal{I} s'annule dans l'équation (2.2). Ensuite, du fait du caractère discret du spectre d'émission du vidéo-projecteur $\mathcal{E}(\lambda, p) = 0$ excepté en un nombre fini de longueurs d'onde, l'opérateur d'intégration est également remplacé par une somme finie dans (2.1). Ainsi, l'image g_k du canal k provenant de l'acquisition de l'image projetée dans la scène g devient :

$$g_k = \mathcal{L}(\lambda_k)\mathcal{R}(\lambda_k)\mathcal{S}_k(\lambda_k). \quad (2.7)$$

Aucune hypothèse n'est faite sur la fonction de réponse qui relie l'image d'entrée f_k au flux émis

5. Cette technologie est transmissive et est fondée sur de minuscules écrans LCD transparents (de 0,55" à 0,9")

6. La technologie DLP est de nature réflexive et est fondée sur un milliers de miroirs mobiles minuscules



FIGURE 2.3 – Exemple de compensation géométrique de vidéos projetées [SG16]. La méthode fait coopérer d’une part un appariement de points d’intérêt couleur pour compenser les fortes distorsions, d’autre part une approche de flot optique pour corriger les faibles déformations entre deux images successives. Elle opère l’appariement entre la vidéo de référence (en haut à gauche) et la vidéo observée. Dans ce travail, le système caméra-projecteur est mobile et le contenu de la vidéo est dynamique. À droite, on compare la compensation obtenue avec la compensation issue de la vérité terrain (*ground-truth* GT).

par le vidéo-projecteur, hormis le fait qu’il s’agit d’une fonction monotone F_k :

$$\mathcal{L}(\lambda_k) = F_k(f_k) \quad (2.8)$$

Finalement, sous ces hypothèses, l’image s’exprime par :

$$g_k = F_k(f_k)\mathcal{R}(\lambda_k)\mathcal{S}_k(\lambda_k) \quad (2.9)$$

Sur la base de ce modèle, nous avons montré dans [SGJ16] que l’égalisation d’histogramme est un invariant photométrique lorsque la surface de projection est spatialement uniforme en couleur puisque le rang des couleurs est préservé. Dans le cas non-uniforme, le rang est préservé localement dans des cas bien précis : soit lorsque la région (image ou surface) est de couleur uniforme, soit lorsque les couleurs de l’image et celles de la surface varient dans le même sens.

Afin de comparer les descripteurs et vérifier ces hypothèses, nous avons évalué dans [SGJ13, SGJ16] le taux d’appariement [vdSGS10] et la précision de la compensation géométrique (réalisée ici par RANSAC) en s’appuyant sur une vérité terrain. Un grand nombre d’images a été testé, à la fois des images de synthèse simulant des variations photométriques et géométriques et des images réelles produites par le système projecteur-caméra. Ces expériences ont montré que les descripteurs fondés sur HSV⁷, *retinex* ou ceux utilisant la luminance produisent des résultats insuffisants quelles que soient les conditions d’acquisition. Les résultats obtenus avec égalisation d’histogramme sont nettement supérieurs en termes de taux d’appariement⁸ tout en montrant la plus

7. HSV est l’espace perceptuel Teinte Saturation Intensité, ou *Hue-Saturation-Value* en anglais

8. En termes de taux d’appariement, ce descripteur est classé avec un rang moyen de 1.38 parmi les 7 descripteurs testés.

faible erreur de compensation⁹. Enfin, nous avons pu observer que les performances du descripteur fondé sur $O_1O_2O_3$, pourtant très bonnes dans le cas d'applications plus classiques d'appariement d'images [vdSGS10], sont ici fortement dépendantes du contenu de l'image et s'écroulent dès que les images sont peu texturées.

2.3.2 Descripteurs binaires

De nombreux descripteurs binaires ont été définis ces dernières années, comme FAST [RPD10], BRISK [LCS11] et FREAK [AOV12], et plébiscités pour leur rapidité de calcul et leur bon pouvoir de séparabilité. De façon assez similaire, les Motifs Binaires Locaux LBP (pour *Local Binary Patterns*) sont des descripteurs de texture [OPH94] qui se définissent à partir de différences de luminances locales binarisées. Étant donné que ces primitives sont simples à calculer tout en s'avérant aussi discriminantes et invariantes que les descripteurs couleur précédents, elles rivalisent avec les méthodes de bancs de filtres de type Gabor. L'opérateur consiste à analyser un voisinage circulaire (de rayon R paramétrable) autour d'un pixel central de niveau de gris g_c . Cette valeur sert de seuil pour chacun de ses voisins de valeur g_p de sorte que le signe $s(g_p - g_c)$ est codé par « 1 » si $g_p > g_c$, « 0 » sinon. Enfin, chaque bit dispose d'un poids 2^P en fonction de sa position dans le voisinage, et l'on peut en déduire une valeur décimale $LBP_{P,R}$:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p. \quad (2.10)$$

Il s'agit donc d'une représentation à la fois compacte et peu sensible aux changements de contraste. Après rotations jusqu'au code minimum, elle s'avère également invariante aux changements de rotation.

Choi [CPR10] a proposé une version couleur marginale en calculant un motif LBP par composante. De façon analogue, [ZBC13] exploite les invariants couleur pixelliques évoqués dans la section 2.2 dans une approche multi-échelle avec réduction de dimensionnalité. Une approche plus vectorielle, appelée LCVBP pour *Local Color Vector Binary Patterns* [LCRP12] définit les LBP couleur sur deux composantes, l'angle et la norme. Enfin Ledoux et al. [LLM16] ont récemment proposé une comparaison de différents ordres couleur pour la définition de nouveaux LBP.

Les ULBP pour *Uniform LBP* [OPM02a] correspondent aux LBP dont le code présente au maximum deux transitions, par exemple 00000000 et 01110000. Il a été montré de façon expérimentale que les ULBP se retrouvent dans la plupart des textures naturelles et ignorent les motifs plus chaotiques (et souvent associés au bruit). Leur représentation s'avère naturellement plus compacte et leur histogramme moins épars. La figure 2.4 montre les 56 différents motifs de ULBP. Nous avons proposé dans [RGL13b] une représentation plus stable vis-à-vis des faibles variations d'orientation (voir la figure 2.5). Les ULBP sont décrits par deux angles θ_0 et θ_1 marqués sur la figure par les flèches incurvées. Puisque les calculs sont faits sur des quantités angulaires, les angles sont convertis en points sur le cercle unité, et θ_0 est représenté par $(\cos\theta_0, \sin\theta_0)$. Le vecteur :

$$\alpha(\theta_0, \theta_1) = [\cos(\theta_0) \sin(\theta_0) \cos(\theta_1) \sin(\theta_1)] \quad (2.11)$$

décrit alors de manière unique chaque *ULBP* comme une fonction de (θ_0, θ_1) . Un exemple de calcul de $\mathbf{v}(\theta_0, \theta_1)$ est fourni sur la figure 2.4(b) où la valeur décimale LBP (bit string 00111111) est représentée par les angles $\theta_0 = \pi/8$ et $\theta_1 = 5\pi/8$. Les 56 résultats de la fonction α peuvent être précalculés et stockés dans une *Look-Up Table* pour être utilisés en-ligne.

Nous avons proposé 4 extensions couleur de ces ELBP (pour *Enhanced-LBP*) dans [GyTL17]. La première d'entre elles (ELBP_a) calcule un ELBP marginalement sur chaque composante tandis

9. En termes de précision de la compensation géométrique, le rang moyen de ce descripteur est de 1.11 sur les 4 descripteurs testés.

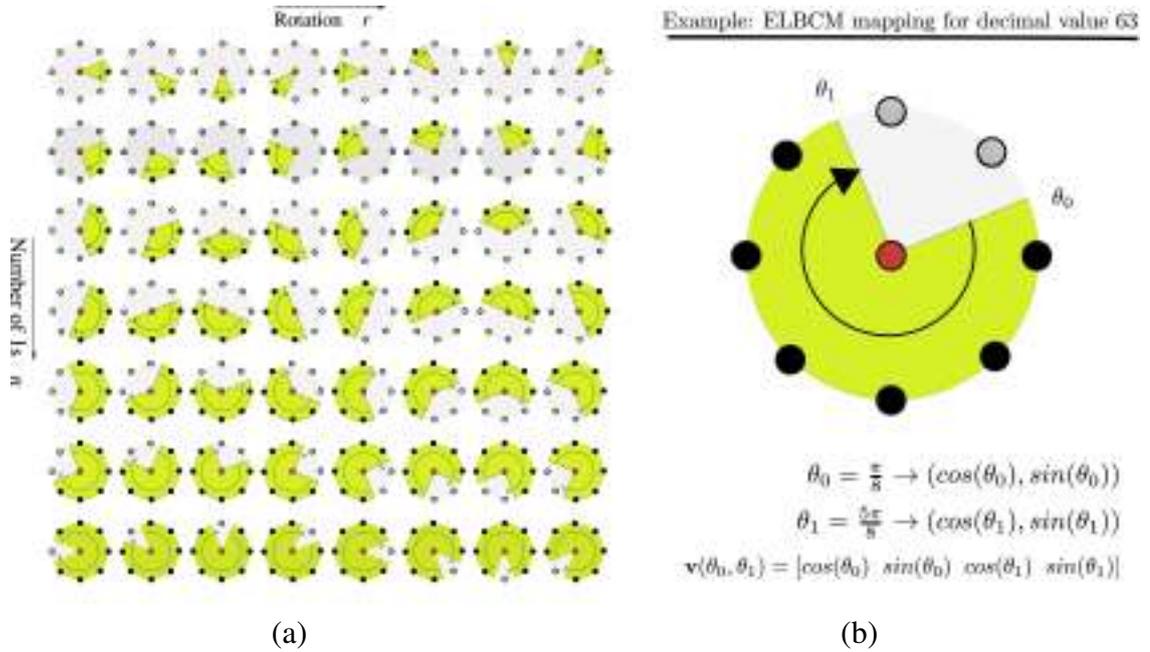


FIGURE 2.4 – (a) Ensemble des 56 différents motifs *LBP* uniformes dans un voisinage $(8, R)$. Pour chaque motif, les flèches indiquent le début et la fin des angles (θ_0 et θ_1). (b) Le motif uniforme *LBP* est défini par deux angles θ_0 et θ_1 et est décrit par $\mathbf{v}(\theta_0, \theta_1) = [\cos(\theta_0) \ \sin(\theta_0) \ \cos(\theta_1) \ \sin(\theta_1)]$. Le *LBP* de valeur décimale 63 correspond au vecteur $\mathbf{v}(\pi/8, 5\pi/8) = [\cos(\pi/8) \ \sin(\pi/8) \ \cos(5\pi/8) \ \sin(5\pi/8)]$.

que les motifs ELBPb à ELBPd sont des versions vectorielles qui se différencient par la couleur choisie comme référence et le mode de comparaison (en remplacement de la fonction $s(g_p - g_c)$). Les bases d'images *ALOIColor* et *ALOIAngle* [GBS05] ont servi à évaluer la séparabilité des descripteurs sous différents illuminants. Pour les 20 premiers objets de la base, les corrélations intra et interclasse \mathcal{C}_{nn} et \mathcal{C}_{nm} sont calculées et moyennées sur RGB et le rapport permet d'évaluer le compromis entre invariance et séparabilité. Les résultats réunis dans le tableau 2.1 montrent des rapports plus élevés pour les descripteurs proposés, avec une corrélation intra classe similaire mais une plus faible corrélation inter classe, donc une meilleure séparabilité. L'apport en séparabilité est plus marqué dans le cas d'un illuminant coloré.

TABLE 2.1 – Valeurs de corrélation inter et intra classe pour les descripteurs *LBP* représentés en RGB ($\times 10^{-2}$) sur la base d'images *ALOI* ainsi que leurs quotients.

	MOTIF LBP								
	LBP	uLBP	rorLBP	nriLBP	LCVBP	ELBPa	ELBPb	ELBPc	ELBP
ALOI COLOR (20 PREMIERS OBJETS)									
\mathcal{C}_{nn}	73.60	65.24	73.02	66.64	27.02	64.46	71.37	71.40	71.70
\mathcal{C}_{nm}	1.22	1.42	2.48	1.37	0.4	0.85	0.61	0.74	0.60
$\mathcal{C}_{nn}/\mathcal{C}_{nm}$	60.33	45.94	29.44	48.64	67.65	75.84	117	96.5	119.5
ALOI ANGLE (20 PREMIERS OBJETS)									
\mathcal{C}_{nn}	62.43	62.08	88.93	61.94	21.50	51.89	49.36	48.46	51.43
\mathcal{C}_{nm}	1.02	1.44	2.27	1.42	0.36	0.70	0.47	0.45	0.58
$\mathcal{C}_{nn}/\mathcal{C}_{nm}$	61.21	43.11	39.18	43.62	59.72	74.13	105.02	107.69	88.67

Ainsi, nous avons proposé une variante des motifs *LBP*, calculée à partir des angles formés par les *ULBP*, ainsi que plusieurs extensions en couleur. Ces descripteurs binaires offrent une bonne

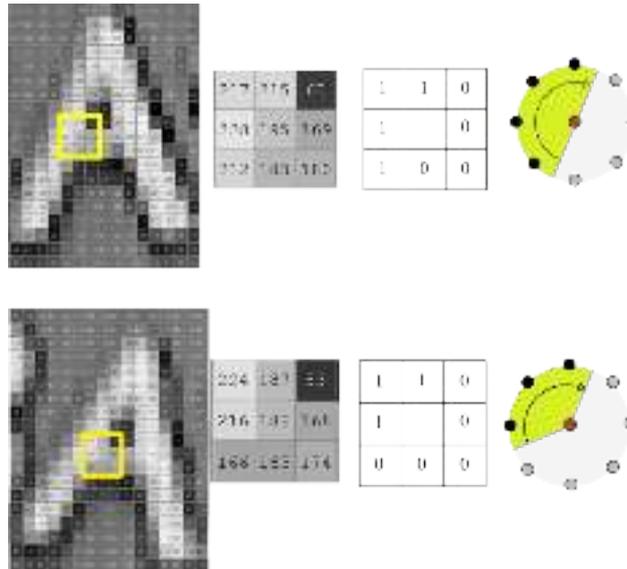


FIGURE 2.5 – Exemple d’instabilité liée à la représentation LBP dans le cas de faibles rotations. Dans cet exemple, les valeurs LBP passent de 195 à 131 (une décroissance de 32,8 %). La représentation proposée est un peu plus stable $3\pi/8$ au lieu de $4\pi/8$ soit une modification de 25 %).

stabilité vis-à-vis des distorsions de la texture. Ces LBP seront exploités dans le chapitre suivant pour la définition d’un descripteur d’apparence.

2.4 Cartes topographiques

Au delà de l’invariance définie localement à partir d’un voisinage de pixels, les ensembles et lignes de niveau (*level-sets* et *level-lines*) introduits dans [CCM99] constituent des primitives visuelles globales qui s’avèrent robustes aux variations uniformes de contraste. Soit $I(p)$ l’intensité au pixel p . I peut être décomposée en *ensembles de niveau supérieurs* (*upper level sets* en anglais) \mathcal{N}^u ou en *ensembles de niveau inférieurs* (*lower level sets* en anglais) \mathcal{N}^l :

$$\mathcal{N}^u(\epsilon) = \{p, I(p) \geq \epsilon\} \quad \mathcal{N}^l(\epsilon) = \{p, I(p) \leq \epsilon\} \quad (2.12)$$

où ϵ se réfère ici au niveau considéré : $\epsilon \in [0 \dots 2^n - 1]$, pour une image codée sur n bits. Les ensembles supérieurs (resp. inférieurs) définis à un niveau ϵ sont par définition inclus dans les ensembles du niveau inférieurs (resp. supérieurs). Les *lignes de niveaux*, qui sont les frontières des ensembles de niveau, forment un ensemble de courbes de Jordan qui produisent une description exhaustive ou complète de l’image. L’ensemble des lignes de niveau est appelé la *carte topographique* et forme un arbre d’inclusion.

Cartes topographiques couleur. La définition de lignes couleur nécessite le respect des mêmes propriétés que celles des lignes de niveau à savoir l’exhaustivité, l’inclusion et l’invariance au contraste. Or, alors que l’échelle de gris est totalement ordonnée, il n’en est pas de même pour les couleurs. Dans [CF00] et [CCM02], les auteurs proposent de calculer les lignes dans l’espace HSV, dont les composantes sont moins corrélées que RGB et plus conformes aux propriétés du SVH. Les couleurs sont totalement ordonnées de manière lexicographique en favorisant l’intensité puis la teinte et la saturation, dans l’objectif d’imiter les lois du SVH. Dans [CCM02], les auteurs calculent la carte topographique de luminance et la considèrent comme étant constante par morceaux, chaque composante connexe héritant d’une saturation et d’une teinte moyenne. Bien que ces deux approches prennent effectivement en compte les caractéristiques du système visuel humain,

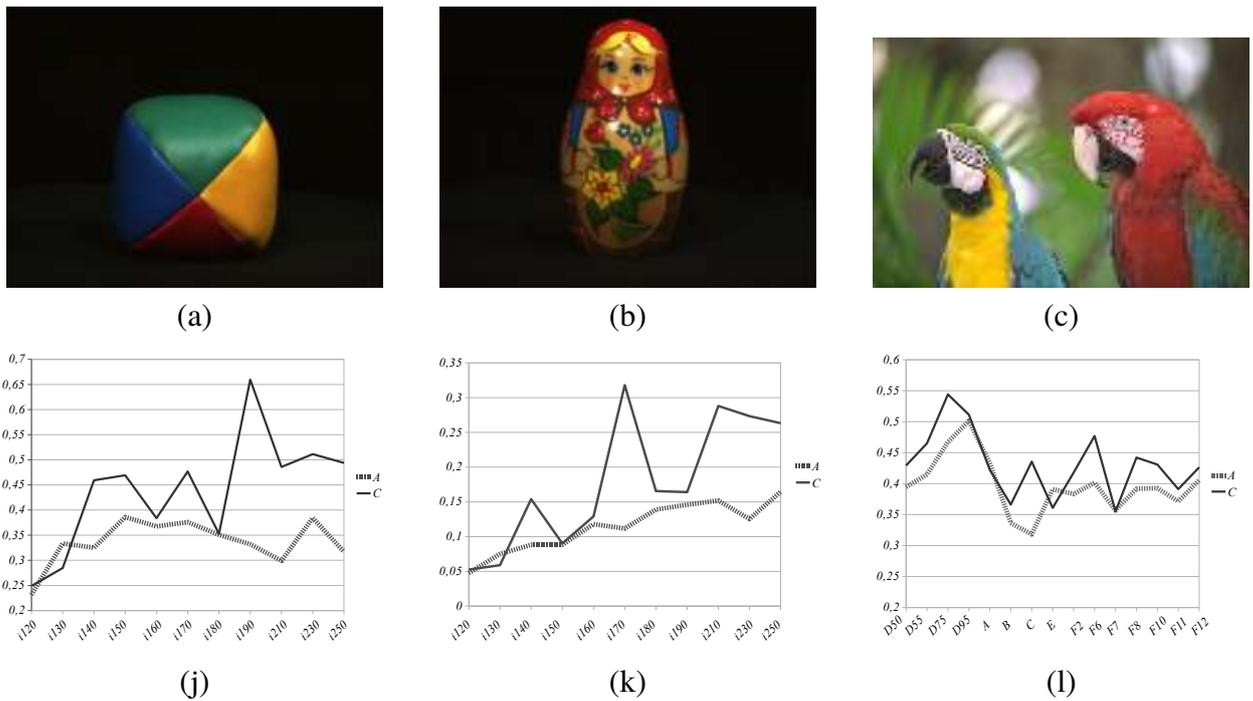


FIGURE 2.6 – Stabilité des lignes de niveaux vis-à-vis des variations d’illumination. On considère 16 niveaux de luminance ϵ . (a) (b) (c) Images initiales. (d) (e) (f) Pourcentage de lignes stables. Les lignes en pointillé (A) correspondent aux résultats obtenus avec [CF00] tandis que ceux en trait plein (C) sont ceux de l’approche que nous avons proposée.

les spécificités de l’espace HSV ne sont pas considérées, en particulier la mauvaise définition de la teinte pour de faibles saturations. Définir les ensembles de niveaux directement dans l’espace RGB est une solution que nous avons proposée pour attaquer ce problème de sensibilité au bruit, et que nous détaillons maintenant.

Carte topographique RGB. Comme précisé dans la section 2.1.2, la plupart des couleurs des images naturelles sont localisées dans l’espace RGB le long de vecteurs de réflexion diffuse c_b . Alors que les ensembles de niveaux de gris sont extraits le long de l’axe achromatique de l’espace RGB, nous proposons d’extraire les ensembles et lignes couleur le long de chaque vecteur de réflexion diffuse c_b révélé par l’image. Contrairement aux propositions précédentes [CF00, CCM02], l’approche que nous avons développée dans [GZ08] ne requiert aucune conversion couleur et ne favorise pas la luminance par rapport aux autres composantes. La première étape de l’algorithme capture les couleurs en fonction de leur distance au noir sans distinction entre les directions de couleurs intrinsèques c_b . Les ensembles de niveaux sont ensuite définis indépendamment le long de chaque vecteur. Toutefois, leur mise en œuvre est relativement lourde car elle nécessite le calcul d’un arbre d’inclusions défini à partir de plusieurs passes d’étiquetage en composantes connexes, à la fois dans l’image et dans l’espace couleur.

Carte topographique dichromatique dans $(\theta\phi\rho)$. Partant de l’espace de représentation sphérique (ρ, θ, ϕ) (voir la figure 2.1 (c)), les ensembles de niveaux ϵ sont définis selon l’axe ρ . À chaque niveau, un histogramme 2D $H_\epsilon(\theta, \phi)$ est calculé. D’après le modèle dichromatique, une forte densité d’une couleur dans cet histogramme doit révéler un vecteur c_b , associé à un ou plusieurs objets de l’image. Les modes sont extraits à chaque niveau puis fusionnés pour déduire finalement l’ensemble des vecteurs de réflexion diffuse. Chaque pixel de l’image est alors agrégé à un ensemble couleur défini par une couleur intrinsèque et un niveau ρ .

Nous avons pu montrer que ces lignes couleur [GZ11] montrent une bonne invariance vis-à-vis de certains types de changements d’illumination, ceux qui impliquent une modification d’échelle

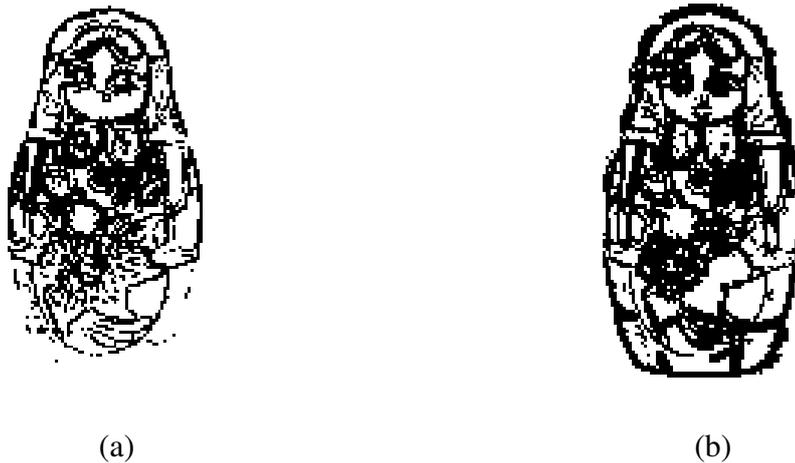


FIGURE 2.7 – Portions de lignes stables sur les 11 images de la poupée russe. (a) approche proposée par Coll et al. [CF00]. (b) approche que nous avons proposée dans [GZ11].

des vecteurs diffus ou des rotations dans l'espace couleur. En prenant l'exemple de trois objets vus sous 11 illuminants différents, la figure 2.6 compare les taux de lignes conservées d'une part par l'approche [CF00] (en gris clair), d'autre part par l'approche que nous avons proposée (en gris foncé). Les deux images de la figure 2.7 représentent en noir les portions de lignes préservées. Les cartes topographiques que nous avons définies permettent de mieux conserver la morphologie de l'objet.

Appariement d'ensembles de niveau. Dans la thèse de Noppon Lertchuwongsa que nous avons co-encadrée avec Bertrand Zavidovique [LGZ11, Ler11], trois méthodes d'appariement stéréoscopiques ont été proposées aux cartes topographiques et aux régions. La première propose un calcul de disparité éparse en utilisant une distance de Hausdorff comme mesure de dissimilarité. La disparité est ensuite raffinée par un appariement local afin qu'une primitive ligne puisse hériter de plusieurs valeurs de disparité. Les éventuelles occultations et les déformations des lignes dues aux changements de point de vue sont ainsi mieux prises en compte. L'utilisation des lignes couleur par rapport aux lignes de niveau classiques a permis d'améliorer la précision de la carte de disparité grâce à un meilleur appariement des primitives. De plus le nombre de lignes couleur est réduit, ce qui a permis d'accélérer le calcul de la carte de disparité. Les deux autres approches fournissent une disparité dense à partir d'ensembles couleur, soit par coopération entre une mise en correspondance éparse et une segmentation en région [LGZ12]. La figure 2.8 montre un exemple de résultat pour chacune de ces trois méthodes.

2.5 Fiabilité de la couleur

Les invariants couleur évoqués dans la Section 2.2 supposent des couleurs suffisamment saturées. Dans le cas contraire, ils réduisent la séparabilité entre les couleurs tout en restant plus coûteux en termes de temps de calcul. Ainsi, leur fiabilité varie à la fois spatialement dans l'image en fonction des caractéristiques de la scène, et temporellement en fonction des changements d'illumination.

En partant d'une approche de suivi de points de type KLT, nous avons fait coopérer l'information classique de luminance avec un invariant couleur local (défini dans un voisinage \mathcal{W} de taille 3×3) à trois composantes défini pour être stable malgré des changements de couleur d'illuminant [Gou07]. Une fonction de pertinence de la teinte est définie comme une fonction croissante du niveau de saturation, comme l'avait précédemment défini Carron [CL94] pour le calcul de gradient couleur.

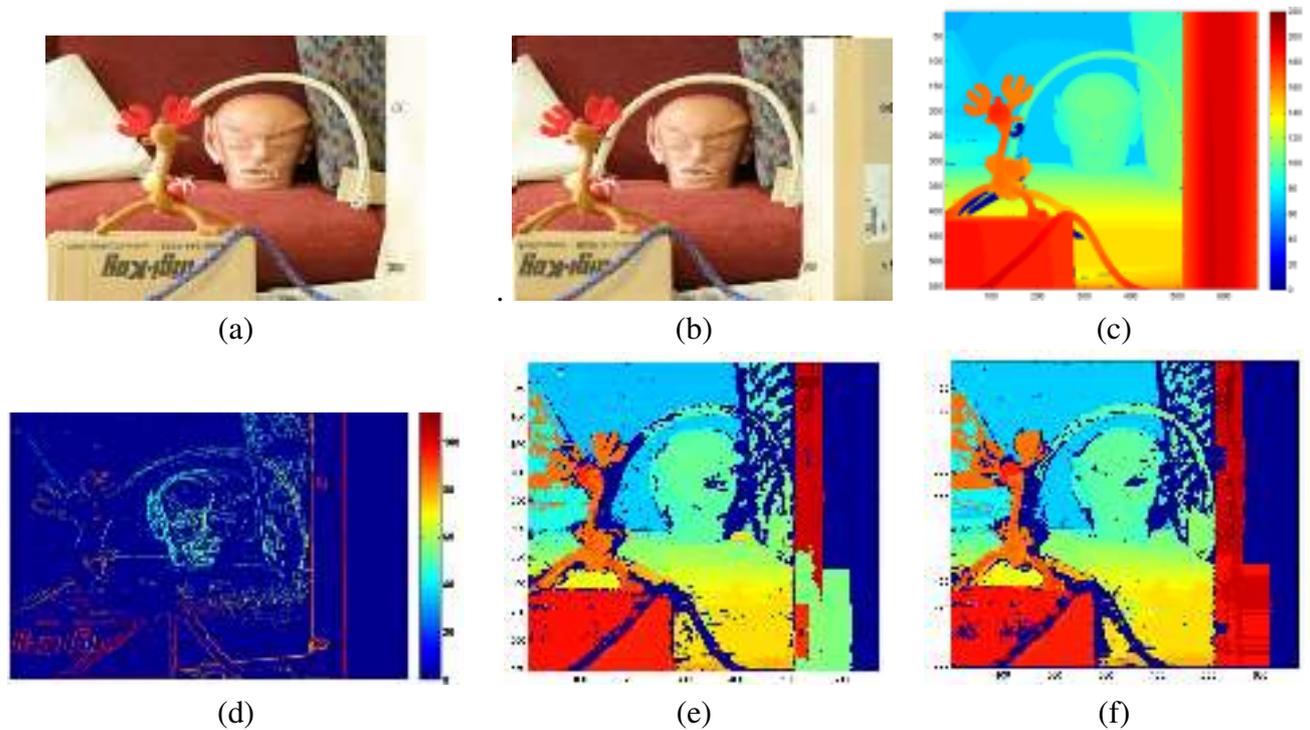


FIGURE 2.8 – Exemple de cartes de disparités calculées à partir de lignes de niveaux sur l’image Reindeer. (a) et (b) : images gauche et droite; (c) vérité terrain, (d) carte de disparité éparse obtenue par transformée de Hausdorff modifiée, (e) méthode *Shape of the Disparity Histogram*, (f) Enrichissement de la carte de disparité par une segmentation en régions.

Servant de fonction de mélange entre les deux attributs couleur, elle alloue une importance plus grande à l’invariant lorsqu’il est pertinent et à la luminance dans le cas contraire.

Afin d’en accélérer le calcul, l’invariant a été remplacé par une représentation définie sur un seul canal au lieu de trois [RGL11] $m = \max(r, g, b)$. Il représente en tout point la valeur la plus saturée parmi les trois composantes normalisées rgb . À l’instar de la plupart des invariants couleur, m réduit la séparabilité entre les couleurs puisque toutes les nuances de gris sont confondues ($r = g = b = \frac{1}{3}$) et deux couleurs différentes peuvent avoir une valeur m identique. Nous avons fait l’hypothèse que ce cas de figure est peu probable dans les images naturelles. En guise d’exemple, la figure 2.9 montre quelques images de la séquence *Cardgame* provenant de la base de donnée ALOI¹⁰. La première ligne montre des images *RGB*, avec des variations d’illumination nettement visibles. L’invariant m et la teinte H sont montrés sur les 2^e et 3^e lignes respectivement. Comme attendu, les variations photométriques ne sont plus visibles mais le contraste est dégradé.

En guise d’évaluation, nous avons comparé pour différentes séquences d’images la robustesse (nombre de points correctement suivis) et le temps de calcul de KLT pour différents attributs : l’intensité I avec normalisation photométrique, *RGB*, *rgb*, *HSV*, ainsi que les composantes seules H et m . Pour finir, nous avons comparé les coopérations $I + H$ et $I + m$ avec fonction de pertinence. Les résultats que nous avons obtenus dans la thèse d’A. Romero [RGL11] ont montré qu’utilisé seul, l’invariant m fait échouer le suivi lorsque la saturation est faible tandis que la méthode $I + m$ est capable de contrer cette perturbation. En effet, elle permet de maintenir les performances du suivi dans le cas de séquences aux couleurs saturées tout en améliorant significativement les résultats dans les cas de faibles saturations. D’autre part, utiliser l’ensemble des trois composantes *rgb* n’améliore pas forcément le suivi par rapport à m mais augmente toujours significativement les temps de calcul.

10. <http://staff.science.uva.nl/~aloi/>

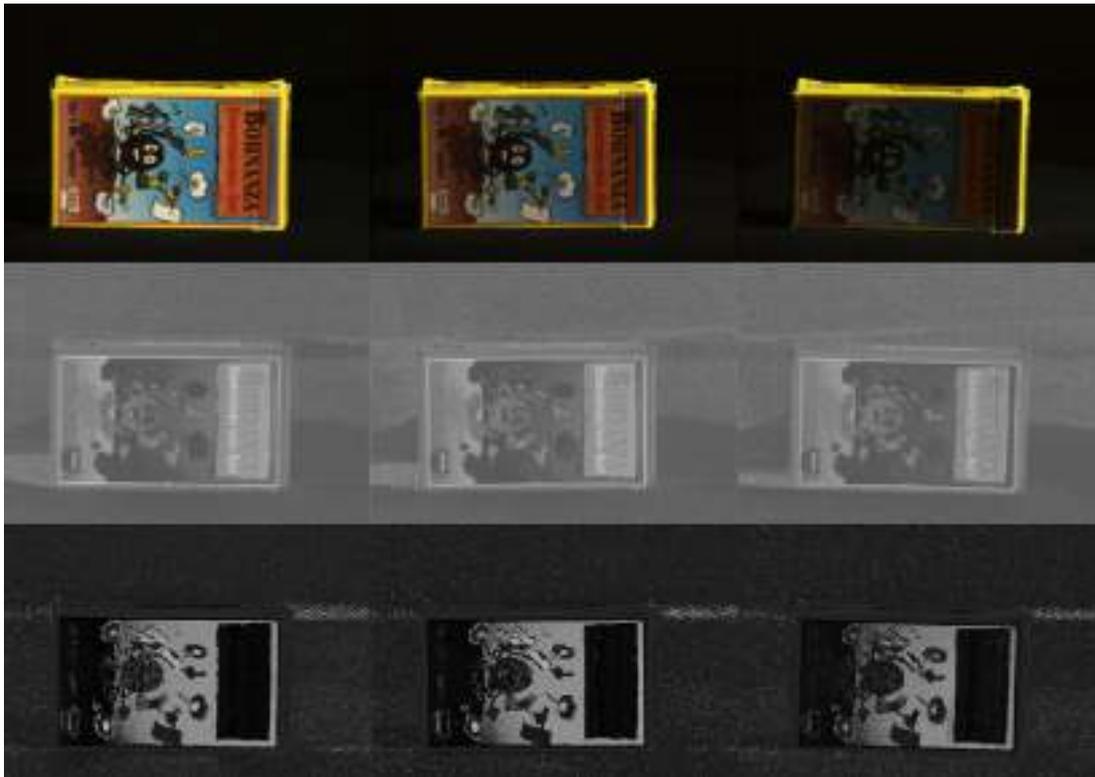


FIGURE 2.9 – De haut en bas : images *RGB* initiales; invariant m ; Teinte (représentée artificiellement dans un intervalle $[0-255]$ même s’il s’agit d’une valeur angulaire. Par conséquent une partie du bruit est due à la représentation et pas seulement aux singularités de la teinte).

Enfin, nous avons montré la pertinence du choix de m comparé à la teinte pour l’ensemble des séquences de test.

2.6 Effacer la couleur

Vouloir effacer la couleur peut paraître insolite à première vue. Nous parlons ici de techniques de réalité augmentée spatiale qui utilisent des systèmes projecteur-caméra (ProCams), et où le contenu virtuel projeté est adapté en fonction de la scène, considérée ici à la fois colorée, non-planaire et même dynamique. C’est le sujet abordé dans la thèse Panagiotis-Alexandros Bokaris [BGJC14, Bok16] que nous avons co-encadrée avec Christian Jacquemin (LIMSI). Par des techniques de compensation photométrique, il s’agit de faire en sorte que le spectateur observe la même image que celle qui apparaîtrait sur une surface blanche.

Notre système ProCam est composé d’un vidéo-projecteur, d’une caméra RGB de bonne résolution spatiale et spectrale, et enfin d’un capteur RGB+D. En premier lieu, le système est calibré colorimétriquement et géométriquement. Pour ce qui est du calibrage géométrique, le projecteur envoie dans la scène une série de motifs de lumière structurée en niveaux de gris. Afin de couvrir l’ensemble du volume 3D qui nous intéresse, un objet plan coloré (bleu en l’occurrence) sert de surface de projection et est déplacé en différentes positions, produisant une série d’images de calibrage. Celles-ci sont acquises simultanément par le capteur RGB et la caméra RGB+D. Les correspondances sont cherchées entre les pixels de chaque plan image et leurs coordonnées 3D dans le volume de projection. Cette étape qui peut paraître lourde à première vue, ne prend en fait qu’une minute et n’est réalisée qu’une seule fois hors-ligne.

À partir de ce système, une méthode de compensation couleur a été définie pour des surfaces dynamiques. Le caractère dynamique implique de limiter le nombre de motifs de calibrage cou-

leur nécessaires à la compensation. De plus, la compensation doit être suffisamment précise pour éviter des aberrations trop visibles. Enfin, le code doit être suffisamment rapide pour s'ajuster aux mouvements de la scène. Après étude de l'état de l'art, nous n'avons identifié que deux approches comparables. Fujii et al. [FGN05] proposent une méthode hybride combinant un calibrage photométrique hors-ligne et un contrôle dynamique qui nécessite la projection de 4 motifs de calibrage. La méthode de Bimber et al. [BR05a] nécessite deux images de calibrage, mais ne prend pas en compte les intercorrélations entre les composantes couleur. Ces deux approches posent l'hypothèse d'une réflectance constante et d'un éclairage ambiant constant.

Sous les mêmes hypothèses, la méthode que nous avons proposée dans le cadre de cette thèse ne nécessite qu'une image de calibrage pour estimer la réflectance spectrale de la surface de projection et ainsi en supprimer la contribution visuelle. Cela est rendu possible par l'apprentissage hors-ligne des relations réflectances-couleurs d'un ensemble de surfaces colorées choisies pour couvrir le spectre du visible. En se plaçant sous l'illuminant de référence, les réflectances sont mesurées à l'aide d'un spectro-radiomètre et les coordonnées couleur à l'aide de la caméra utilisée par la suite pour la compensation. Ainsi, face à une surface de projection inconnue, les nouvelles réflectances sont déduites par minimisation d'une fonction de coût impliquant les réflectances préalablement apprises.

Cette méthode a été entièrement implémentée sur GPU pour permettre une compensation en temps-réel. Nous avons confronté cette approche aux méthodes [FGN05, BR05a] en évaluant les mesures d'erreurs (RMS¹¹ et S-CIELAB¹²) entre l'image initiale et l'image observée après compensation. Sur les six surfaces colorées et les six images de référence testées, les résultats sont en faveur de notre approche. En guise d'exemple, la figure 2.10 compare les résultats de compensation pour les trois méthodes considérées.

Le potentiel artistique des techniques de réalité augmentée proposées a été exploré à travers l'installation artistique Gardien du Temple, que nous avons réalisée en collaboration avec la metteur en scène Véronique Caye, qui était par ailleurs associée au suivi des travaux de thèse de P-A. Bokaris. La révélation et l'effacement ont toujours été un terrain d'expérimentation commun aux artistes et aux scientifiques. Cette installation audiovisuelle et interactive utilise les techniques de réalité augmentée pour effacer la présence d'un poème écrit sur un tapis, invitant le spectateur à le révéler. Des performances réalisées par l'artiste circassienne Ashtar Muallem (voir Fig.2.11) permettaient de donner vie à ce texte. Dans de telles applications, la précision et la robustesse des méthodes utilisées sont cruciales. Cette installation artistique a donc permis de tester et d'évaluer les travaux de cette thèse dans un contexte collaboratif et exigeant [BGC⁺17].

2.7 Conclusion

Définir des primitives visuelles couleur, utiliser les invariants couleur sans réduire les performances des traitements, estimer les variations photométriques pour les compenser, voilà trois des objectifs qui ont guidé nos réflexions autour de la couleur. Les approches proposées se situent au tout premier niveau de la chaîne de traitement puisqu'elles se réfèrent à la couleur d'un pixel, à celle d'un voisinage ou d'un ensemble connexe. L'objectif visé est à chaque fois un appariement en vue d'une tâche de plus haut niveau : la reconnaissance de motifs, le suivi de points d'intérêt, la compensation couleur, le calcul d'une carte de disparité.

Le chapitre qui suit nous amène au niveau suivant de la chaîne de traitement. À partir de primitives visuelles de bas-niveau telles que celles décrites dans ce chapitre, vient la notion d'objet. De même que nous avons pris ici la couleur comme fil rouge, l'apparence est mise au cœur de la

11. Root Mean Square error

12. La distance couleur S-CIELAB correspond à une extension de la distance ΔE , où l'image est préalablement filtrée par des filtres gaussiens de façon à prendre en compte la sensibilité au contraste.

réflexion.

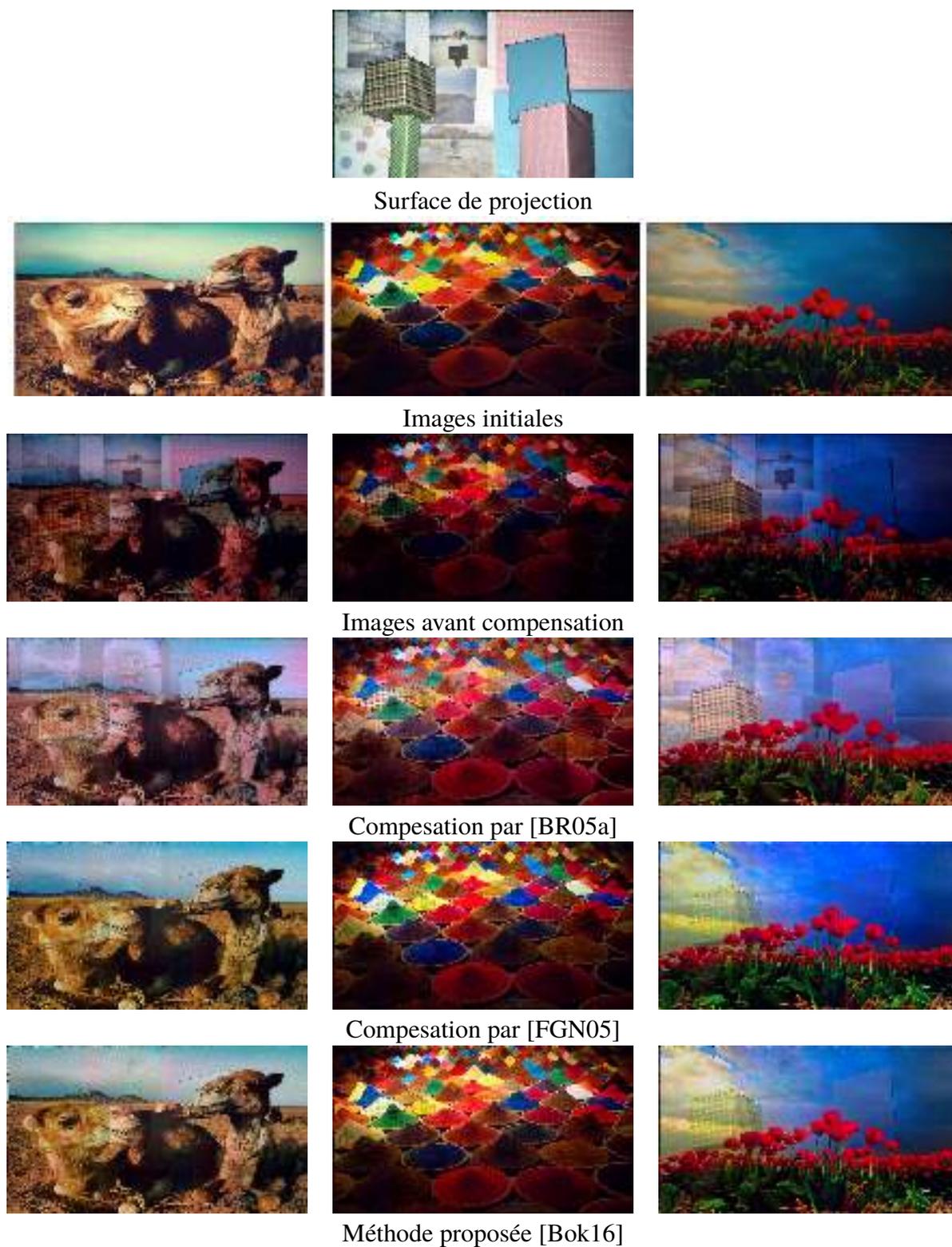


FIGURE 2.10 – Comparaison des résultats de compensation couleur par vidéo-projection.

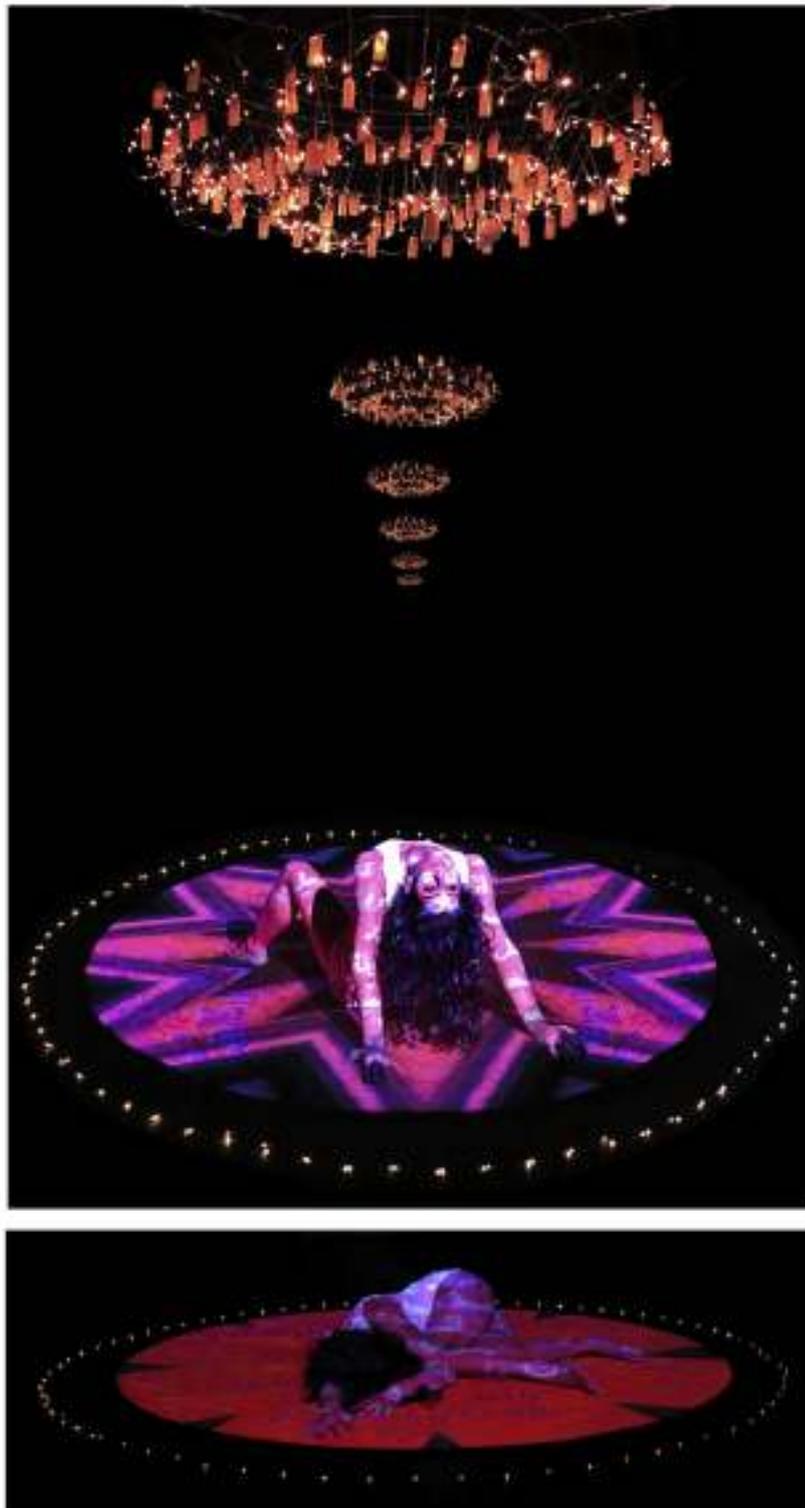


FIGURE 2.11 – Image de l'installation-performance Gardien Du Temple prise au festival CURIOSITAS 2015.

Chapitre 3

Apparence

LA reconnaissance d'images ou d'objets peut être abordée suivant deux paradigmes différents, d'une part la recherche de catégories génériques (des piétons, des voitures), d'autre part l'identification d'instances d'un objet cible bien particulier (*Où est Charlie ?*). Dans le premier cas (celui des approches dites *génératives*), le modèle est appris au préalable à partir d'un ensemble d'images. Dans le second cas (celui des approches dites *descriptives*), les descripteurs peuvent être spécifiques à l'objet. Il en est de même pour les approches de suivi temporel. En effet, des objets génériques peuvent être détectés indépendamment dans chaque image de la vidéo puis appariés entre deux prises de vue. Une autre stratégie consiste à détecter l'objet dans la première image de la vidéo puis à retrouver cette cible particulière dans la suite de la séquence.

Quoi qu'il en soit, le succès d'un descripteur dépend de sa capacité à assurer une bonne séparabilité vis-à-vis des autres régions présentes (soit dans la base de données, soit dans la scène courante) afin d'éviter les possibles ambiguïtés d'appariement, tout en restant stable face aux changements, que ceux-ci soient liés à l'objet ou dûs à des facteurs externes. Le choix se fonde généralement sur un compromis entre ces deux aspects (séparabilité et stabilité).

La partie 3.1 de ce chapitre propose une discussion sur le choix d'une représentation d'objet. Cette problématique étant extrêmement vaste [GL11, DSTR11] et en perpétuel mouvement, elle propose des voies de réflexion et explique les choix de modèles que nous avons effectués. Les parties 3.2 et 3.3 présentent des travaux que nous avons conduits autour des descripteurs par distributions de primitives et des matrices de covariance. Enfin, la partie 3.4 s'intéresse à la coopération des descripteurs.

3.1 Le choix d'une représentation

Du fait du volume considérable de données multimédia désormais accessibles, et des performances de calcul accrues, les méthodes d'apprentissage profond, telles que les réseaux de neurones convolutionnels [LBBH98, BCV13] suscitent un engouement exceptionnel depuis le concours *ImageNet* en 2012 [KSH12]. L'apprentissage de dictionnaires est également un domaine de recherche très dynamique, avec les approches par sacs de mots visuels [SZ03, CDF⁺04, EPLE⁺17] (*Bag of Visual Words*), les représentations parcimonieuses (*Sparse Coding*) [BLM15, FB17], les noyaux de Fisher [PD07, Cin14] ou encore le descripteur VLAD [JDSP10].

Alors que ces approches d'apprentissage de représentations deviennent prédominantes dans le cas de la recherche de catégories d'objets et la fouille de données, les descripteurs image dits *classiques* se sont avérés pertinents pour les tâches de ré-identification et de suivi de cibles auxquelles nous nous sommes attelés, dans des contextes applicatifs où les contraintes temps-réel étaient extrêmement fortes. Déjà dans nos travaux de thèse (2002-2005) déroulée dans un contexte applicatif de traçabilité dans l'agro-alimentaire, il s'agissait de suivre un marquage à l'encre qui présentait

des spécularités importantes. Le projet LOVE¹ (2007-2009) nous a ensuite amenés à nous intéresser aux objets de type véhicules et piétons à partir de caméras embarquées dans le véhicule. Dans le projet SPY² (2010-2013), il s'agissait de définir des systèmes de surveillance mobiles pour la sécurité publique, pouvant être embarqués sur le véhicule de PPSL³ ou déployés rapidement sur un site à surveiller. Les caméras étaient fournies par la société Eolcam-Les Ulis, et disposaient de peu de capacité de calcul et de mémorisation. Ces travaux ont également été menés au cours de deux thèses, celles de Florence Laguzet (2009-2013) et Andres Romero Mier y Terán (2010-2013), co-encadrées avec Lionel Lacassagne, qui est spécialisé en Adéquation-Algorithmes-Architecture. Un travail important ayant été mené sur ces aspects, nous en montreront quelques résultats.

En dehors de ces contraintes applicatives, le choix d'un descripteur dépend d'une part de la nature de l'objet à caractériser, tant dans son apparence que dans sa dynamique, d'autre part de l'environnement et de ses possibles variations.

3.1.1 Critères liés à l'objet

Dans [YJS06], les auteurs suggèrent neuf possibilités de représentation géométrique de l'objet allant du simple centre d'inertie au squelette articulé. À ces représentations dans le domaine spatial se rajoutent des représentations dans les espaces des caractéristiques, par exemple dans le domaine spectral (couleur) ou fréquentiel (gradient, texture) dont la pertinence est liée à la nature de l'objet, selon qu'il s'agit d'un objet fortement structuré (un véhicule par exemple) ou non (les piétons).

Les **objets structurés** (comme ceux de la figure 3.1) sont généralement manufacturés et leur forme est pour la plupart géométrique et identifiable, en particulier lorsqu'ils évoluent dans un environnement naturel [Leo12]. Ils se distinguent par la présence de coins et de points d'intérêt [HS88], définis en couleur ou en binaire [LCS11], des attributs de forme [BMP01, LVAN12] peuvent être calculés à partir de leurs frontières [YPC⁺16] ou de segments de contours [TFD13, CLK16]. La plupart des objets physiques rigides possèdent six degrés de liberté (translation et rotation) et se déplacent dans la scène en conservant leur forme (comme les objets de la figure 3.1). Ce déplacement dans la scène produit dans l'image des mouvements de type translationnel, affine ou homographique, vis-à-vis desquels la plupart des points d'intérêt [HS88, LK⁺81, TK91, Low04, Gou07, GCFMT12] sont invariants.

Les **objets peu structurés** (comme ceux de la figure 3.2) se distinguent plutôt par leur couleur, leur texture ou la combinaison des deux [Pen15] ainsi que par un contour irrégulier. Les objets peu structurés auxquels nous nous sommes intéressés étaient des objets déformables. Tous les points 3D ne suivent pas la même trajectoire, les distorsions ne sont pas toutes modélisables aisément, donc rarement prévisibles dans le temps. Si l'apparence globale de l'objet est considérée (et non son comportement interne), une représentation frustrée de l'objet peut suffire : ils peuvent par exemple être assimilés à un simple plan texturé et coloré se déplaçant par similitude. Là encore, les histogrammes couleur, lorsqu'ils sont normalisés par rapport à la taille de l'objet, sont invariants vis-à-vis des changements d'échelle et des rotations puisqu'ils ne prennent pas en compte l'agencement spatial des pixels. Ces objets peuvent être vus comme un ensemble de points homogènes en terme de couleur et/ou de luminance représentés par des distributions classiques [CRM03, FZ04, KAR13] des mélanges de gaussiennes [LLP06], ou bien comme un ensemble de régions uniformément colorées comme des superpixels [LYYH16]. Les distributions classiques négligent toute information liée à l'agencement spatial des pixels de l'objet. Ce défaut peut toutefois être atténué par l'utilisation de structures de données mixtes intégrant à la fois des informations photométriques et spatiales, comme les spatiogrammes [BR05b] ou les histogrammes projetés

1. Projet ANR Predit, pôle de compétitivité Systematic, Logiciel d'Observation des Vulnérables
2. Projet européen ITEA2, Surveillance imProved sYstem
3. PPSL est le Pôle Pilote de Sécurité Locale, situé à Élancourt, et inauguré en 2010.



FIGURE 3.1 – Objets structurés au mouvement rigide



FIGURE 3.2 – Objets peu structurés au mouvement non-rigide

FIGURE 3.3 – Exemple d'objets dont le mouvement est non-rigide ou rigide (ces images viennent de différentes sources [GBS05, oW09, MFTM01]).

[GLL10b]. Ils peuvent être associés à un descripteur de forme, de contours [Zen86, GS99] possiblement déformables (contours actifs, *snakes*) [KWT88, TS92, CKS97, PD00, CV01] ou de silhouette [HHD00, YCHC10]. On peut envisager des approches fréquentielles et spatio-fréquentielles [FS89], comme les ondelettes de Haar ou leurs déclinaisons [VJ04], les motifs LBP [OPM02b, ZL11, NVC12, ZBC13, RGL13b, LP16, LLM16] et les représentations statistiques telles que les matrices de co-occurrence, les corrélogrammes [Zt09] ou les spatiogrammes [BR05b]. Enfin, les descripteurs de région par matrice de covariance [PTM06] [TPM06, PTM06, LW12] ont montré de très bonnes performances dans le cas de la détection, la ré-identification et le suivi de piétons [TPM08, YO11, TPM08, PTM06, BCBT11, RGL13a, BB14], qui correspondent à des objets par nature déformables et d'apparence variable. Nous aurons l'occasion d'y revenir dans la partie 3.3.

3.1.2 Critères externes à l'objet

La pertinence d'un descripteur dépend également des perturbations que l'objet est susceptible de rencontrer dans un type de scène donné comme des variations géométriques ou photométriques (voir Fig. 3.4), des occultations, et un arrière-plan variable. Lorsqu'une base d'images est disponible, les méthodes de représentation par apprentissage permettent d'exploiter des classes d'objets pour lesquelles ces variabilités intra-classe sont prises en compte. Lorsqu'il s'agit de détecter un objet singulier, dont une seule instance est disponible, il convient d'exploiter une représentation invariante.

Certains modèles d'objets sont définis pour être invariants vis-à-vis des variations photométriques comme les approches fondées sur le gradient de luminance telles que les descripteurs HOG [DT05] et SIFT [vdSGS10] ou encore les descripteurs impliquant des lignes de niveau [CCM99] dont la position reste stable si le seuil de détection est automatique et adaptatif. En revanche, les histogrammes s'avèrent extrêmement sensibles à ces variations à moins d'être normalisés ou égalisés [FHST05], ou d'être calculés à partir d'invariants couleur [GS99].

Par ailleurs, l'objet peut tout simplement être dissimulé de manière partielle ou totale par un autre élément de la scène ou par lui-même (voir Fig.3.5). Une réponse à l'occultation partielle peut consister à attribuer une confiance moindre aux points de l'objet situés sur sa périphérie, qui



FIGURE 3.4 – Variations photométriques



FIGURE 3.5 – Personne occultée, et avec une grande variété de poses.



FIGURE 3.6 – Piéton marchant devant un fond encombré et variable.

est la première zone impactée par l’occultation [CRM03]. L’objet peut également être vu comme un ensemble de fragments [EDB12] ou de points [VM11] pouvant être occultés individuellement sans dégrader la totalité de la représentation de l’objet. L’appariement se fait alors sur les données restantes.

D’autres variations sont directement imputables au capteur et aux paramètres d’acquisition et de transmission de la vidéo. Il peut s’agir d’apparition de flou due à un défaut de mise au point ou au mouvement. Les détecteurs de points d’intérêt comme SIFT et SURF [BETV08] sont robustes au flou, dans une certaine mesure, car leurs descripteurs incluent artificiellement des caractéristiques issues de la dégradation de leur zone d’intérêt pour différentes résolutions inférieures et pour différents niveaux de flou.

Enfin, l’environnement autour de l’objet dans l’image peut nuire à sa détection, comme le montrent les exemples de la figure 3.6. Des ambiguïtés d’appariement peuvent apparaître lorsque le fond possède des caractéristiques proches de celles de l’objet. Différentes stratégies permettent de répondre à ce problème de *distraction*, suivant les données disponibles (le nombre d’instances de l’objet) et l’objectif visé (suivi temporel ou identification). Il peut s’agir simplement de segmenter l’objet du fond [SMYS09] ou de raisonner à partir de primitives éparses [BPB07, SZT12, RM16]. Les ambiguïtés peuvent être levées en exploitant un ensemble de modèles de l’objet pris sous différents points de vue [Lei12] ou encore en apprenant en-ligne les informations d’apparence du fond au cours du suivi d’un objet [YO11, TSG⁺14, KLCR15].

3.1.3 Discussion

Dans le cadre du projet LOVE, les objets (véhicules et piétons) étaient vus comme des obstacles et détectés à partir d’histogrammes de disparité (la *UV-disparité* [LAT02, GPV10]). Dans le projet SPY, les objets étaient définis comme des objets en mouvement et détectés par soustraction de fond. Ces méthodes de détection ne fournissent ainsi aucune connaissance *a priori* (sur la structure, la texture, la couleur) pouvant être utile pour le choix d’une représentation pertinente pour l’objet détecté. Dans ce contexte, nous avons exploré trois stratégies différentes :

- le choix de descripteurs favorisant la rapidité au détriment à la fois de la séparabilité et de la stabilité vis-à-vis des perturbations. Nous avons ensuite proposé des améliorations de ces

descripteurs ;

- l'utilisation de descripteurs moins rapides mais plus discriminants, en mettant une énergie importante sur leur accélération et sur le choix des attributs exploités ;
- l'établissement d'une coopération entre deux types de descripteurs, un premier descripteur rapide pour un suivi standard, un second descripteur plus discriminant qui est exécuté uniquement lorsque des perturbations sont détectées. Cette stratégie permet de prendre en compte la variation du contexte d'acquisition.

Pour ce faire, nous nous sommes intéressés à trois types de représentation : les distributions couleur dans le cadre du suivi *Mean-Shift*, un ensemble de points par la méthode *Flock of Tracker* et enfin les descripteurs par matrice de covariance.

L'utilisation de distributions couleur pour représenter l'objet offre plusieurs avantages : 1) elles assurent une bonne robustesse vis-à-vis des variations géométriques de l'objet ; 2) l'utilisation de la couleur permet une bonne unicité de la représentation. 3) ces distributions existent pour tous les types d'objets, ce qui n'est pas le cas de bien d'autres primitives à base de texture ou de points d'intérêt ; 4) il est possible de les calculer à partir d'invariants couleur, ajoutant ainsi une certaine insensibilité aux variations photométriques. Cependant, la bonne résistance face aux variations de l'objet s'obtient souvent au prix d'une perte de séparabilité. Ce pouvoir discriminant peut être accru par l'utilisation de représentations spatio-couleur. Pour contrer les perturbations, nous avons proposé une méthode de sélection et de commutation des attributs couleur. Cela fera l'objet de la section 3.2.

Enfin, le descripteur de régions par covariance proposé par Tuzel *et al.* [TPM06] combine de façon très compacte les informations liées à des primitives visuelles potentiellement très variées, allant de la simple luminance au flot optique en exploitant leurs intercorrélations. Ce descripteur s'adresse à tous types d'objets mais s'avère plus gourmand en temps de calcul. Comme son pouvoir discriminant est important, il peut être exploité pour retrouver un objet perdu, voire s'élargir au problème de ré-identification d'objets. Nous aborderons ce descripteur dans la partie 3.3.

Comme nous l'avons vu dans la section précédente, un ensemble de points s'avère particulièrement approprié pour représenter des objets structurés au mouvement rigide, tels que des véhicules. Une partie de ces points peut être occultée sans nuire à l'appariement global. De plus, nos travaux sur l'invariance couleur nous ont permis de définir des méthodes de suivi de points robustes aux changements photométriques. Ce type de représentation est exploité dans la Section 3.4 dans le cadre d'une coopération avec le descripteur par covariance.

3.2 Descripteurs par distributions de primitives

Nous avons exploité des représentations par distribution dans le cadre du suivi d'objet dans la thèse de Florence Laguzet [LGLE11]. Les distributions peuvent être paramétrées sous la forme d'un mélange de gaussiennes, qui sont alors mises à jour de façon adaptative au cours du temps [MRG99, PD02] afin de rendre le modèle résistant face aux perturbations évoquées précédemment dans la section 3.1.2. Les distributions couleur peuvent également être utilisées sous leur forme non-paramétrée comme dans les méthodes basées noyau, telles que *Mean-Shift* [CRM03]. Le problème de faible séparabilité objet-fond a donné lieu à plusieurs contributions : la comparaison des différentes mesures de similarité [CRL05], la séparation objet/fond [SMYS09], l'amélioration de la représentation de l'objet [GLL10b, GLL10a], l'association avec des approches locales [BPB07] ou encore avec des *points-clés* [SZT12]. Dans [Lei12], plusieurs histogrammes de l'objet à différents points de vue sont utilisés pour fiabiliser la modélisation.

Deux approches sont évoquées ici. La première (section 3.2.1) utilise une représentation spatio-colorimétrique de l'objet afin d'améliorer le pouvoir discriminant de l'histogramme. La seconde (section 3.2.2) opère un changement de primitives couleur dès qu'une perte de séparabilité ob-

jet/fond est observée.

3.2.1 Degré de connexité couleur

Le degré de connexité proposé initialement pour la classification couleur [Fon01, Gou08], a été exploité pour le suivi d'objet par *Mean-Shift*. L'espace couleur est partitionné en cubes, dans lesquels le degré de connexité est défini comme la somme des probabilités de co-occurrence de couleurs normalisée par la somme des probabilités. Ce degré est élevé si l'intervalle correspond à des composantes connexes dans l'image c'est-à-dire à une classe significative en termes de connexité. Il est maximal lorsque toutes les couleurs appartiennent à une même composante et diminue avec le nombre de régions. Contrairement au corrélogramme ou à l'histogramme, une petite région homogène peut créer une forte valeur dans la distribution. Les figures Fig.3.7(a) à 3.7(c) montrent trois images synthétiques de taille 16×16 avec 4 couleurs équiprobables mais dont les degrés de connexité sont différents ((a) : 160, (b) : 80, (c) : 12).

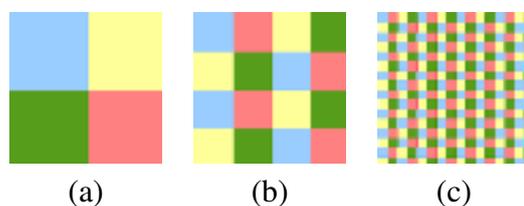


FIGURE 3.7 – Illustration du degré de connexité. Les images (a) à (c) montrent trois images couleur de taille 16×16 avec 4 couleurs équiprobables mais des valeurs de degré de connexité très différentes : (a) 160, (b) 80, (c) 12.

Afin d'améliorer le pouvoir discriminant de l'algorithme de suivi *Mean-Shift*, nous avons modifié l'algorithme en remplaçant les distributions classiques par les distributions de degrés de connexité [GLL10a]. L'objet est préalablement détecté dans la première image de la séquence et est représenté par sa boîte englobante sur laquelle est calculée la distribution 3D des degrés de connexité. Chaque probabilité est pondérée selon la position du pixel sur l'objet par un noyau d'Epanechnikov afin de réduire l'importance de la périphérie par rapport aux pixels centraux. Le suivi *Mean-Shift* consiste à déterminer de façon itérative, par descente de gradient, la position dans la nouvelle image qui maximise la similarité de Bhattacharyya ρ entre la distribution de la cible et celle du candidat (l'objet recherché). Pour ce calcul, la contribution de chaque pixel du candidat est pondérée selon sa similarité avec la distribution de la cible. La qualité de cette matrice de poids w s'avère déterminante pour la bonne convergence du suivi, ces poids devant être bien répartis sur le candidat tout en étant suffisamment élevés. Une fois la position identifiée, le processus est itéré dans le domaine des échelles.

La robustesse de l'approche a été comparée à celle issue d'une distribution couleur RGB sur différentes séquences de véhicules ou de piétons. Les résultats ont montré que les degrés de connexité fiabilisaient le suivi sur les séquences testées, les objets étant suivis plus longtemps dans la séquence, ou avec plus de précision. Le suivi à l'aide des degrés de connexité peut s'avérer plus gourmand en temps de calcul (20 % à 32% de temps de traitement en plus selon la cible suivie).

3.2.2 Sélection automatique d'attributs couleur

Le choix de l'espace couleur a un impact substantiel sur la qualité des résultats. Prenons l'exemple de la figure 3.8 [KMM10] qui présente les images à différents instants t d'un panda marchant dans son enclos filmé avec une vue légèrement plongeante. Cette vidéo présente plusieurs difficultés pour le suivi. Tout d'abord, la cible est déformable : vu de côté au début de la séquence ($t = 110$), le



FIGURE 3.8 – Résultats du suivi *Mean-Shift* sur la séquence panda [KMM10], pour différentes représentations couleur et en utilisant des histogrammes de taille 16^3 .

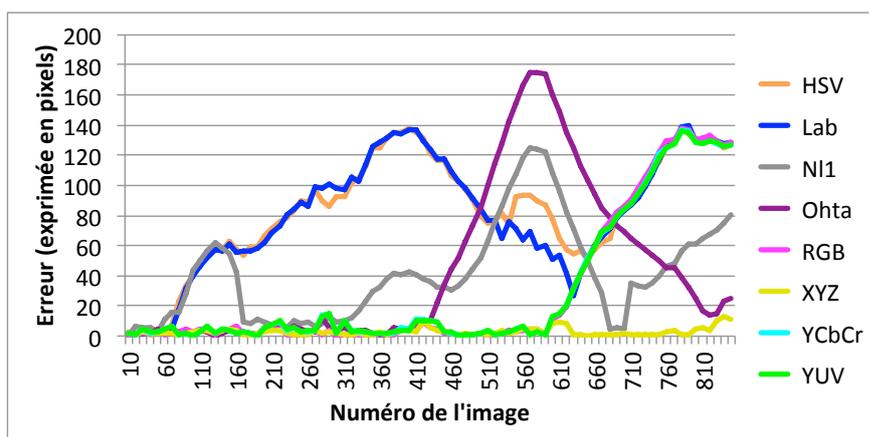


FIGURE 3.9 – Erreur de suivi (distance pixellique entre la détection et la vérité terrain) pour huit espaces couleurs.

panda se retourne ensuite pour s'éloigner vers le coin de l'enclos ($t = 710$). Durant la séquence, il passe derrière un arbre, ce qui le dissimule partiellement ($t = 610$). Il y a également un changement de zoom de la caméra. Les différentes boîtes englobantes de la figure 3.8 correspondent aux résultats du suivi *Mean-Shift* pour différents espaces couleur et en considérant des distributions 3D de taille 16^3 . La figure 3.9 affiche les erreurs de suivi obtenues par rapport à une vérité terrain. Nous avons pris en compte les espaces *RGB* et *XYZ*, les composantes normalisées *rgb*, puis un certain nombre d'espaces luminance-chrominance tels que *HSV*, *YCbCr*, *YUV*, *CIE L*a*b** ainsi que l'espace des composantes indépendantes de *Ohta*. Finalement, seule la représentation *XYZ* permet de suivre l'objet jusqu'à la fin de la séquence.

Dans les travaux de thèse de Florence Laguzet que nous avons co-encadrée avec Lionel Lacasagne [LGL11, Lag13] deux problèmes ont été abordés : 1) la sélection automatique de l'espace couleur permettant une meilleure séparabilité entre l'objet et le fond ; 2) le changement automatique de représentation couleur lorsque la séparabilité objet/fond diminue au cours du temps.



FIGURE 3.10 – Exemple de répartition des poids w (après seuillage) sur le panda.

Choix de l'espace couleur. Premièrement, pour exacerber la distinction objet/fond, les couleurs de la cible sont pondérées par le rapport log-vraisemblance objet/fond. L'espace couleur est sélectionné à l'initialisation du suivi, après avoir opéré une première itération du suivi entre les deux premières images de la séquence. L'espace couleur sélectionné est celui qui maximise la similarité entre le candidat et la cible tout en minimisant la similarité entre le modèle et l'arrière-plan. La normalisation par log-vraisemblance peut toutefois éliminer les poids w (se reporter à la section 3.2.1 qui résume l'algorithme *MS*), qui sont pourtant déterminants pour la qualité de la descente de gradient du *Mean-Shift*. Sur la figure 3.10 qui montre les poids obtenus sur le panda pour les différents espaces couleur, on remarque que les poids associés à *rgb* ne permettent pas de représenter la globalité de la cible. Un second critère vise donc à mesurer la qualité des pondérations w en termes de distribution spatiale et d'amplitude, ces poids devant être uniformément répartis sur les pixels associés à l'objet. Ainsi, dans la séquence du panda, l'espace couleur *rgb* est écarté d'emblée, et c'est l'espace $Y C_b C_r$ qui est finalement sélectionné. Le critère de décision regroupe ainsi les deux éléments clés de la réussite du suivi à savoir la séparabilité objet/fond et la qualité des poids w , en profitant des données directement fournies par la méthode.

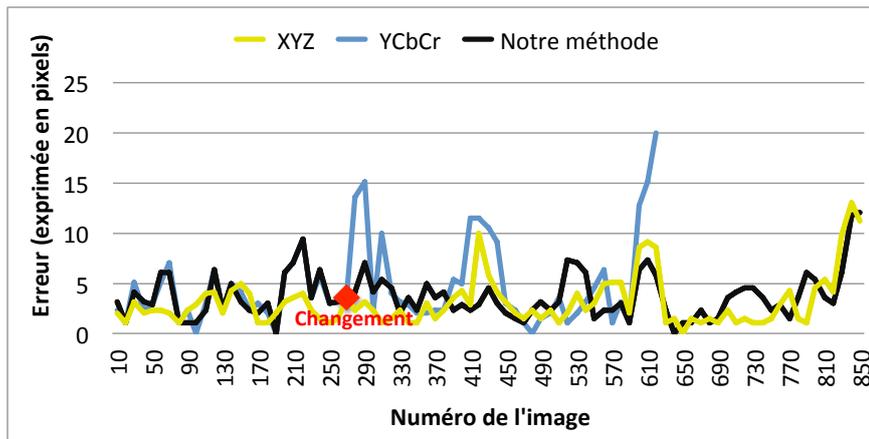


FIGURE 3.11 – Évolution de l'erreur de localisation de la cible par rapport à la réalité terrain pour des histogrammes de taille 16^3 .

Changement d'attributs. Au cours du suivi, la représentation choisie initialement peut s'avérer moins discriminante à cause des occultations partielles ou des variations de fond. Un changement de représentation est requis lorsque la similarité objet/fond décroît ou lorsque la qualité des poids w s'amointrit. Des critères assez proches de ceux utilisés pour la sélection sont ainsi mesurés à chaque instant. La figure 3.11 montre l'évolution de l'erreur de localisation obtenue pour les deux espaces XYZ et $Y C_b C_r$ et par la méthode de commutation (en noir) pour la séquence du panda. Dans les premières images de la séquence, l'espace couleur $Y C_b C_r$ est sélectionné puis un changement est préconisé à $t = 271$ car la représentation de la cible s'appauvrit. Finalement, le panda est correctement suivi jusqu'à la fin de la séquence à l'aide de l'espace XYZ .

Accélération du code. L'algorithme est intéressant du point de vue des recherches en Adéquation Algorithme Architecture car il se décompose en différentes fonctions (quantification, conver-

TABLE 3.1 – Temps d’exécution en ms/image obtenus pour le suivi de deux cibles, de tailles respectives 191×73 pixels (Vidéo 1) et 33×41 (Vidéo 2).

Transformations	Mono thread				SIMD + OpenMP			
	Scalaire		SIMD		2 threads		4 threads	
Taille d’histogramme	16	32	16	32	16	32	16	32
Vidéo 1	31,5	45,2	16,5	21,5	11,8	14,7	8,9	11,3
Vidéo 2	5,6	18,6	2,7	7,2	2,2	5,0	1,8	3,8

sion d’espaces couleur, calcul d’histogramme) de natures différentes, faisant appel à l’entrelacement des données, à des calculs flottants et entiers, à des algorithmes réguliers (accès de valeurs stockées de façon contiguë en mémoire) ou irréguliers. Le tableau 3.1 présente, en millisecondes, les temps d’exécution sur le quad-cœurs Conroe⁴, en scalaire et avec les différentes transformations effectuées. La parallélisation des données par l’utilisation d’instructions SIMD⁵ permet un gain $\times 2$. La parallélisation des tâches par l’utilisation de 4 threads offre un gain supplémentaire $\times 2$.

Cette première section a permis d’aborder les travaux réalisés autour d’une approche de suivi d’objets rapide exploitant une description fondée sur les distributions couleur ou spatio-couleur. Un mécanisme de sélection automatique des attributs couleur a été défini en partant des indicateurs directement fournis par l’approche. La section qui suit s’intéresse aux descripteurs par covariance de caractéristiques.

3.3 Descripteurs par covariance

Soit I une image en niveaux de gris ou vectorielle (couleur, multispectrale, infra-rouge, profondeur, etc.), de taille $W \times H$ et soit F l’image de taille $W \times H \times d$ rassemblant les d primitives visuelles calculées en chaque point de I :

$$F(x, y) = \phi(I, x, y) \quad (3.1)$$

Ici, ϕ représente le passage d’un pixel à son vecteur de dimension d comprenant les primitives visuelles telles que l’intensité, la texture, la couleur les gradients, les réponses de filtres. Soit $\{\mathbf{z}_k\}_{k=1 \dots n}$ le vecteur associé en chaque point dans la région $R \subset F$ de size n . Cette région est représentée par une matrice de covariance de taille $d \times d$:

$$\mathbf{C}_R = \frac{1}{n-1} \sum_{k=1}^n (\mathbf{z}_k - \boldsymbol{\mu})(\mathbf{z}_k - \boldsymbol{\mu})^T \quad (3.2)$$

où $\boldsymbol{\mu}$ est le vecteur descripteur moyen dans la région. Ces matrices sont compactes avec $(d^2 + d)/2$ éléments différents, et cette dimension est indépendante de la taille de l’objet.

Suivant les primitives prises en compte, ce descripteur peut fournir une bonne invariance par rapport aux changements d’illumination, d’échelle et de rotation. Son calcul peut être accéléré par l’utilisation d’images intégrales et d’instructions spécifiques [LRG⁺15], de sorte qu’il devienne lui aussi indépendant de la taille de la région à décrire. Forts de ces qualités, ces descripteurs ont prouvé leur succès dans de nombreuses tâches : la détection d’objets [TPM06] ou de personnes [TPM08, YO11], la classification de texture [TTL09, LW12, HSWL12, SBB15], le suivi [PTM06, RGL13a],

4. Processeur Q6600, Fréquence 2,4GHz, Cache L2 de taille 2×4 Mo, registres de type SSE de taille 128 bits, version SSE3.

5. Single Instruction Multiple Data

la ré-identification [BB14], la reconnaissance de visages [PYL08] ou d'actions [GIK10, HSWL12]. Dans ces travaux, plusieurs réflexions ont été abordées : le nombre et la nature des primitives qui viennent nourrir le descripteur ainsi que la manière dont ils sont introduits ; l'arrangement spatial des descripteurs sur l'objet à décrire et les poids alloués à chacun d'entre eux ; pour finir, l'étude sur les métriques dans la variété riemannienne à laquelle ces matrices appartiennent. Cet espace peut être vu intuitivement comme une surface continue résidant dans un espace de grande dimension. La difficulté est alors de définir des distances et des mesures de similarité dans ce type d'espaces. Plusieurs approches ont été proposées parmi lesquelles les métriques fondées sur l'algèbre de Lie [FM99, PTM06], la projection sur le plan tangent à la matrice identité [GIK10], les variétés de Grassmann et des *Reproducing Kernel Hilbert Spaces* (RKHS) [HSWL12] ou encore en l'utilisation d'une transformation Log-Euclidienne [LW12].

Nos travaux liés aux matrices de covariance se sont focalisés sur l'utilisation des primitives visuelles pour atteindre un bon compromis entre la compacité de la représentation et sa séparabilité. Nous nous intéressons ensuite à l'utilisation de ce descripteur dans le cadre du suivi d'objets multiples.

3.3.1 Covariance et texture : le descripteur ELBCM

Généralement le vecteur de primitives \mathbf{z}_k intègre des informations spatiales, couleur (ou intensité) et texture (ou gradient), et plus rarement des informations dynamiques [SSHL13]. Dans les travaux d'origine [TPM06], les primitives sont les coordonnées spatiales (x, y) , la luminance et ses dérivées premières et secondes, ce qui aboutit à un descripteur 7×7 . De nombreuses extensions ont été proposées pour améliorer le pouvoir discriminant de la représentation ou pour en réduire la taille.

En termes de texture, deux types de primitives sont particulièrement exploitées : les réponses de filtres de Gabor à l'instar du descripteur GRCM [PYL08], et les Local Binary Patterns (LBP) abordés précédemment dans la Section 2.3.2. Citons en particulier le *Local Binary Covariance Matrix* (LBCM) [GR11] ainsi que le travail [ZL11] qui utilise conjointement les filtres et les LBP dans le descripteur *Gabor-LBP based Region Covariance Descriptor* (GLRCD) pour la ré-identification. Que ce soit pour les dérivées ou la texture, ces caractéristiques sont généralement calculées à partir de la luminance à l'exception du descripteur MRCG (pour *Mean Riemannian Covariance Grid*) [BCBT12] qui concatène la norme et l'orientation de chaque canal RGB.

Si certaines applications comme la reconnaissance faciale [YO11] ou la détection de piétons, ne requièrent pas l'usage de la couleur, d'autres optent pour HSV [MSJ12], *CIE L*a*b** [HBRB11] ou encore des invariants couleur [ZBC13].

Puisque la diversité des combinaisons possibles rend difficile le choix de la meilleure configuration pour une application donnée, Bak et al. [BB14] ont proposé l'approche COSMATI (CORrelation-based Selection of covariance MATRices). Cette méthode permet d'identifier les 4 caractéristiques visuelles les plus discriminantes parmi 10, en exploitant des mesures de corrélation inter- et intra-classe. Pour la détection d'objets ou la ré-identification de personnes, la priorité est généralement mise sur le pouvoir discriminant du descripteur plutôt que sur sa compacité, au détriment du temps d'exécution. En guise d'exemple, le descripteur BIF (*Bio-Inspired Features* [MSJ12]) utilise 128 cartes de réponses de filtres de Gabor qui sont ensuite partitionnées en un ensemble de petites régions rectangulaires sur lesquelles un opérateur de *max pooling* est utilisé. La région est alors décrite par l'ensemble des matrices de covariance calculées sur ces régions.

Nous nous intéressons ici à l'intégration des primitives de texture et de couleur dans le descripteur par covariance dans un objectif de suivi et de ré-identification d'objets. La figure 3.12 permet d'illustrer le problème de ré-identification, où les objets sont détectés et suivis dans une première prise de vue à un instant donné. Leurs descripteurs sont stockés pour une future ré-identification à partir d'une deuxième prise de vue.

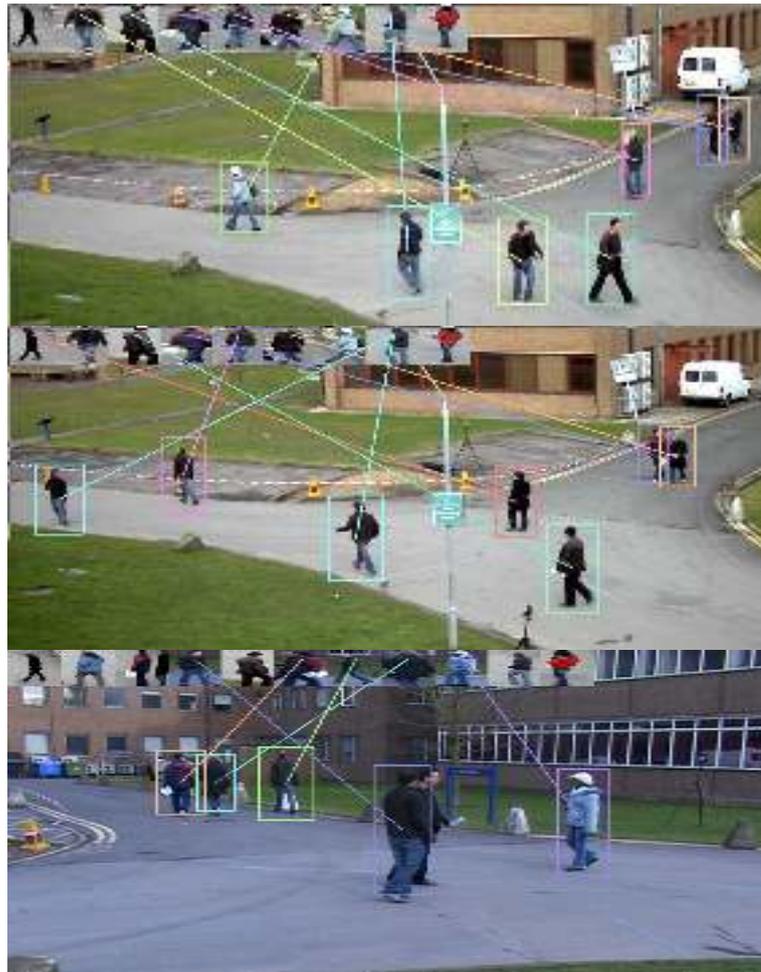


FIGURE 3.12 – Exemple de ré-identification d’objets dans la base d’images TUDCampus. Les objets sont suivis dans le champs d’une première caméra (les 2 premières images) et leur apparence est stockée en mémoire pour une prochaine ré-identification, soit dans le même champ de vision, soit à partir d’un autre point de vue (image du bas).

Les motifs LBP peuvent être intégrés de différentes façons. Les GLRCD [ZL11] exploitent directement les valeurs décimales des LBP de façon très compacte mais le descripteur s’avère instable pour les rotations. De plus, les opérations arithmétiques sur ces valeurs ont peu de sens en termes de texture. Les descripteurs LBCM [GR11] utilisent les chaînes binaires où chaque bit constitue l’une des composantes du vecteur ce qui aboutit à une représentation plus stable mais moins compacte. La variance LBP peut également être introduite [RGL13a], ce qui aboutit à un descripteur compact mais moins discriminant. Toutefois, associé à la couleur, ses résultats sont tout à fait acceptables pour le suivi d’objet, puisque la cohérence temporelle permet de réduire l’espace de recherche [RGL13a].

Les descripteurs que nous avons proposés dans la thèse d’Andres Romero [RGL13b] les ELBCM (pour *Enhanced LBP-based Covariance Matrix*), utilisent les angles formés par les ULBP uniformes mentionnés dans la section 2.3.2 et exploitent quatre mesures s trigonométriques, où θ_0 et θ_1 décrivent les angles de début et de fin des plages uniformes du LBP, comme l’illustre la figure 2.4. Ainsi, le vecteur de primitives utilisé pour la définition du descripteur ELBCM est décrit de la manière suivante :

$$\phi(I, x, y) = [\mathbf{S} \quad \mathbf{C} \quad \mathbf{T}] \quad (3.3)$$

où \mathbf{S} , \mathbf{C} et \mathbf{T} correspondent respectivement aux vecteurs de caractéristiques spatiales, couleur et texturales. Pour les ELBCM proposés, \mathbf{T} correspond aux ELBP de l’équation (2.11). Après la thèse

d’A. Romero, nous avons récemment poursuivi les développements du descripteur en l’étendant à la couleur [GyTL17]. Nous avons pu montrer de bonnes performances dans le cadre de la reconnaissance de textures, de visages et en ré-identification de personnes sur la base de données *CAVIAR4REID* [CCS⁺11]. En guise d’exemple, le tableau 3.2 montre une évaluation de plusieurs vecteurs descripteurs en reconnaissance de texture. L’information spatiale est fournie soit par les coordonnées soit par la distance au centre de la région à décrire, ce qui apporte à la fois une meilleure compacité et une plus grande invariance en rotation. Les espaces couleur suivants ont été testés : la luminance I , RGB , rgI (r et g étant les composantes normalisées et I la luminance), le modèle couleur gaussien, HSV , RGB avec égalisation d’histogramme (HE), l’espace de composantes opposées Opp et sa version normalisée $nOpp$. Pour finir, m est défini en chaque pixel par $\max(r, g, b)$ (voir section 2.5). Notons que m , rg , Opp , $nOpp$ sont mal définis pour des couleurs peu saturées. Pour ce qui concerne les gradients, plusieurs types de dérivées sont utilisées, définies en luminance ou en couleur. Enfin, les différentes variantes de descripteurs par covariance fondés sur les filtres de Gabor et les LBP sont testées, en particulier les variantes ELBCM présentées précédemment.

La base d’images *KTH_TIPS* [HCFE04], dont la figure 3.13 montre quelques exemples, est utilisée ici pour illustrer les résultats. Elle comprend 81 types de texture sous forme d’images de taille $\simeq 200 \times 200$ pixels. Pour chaque type de texture, des images sont prises à différentes échelles, sous différentes illuminations et poses. Pour chaque classe, nous avons sélectionné 40 images de façon aléatoire pour constituer la base d’apprentissage tandis que les 41 restantes sont gardées pour le test. Une classification KNN est utilisée avec $k = 5$. Chaque texture est décrite par une seule matrice de covariance et chaque combinaison du tableau 3.2 est testée, avec des caractéristiques de texture calculées sur la luminance. Les taux de reconnaissance affichés sur la figure 3.14 sont les résultats moyens obtenus pour chaque caractéristique indépendamment. Par exemple, le résultat noté xy correspond à la moyenne des résultats associés à tous les descripteurs qui mettent en jeu xy , quels que soient \mathbf{C} et \mathbf{T} . ELBP obtient le meilleur taux de reconnaissance. Concernant la couleur, les meilleurs résultats sont obtenus pour $\mathbf{C} = rgI$ et RGB avec égalisation d’histogramme (HE). Comme l’attestent les faibles taux obtenus par la luminance seule, la couleur a clairement un impact positif sur les performances, à l’exception des espaces HSV , $nOpp$ et m , trop sensibles aux faibles saturations.

Nous avons également réalisé une évaluation en calculant la texture à partir des composantes couleur. La Figure 3.14 (b) montre les résultats obtenus pour rgI . Les barres jaunes correspondent à des LBP existants, les barres orange aux variantes que nous avons proposées et qui ont été abordées dans la Section 2.3.2. La matrice de covariance ELBP a permis d’obtenir les meilleurs résultats, et cela est resté vrai pour la plupart des combinaisons qui ont été testées dans nos travaux. Les autres versions ELBP figurent dans le haut du classement, avec des taux de reconnaissance dépassant toujours 93%.

3.3.2 Suivi d’objets multiples

Le suivi multi-objet consiste ici à détecter plusieurs objets dans chaque image d’une vidéo et à leur associer une trajectoire distincte. À cause des occultations, des disparitions, de l’apparition de faux positifs, les trajectoires peuvent dériver et se fragmenter. Andriyenko et Schindler proposent, dans [AS11, ASR12], l’utilisation de fonctions d’énergie qui évaluent le coût associé à un ensemble de positions en considérant à la fois les propriétés dynamiques des cibles comme la vitesse linéaire et angulaire, la persistance de la trajectoire en prévenant et en pénalisant les déplacements inattendus, et enfin les détections et les disparitions de cible au-delà de la zone d’intérêt. La possibilité d’inclure des informations relatives à l’apparence dans la fonction d’énergie a été laissée ouverte dans leurs travaux [AS11, ASR12]. Dans le cadre de la thèse d’Andres Romero Mier y Téran, nous avons enrichi l’approche en améliorant sa robustesse grâce à la représentation des cibles par les matrices de covariance.

TABLE 3.2 – Caractéristiques utilisées dans le vecteur descripteur.

Space S	
xy	$[x, y]$
d	$\sqrt{(x - x_0)^2 + (y - y_0)^2}$
Color C	
<i>I</i>	Luminance <i>I</i>
<i>RGB</i>	$[R, G, B]$
<i>HSV</i>	$[H, S, V]$
<i>Gaussian</i>	$[E, E_\lambda, E_{\lambda\lambda}]$
<i>m</i>	$\max(r, g, b)$ [RGL13b]
<i>EH</i>	RGB avec égalisation d'histogramme
<i>rgI</i>	composantes RG normalisée + I
<i>Opp</i>	composantes couleur opposées
<i>nOpp</i>	idem normalisées
Texture T	
<i>Ix, Iy</i>	gradient en <i>x</i> et <i>y</i> $[I_x, I_y]$
<i>n, a</i>	norme et orientation du gradient $[\nabla^I = \sqrt{I_x^2 + I_y^2}, \phi^I = \arctan \frac{I_y}{I_x}]$
<i>Ix, Iy, Ixx, Iyy</i>	dérivées du 1er et 2nd ordre $[I_x, I_y, I_{xx}, I_{yy}]$
<i>Di Zenzo</i>	gradient vectoriel de [Zen86]
<i>Cx, Cy</i>	gradients couleur marginaux $[C1_x, C1_y, C2_x, C2_y, C3_x, C3_y]$
<i>Gab uv</i>	réponses du filtre de Gabor pour les premiers niveaux d'échelle <i>u</i> et d'orientation <i>v</i> [PYL08] $[g_{00}(x, y) \dots, g_{(u-1)(v-1)}]$
<i>LBP</i>	valeur décimale des LBP $\in [0, 255]$ pour les $LBP_{8,1}$
<i>Gab LBP</i>	14 réponses de filtres de Gabor + LBP [ZL11]
<i>uLBP</i>	LBP uniformes $\in [0, 9]$ pour $LBP_{8,1}$
<i>nriLBP</i>	LBP non-invariant en rotation $\in [0, 58]$ pour $LBP_{8,1}$
<i>ror LBP</i>	LBP invariant en rotation $\in [0, 255]$ pour $LBP_{8,1}$
<i>V</i>	variance LBP [RGL13a]
<i>pLBPb</i>	LBP couleur utilisant la couleur moyenne de la région comme vecteur de référence
<i>pLBPc</i>	LBP couleur utilisant un ordonnancement réduit utilisant l'origine de l'espace couleur comme point de référence
<i>pLBPd</i>	LBP couleur défini par ordonnancement après projection sur la couleur du pixel central
<i>ELBP_a</i>	ELBP marginal : $\alpha_i(nriLBP_i) \forall \text{ canal } i = 1..k$
<i>ELBP_b</i>	ELBP défini à partir de pLBP _b : $\alpha(pLBP_b)$
<i>ELBP_c</i>	ELBP défini à partir de pLBP _c : $\alpha(pLBP_c)$
<i>ELBP_d</i>	ELBP défini à partir de pLBP _d : $\alpha(pLBP_d)$



FIGURE 3.13 – Quelques exemples de la base KTH_TIPS_Color.

Image par image, l’algorithme évalue toutes les correspondances possibles entre les détections $\mathbf{d}_i \in D_t$ et les pistes $\mathbf{p}_j \in P_t$, qui sont créées ou détruites tout au long de la séquence. Finalement, l’algorithme génère une cartographie dynamique contenant l’ensemble des couples $M_T = \{(\mathbf{d}_i, \mathbf{p}_j)\}$ qui minimise une fonction d’énergie discrète :

$$E(\mathbf{d}_i, \mathbf{p}_j) = \underbrace{E_{det}}_{\text{confiance de détection}} + \underbrace{E_{dyn}}_{\text{modèle dynamique}} + \underbrace{E_{add}}_{\text{modèle de persistance}} + \underbrace{E_{app}}_{\text{similarité d'apparence}} . \quad (3.4)$$

Cette fonction d’énergie $E(\mathbf{d}_i, \mathbf{p}_j)$ privilégie les configurations plausibles et pénalise celles qui sont incohérentes. Notre fonctionnelle d’énergie comprend la plupart des termes utilisés dans [AS11]. Le terme de détection E_{det} est fondé sur la confiance de détection en partant des données image. Le terme E_{dyn} modélise la dynamique des objets (vitesses linéaire et angulaire). Le terme E_{add} pénalise les créations (ou disparitions) de nouvelles pistes et favorise les trajectoires persistantes et continues. Lorsque des correspondances sont trouvées entre les détections et les pistes, ce terme considère également l’hypothèse d’un nouvel objet entrant dans la scène. Cette hypothèse est pénalisée de façon proportionnelle à la distance euclidienne entre la détection et les frontières de la zone d’analyse. Finalement le terme d’apparence E_{app} correspond à notre contribution publiée dans [RGL13a]. Il mesure la similarité entre la cible à l’emplacement actuel et le modèle de covariance associé aux trajectoires \mathbf{p}_j . Chaque détection $\mathbf{d}_i \in D_t$ possède un descripteur par covariance $\hat{\mathbf{C}}_i$ qui est comparé à chaque $\mathbf{p}_j \in P_t$ en utilisant leur matrice « représentative » $\bar{\mathbf{C}}_j$. Dans le cas d’une forte similarité, la fonction d’énergie obtient une valeur négative sinon la fonction d’énergie augmente proportionnellement avec la divergence entre les descripteurs. La divergence choisie est TBD (pour *Total Bregman Divergence*) proposée par [LVAN12, CSBP11, CSBP13] et la matrice représentative $\bar{\mathbf{C}}_j$ est la norme l_1 TBD (t-center). Comme évoqué précédemment, les matrices de covariance appartiennent à une variété riemannienne formée par l’ensemble des matrices définies positives, pour laquelle de nombreuses distances et divergences ont été proposées, telles que *MRC* (pour *Mean Riemannian Covariance*). La divergence TBD a l’avantage d’être peu coûteuse en temps de calcul par rapport aux distances riemanniennes dans le sens où elle ne requiert que des multiplications de matrices et le calcul de déterminants, contrairement aux métriques nécessitant le calcul des valeurs propres. Concernant la minimisation de l’équation (3.4), nous utilisons l’algorithme de Munkres [Mun57] (la méthode Hongroise) utilisé par ailleurs dans [JA09].

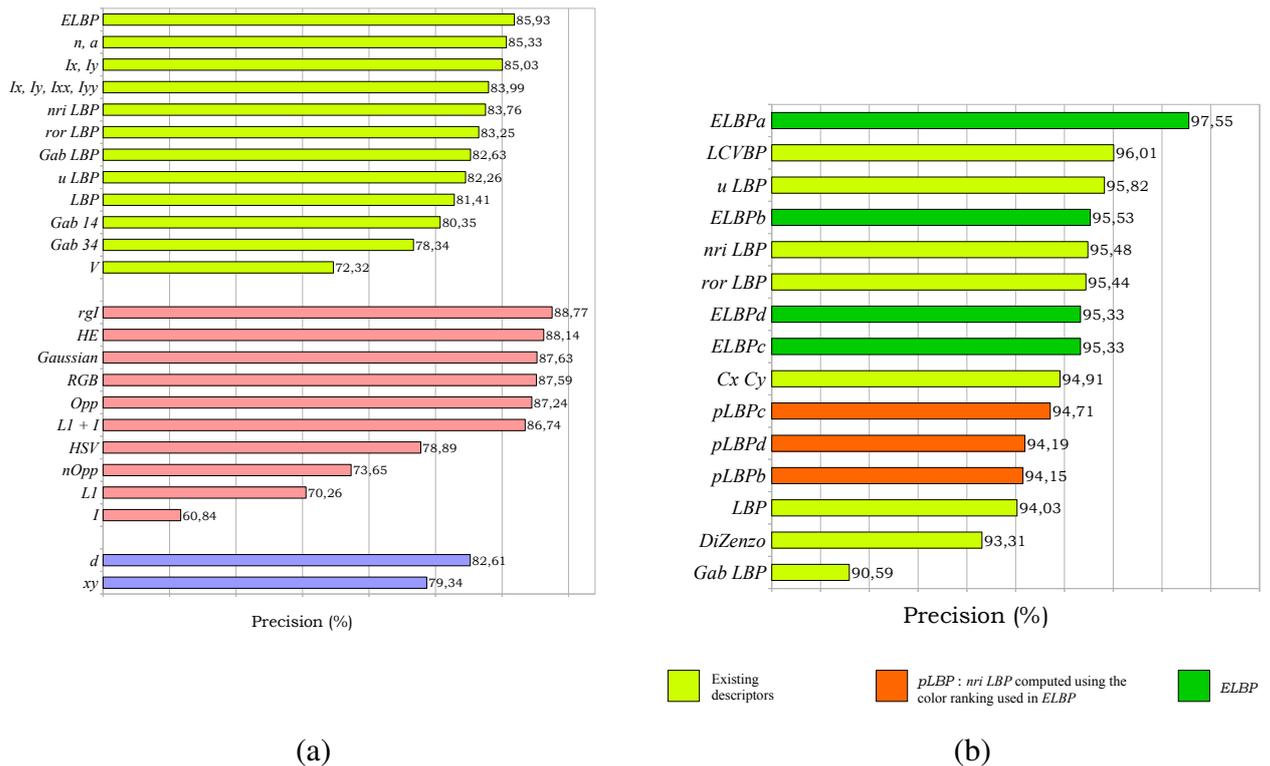


FIGURE 3.14 – (a) Taux de reconnaissance moyen (%) pour chaque caractéristique sur la base d’images *KTH_TIPS* (texture **T** en jaune, couleur **C** en orange et localisation spatiale **S** en bleu). Les composantes de texture sont calculées à partir de la luminance. (b) Taux de classification en utilisant la couleur (RGB ou *rgI*) pour le calcul des composantes de texture.

Nous avons évalué l’algorithme en utilisant les métriques proposées dans [SBB⁺07] et [LHN09] et en partant de bases de données très largement utilisées dans la communauté (PETS 2009-S2L1-V1⁶, la base TUD⁷ avec les séquences TUD-Campus, TUD-Crossings et TUD-Stadtmitte), ainsi que des séquences enregistrées au laboratoire IEF. Certaines séquences sont visibles en-ligne⁸. Les figures 3.15 et 3.16 sont illustratives des séquences utilisées (base PETS2009 et TUD respectivement). Sur la base de ces expériences, les résultats ont montré que notre approche de suivi n’est pas aussi précise que celles proposées par [MRS13] ou [BFTF11], mais elle a l’avantage de travailler à la volée, en estimant les trajectoires image par image. *A contrario*, dans les méthodes proposées dans [BFTF11, MRS13], le suivi nécessite l’ensemble des images de la vidéo pour estimer le nombre de trajectoires optimal et leur tracé le plus probable.

3.4 Coopération de descripteurs

Aucune des méthodes de suivi n’est applicable de manière universelle, chaque algorithme ayant ses avantages et ses limites. Les performances sont bien souvent dépendantes de l’application et il n’existe pas d’algorithme qui puisse être considéré comme *supérieur* aux autres. Quelques-uns des critères à prendre en compte dans la définition du descripteur et de son appariement sont ceux que nous avons évoqués dans la partie 3.1.1. Or, l’objet n’est pas forcément connu *a priori* lorsque les objets sont détectés en utilisant des algorithmes de soustraction de fond et non une reconnaissance.

6. <http://www.cvg.rdg.ac.uk/PETS2009>

7. <http://www.mis.tu-darmstadt.de/node/428>

8. <http://andresromeromier.wikispaces.com/>



FIGURE 3.15 – Images issues de la séquence *PETS2009-S2-L1-Time_12-34-View_001* qui montre des personnes qui se croisent et s’occulent. La minimisation d’énergie utilisant l’apparence permet de suivre les trajectoires et maintenir l’identité des objets.

En combinant plusieurs descripteurs, en les faisant coopérer, l’on peut imaginer pouvoir répondre à un plus grand nombre de difficultés. L’objectif est d’obtenir un bon équilibre entre la robustesse et la précision tout en conservant des besoins en puissance de calcul aussi faibles que possible afin de maintenir l’exécution en temps réel. Nous avons proposé deux types de coopération : points-covariance (section 3.4.1) et distribution-covariance (section 3.4.2).

3.4.1 Coopération Points-Covariance

Afin de répondre au problème de suivi d’objet sur de longues durées, nous avons proposé une approche fondée sur le suivi *FoT* [VM11] qui analyse de manière robuste le flot optique associé aux points de l’objet. D’autre part, nous utilisons le descripteur par covariance pour modéliser l’apparence globale (couleur et texture).

L’approche *FoT* estime de manière robuste le mouvement apparent de l’objet et permet théoriquement de résister aux occultations partielles allant jusqu’à 50% de masquage [VM11]. La restriction principale de cette méthode est que l’objet doit montrer un nombre suffisant de points pour que le suivi par *KLT* soit fiable. Elle s’avère pertinente pour suivre des objets dont le mouvement est spatialement uniforme, comme des objets rigides.

Le suivi par covariance *CT* apparaît comme une alternative adéquate. Contrairement à *FoT*, de

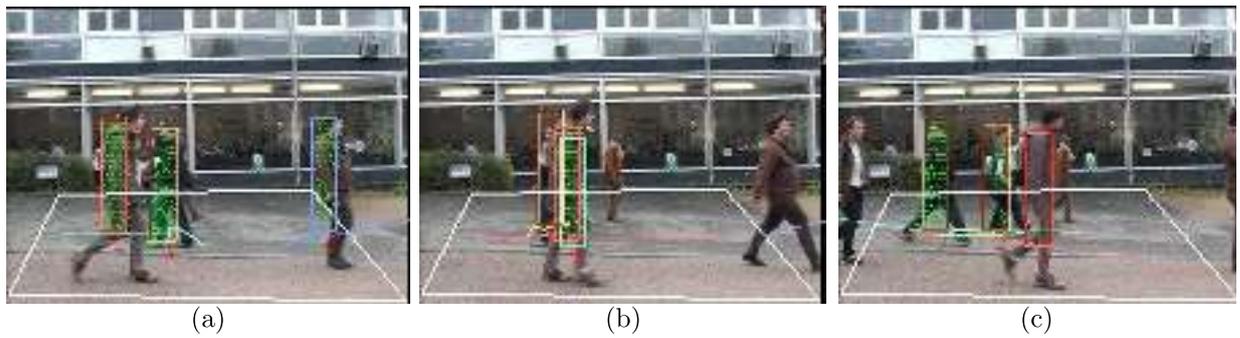
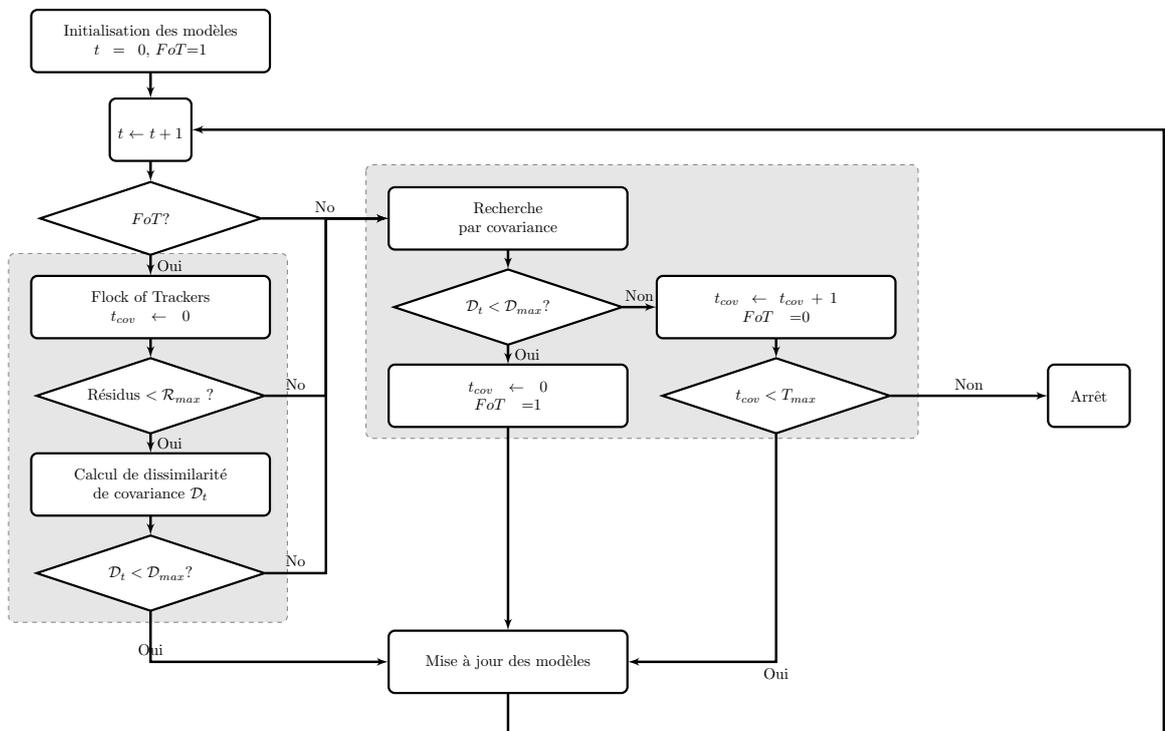


FIGURE 3.16 – Exemple de gestion des occultations sur la séquence TUD-campus sequence.

bonnes performances sont atteintes même pour des objets faiblement texturés ou pour des objets au mouvement non-rigide. Sa principale faiblesse réside dans le temps d'exécution élevé et ceci devient plus important lorsque la recherche de la nouvelle position de l'objet se fait de manière exhaustive dans l'image. Notons toutefois que CT peut être accélérée soit par transformation du code, par l'utilisation d'images intégrales, d'instructions dédiées (SIMD) et de parallélisation, comme nous l'avons effectué dans [RLHZG13, RLGHZ14] ou par un couplage avec d'autres méthodes de suivi telles que le filtrage particulaire [LLS10] pour limiter l'espace de recherche.

L'algorithme proposé est décrit graphiquement par la figure 3.17. FoT est exécuté en premier lieu ($FoT = 1$). L'analyse de la variance des vecteurs de mouvement permet d'évaluer la qualité de l'appariement. FoT est exécuté tant que les résidus du suivi KLT restent inférieurs à une valeur seuil \mathcal{R}_{max} ou que la dissimilarité \mathcal{D}_t entre deux matrices de covariance successives est inférieure à une valeur \mathcal{D}_{max} . Dans le cas où FoT a échoué (à cause d'une occultation par exemple), CT est exécuté pour retrouver l'objet pendant une période de temps T_{max} . Si l'objet n'est pas retrouvé après cette période, il est considéré comme définitivement perdu.

FIGURE 3.17 – Mécanisme de coopération points-covariance ou $FoT+CT$.

3.4.2 Coopération Distribution - Covariance

Les approches *Mean-Shift* (*MS*) et *CT* vues précédemment, possèdent également des propriétés complémentaires. *MS* est très rapide mais son pouvoir de séparation est relativement faible, ne permettant pas de faire face aux occultations et aux déplacements importants de l'objet dans l'image. Une association avec *CT* doit permettre de pallier cette faiblesse. Ici aussi, la dissimilarité entre matrices de covariance est exploitée comme indicateur de précision de *MS*, en jugeant *a posteriori* de la réussite de *MS*.

Le tableau 3.3 montre les résultats obtenus pour quatre séquences. *Pedxing* et *Panda* sont des séquences où la caméra est stable tandis que les séquences *Motocross* et *Carchase* correspondent à des vidéos prises d'une caméra mobile (soit fixée sur le cadre d'une moto, soit sur un hélicoptère). Le tableau présente pour chacune d'elles le nombre de cycles d'horloge par point (*cpp*) utilisé durant toute la durée de la séquence par chacune des approches *MS* et *CT* ainsi que le pourcentage d'images (*%im*) de la vidéo pour lequel l'un ou l'autre des descripteurs est utilisé. Dans les séquences où la caméra est stable (*Pedxing* et *Panda*), *MS* est exécuté la plupart du temps étant donné que le mouvement est régulier. Lorsque la caméra est placée sur un porteur mobile (*Motocross* et *Carchase*), *CT* est exécuté plus fréquemment, du fait du mouvement erratique des cibles dans l'image. Ainsi, la coopération permet d'une part de réduire le temps d'exécution lorsque le mouvement est régulier (en utilisant un descripteur rapide comme *MS*), d'autre part de fiabiliser le suivi en faisant appel à un descripteur plus coûteux mais plus discriminant.

TABLE 3.3 – Analyse de la procédure de coopération *MS+CT*. Ces résultats ont été obtenus sur un processeur *Nehalem* (X5550, 2×4 cœurs, 2.66 GHz, instructions SIMD de type SSE 4.2, RAM de type DDR3 PC3-8500 (1,066 MHz)). Le tableau montre pour chaque séquence : la taille de l'objet suivi à l'initialisation ; le pourcentage d'images (*%im*) pour lequel les algorithmes *MS* et *CT* sont exécutés ainsi que le nombre total de cycles d'horloge par point (*cpp*) pendant toute la durée de la séquence.

Séquence	Taille objet	# im	MS		CT		Temps [ms]/im
			<i>cpp</i> ($\times 10^6$)	<i>%im</i>	<i>cpp</i> ($\times 10^6$)	<i>%im</i>	
<i>Pedxing</i>	93 × 35	189	7.20	70	60.59	30	10.1
<i>Panda</i>	23 × 28	940	2.48	88	8.74	12	1.6
<i>Motocross</i>	64 × 47	2665	5.07	12	30.04	88	10.6
<i>Carchase</i>	45 × 97	2999	0.99	13	19.19	87	7.3

3.5 Conclusion

L'apparence d'un objet peut être décrite de différentes façons, allant de la distribution couleur aux représentations parcimonieuses, en passant par les contours déformables. Nous avons abordé ici deux types de contributions sur la représentation de l'apparence, qui se sont inspirées des approches de suivi *Mean-Shift* ou des descripteurs de régions par covariance. En profitant des données fournies par les méthodes d'appariement, il est possible d'évaluer la qualité des résultats et d'en déduire des indicateurs qui peuvent être scrutés pour prendre une décision telle que la commutation vers un autre modèle de représentation. Cela mène à des méthodes plus robustes vis-à-vis des perturbations susceptibles de modifier l'apparence d'un objet, tout en montrant des temps de calcul compétitifs.

Alors que ce chapitre considérait le mouvement comme une information fournie par un appariement temporel, le chapitre suivant aborde le mouvement comme un attribut d'entrée qu'il s'agit d'analyser et de caractériser.

Chapitre 4

Mouvement

MES premiers travaux de recherche se sont principalement focalisés sur l'analyse d'images statiques et la couleur en considérant le mouvement soit comme un obstacle, soit comme un résultat : soit un obstacle que j'ai cherché à contourner en m'intéressant aux invariants géométriques et couleur ; soit un résultat produit par un algorithme d'appariement de descripteurs d'apparence. Alors que la notion de mouvement a longtemps sous-tendu mes réflexions, elle est plus récemment devenue pour moi une primitive d'entrée à décrire et à analyser. Ces travaux se sont développés dans le cadre de projets collaboratifs tels que SPY¹ (2010-2013), pour la conception d'un système de surveillance, puis MAPOCA² (2014-2016) sur la reconnaissance d'activités humaines, et enfin par le co-encadrement des travaux de thèse de Tan Khoa Mai³ autour de l'analyse de scènes en mouvement. Nous ne ferons qu'effleurer ici les travaux encore récents autour de l'analyse de vidéos de langue des signes⁴, pourtant pleinement inscrits dans cette thématique du mouvement.

La plupart des tâches d'analyse de mouvement partent d'un calcul de flot optique (mouvement apparent des pixels) ou de trajectoires obtenues après une poursuite d'objets ou de points. Nous abordons ici les travaux réalisés sur l'estimation d'une carte de mouvement (section 4.1), son utilisation pour l'analyse de scènes (section 4.2), puis les structures de représentation spatio-temporelles qui ouvrent la voie à la reconnaissance d'activités (section 4.3).

4.1 Estimer le mouvement

Lorsque l'on évoque le mouvement, il s'agit bien sûr d'un mouvement apparent dans l'image, qui provient soit du déplacement 3D d'objets dans la scène, soit du mouvement de la caméra (égo-mouvement). Il peut être vu soit comme un champ de vecteurs de vitesse, le flot optique, soit comme un ensemble de trajectoires, qui peuvent être denses ou semi-denses.

4.1.1 Le flot optique

À partir d'un modèle de variation de la luminance des points image au cours du temps, estimer le flot optique consiste à déterminer le vecteur de déplacement $\mathbf{u} = (u, v)$ de ces points entre deux images temporellement proches. Partant du modèle simplifié qu'est celui de la conservation de l'illumination $I_t(\mathbf{x}) = I_{t+1}(\mathbf{x} + \mathbf{u})$, le problème revient à résoudre une équation à deux inconnues :

-
1. Projet Européen ITEA2 SPY : Surveillance imProved sYstem
 2. Projet de bourse post-doctorale de 18 mois MAPOCA pour Modélisation des Activités par Processus ponctuels et Covariance de cAractéristiques. Dans ce projet il s'agissait de comparer les processus ponctuels à l'ENSTA ParisTech avec Antoine Manzanera et les matrices de covariance pour l'analyse d'activités humaines.
 3. La thèse de Tan Khoa Mai *Vers un système de vision opportuniste pour l'analyse de scènes complexes* est co-encadrée avec Samia Bouchafa (IBISC) depuis octobre 2015
 4. Il s'agit d'une collaboration interne avec Annelies Braffort du groupe ILES au LIMSI.

$u\nabla I + I_t = 0$ où ∇I et I_t correspondent respectivement aux dérivées spatiales et temporelles de l'image.

L'estimation du flot optique est donc un problème mal posé (*aperture problem*) dont la résolution passe par l'adoption d'hypothèses telles que l'uniformité locale ou la régularité du mouvement. Les approches globales, comme [HS81], estiment le flot de façon dense dans l'image par optimisation d'une fonction d'énergie intégrant une contrainte de lissage de la carte de flot. Les variantes proposées dans la littérature diffèrent par la forme de la fonction de coût, par la méthode d'optimisation [BA96], ou encore par les primitives visuelles : couleur, gradient, texture [XJM10, BM11, WRHS13, KLL13]. Certains travaux essaient simultanément de segmenter les objets et d'estimer le flot optique [SRB10, YL15] au prix d'une plus grande complexité de calcul.

Les approches locales dont KLT [LK⁺81] est fondatrice, estiment le flot à partir d'un bloc de pixels, généralement de façon éparse. Elles font l'hypothèse que le bloc de pixels porte un seul et même mouvement. La méthode KLT nécessite l'inversion de la matrice de tenseur de structure. Puisque cette matrice peut s'avérer mal conditionnée sur les zones de couleur uniforme, un critère de force des coins (*cornerness*) permet d'éviter le calcul en ces points. De nombreux auteurs ont proposé des variantes de cette approche, comme [BA96, BHS98] où la fonction de coût quadratique est remplacée par une fonction robuste, ou encore [Far00, BA96] qui intègrent la couleur. D'autres travaux se concentrent sur l'analyse du tenseur de structure [NG98, BW02, MN01, LCR03].

Auto-évaluation de la qualité du flot. Partant de l'approche KLT, nous avons proposé une méthode d'estimation du flot optique capable de s'auto-évaluer puis de se corriger [MGB17b]. Dans un premier temps, une analyse simple de l'image fournit la variance de couleur ainsi que le fameux critère *cornerness*. Il s'agit d'un premier indicateur de fiabilité qui permet d'atténuer l'importance des zones uniformes pour lesquelles l'estimation de flot optique s'avère peu fiable. Ensuite, une première itération d'estimation du flot fournit deux indicateurs supplémentaires : l'écart du résultat par rapport à la droite de contrainte du flot optique et, d'autre part la variance locale des valeurs de flot dans un voisinage 3×3 . La figure 4.1 permet d'illustrer ces critères à partir de l'image Venus de la base de données Middlebury⁵. Ces critères d'évaluation nous permettent de déterminer en chaque pixel un indice de qualité du résultat de flot optique.

Correction itérative. Cet indice est ensuite exploité dans un processus itératif de correction du flot optique. Cette correction part de l'hypothèse réaliste selon laquelle le flot optique est continu entre deux pixels spatialement proches et de caractéristiques (couleur, texture) similaires. Les résultats fiables sont propagés vers les valeurs moins fiables en tirant profit des informations locales telles que les similarités couleur. Le flot est corrigé en chaque point en fonction des caractéristiques de son voisinage, l'influence d'un voisin étant d'autant plus importante que celui-ci : 1) est proche en termes de couleur, 2) est proche spatialement et 3) possède un indice de qualité du flot élevé. À chaque itération, les valeurs de flot ainsi que les indices de qualité sont mis à jour. Le processus s'arrête lorsque la qualité cesse d'évoluer. La figure 4.2 fournit deux exemples de résultats obtenus en les comparant à des méthodes existantes.

5. La base de données Middlebury fournit quelques images et les résultats de flot optique : <http://vision.middlebury.edu/flow/data/>

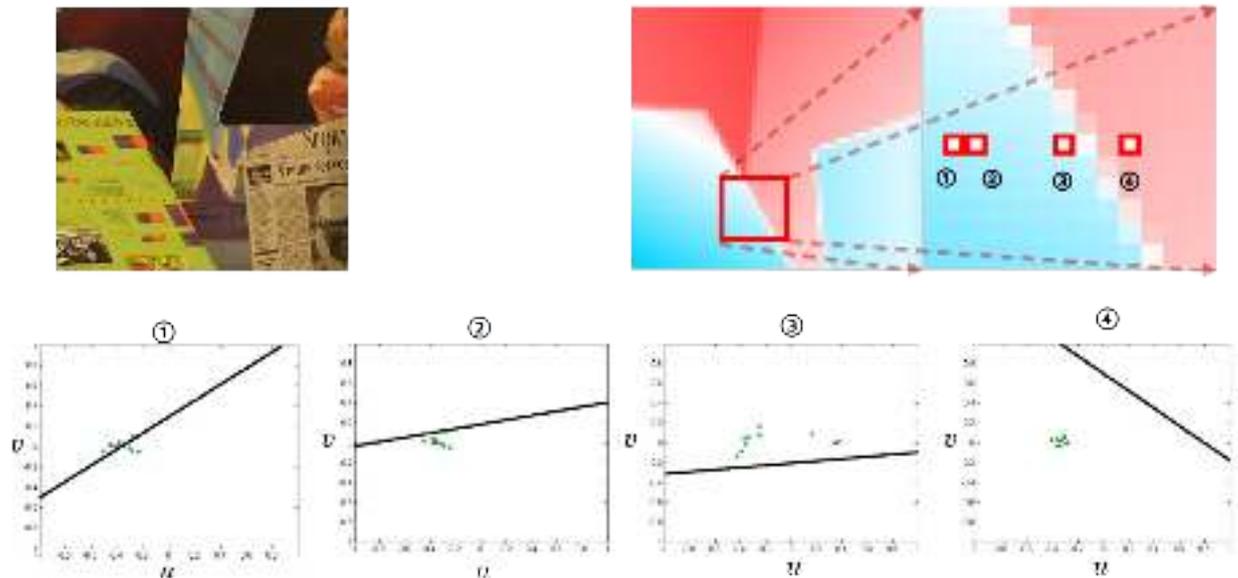


FIGURE 4.1 – Évaluation de la qualité du flot optique. Image Venus de la base Middlebury (en haut à gauche) et carte de flot optique issue de la vérité terrain (en haut au centre). En quatre zones différentes (indiquées par un chiffre de 1 à 4), les valeurs (u, v) de flot optique KLT sont collectées dans un voisinage de 3×3 pixels. Pour chacune de ces zones, les figures du bas représentent les valeurs de flot estimées en vert, la vérité terrain en rouge et la droite de contrainte de luminosité. Dans les zones texturées et localisées sur une surface plane (1 et 2), la variance du flot est faible et les résultats sont proches de la ligne de contrainte. Dans les zones situées à proximité d’une frontière entre deux objets, soit la variance est élevée (point 3), soit les points sont éloignés de la ligne de contrainte (point 4). Ces deux critères sont utilisés conjointement dans notre approche pour évaluer la qualité du flot.

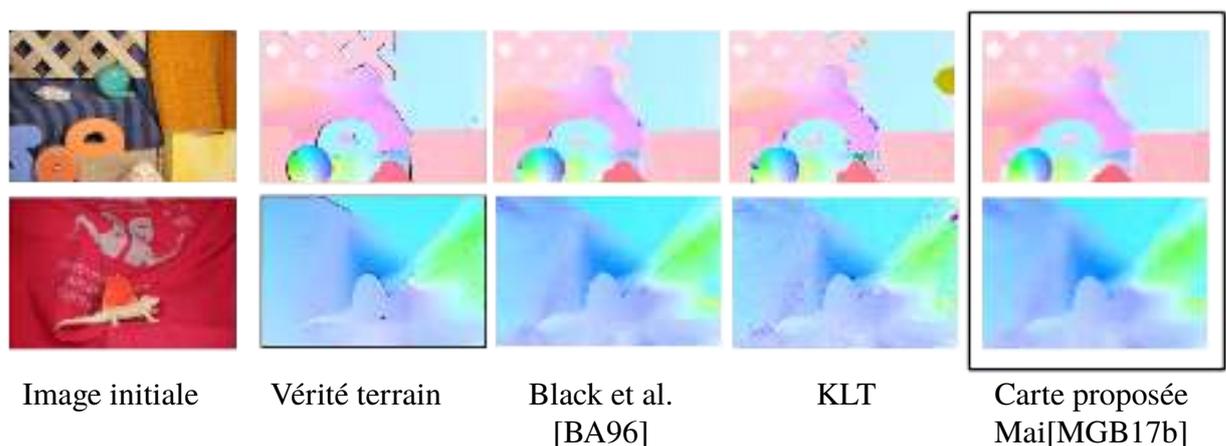


FIGURE 4.2 – Exemples de résultats de cartes de flot optique.

Évaluation. Les résultats de flot optique ont été évalués de façon assez classique par le biais de deux mesures calculées à partir des données de vérité terrain : l’erreur angulaire AAE (pour *Average Angular Error* en anglais) et l’erreur de norme EPE (pour *End Point Error*). Deux conclusions se dégagent. D’une part, les mesures de qualité proposées sont cohérentes avec la mesure d’erreur puisque la mesure de qualité est effectivement élevée aux points pour lesquels l’erreur de flot est faible. D’autre part, la phase de correction des flots permet de réduire significativement l’amplitude d’erreurs. Enfin, les résultats obtenus sont meilleurs que les méthodes classiques [HS81, LK⁺81] et

proches de ceux des approches récentes de l'état de l'art. Étant donné la multiplicité des méthodes de calcul de flot optique, un site d'évaluation a été mis en place à Middlebury⁶. Nous avons soumis nos résultats et avons obtenu le rang de 85 début 2017. Nous avons choisi d'attendre les prochaines améliorations de l'algorithme pour publier ce résultat sur le site.

4.1.2 Les trajectoires

Le mouvement dans l'image peut également être vu comme un ensemble de trajectoires définies à partir d'une méthode de suivi de points ou de flot optique. Nous nous sommes intéressés aux trajectoires denses et semi-denses et les avons exploitées pour la reconnaissance d'actions. Les trajectoires denses proposées dans [WKSL11] sont extraites à partir d'un flot optique dense filtré temporellement par un filtre médian à plusieurs niveaux d'échelles spatiales. Un critère local permet d'écarter à la fois les trajectoires peu porteuses d'information et celles dont le mouvement est jugé trop brutal pour être fiable, à savoir celles dont l'amplitude dépasse 70% de l'amplitude moyenne de la trajectoire. Cette approche a été améliorée par les mêmes auteurs [WS13] en compensant, le cas échéant, l'égomouvement estimé par homographie. L'appariement temporel est réalisé par estimation du flot optique et appariement de descripteurs SURF. Enfin, les trajectoires semi-denses [GM12] correspondent aux trajectoires de points-clés extraits de l'image. Un descripteur multi-échelle est associé à chaque point et est mis en correspondance entre les images successives de la séquence. Chaque point est également associé à un vecteur d'état relatif à son apparence et à sa cinématique, qui est mis à jour et filtré à chaque instant.

4.2 Analyser les scènes en mouvement

Les cartes de mouvement, qu'elles soient binaires ou vectorielles, servent à fournir une première interprétation sommaire de la scène considérée, par exemple en détectant les plans dominants d'une scène structurée (section 4.2.1) ou encore en construisant des *cartes de normalité* du mouvement desquelles se déduisent les anomalies (section 4.2.2).

4.2.1 Extraire la structure de la scène

Nous nous plaçons ici dans le contexte d'une caméra se déplaçant uniquement en translation suivant l'axe optique (Z). La méthode appelée *c-vélocité* consiste à exploiter la relation linéaire existant alors entre la norme w des vecteurs du flot optique et la valeur $c = \sqrt{x^2 + y^2}$ qui dépend seulement de la position (x, y) du point dans l'image. Cette méthode permet de simplifier l'extraction de plans 3D en exploitant une technique cumulative. Proposée par Samia Bouchafa [BZ12] au cours de sa participation au projet LOVE, cette méthode est reprise actuellement dans les travaux de thèse de Tan Khoa Mai que nous co-encadrons ensemble depuis octobre 2015.

Alors que l'espace de vote *c-vélocité* exploite exclusivement la norme des vecteurs de flot sans tirer profit de leurs orientations, la nouvelle approche appelée *uv-vélocité* [MGB17a] permet d'intégrer pleinement ces deux informations. Cette méthode ne nécessite plus le calcul et la discrétisation de l'axe c . Même si deux espaces de votes sont désormais nécessaires au lieu d'un seul, les traitements sont fortement allégés. Les composantes de mouvement (u, v) d'un plan peuvent s'exprimer en fonction de la norme du plan au point considéré et les composantes de translation et de rotation de la caméra. Les hypothèses posées initialement par la *c-vélocité* ont été maintenues pour simplifier les équations de mouvement. Tout d'abord, la caméra se déplace en ligne droite suivant l'axe optique. Cette hypothèse, qui nous satisfait pour les premières étapes de notre réflexion, permet de négliger deux composantes de translation ainsi que les trois composantes de rotation. Ensuite,

6. <http://vision.middlebury.edu/flow/eval/results/results-e1.php>

en nous plaçant dans un monde de Manhattan, par exemple une scène urbaine ou l'intérieur d'un bâtiment, nous supposons la présence d'un sol (une route par exemple), de plans latéraux (les murs ou les bâtiments), et éventuellement de plans frontaux (obstacles, bâtiments, murs). Cela permet de définir trois types de vecteurs de normale, à savoir $[0,1,0]^T$ (pour les plans horizontaux), $[0,0,1]^T$ (plans frontaux) et $[1,0,0]^T$ (plans latéraux) dans le repère associé à la caméra. Tous calculs faits, cela aboutit à trois ensembles de relations simples rassemblées dans le tableau 4.1 qui expriment le mouvement (u, v) en chaque point en fonction des coordonnées image (x, y) , de la composante de translation de la caméra T_Z suivant l'axe optique, la focale f , les coordonnées du point d'expansion des flots (x_{FOE}, y_{FOE}) (FOE pour *Focus Of Expansion*) qui est préalablement calculé et enfin les distances des plans par rapport à l'origine d_h , d_f et d_l . Il en ressort que les composantes u et v s'expriment soit par une fonction linéaire soit par une fonction quadratique des coordonnées (x, y) .

TABLE 4.1 – Les équations de mouvement 2D pour 3 modèles de plan : horizontal, frontal et latéral.

Plan horizontal	Plan frontal	Plan latéral
$u = \frac{T_Z}{fd_h} y (x - x_{FOE})$	$u = \frac{T_Z}{d_f}(x - x_{FOE})$	$u = \frac{T_Z}{fd_l} x (x - x_{FOE})$
$v = \frac{T_Z}{fd_h} y (y - y_{FOE})$	$v = \frac{T_Z}{d_f}(y - y_{FOE})$	$v = \frac{T_Z}{fd_l} x (y - y_{FOE})$

Deux espaces de vote sont créés, la u -vitesse et la v -vitesse, de tailles respectives $u_{max} \times W$ et $H \times v_{max}$ pour une image de taille $H \times W$. La figure 4.3 montre un exemple de calcul de v -vitesse. Chaque ligne de la v -vitesse (cf. figure 4.3(b)) correspond à un histogramme des composantes v du flot optique calculé sur cette ligne dans l'image. Puisqu'une expression quadratique relie v et y (voir tableau 4.1), une parabole apparaît dans l'espace v -vitesse. Elle correspond aux pixels appartenant au plan horizontal, à savoir la route. Ensuite, la ligne droite dans l'espace v -vitesse témoigne de la présence d'un plan frontal (ici un véhicule). Une transformation de Hough classique est utilisée pour détecter les paraboles et les lignes droites correspondant aux différents plans 3D présents dans la scène. Le concept de la uv -vitesse a été publié dans [MGB17a]. Les premières évaluations nous ont permis de vérifier que la précision de détection reste identique à celle obtenue par la c -vitesse, avec un temps de traitement réduit. Une évaluation plus poussée doit être réalisée à court terme. Dans la suite de ces travaux, nous envisageons d'étendre la réflexion à d'autres types de mouvement de la caméra, en étudiant en particulier l'impact des rotations sur les résultats de uv -vitesse.

4.2.2 Détection d'anomalies

Le terme d'anomalie revêt plusieurs sens et peut se caractériser dans l'image à différents niveaux : celui du pixel, celui de l'apparence des objets ou encore de leur comportement. Quoi qu'il en soit, il s'agit d'un écart par rapport à une normalité préalablement apprise. Dans le cadre du projet européen SPY sur la sécurité publique, nous avons proposé deux approches de détection d'anomalies [TGL12b, CBAGL16].

La méthode développée dans le cadre du post-doctorat de Hedi Tabia [TGL12b] définit l'anomalie au niveau du pixel par analyse du flot optique KLT, en modélisant la normalité de façon probabiliste. Cette méthode est particulièrement adaptée aux scènes de foule pour lesquelles la segmentation de l'image en objets est délicate [ABF06, AS07]. Une soustraction du fond est préalablement opérée afin de limiter l'analyse aux pixels en mouvement. Le comportement de chaque pixel est observé durant une certaine période de temps, appelée *historique de comportement*. Les densités

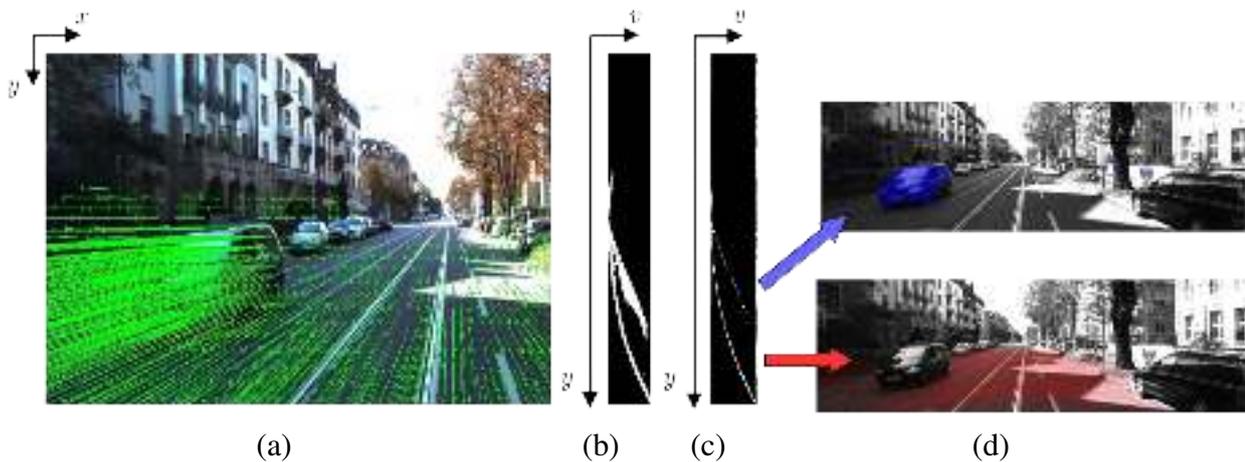


FIGURE 4.3 – Principe de la uv -vélocité. (a) : image extraite de la base KITTI et flot optique fourni par la vérité terrain. (b) Espace de vote v -vélocité avant filtrage et (c) après filtrage. La parabole correspond aux vecteurs de flot optique associés au plan de la route, tandis que la droite correspond au plan frontal (la voiture située sur la chaussée). (d) Détection des points associés au plan frontal en bleu (ici la voiture) et des points associés au plan horizontal en rouge (ici la route)

de probabilités de l’amplitude et de l’orientation du flot optique sont modélisées par un mélange de gaussiennes. Un pixel est considéré comme *normal* lorsque ses valeurs sont cohérentes avec le modèle appris et *anormal* dans le cas contraire. L’approche a été évaluée sur la base de données proposée par Mehran et al. [MOS09] composée de 20 vidéos : 12 vidéos présentant des situations de foules normales et 8 vidéos présentant des scènes d’évasion. La méthode détecte moins de 5 % de pixels anormaux dans les vidéos montrant des situations normales, tandis que les situations anormales produisent plus de 20 % de pixels anormaux.

Ces travaux ont été continués lors du stage post-doctoral de Ahmed Chamseddine [CBAGL16] (financé également dans le cadre du projet SPY) durant lequel nous avons proposé un système modulaire qui, au-delà de la détection d’événements anormaux, est capable de les catégoriser. Des modèles de normalité complémentaires sont définis à la fois globalement dans l’image et localement dans des blocs de pixels. Trois catégories de caractéristiques ont été prises en compte : 1) l’évolution spatio-temporelle du mouvement binaire issu d’une soustraction du fond ; 2) le flot optique extrait à partir d’une approche KLT pyramidale ; 3) les dérivées spatiales du mouvement. Au niveau local, un mélange de gaussiennes est, là encore, exploité pour modéliser la normalité de mouvement en chaque bloc de pixels. Une analyse en composantes principales permet d’en réduire la dimensionnalité. La figure 4.4 montre des résultats obtenus sur la base UMN⁷ et la figure 4.5 la base UCSD [LMV13a] qui montre des vidéos de piétons marchant sur une place. Enfin, la figure 4.6 permet de comparer les résultats de l’approche avec des méthodes issues de la littérature.

L’approche proposée permet d’analyser qualitativement l’activité à partir d’un ensemble d’histogrammes hybrides prenant en compte à la fois l’orientation du flot optique et celle du gradient temporel ainsi que les statistiques des pixels en mouvement. Une signature binaire de 13 éléments est produite à partir de ces caractéristiques pour une classification.

4.3 Décrire le mouvement

Dans les applications de surveillance et d’interaction homme-machine, il est souvent nécessaire d’avoir recours à des représentations spatio-temporelles des images et des objets, afin de détecter des

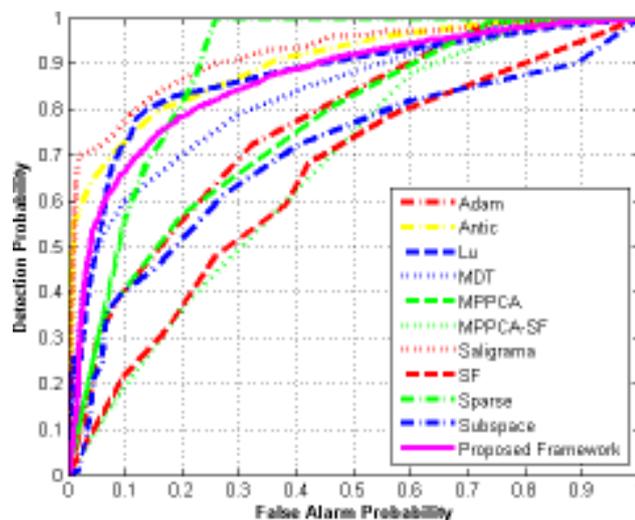
7. Unusual Crowd Activity Dataset : <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>



FIGURE 4.4 – Exemples de détections d’anomalies sur des images extraites de la base de données UMN.



FIGURE 4.5 – Exemples de détections d’anomalies sur des images extraites de la base de données UCSD.



Méthode		Principe de la méthode
Adam	[ARSR08]	Real-time detection of unusual events based on multiple local monitors which collect low-level statistics.
Antic	[AO11]	Scene parsing : localization of the abnormalities using statistical inference.
Lu	[LSJ13]	Fast sparse combination learning.
MDT	[LMV13b]	Mixture of Dynamic Textures models.
SF	[MOS09]	Social force models.
MPPCA	[KG09]	Space-Time Markov Random Field. Atomic motion patterns are learnt via a Mixture of Probabilistic Principal Component Analyzers.
MPPCA-SF	[LMV13b, MOS09, KG09]	Idem where MPPCA is combined with SF Normalized flow.
Saligrama	[SC12]	Some empirical rules are used to fuse local statistics across spatio-temporal locations and scales and produce a composite score for a video segment.
Sparse	[YYL11]	Sparse reconstruction cost over a normal dictionary.
Subspace	[EV09]	Subspace clustering based on sparse representation.

FIGURE 4.6 – Détection d’anomalies sur la base UCSD (scène 1). Comparaison des courbes ROC de l’approche proposée dans [CBAGL16] avec celles des méthodes issues de l’état de l’art.

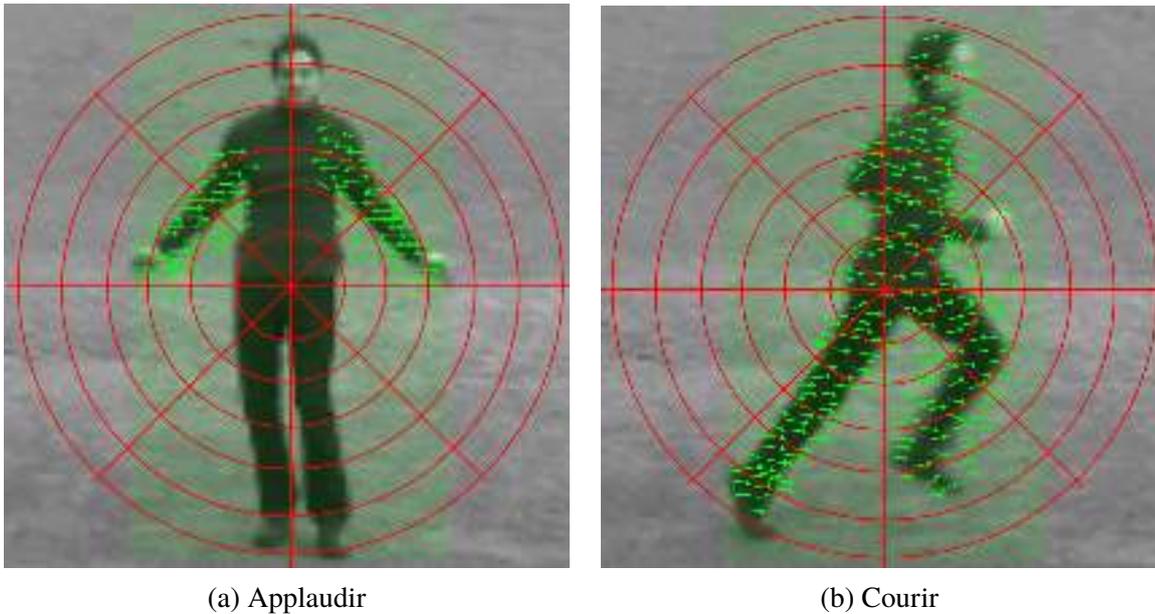


FIGURE 4.7 – Vecteurs de flot optique pour deux vidéos d’action humaine : applaudir et courir (base de données KTH).

anomalies, reconnaître des activités humaines, ou encore des gestes. Comme pour les descripteurs d’apparence abordés dans le chapitre précédent, un descripteur de mouvement est d’autant plus efficace qu’il est compact, invariant et discriminant.

Dans le cadre des projets SPY et MAPOCA, nous nous sommes principalement intéressés à l’analyse des activités humaines. La reconnaissance d’activités peut passer par l’identification d’actions simples du quotidien telles que marcher, courir, sauter. La littérature scientifique présente de nombreuses analyses de l’état de l’art sur les descripteurs de mouvement [Pop10, TCSU08]. Notamment les travaux [AS10, JJB13] proposent un ensemble de primitives cinématiques dérivées du flot optique. Dollár et al. [DRCB05] ont développé un système de reconnaissance fondé sur des primitives spatio-temporelles locales et fonctionnant sur des vidéos acquises avec une caméra en mouvement. De nombreuses approches sont évaluées sur la base de données KTH [LL04]⁸ qui présente six types d’actions (marcher, faire du jogging, courir, boxer, agiter les bras et applaudir) réalisées plusieurs fois par 25 sujets, dans des scénarios d’intérieur et d’extérieur et à différentes échelles d’observation.

4.3.1 Histogrammes de mouvement en représentation polaire

Dans les travaux de post-doctorat d’Hedi Tabia [TGL12a] dans le cadre du projet SPY, nous avons proposé une méthode de reconnaissance à partir d’histogrammes concentriques, d’une façon similaire au descripteur de forme *Shape Context* [BMP01]. Il s’agit tout d’abord de détecter les personnes dans la vidéo puis de construire des histogrammes de flot optique centrés sur ces personnes. La figure 4.7 permet d’illustrer le principe de ce descripteur qui combine les informations de forme et de mouvement. Nous considérons un repère de coordonnées dont l’origine est le centre d’inertie de la forme humaine préalablement détectée. La région est décomposée en 48 zones (6 amplitudes différentes \times 8 directions) selon un système de coordonnées polaires. De façon similaire à [ICD08], un histogramme de mouvement à 4 cellules est construit en chacune de ces portions d’images, une cellule comptant les flots pour 4 orientations données : 0 , $\pi/2$, $\pi/4$ et $3\pi/4$. L’ensemble de ces 48 histogrammes forme ainsi un descripteur de 192 valeurs.

8. KTH est disponible sur le web : <http://www.nada.kth.se/cvap/actions/>

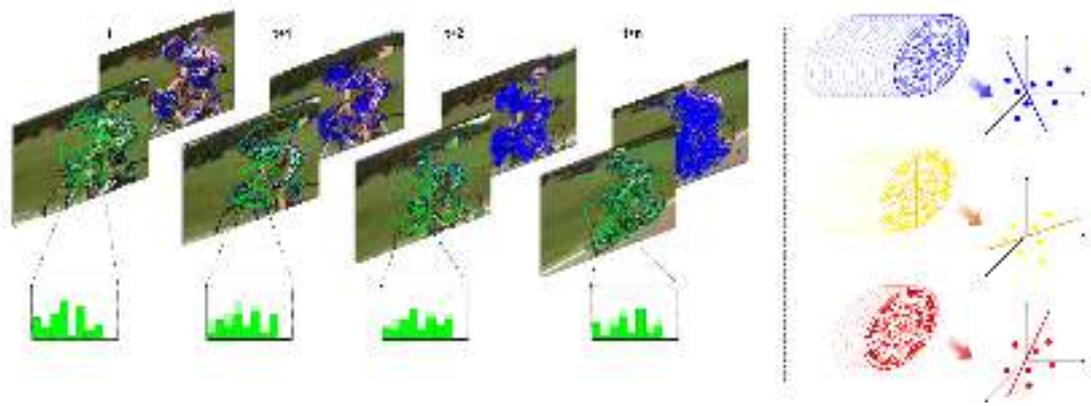


FIGURE 4.8 – Caractérisation de l'activité humaine par histogrammes centrés sur l'action et classification SVM [CaMGN16].

Sur la base de ce descripteur, une approche par sac de mots est utilisée. Une classification en k classes (par k -means, en utilisant une distance χ^2) permet d'assigner les histogrammes calculés à un alphabet de taille k . Cet alphabet sert à construire une représentation discriminante à partir de laquelle l'action humaine est décrite. Par analogie avec les mots-clés (*keywords*), ces vecteurs sont appelés des mouvements-clés (*keymotions*). Une classification est enfin effectuée en considérant ces sacs de mouvements-clés comme des vecteurs à classifier. Après avoir affecté chaque histogramme de mouvement à la classe la plus proche, le problème revient à de l'apprentissage supervisé multi-classe. Une classification naïve Bayésienne [Lew98] ainsi que les SVM (*Support Vector Machine*) [Vap98] ont été testés. Les expériences réalisées sur la base KTH ont permis d'atteindre un taux de reconnaissance de 82.36 % par SVM et 75.83 % par classification bayésienne. Par ailleurs, la classification par SVM s'avère plus performante que les approches *Space-time interest Points* [LMSR08] (80%), *Velocity histories* [MPK09] (74%) et *Spatio-temporal Cuboids* [DRCB05] (66%).

Cette idée de descripteur centré sur l'action a été reprise dans le cadre du projet MAPOCA [CaMGN16] (voir Fig. 4.8), avec un schéma de classification par SVM. Les primitives de mouvement sont ici les trajectoires denses et semi-denses décrites dans la section 4.1.2. Celles-ci sont caractérisées par leur âge (le nombre d'images depuis leur naissance), la norme et la direction de leur vitesse ainsi que leur courbure. Les histogrammes concentriques sont mis à jour temporellement par moyenne récursive et non plus recalculés pour chaque nouvelle image. Cela apporte une plus grande flexibilité de la reconnaissance des actions vis-à-vis de leur durée. Enfin, plusieurs échelles temporelles sont prises en compte dans le calcul des histogrammes. Les résultats de reconnaissance atteignent 94 % sur la base de KTH et 88.7 % sur la base *UT-Interaction*. Cette dernière est constituée de 120 vidéos [RA10] qui montrent six différents types d'interactions humaines.

4.3.2 Représentation par mélange de gaussiennes

Dans le cadre du projet MAPOCA, nous avons proposé dans [MMGB15] (voir la figure 4.9) de caractériser les trajectoires par leur cinématique (courbure θ , vitesse s , torsion k , etc) et de considérer chacune de ces caractéristiques comme une variable aléatoire respectant un modèle de mélange de K gaussiennes (*MoG*). Les moyennes des densités sont initialisées à l'aide des premières trajectoires. Ces distributions sont ensuite mises à jour, voire éliminées, à l'arrivée de nouvelles variables. Afin d'intégrer la dimension spatiale du mouvement à différents niveaux d'analyse, le calcul des *MoG* est fait à partir de portions d'images extraites par un partitionnement progressif de l'image. Cette approche a été employée dans un tout premier travail de collaboration informelle avec Annelies Braffort au LIMSI sur l'aide à l'annotation de vidéos de langue des signes. Il s'agissait ici de reconnaître les dates dans les vidéos, en particulier les jours et les mois. L'évaluation a été faite sur

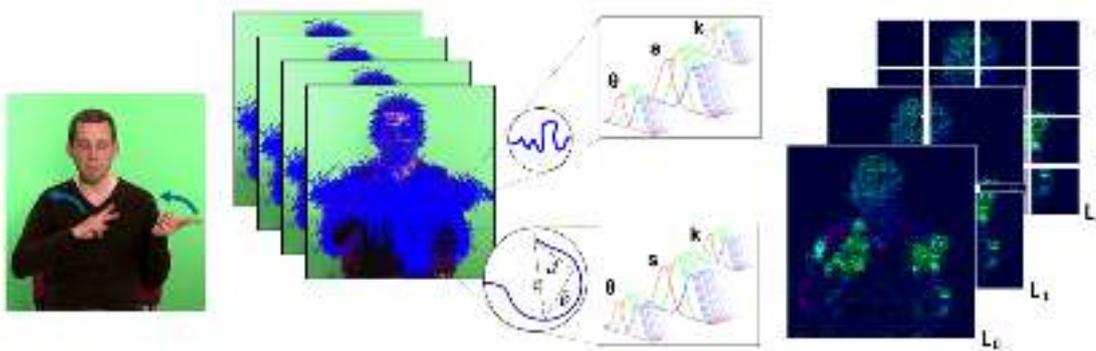


FIGURE 4.9 – Analyse de mouvement par mélange de gaussienne et représentation pyramidale [MMGB15].

une base de données interne au LIMSI, qui a donné 74 % de bonne reconnaissance pour les jours et 72 % pour les mois. Par ailleurs, sur la base KTH, un taux de reconnaissance de 92.23% a été atteint et 90.3% sur *UT-Interaction*.

4.4 Conclusion

Ce chapitre a abordé le mouvement comme une primitive de bas niveau dont l'analyse permet d'extraire des informations sur la géométrie de la scène, ou tout simplement de détecter des anomalies dans les vidéos. En premier lieu, nous avons proposé une méthode de flot optique qui, partant d'une approche classique, estime la fiabilité des résultats *a posteriori* et les corrige en propageant les flots les plus pertinents en exploitant des hypothèses d'homogénéité locale en termes de couleur. Nous avons ensuite proposé des descripteurs de mouvement centrés sur l'action. Ces descripteurs de mouvement ont été utilisés dans le cadre la reconnaissance d'activités humaines, que ce soit dans leurs activités individuelles, ou leurs interactions comme dans la communication visuo-gestuelle. Nous avons introduit ici les premières étapes d'un travail que nous souhaitons développer plus largement dans les années à venir. Les travaux abordés dans ce chapitre autour du mouvement sont ceux pour lesquels les perspectives de travail sont les plus nombreuses. Ces perspectives font l'objet du chapitre suivant.

Chapitre 5

Conclusion

Après des études d'ingénieur, c'est le goût du traitement d'image qui m'a amenée à la recherche. Ma thèse s'est déroulée dans un laboratoire d'agro-alimentaire, en commençant sur le thème de la segmentation couleur puis en se ré-orientant¹ vers le suivi de points d'intérêt pour l'asservissement visuel. Arrivée en 2006 au laboratoire IEF, j'ai mené mon travail dans un département spécialisé en architecture et en vision autour des systèmes autonomes. En rejoignant le LIMSI en 2013², j'ai découvert de nouvelles problématiques autour de l'interaction, de la réalité augmentée et des applications art-science qui constituent un terreau fertile pour mes réflexions autour de la couleur, de l'apparence et du mouvement.

Ce parcours professionnel a bien entendu influencé mon travail, faisant de moi une spécialiste du traitement d'image, capable d'appliquer ces techniques à des thèmes scientifiques complexes tels que la Réalité Augmentée ou la vidéo-surveillance, à défaut d'être la spécialiste d'une méthodologie bien spécifique et à laquelle la communauté pourrait facilement m'associer. À différents niveaux de la chaîne de vision, de la couleur au mouvement, j'ai cherché à définir des primitives et des descripteurs image (du pixel à l'action) dans l'objectif de satisfaire les critères d'unicité, d'invariance, de rapidité, de compacité, tout en les intégrant dans des tâches d'appariement. Mes travaux actuels et leurs perspectives s'orientent vers l'analyse de scènes et d'activités humaines (section 5.1), avec des réflexions autour de quatre thèmes abordés respectivement dans les sections 5.2 à 5.3 : la pluralité des descripteurs, la temporalité, le contexte et la reconfigurabilité.

5.1 Applications

Par « analyse de scènes », j'entends les premières étapes de l'interprétation, telles que la détection d'anomalies (comme évoqué dans le chapitre 4), et la détermination de la structure géométrique de la scène, comme nous le développons actuellement autour de la *uv*-vitesse. Ces deux sujets font respectivement l'objet des thèses d'Adrien Arnaud (thèse CIFRE co-encadrée avec Mehdi Ammi du LIMSI depuis 2016) et de Tan Khoa Mai (thèse sur contrat doctoral ED STIC Paris Saclay, co-encadrée depuis 2015 avec Samia Bouchafa à l'IBISC). Dans la thèse d'Adrien Arnaud, nous nous intéressons à la construction automatique de maquettes 3D d'intérieurs à partir d'un simple scan réalisé à l'aide d'une tablette équipée d'un capteur RGB+D et de capteurs proprioceptifs. Les applications visées sont la rénovation de logements pour les particuliers, ou encore l'évaluation des performances énergétiques. Concernant la thèse de Tan Khoa Mai, il s'agira, là aussi, de se positionner dans un cadre de caméra en mouvement et d'extraire la structure de la scène à partir du flot optique.

1. après la mutation de mon encadrante à l'ADEME
2. Suite à des restructurations au sein de l'IEF

En termes d'analyse des activités, nous nous intéresserons tout particulièrement à la communication visuo-gestuelle, à commencer par la Langue des Signes Française (LSF). Dans ce cadre d'étude, nous comptons développer des approches pouvant se généraliser et s'étendre aisément vers d'autres applications telles que l'assistance et le maintien des personnes à domicile en profitant des bases de données ouvertes comme celles du challenge THUMOS³. Ce sera l'objet de la thèse de Hussein Chaaban, co-encadrée avec Annelies Braffort (LIMSI) depuis octobre 2017. Notons que l'analyse de scènes et l'analyse d'activités humaines sont indissociables, dans le sens où la scène peut soit se caractériser par la présence d'activités humaines, soit contraindre le déroulement des activités et constituer un contexte pour l'action.

Nos travaux sur l'analyse de vidéos de LSF ont débuté en 2015 dans le cadre du post-doctorat de Fabio Martinez (financement sur projet Digiteo MAPOCA), par une première collaboration avec Annelies Braffort au LIMSI. Avec plusieurs membres du groupe ILES (Information Langue Écrite et Signée), elle travaille sur l'élaboration, l'analyse et l'annotation de lexiques et de corpus de LSF. Ces études s'avèrent précieuses pour les recherches en sciences du mouvement et en linguistique de la LSF. Notre collaboration a pour objectif de mettre en place une pré-annotation automatique des vidéos à l'attention des experts en LSF. Cette première ébauche n'aura plus qu'à être validée, modifiée ou complétée par l'expert, qui disposera de temps supplémentaire pour l'analyse fine de la vidéo. L'annotation se fera de manière incrémentale, en commençant par une caractérisation du degré illustratif du signe (ou le degré d'iconicité). En effet, un signe ne provient pas forcément d'un lexique standard mais peut *montrer à voir*, par exemple à l'aide de transfert (de forme/taille, de personnes, de situation) qui s'inspirent d'une expérience réelle ou imaginaire et tentent de la reproduire dans l'espace de signation. Pour faire une analogie avec la langue vocale, ce *montrer à voir* s'exprime par le geste (« un poisson gros comme ça ») ou la prosodie (pour imiter un personnage). Après cette première classification des signes, nous nous attacherons à définir des méthodes adaptées pour chacune de ces deux classes, pour aller vers une caractérisation plus riche. La segmentation des vidéos, c'est-à-dire la séparation des signes et de leurs transitions, fait également partie des objectifs à moyen terme. Enfin, nous nous pencherons sur la recherche d'un signe dans une vidéo. À terme, ces travaux devraient mener vers l'interaction par le geste et le développement de prototypes d'applications pédagogiques pour les communautés de Sourds. Ces travaux ont commencé dans le cadre d'un projet interne au LIMSI (*Action Incitative* 2015-16, stage M2 de Kenza Benseghir) et ont continué au cours du projet exploratoire de l'Institut Carnot (et le stage M2 de Hussein Chaaban) qui nous a permis de mettre en place la détection de la plupart des articulateurs impliqués dans la langue des signes : tête, bouche, yeux (orientation et clignement), sourcils, mains, buste. La thèse de Valentin Belissen commencée en octobre 2017, en co-encadrement au LIMSI avec Annelies Braffort, devrait permettre d'avancer sur ces problématiques.

D'autre part, je compte maintenir une activité art-science régulière, dans l'idée de rencontrer le public hors des murs du laboratoire, et leur faire vivre une expérience en lien avec les résultats issus de nos recherches. Cela se fera par le dialogue et les réflexions avec les artistes et prendra la forme d'installations, de préférence interactives. Dans cette réflexion, l'observateur est situé au cœur du processus d'analyse, de rendu et d'interaction. Par exemple, dans l'approche HTRI (issue d'une collaboration art-sciences avec Bertrand Planes et Christian Jacquemin) évoquée à la section 2.1.3, nous avons exploité la saillance visuelle couleur et temporelle de manière à fournir une image cohérente avec les propriétés du système visuel humain. À l'inverse, dans l'installation *Gardien du Temple*⁴, montrée au festival CURIOSITas en 2015, un texte est camouflé par vidéo-projection, et donc dissimulé de l'observateur. Enfin, j'ai tout récemment entamé une nouvelle réflexion avec Véronique Caye sur son projet *Vera Icona (Les invisibles)*⁵ qui traite du sujet de l'image qui n'est pas

3. <http://crcv.ucf.edu/THUMOS14/home.html>

4. projet art-science 2014-15 soutenu par la Diagonale Paris-Saclay, monté dans le cadre de la thèse de P-A. Bokaris, et en collaboration avec l'artiste V. Caye, C. Jacquemin au LIMSI et avec J-M. Chomaz au LadHyX.

5. projet art-science 2016-17 soutenu par la Diagonale Paris-Saclay, collaboration avec l'artiste V. Caye, P-A. Bo-

faite de la main de l'homme, et plus précisément de la représentation des visages comme instrument politique. Le projet en propose une métaphore contemporaine à travers une installation interactive sur le thème de la vidéo-surveillance, où l'image du visiteur sera dérobée à son insu. Les premières étapes du développement ont été réalisées par Toky Rakotomalala lors de son stage (niveau 4^e année d'ingénieur) au cours de l'été 2017. Une résidence d'artistes est prévue en novembre 2017 et devrait se conclure par la première exposition au public.

Enfin, certaines applications nécessitent une rapidité de traitement, en particulier pour les applications embarquées. Dès que de nouveaux descripteurs et méthodes seront proposées, il sera intéressant d'évaluer leur performances et leur portage sur les architectures parallèles, en collaboration avec Lionel Lacassagne de l'équipe ALSOC au LIP6.

5.2 Pluralité des descripteurs

La plupart des descripteurs évoqués dans ce rapport sont liés à un objet de la scène et l'objet est modélisé de façon globale. Or, la modélisation peut s'enrichir par un ensemble de descripteurs fragmentaires, soit répartis uniformément ou de façon contextuelle, et à différents niveaux d'échelle. Cette *pluralité spatiale* répond en particulier aux tâches de vision par ordinateur qui requièrent un fort pouvoir discriminant des représentations comme la ré-identification de personnes, ou la reconnaissance d'objets⁶. Au delà des descripteurs d'objets, décrire la relation entre les objets s'avère porteuse de sens, notamment dans le cadre du suivi multi objets. Modéliser les articulations entre objets, leurs relations spatiales et leurs dynamiques mutuelles peut être un pas vers une meilleure compréhension d'une activité humaine. Cette réflexion prend corps dans les travaux menés actuellement sur l'annotation de vidéos de LSF, où le sens provient de la dynamique conjointe de plusieurs articulateurs. Dans un premier temps, il s'agira d'identifier, par analyse statistique, les articulateurs (parmi les mains, la tête, les clignements des yeux, les haussements de sourcils, l'orientation du regard, la labialisation) les plus informatifs pour la réalisation de cette tâche afin de leur attribuer une importance plus forte dans le processus de classification. Nous comparerons ces descripteurs dédiés avec des descripteurs plus génériques, à commencer par les descripteurs par covariance. Un objet peut également exploiter plusieurs descripteurs au cours du temps (*pluralité temporelle*), qui peuvent prendre successivement la main en fonction d'un contexte, comme les travaux de la section 3.4 ont pu en montrer un exemple. Il ne s'agit pas de calculer simultanément l'ensemble de ces modèles mais de les faire commuter afin de garantir une bonne fiabilité sans accroître les temps de calcul.

5.3 Temporalité d'une action

La section 4.3 a décrit quelques descripteurs d'actions. Nous avons vu qu'avec l'utilisation de mélanges de gaussiennes, le descripteur était mis à jour à la volée à l'aide du calcul récursif des statistiques. Plusieurs temporalités (ou échelles temporelles) étaient prises en compte à l'aide de fenêtres glissantes. Néanmoins, la sélection d'une temporalité optimale reste un problème largement ouvert, tant elle est variable d'un individu à l'autre et d'une action à l'autre. La plupart des travaux émettent des *a priori* sur la durée des actions ou considèrent plusieurs échelles temporelles, comme dans [LQY⁺16], où plusieurs granularités d'analyse sont prises en compte à l'aide de différents CNN⁷. D'autres, comme dans [YRMFF16], partent de l'idée que seules quelques images de la vidéos suffisent à définir la localisation et la nature de l'action.

karis

6. Les points d'intérêt SIFT sont un exemple concret de descripteurs fragmentaires multi-échelle et dont l'intérêt n'est plus à démontrer

7. Convolutional Neural Networks

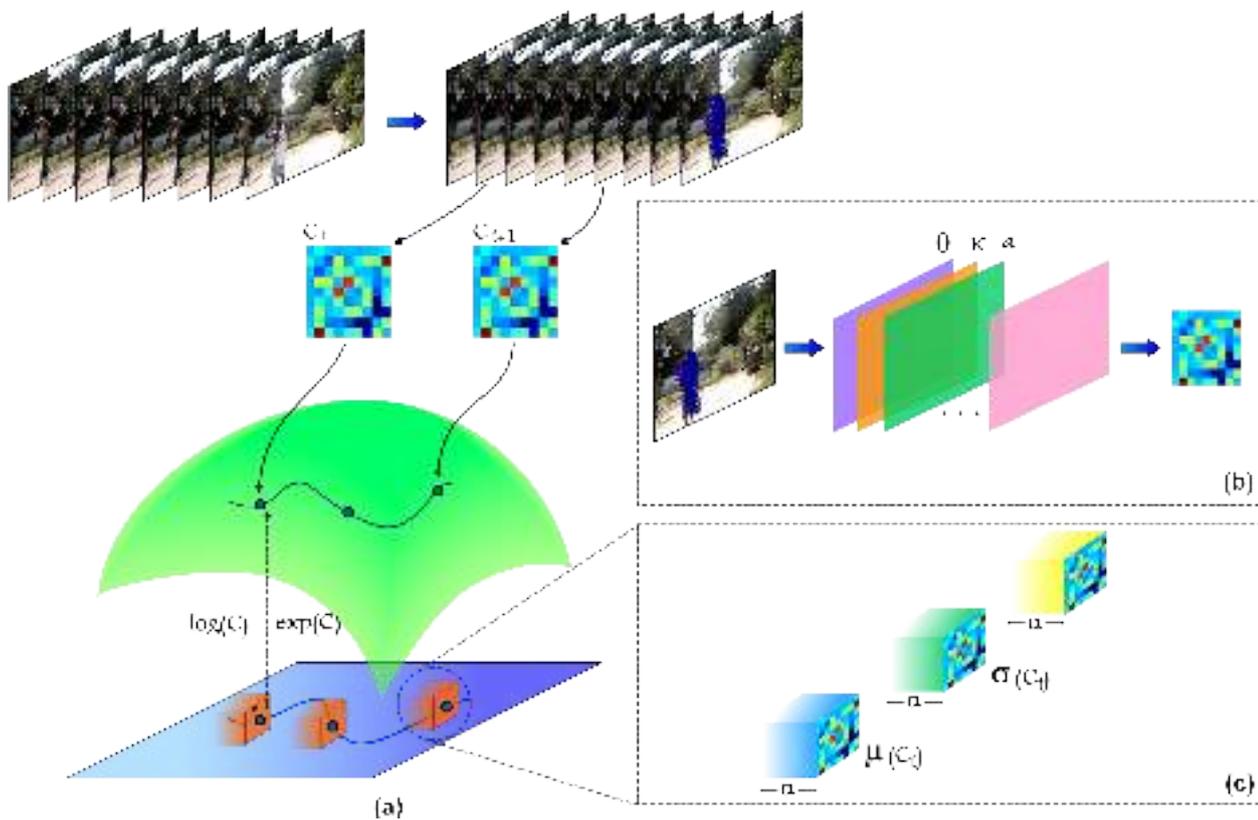


FIGURE 5.1 – Principe de l’analyse d’activités par matrice de covariance. (a) Les matrices de covariance sont calculées dans chaque image de la séquence. Elles partent des trajectoires semi-denses pour lesquelles sont calculées des caractéristiques cinématiques (b). Les matrices de covariance forment une trajectoire dans l’espace riemannien. Après transformation vers un espace euclidien (le plan tangent), les statistiques peuvent être calculées (c) .

Nos travaux passés nous avaient permis de constater le fort potentiel des descripteurs par covariance dans le cas de l’apparence (cf. Section 3.3.1). Dans le cadre du stage post-doctoral de Fabio Martinez, nous avons étendu le travail à la reconnaissance d’actions, en se penchant sur leur mise à jour temporelle. À partir des trajectoires denses extraites de la vidéo, différents attributs cinématiques sont combinés dans une matrice de covariance. La séquence temporelle de ces descripteurs forme une trajectoire dans la variété riemannienne, que nous supposons caractéristique de l’activité (cf. figure 5.1). Nous allons proposer un ensemble de statistiques mises à jour de façon récursive pour prendre en compte l’historique de l’activité. Deux stratégies sont envisagées : le calcul des statistiques le long de la trajectoire formée par les matrices de covariance dans l’espace riemannien et la projection des matrices de covariance vers un espace linéaire, pour permettre ensuite une classification par SVM. Ces descripteurs d’activité sont génériques dans le sens où ils peuvent décrire une dynamique locale associée à des gestes (pour la LSF par exemple) ou plus globale en décrivant les activités humaines au sens large. Le caractère récursif de la mise à jour autorise le traitement des vidéos à la volée et permet une reconnaissance des actions sur des vidéos qui ne sont pas segmentées au préalable (*untrimmed videos*).

Cette idée sera également exploitée dans le cadre de l’annotation de vidéos de LSF. Là aussi se pose la question de la temporalité et, avec elle, celle de la segmentation des signes et des transitions. Une application sur tablette devrait permettre à l’annotateur de sélectionner un signe de son choix en cliquant simplement sur l’une des images de la vidéo. Ainsi la segmentation du signe sera faite de façon relative autour de cette image, par recherche de la transition précédente et de la transition suivante.

5.4 Modéliser le contexte

Introduire des informations de contexte dans les méthodes d'analyse d'image devrait en faciliter l'interprétation. D'un point de vue du traitement d'information, le terme *contexte* s'avère extrêmement flou et difficile à définir. L'article [SGL15] en regroupe quelques définitions, tout en dressant un état de l'art des stratégies utilisées en traitement de l'information. De manière générale, une caractéristique est contextuelle pour une entité si elle la contraint et est susceptible de l'affecter. Dans ce sens, en vision par ordinateur, le contexte se décline à plusieurs niveaux, partant de celui du pixel vers une information plus sémantique. Dans [DHH⁺09], dix catégories de contextes sont identifiées pour le problème de détection d'objets.

Dans les approches de segmentation d'images ou de suivi, le contexte est généralement défini comme le simple voisinage direct de l'objet d'intérêt (un contexte spatial) caractérisé par son apparence, par exemple les superpixels voisins dans [FMPL16] ou encore l'entourage de l'objet à suivre, en prenant en compte la présence de distracteurs et de supporters [DVM11]. Le contexte peut-être appris au cours de la vidéo à l'aide de représentations parcimonieuses et de *Bags of Words* comme dans l'approche CEST (*Contexte-aware Exclusive Sparse Tracker*) [ZGL⁺16].

Le contexte peut être géométrique et correspondre à des informations de structure 3D de la scène. Dans la thèse d'Adrien Arnaud (commencée en janvier 2016) évoquée dans la section 5.1, nous envisageons de détecter les éléments clés comme les ouvrants ou les éléments de chauffage, en s'aidant à la fois du contexte, de l'apparence et de la géométrie et en introduisant des connaissances *a priori* sur la localisation des éléments dans une pièce. Dans les travaux de thèse de Tan Khoa Mai, les informations de structure apportées par l'analyse de la *uv* vitesse doivent être prises en compte dans le calcul de la carte de vitesse calculée dans l'image suivante de la vidéo.

Ensuite, si l'on s'intéresse au déplacement de personnes, leur comportement dynamique dépend de leur environnement, et en particulier de la présence d'autres individus (un contexte social) ou d'obstacles. Dans le modèle de force sociale (*Social Force Model*), utilisé notamment pour l'analyse de foules [MOS09, PESvG09], le mouvement de personnes est expliqué par deux forces, l'une exprimant la volonté propre de l'individu, la seconde exprimant l'interaction. Cette dernière se décompose en une force de répulsion qui maintient les individus à une certaine distance l'un de l'autre, et une force d'environnement qui prend en compte la présence d'obstacles physiques. Dans [LTFK⁺14], les interactions sont modélisées par des *interaction feature strings* qui combinent des primitives de bas niveau comme l'apparence et le mouvement ainsi que les interactions physiques entre les objets.

La prise en compte du contexte est une problématique que nous avons déjà rencontrée dans des travaux antérieurs : pour la détection d'anomalies dans les vidéos où le contexte était déterminé par une carte des mouvements « normaux » ; pour le suivi d'objets où des informations de contexte étaient injectées soit *via* une mesure de distinction objet/fond, soit par une détection d'occultation. Enfin, le projet de l'Institut Farman *Context-Aware Tracking* (collaboration 2017-18 avec S. Rodriguez, SATIE) s'inscrit pleinement dans cette problématique. Il s'agira de modéliser les informations contextuelles à partir de critères temporels (la dynamique des objets) et spatiaux (la localisation par rapport aux éléments de la scène). Nous évaluerons l'impact de ces informations sur le suivi multi-objets.

5.5 Reconfigurabilité

Devant la variété des contraintes induites par les applications, les approches d'analyse de séquences d'images peinent à être suffisamment génériques car elles sont souvent proposées en réponse à une/plusieurs contraintes (perturbations rencontrées, caractéristiques des capteurs, temps d'exécution) ou un/plusieurs contextes donnés.

Certaines méthodes de suivi d'objets répondent à ce problème, en proposant des modèles d'apparence adaptatifs [SCS12], où la plasticité des modèles permet d'en améliorer la stabilité. Ces approches se fondent généralement sur des classifications discriminantes objet/fond et partent d'un jeu de primitives visuelles qui reste le même tout au long de la vidéo mais dont le paramétrage peut être modifié, par exemple à travers des approches par ACP, par mélange de gaussiennes, par SVM, par représentations parcimonieuses [JLY12], par *boosting*, ou encore par *Random Forest*. Peu de travaux concernent l'évaluation automatique et la sélection d'un ensemble de primitives visuelles à la volée, et la modification des primitives au cours du temps.

À partir des attributs de bas niveau disponibles (couleur, géométrie, profondeur, normales aux surfaces, voisinage), nous souhaitons proposer des mesures de qualité ou de pertinence des attributs et des méthodes, en exploitant le maximum de données fournies par l'algorithme lui-même. Nous l'avons fait pour le suivi de points qui privilégie la couleur uniquement lorsqu'elle est jugée pertinente et nous souhaitons étendre la réflexion à d'autres catégories de primitives. Ensuite, pour une meilleure fiabilité, les méthodes devraient être capables de s'auto-évaluer, soit pour se corriger, soit pour faire appel à une approche *a priori* plus fiable, comme cela a été initié dans les approches décrites à la section 3.4. Ces pistes de travail sont abordées actuellement dans les travaux de thèse de Tan Khoa Mai (sur le flot optique et la détection de plans 3D par UV-vélocité), où les approches proposées sont capables d'évaluer la fiabilité du flot optique estimé et de corriger les résultats de façon itérative. Nous souhaitons développer un système *opportuniste* dans le sens où il exploite les attributs disponibles selon leur qualité.

Bibliographie

- [ABF06] E. L. Andrade, S. Blunsden, and R. B. Fisher. Hidden markov models for optical flow analysis in crowds. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 1, pages 460–463, 2006.
- [ACC⁺16] A. Arnaud, J-B. Corrége, C. Clavel, M. Gouiffès, and M. Ammi. Exploration of virtual environments on tablet : Comparison between tactile and tangible interaction techniques. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction, ICMI 2016*, pages 357–361, New York, NY, USA, 2016. ACM.
- [ACGA16] A. Arnaud, J. Christophe, M. Gouiffes, and M. Ammi. 3d reconstruction of indoor building environments with new generation of tablets. In *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology, VRST '16*, pages 187–190, New York, NY, USA, 2016. ACM.
- [AO11] B. Antic and B. Ommer. Video parsing for abnormality detection. In *International Conference on Computer Vision*, pages 2415–2422, 2011.
- [AOV12] A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK : Fast Retina Keypoint. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
- [ARSR08] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3) :555–560, 2008.
- [AS07] S. Ali and M. Shah. A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6, June 2007.
- [AS10] S. Ali and M. Shah. Human Action Recognition in Videos Using Kinematic Features and Multiple Instance Learning. *PAMI*, 32 :288–303, February 2010.
- [AS11] A. Andriyenko and K. Schindler. Multi-target tracking by continuous energy minimization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1265–1272. IEEE, 2011.
- [ASR12] A. Andriyenko, K. Schindler, and S. Roth. Discrete-continuous optimization for multi-target tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1926–1933, 2012.
- [BA96] M. J. Black and P. Anandan. The Robust Estimation of Multiple Motions : Parametric and Piecewise-Smooth Flow Fields. *Computer Vision and Image Understanding*, 63(1) :75–104, January 1996.
- [BB14] S. Bık and F. Brémond. *Re-identification by Covariance Descriptors*, pages 71–91. Springer London, London, 2014.
- [BCBT11] S. Bık, E. Corvee, F. Brémond, and M. Thonnat. Multiple-shot human re-identification by Mean Riemannian Covariance Grid. In *Advanced Video and Signal-Based Surveillance (AVSS), IEEE International Conference on*, pages 179–184. IEEE, 2011.

- [BCBT12] S. Bağ, E. Corvee, F. Brémond, and M. Thonnat. Boosted human re-identification using Riemannian manifolds. *Image and Vision Computing*, 30(6) :443–452, 2012.
- [BCV13] Y. Bengio, A. Courville, and P. Vincent. Representation Learning : A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8) :1798–1828, Aug 2013.
- [Ben07] A. Benoit. *The human visual system as a complete solution for image processing*. Theses, Institut National Polytechnique de Grenoble - INPG, February 2007.
- [BETV08] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up Robust Features (SURF). *Computer Vision and Image Understanding : CVIU*, 110(3) :346–359, June 2008.
- [BFTF11] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(9) :1806–1819, 2011.
- [BGC⁺17] P-A. Bokaris, M. Gouiffès, V. Caye, C. Jacquemin, and J-M. Chomaz. Gardien du temple : concealment and revelation. an interactive installation involving poetry, performance and spatial augmented reality. *Leonardo MIT Press Journal*, 2017.
- [BGJC14] P-A. Bokaris, M. Gouiffès, C. Jacquemin, and J-M. Chomaz. Photometric compensation to dynamic surfaces in a projector-camera system. In *European Conference on Computer Vision*, pages 283–296, 2014.
- [BHS98] A. Bab-Hadiashar and D. Suter. Robust Optic Flow Computation. *International Journal on Computer Vision*, 29(1) :59–77, August 1998.
- [BLM15] Q. Barthélemy, A. Larue, and J.I. Mars. Color Sparse Representations for Image Processing : Review, Models, and Prospects. *IEEE Transactions on Image Processing*, pages 1–12, 2015.
- [BM04] S. Baker and I. Matthews. Lucas-kanade 20 years on : A unifying framework. *International Journal on Computer Vision*, 56(3) :221–255, Feb 2004.
- [BM11] T. Brox and J. Malik. Large Displacement Optical Flow : Descriptor Matching in Variational Motion Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3) :500–513, March 2011.
- [BMP01] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 :509–522, 2001.
- [Bok16] P-A. Bokaris. *Video-projected augmented reality : Photometric compensation for static and dynamic concealment*. PhD thesis, Université Paris Saclay, 2016.
- [BPB07] R. V. Babu, P. Pérez, and P. Bouthemy. Robust tracking with motion estimation and local kernel-based color modeling. *IVC*, 25(8), 2007.
- [BR05a] O. Bimber and R. Raskar. *Spatial Augmented Reality : Merging Real and Virtual Worlds*. A. K. Peters, Ltd., Natick, MA, USA, 2005.
- [BR05b] S.T. Birchfield and S. Rangarajan. Spatiograms versus Histograms for region-based tracking. In *Computer Vision and Pattern Recognition*, pages 1158–1163, 2005.
- [BW02] T. Brox and J. Weickert. Nonlinear matrix diffusion for optic flow estimation. In *Pattern Recognition (Proc. DAGM)*, volume 2449 of *Lecture Notes in Computer Science*, pages 446–453. Springer, Sept. 2002.
- [BZ12] S. Bouchafa and B. Zavidovique. c-Velocity : A Flow-Cumulating Uncalibrated Approach for 3d Plane Detection. *International Journal of Computer Vision*, 97(2) :148–166, April 2012.

-
- [CaMGN16] F. Martínez Carillo, A. Manzanera ad M. Gouiffès, and T. P. Nguyen. Action-centric polar representation of motion trajectories for online action recognition. In *International Conference on Computer Vision Theory and Applications*, pages 442–448, 2016.
- [CBAGL16] A. Chamseddine Ben Abdallah, M. Gouiffès, and L. Lacassagne. A Modular System for Global and Local Abnormal Event Detection and Categorization in Videos. *Machine Vision and Applications*, 27(4) :463–481, February 2016.
- [CCM99] V. Caselles, B. Coll, and J-M. Morel. Topographic maps and local contrast change in natural images. *International Journal on Computer Vision*, 33(1) :5–27, 1999.
- [CCM02] V. Caselles, B. Coll, and J-M. Morel. Geometry and color in natural images. *Journal of Mathematical Imaging and Vision*, 16(2) :89–105, 2002.
- [CCS⁺11] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom Pictorial Structures for Re-identification. In *British Machine Vision Conference (BMVC)*, 2011.
- [CDF⁺04] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.
- [CF00] B. Coll and J. Froment. Topographic maps of color images. In *IEEE International Conference on Pattern Recognition*, Barcelona, 2000.
- [Cin14] R. G. Cinbis. *Fisher kernel based models for image classification and object localization*. Thèse, Université de Grenoble, July 2014.
- [CKS97] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International journal of computer vision*, 22(1) :61–79, 1997.
- [CL94] T. Carron and P. Lambert. Color edge detector using jointly hue, saturation and intensity. In *IEEE International Conference on Image Processing*, volume 3, pages 977–981. IEEE, 1994.
- [CLK16] J. Chan, J. A. Lee, and Q. Kemao. BORDER : An Oriented Rectangles Approach to Texture-Less Object Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2855–2863, June 2016.
- [CPR10] J.Y. Choi, K.N. Plataniotis, and Y.M. Ro. Using colour Local Binary Pattern features for face recognition. In *ICIP*, pages 4541–4544, Sept 2010.
- [CRL05] C. Yang, R. Duraiswami, and L. Davis. Efficient Mean-Shift tracking via a new similarity measure. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 176–183, 2005.
- [CRM03] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 :564–577, 2003.
- [CSBP11] A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos. Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet Divergence. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2399–2406. IEEE, 2011.
- [CSBP13] A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos. Jensen-bregman LogDet divergence with application to efficient similarity search for covariance matrices. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(9) :2161–2174, 2013.
- [CV01] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2) :266–277, 2001.
-

- [DHH⁺09] S. K. Divvala, D. Hoiem, J. H. Hays, A. A. Efros, and M. Hebert. An empirical study of context in object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1271–1278, June 2009.
- [DKMB09] B. De Decker, J. Kautz, T. Mertens, and P. Bekaert. Capturing multiple illumination conditions using time and color multiplexing. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2536–2543, June 2009.
- [DRCB05] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 65 – 72, October 2005.
- [DSTR11] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher. Appearance-based person reidentification in camera networks : problem overview and current approaches. *Journal of Ambient Intelligence and Humanized Computing*, 2(2) :127–151, 2011.
- [DT05] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *CVPR*, pages 886–893, 2005.
- [DVM11] T. B. Dinh, N. Vo, and G. Medioni. Context tracker : Exploring supporters and distracters in unconstrained environments. In *CVPR 2011*, pages 1177–1184, June 2011.
- [DZR⁺08] J. Dehos, E. Zeghers, C. Renaud, F. Rousselle, and L. Sarry. Radiometric compensation for a low-cost immersive projection system. In *Proc. of ACM symposium on Virtual reality softw. and techno., VRST '08*, pages 130–133, New York, NY, USA, 2008. ACM.
- [EDB12] E. Erdem, S. Dubuisson, and I. Bloch. Fragments based tracking with adaptive cue integration. *Computer Vision and Image Understanding*, 2012.
- [EPLE⁺17] H. J. Escalante, V. Ponce-López, S. Escalera, X. Baró, A. Morales-Reyes, and J. Martínez-Carranza. Evolving weighting schemes for the bag of visual words. *Neural Comput. Appl.*, 28(5) :925–939, May 2017.
- [EV09] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2797, 2009.
- [Far00] G. Farneback. Fast and accurate motion estimation using orientation tensors and parametric motion models. In *IEEE International Conference on Pattern Recognition*, volume 1, pages 135–139, 2000.
- [FB17] L. Fagot-Bouquet. *Representing to follow : Exploitation of parsimonious representations for multi-object tracking*. Thèse, Université Toulouse III Paul Sabatier (UT3 Paul Sabatier), March 2017.
- [FCM00] B. Funt, F. Ciurea, and J. Mccann. Retinex in Matlab. *Journal of Electronic Imaging*, pages 112–121, 2000.
- [FGN05] K Fujii, M.D. Grossberg, and S.K. Nayar. A projectorcamera system with real-time photometric adaptation for dynamic environments. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 814–821, 2005.
- [FHST05] G. D. Finlayson, S. D. Hordley, G. Schaefer, and G. Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38 :179–190, 2005.
- [FM99] W. Forstner and B Moonen. A metric for covariance matrices. Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University, 1999.

-
- [FMPL16] H. Fan, X. Mei, D. Prokhorov, and H. Ling. Cross datasets vegetation detection with spatial prior and local context. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, pages 735–740, June 2016.
- [Fon01] M. Fontaine. *Segmentation non supervisée d’images couleur par analyse de la connexité des pixels*. PhD thesis, Université de Lille I, Décembre 2001.
- [FS89] I. Fogel and D. Sagi. Gabor filters as texture discriminator. *Biological cybernetics*, 61(2) :103–113, 1989.
- [FZ04] D. Freedman and T. Zhang. Active contours for tracking distributions. *Image Processing, IEEE Transactions on*, 13(4) :518–526, 2004.
- [GBS05] J.M. Geusebroek, G.J. Burghouts, and A.W.M. Smeulders. The Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61(1) :103–112, 2005.
- [GCFMT12] M. Gouiffès, C. Collewet, C. Fernandez-Maloigne, and A. Trémeau. A study on local photometric models and their application to robust tracking. *Computer Vision and Image Understanding*, 116(8) :896–907, 2012.
- [GGvdWG12] T. Gevers, A Gijsenij, Joost van de Weijer, and J-M Geusebroek. *Color in Computer Vision : Fundamentals and Applications*. John Wiley & Sons, Inc., 2012.
- [GIK10] K. Guo, P. Ishwar, and J. Konrad. Action recognition using sparse representation on covariance manifolds of optical flow. In *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 188–195, 2010.
- [GL11] K. Grauman and B. Leibe. *Visual Object Recognition*. Synthesis Lectures on Artificial Intelligence and Machine Learning Series. Morgan & Claypool, 2011.
- [GLL10a] M. Gouiffès, F. Laguzet, and L. Lacassagne. Color connectedness degree for mean shift tracking. In *ICPR*, Istanbul, 2010.
- [GLL10b] M. Gouiffès, F. Laguzet, and L. Lacassagne. Projection-histograms for mean-shift tracking. In *IEEE ICIP*, pages 4617–4620. IEEE, 2010.
- [GM12] M. Garrigues and A. Manzanera. Real time semi-dense point tracking. In Aurélio Campilho and Mohamed Kamel, editors, *International Conference on Image Analysis and Recognition*, pages 245–252, Berlin, Heidelberg, 2012.
- [Gou05] M. Gouiffès. *Apports de la couleur et des modèles de réflexion pour l’extraction et le suivi de primitives*. PhD thesis, Université de Poitiers, 2005.
- [Gou07] M. Gouiffès. Tracking by combining photometric normalization and color invariants according to their relevance. In *IEEE International Conference on Image Processing*, 2007.
- [Gou08] M. Gouiffès. Adaptive spatio-colorimetric classification. In *European Conference on Colour in Graphics, Imaging, and Vision*, pages 411–416, 2008.
- [GPV10] M. Gouiffès, A. Patri, and M. Vasiliu. Robust obstacles detection and tracking using disparity for car driving assistance. In *IS&T SPIE Electronic Imaging*, volume 7539, pages 7539 – 7539 – 8, 2010.
- [GR11] S. Guo and Q. Ruan. Facial expression recognition using local binary covariance matrices. In *Wireless, Mobile & Multimedia Networks (ICWMMN 2011), 4th IET International Conference on*, pages 237–242, 2011.
- [GS99] T. Gevers and A. W. M. Smeulders. Color-based object recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(3) :453–464, 1999.
- [GyTL17] M. Gouiffès, A. Romero Mier y Terán, and L. Lacassagne. Color enhanced local binary patterns in covariance matrices descriptors (ELBCM). *Journal of Visual Communication and Image Representation*, 49(Supplement C) :447 – 458, 2017.
-

- [GZ08] M. Gouiffès and B. Zavidovique. A color topographic map based on the dichromatic model. *EURASIP Journal on Image and Video Processing*. 14 pages, 2008.
- [GZ09] M. Gouiffès and B. Zavidovique. Segments of color lines : a comparison through a tracking procedure. In *6th ICINCO International Conference on Informatics in Control, Automation and Robotics*, Milano, Italy., 2009.
- [GZ11] M. Gouiffès and B. Zavidovique. Body color sets : a compact and reliable representation of images. *Journal of Visual Communication and Image Representation*, 22 :48–60, 2011.
- [HBRB11] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *Image Analysis*, pages 91–102. Springer, 2011.
- [HCFE04] E. Hayman, B. Caputo, M. Fritz, and J-O Eklundh. On the significance of real-world conditions for material classification. *Computer Vision-ECCV 2004*, pages 253–266, 2004.
- [Hea92] G. Healey. Segmenting images using normalized color. *IEEE Transactions on Systems, Man and Cybernetics*, 22(1) :64–73, 1992.
- [HHD00] I. Haritaoglu, D. Harwood, and L. S. Davis. W^4 : real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8) :809–830, 2000.
- [HS81] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial intelligence*, 17(1) :185–203, 1981.
- [HS88] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.
- [HSWL12] M. T. Harandi, C. Sanderson, A. Wiliem, and B. C. Lovell. Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures. In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, pages 433–439, 2012.
- [ICD08] N. Ikizler, R.G. Cinbis, and P. Duygulu. Human action recognition with line and flow histograms. In *IEEE International Conference on Pattern Recognition*, dec. 2008.
- [JA09] K. Jungling and M. Arens. Detection and tracking of objects with direct integration of perception and expectation. In *International Conference on Computer Vision*, pages 1129–1136. Workshop, 2009.
- [JDSP10] H. Jégou, M. Douze, C. Schmid, and P. Pérez. Aggregating local descriptors into a compact image representation. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3304–3311, June 2010.
- [JFS01] H. Jin, P. Favaro, and S. Soatto. Real-time feature tracking and outlier rejection with changes in illumination. In *ICCV*, pages 684–689, 2001.
- [JJB13] M. Jain, H. Jégou, and P. Bouthemy. Better Exploiting Motion for Better Action Recognition. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2555–2562, June 2013.
- [JLY12] X. Jia, H. Lu, and M. H. Yang. Visual tracking via adaptive structural local sparse appearance model. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1822–1829, June 2012.
- [KAR13] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person reidentification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7) :1622–1634, 2013.

- [KFC10] C. Kanan, A. Flores, and G. W. Cottrell. Color constancy algorithms for object and face recognition. In *Proceedings of the 6th International Conference on Advances in Visual Computing - Volume Part I, ISVC'10*, pages 199–210, Berlin, Heidelberg, 2010. Springer-Verlag.
- [KG09] J. Kim and K. Grauman. Observe locally, infer globally : A space-time mrf for detecting abnormal activities with incremental updates. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2921–2928, 2009.
- [KLCR15] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg. Multiple Hypothesis Tracking Revisited. In *International Conference on Computer Vision, International Conference on Computer Vision*, pages 4696–4704, Washington, DC, USA, 2015.
- [KLL13] T. H. Kim, H. S. Lee, and K. M. Lee. Optical Flow via Locally Adaptive Fusion of Complementary Data Costs. In *International Conference on Computer Vision*, pages 3344–3351, dec 2013.
- [KMM10] Z. Kalal, J. Matas, and K. Mikolajczyk. P-N Learning : Bootstrapping Binary Classifiers by Structural Constraints. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [KSH12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems, NIPS'12*, pages 1097–1105, USA, 2012. Curran Associates Inc.
- [KSK90] G.J. Klinker, S.A. Shafer, and T. Kanade. A physical approach to image understanding. *International Journal of Computer Vision*, 4 :7–38, 1990.
- [KWT88] M. Kass, A. Witkin, and D. Terzopoulos. Snakes : Active contour models. *International journal of computer vision*, 1(4) :321–331, 1988.
- [Lag13] F. Laguzet. *Étude et optimisation d'algorithmes pour le suivi d'objets couleur*. Thèse, Université Paris-Sud, Décembre 2013.
- [LAT02] R. Labayrade, D. Aubert, and J. P. Tarel. Real time obstacle detection in stereovision on non flat road geometry through « v-disparity » representation. In *Intelligent Vehicle Symposium, 2002. IEEE*, volume 2, pages 646–651 vol.2, June 2002.
- [LBBH98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, November 1998.
- [LCR03] H. Liu, R. Chellappa, and A. Rosenfeld. Accurate dense optical flow estimation using adaptive structure tensors and a parametric model. *IEEE Transactions on Image Processing*, 12(10) :1170–1180, October 2003.
- [LCRP12] S. H. Lee, J. Y. Choi, Y. M. Ro, and K.N. Plataniotis. Local Color Vector Binary Patterns From Multichannel Face Images for Face Recognition. *IEEE Transactions on Image Processing*, 21(4) :2347–2353, April 2012.
- [LCS11] S. Leutenegger, M. Chli, and R. Y. Siegwart. BRISK : Binary Robust invariant scalable keypoints. In *International Conference on Computer Vision*, pages 2548–2555, Nov 2011.
- [Lei12] I. Leichter. Mean Shift Trackers with Cross-Bin Metrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4) :695–706, April 2012.
- [Leo12] I. Leonard. *Reconnaissance des objets manufacturés dans des vidéos sous-marine*. PhD thesis, ISEN, Université Bretagne Occidentale, 2012.
- [Ler11] N. Lertchuwongsa. *Color Lines, Regions and Their Stereo Matching*. Thèse de doctorat, Université Paris-Sud, France, Décembre 2011.

- [Lew98] D. D. Lewis. Naive (Bayes) at Forty : The Independence Assumption in Information Retrieval. In *European Conference on Machine Learning*, pages 4–15, 1998.
- [LGL11] F. Laguzet, M. Gouiffès, and L. Lacassagne. Automatic colorspace switching for robust tracking. In *IEEE ICSIPA*, November 2011.
- [LGLE11] F. Laguzet, M. Gouiffès, L. Lacassagne, and D. Etiemble. Automatic color space switching for robust tracking. In *IEEE ICSIPA*, pages 295–300. IEEE, 2011.
- [LGZ11] N. Lertchuwongsa, M. Gouiffès, and B. Zavidovique. Mixed color/level lines and their stereo- matching with a modified hausdorff distance. *Integr. Comput.-Aided Eng.*, 18(2) :107–124, April 2011.
- [LGZ12] N. Lertchuwongsa, M. Gouiffès, and B. Zavidovique. Enhancing a disparity map by color segmentation. *Integr. Comput.-Aided Eng.*, 19(4) :381–397, October 2012.
- [LHN09] Y. Li, C. Huang, and R. Nevatia. Learning to associate : Hybridboosted multi-target tracker for crowded scene. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2953–2960. IEEE, 2009.
- [LK⁺81] B.D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th international joint conference on Artificial intelligence*, 1981.
- [LL04] I. Laptev and T. Lindeberg. Velocity adaptation of space-time interest points. In *IEEE International Conference on Pattern Recognition*, 2004.
- [LLM16] A. Ledoux, O. Losson, and L. Macaire. Color local binary patterns : compact descriptors for texture classification. *Journal of Electronic Imaging*, 25(6) :061404, May 2016.
- [LLP06] A. Lehuger, P. Lechat, and P. Pérez. An adaptive mixture color model for robust visual tracking. In *IEEE International Conference on Image Processing*, pages 573–576, 2006.
- [LLS10] Y. Liu, G. Li, and Z. Shi. Covariance Tracking via Geometric Particle Filtering. *EURASIP Journal on Advances in Signal Processing*, 2010 :1–10, 2010.
- [LMSR08] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *IEEE Conference on Computer Vision and Pattern Recognition*, june 2008.
- [LMV13a] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99) :1–1, 2013.
- [LMV13b] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly Detection and Localization in Crowded Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99) :1–1, 2013.
- [Low04] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision*, 60(2) :91–110, 2004.
- [LP16] X. Li and K. N. Plataniotis. Color texture representation using circular-processing based Hue-LBP for histo-pathology image analysis. In *IEEE International Conference on Image Processing*, pages 3573–3577, Sept 2016.
- [LQY⁺16] Q. Li, Z. Qiu, T. Yao, T. Mei, Y. Rui, and J. Luo. Action recognition by learning deep multi-granular spatio-temporal video representation. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, ICMR '16*, pages 159–166, New York, NY, USA, 2016. ACM.

- [LRG⁺15] F. Laguzet, A. Romero, M. Gouiffès, L. Lacassagne, and D. Etiemble. Color tracking with contextual switching : real-time implementation on cpu. *Journal of Real-Time Image Processing*, 10(2) :403–422, Jun 2015.
- [LSJ13] C. Lu, J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab. In *International Conference on Computer Vision*, pages 2720–2727, Dec 2013.
- [LTFK⁺14] L. Leal-Taixé, M. Fenzi, A. Kuznetsova, B. Rosenhahn, and S. Savarese. Learning an image-based motion context for multiple people tracking. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14*, pages 3542–3549, Washington, DC, USA, 2014. IEEE Computer Society.
- [LVAN12] M. Liu, B.C. Vemuri, S.-I. Amari, and F. Nielsen. Shape retrieval using hierarchical total bregman soft clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12) :2407 –2419, December 2012.
- [LW12] P. Li and Q. Wang. Local log-Euclidean covariance matrix (L2ECM) for image representation and its applications. In *European Conference on Computer Vision*, pages 469–482. Springer, 2012.
- [LYYH16] Y. J. Liu, C. C. Yu, M. J. Yu, and Y. He. SLIC : A Fast Method to Compute Content-Sensitive Superpixels. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 651–659, June 2016.
- [MFTM01] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.
- [MGB17a] T.K. Mai, M. Gouiffès, and S. Bouchafa. Exploiting optical flow field properties for 3d structure identification. In *15th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, Madrid, Spain, 2017.
- [MGB17b] T.K. Mai, M. Gouiffès, and S. Bouchafa. Optical flow refinement using reliable flow propagation. In *International Conference on Computer Vision Theory and Applications*, pages 451–458. INSTICC, SciTePress, 2017.
- [MJ05] T. Moerland and F. Jurie. Learned Color Constancy From Local Correspondences. In *IEEE ICME'05*, pages 820–823, 2005.
- [MMGB15] F. Martínez, A. Manzanera, M. Gouiffès, and A. Braffort. A gaussian mixture representation of gesture kinematics for on-line sign language video annotation. In *Advances in Visual Computing : 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part II*, pages 293–303, Cham, 2015.
- [MN01] M. Middendorf and H. H. Nagel. Estimation and interpretation of discontinuities in optical flow fields. In *International Conference on Computer Vision*, volume 1, pages 178–183 vol.1, 2001.
- [MOS09] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 935–942, June 2009.
- [MPJ13] M. Gouiffès, B. Planes, and C. Jacquemin. HTRI : High time range imaging. *Journal of Visual Communication and Image Representation*, pages 361–372, 2013.
- [MPK09] R. Messing, C. Pal, and H. Kautz. Activity recognition using the velocity histories of tracked keypoints. In *International Conference on Computer Vision*, 2009.
- [MRG99] S. J. McKenna, Y. Raja, and S. Gong. Tracking colour objects using adaptive mixture models. *Image and vision computing*, 17(3) :225–231, 1999.

- [MRS13] A. Milan, S. Roth, and K. Schindler. Continuous energy minimization for multi-target tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2013.
- [MSB13] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a « completely blind » image quality analyzer. *IEEE Signal Processing Letters*, 20(3) :209–212, March 2013.
- [MSJ12] B. Ma, Y. Su, and F. Jurie. BiCov : a novel image representation for person re-identification and face verification. In *British Machine Vision Conference*, 2012.
- [MSP⁺11] M. R. Marnier, R. T. Smith, S. R. Porter, Markus M. Broecker, B. Close, and B. H. Thomas. Large scale spatial augmented reality for design and prototyping. In Borko Furht, editor, *Handbook of Augmented Reality*, pages 231–254. Springer New York, 2011.
- [Mun57] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial & Applied Mathematics*, 5(1) :32–38, 1957.
- [NFT97] S.K. Nayar, X. Fang, and B. Terrance. Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 21(3) :163–186, 1997.
- [NG98] H.-H. Nagel and A. Gehrke. Spatiotemporally adaptive estimation and segmentation of OF-fields. In Hans Burkhardt and Bernd Neumann, editors, *European Conference on Computer Vision*, number 1407 in Lecture Notes in Computer Science, pages 86–102, June 1998. DOI : 10.1007/BFb0054735.
- [NPGB03] S.K. Nayar, H. Peri, M.D. Grossberg, and P.N. Belhumeur. A Projection System with Radiometric Compensation for Screen Imperfections. In *ICCV Workshop PROCAMS*, Oct 2003.
- [NVC12] P.T. Nguyen, N.S. Vu, and A. Caplier. Face recognition using multi-modal binary patterns. In *IEEE International Conference on Pattern Recognition*, 2012.
- [OMC03] S. Obdrzalek, M. Matas, and O. Chum. On the interaction between object recognition and colour constancy. In *IEEE Workshop on Color and Photometric Methods in Computer Vision*, 2003.
- [OPH94] T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Pattern Recognition, 1994. Vol. 1 - Conference A : Computer Vision amp ; Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 582 –585 vol.1, oct 1994.
- [OPM02a] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7) :971–987, 2002.
- [OPM02b] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7) :971–987, 2002.
- [oW09] University of Washington. Privacy enhancing technologies symposium, 2009.
- [PD00] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(3) :266–280, 2000.
- [PD02] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3) :223–247, 2002.

- [PD07] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.
- [Pen15] X. Peng. Combine Color and Shape in Real-time Detection of Texture-less Objects. *Comput. Visions and Image Understanding*, 135(C) :31–48, June 2015.
- [PESvG09] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool. You’ll never walk alone : Modeling social behavior for multi-target tracking. In *2009 IEEE 12th International Conference on Computer Vision*, pages 261–268, Sept 2009.
- [Pop10] R. Poppe. A survey on vision-based human action recognition. *Image Vision Comput.*, 28 :976–990, June 2010.
- [PTM06] F. Porikli, O. Tuzel, and P. Meer. Covariance tracking using model update based on lie algebra. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 728–735. IEEE, 2006.
- [PYL08] Y. Pang, Y. Yuan, and X. Li. Gabor-based region covariance matrices for face recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(7) :989–993, 2008.
- [RA10] M. S. Ryoo and J. K. Aggarwal. UT-Interaction Dataset, ICPR contest on Semantic Description of Human Activities (SDHA). http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html, 2010.
- [RGL11] A. Romero, M. Gouiffès, and L. Lacassagne. Feature points tracking adaptive to saturation. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 277–282, nov. 2011.
- [RGL13a] A. Romero, M. Gouiffès, and Lacassagne L. Total Bregman divergence for multiple object tracking. In *IEEE International Conference on Image Processing*, 2013.
- [RGL13b] A. Romero, M. Gouiffès, and L. Lacassagne. Enhanced Local Binary Covariance Matrices ELBCM for texture analysis and object tracking. In *MIRAGE 2013, Berlin, Germany. ACM International Conference Proceedings Series*. Association for Computing Machinery, 2013.
- [RLGHZ14] A. Romero, L. Lacassagne, M. Gouiffès, and A. Hassan Zahraee. Covariance tracking : architecture optimizations for embedded systems. *EURASIP Journal on Advances in Signal Processing*, page 25, December 2014.
- [RLHZG13] A. Romero, L. Lacassagne, A. Hassan Zahraee, and M. Gouiffès. Real-time covariance tracking algorithm for embedded systems. In *Design and Architectures for Signal and Image Processing (DASIP), 2013 Conference on*. IEEE, 2013.
- [RM16] S. K. Ravindran and A. Mittal. CoMaL : Good Features to Match on Object Boundaries. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 336–345, June 2016.
- [RPD10] E. Rosten, R. Porter, and T. Drummond. Faster and better : A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1) :105–119, 2010.
- [SBB⁺07] R. Stiefelhagen, K. Bernardin, R. Bowers, J. Garofolo, D. Mostefa, and P. Soundararajan. The CLEAR 2006 evaluation. *Multimodal Technologies for Perception of Humans*, 2007.
- [SBB15] S. Said, L. Bombrun, and Y. Berthoumieu. Texture classification using Rao’s distance on the space of covariance matrices. In *Geometric Science of Information*, pages 371–378. Springer, 2015.

- [SC12] V. Saligrama and Zhu Chen. Video anomaly detection based on local statistical aggregates. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2112–2119, 2012.
- [SCG⁺15] A. Setkov, F. Martínez Carillo, M. Gouiffès, C. Jacquemin, M. Vanrell, and R. Baldrich. Dacimpro : A novel database of acquired image projections and its application to object recognition. In *Advances in Visual Computing - 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part II*, pages 463–473, 2015.
- [SCS12] S. Salti, A. Cavallaro, and L. Di Stefano. Adaptive appearance modeling for video tracking : Survey and evaluation. *IEEE Transactions on Image Processing*, 21(10) :4334–4348, Oct 2012.
- [Set15] A. Setkov. *Image et Vision par ordinateur pour la réalité augmentée : Invariance couleur et mise en correspondance pour la définition d’un système caméra-vidéoprojecteur*. PhD thesis, University Paris-Saclay, 2015.
- [SG16] A. Setkov and M. Gouiffès. Geometric compensation of dynamic video projections. In *IEEE International Symposium on Multimedia, ISM 2016, San Jose, CA, USA, December 11-13, 2016*, pages 587–590, 2016.
- [SGJ13] A. Setkov, M. Gouiffès, and C. Jacquemin. Color invariant feature matching for image geometric correction. In *Proceedings of the 6th International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications, MIRAGE ’13*, pages 7 :1–7 :8, New York, NY, USA, 2013. ACM.
- [SGJ16] A. Setkov, M. Gouiffès, and C. Jacquemin. Evaluation of color descriptors for projector-camera systems. *Journal of Visual Communication and Image Representation*, 36 :11 – 27, 2016.
- [SGL15] L. Snidaro, J. García, and J. Llinas. Context-based information fusion : A survey and discussion. *Information Fusion*, 25(Supplement C) :16 – 31, 2015.
- [Sha85] S.A. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10(4) :210–218, 1985.
- [SMYS09] S.Rastegar, M.Bandarabadi, Y.Toopchi, and S.Ghoreishi. Kernel based Object Tracking Using Metric Distance Transform and RVM Classifier. In *AJBAS (3)’09*, pages 2778–2790, 2009.
- [SRB10] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2432–2439. IEEE, June 2010.
- [SSHL13] A. Sanin, C. Sanderson, M. T. Harandi, and B. C. Lovell. Spatio-temporal covariance descriptors for action and gesture recognition. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 103–110. IEEE, 2013.
- [ST94] J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR’94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE, 1994.
- [SZ03] J. Sivic and A. Zisserman. Video Google : A text retrieval approach to object matching in videos. In *IEEE International Conference on Computer Vision*, volume 2, pages 1470–1477, 2003.
- [SZT12] D. Song, B. Zhao, and L. Tang. Mean-Shift algorithm fused with corner feature and color feature for target tracking. *Systems Engineering and Electronics*, 34(1) :199–203, 2012.

- [TCSU08] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine Recognition of Human Activities : A Survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 18, 2008.
- [TFD13] F. Tombari, A. Franchi, and L. Di. BOLD Features to Detect Texture-less Objects. In *2013 IEEE International Conference on Computer Vision*, pages 1265–1272, Dec 2013.
- [TFMB04] A. Trémeau, C. Fernandez-Maloigne, and P. Bonton. *Image numérique couleur : de l'acquisition au traitement*. Dunod, 2004.
- [TGL12a] H. Tabia, M. Gouiffès, and L. Lacassagne. Motion histogram quantification for human action recognition. In *IEEE International Conference on Pattern Recognition*, pages 2404–2407, 2012.
- [TGL12b] H. Tabia, M. Gouiffès, and L. Lacassagne. Motion modeling for abnormal event detection in crowd scenes. In *International Symposium on Signal, Images, Video and Communications*, University of Valenciennes, France, 2012.
- [TK91] C. Tomasi and T. Kanade. *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ., 1991.
- [TPM06] O. Tuzel, F. Porikli, and P. Meer. Region covariance : A fast descriptor for detection and classification. In *European Conference on Computer Vision*, pages 589–600. Springer, 2006.
- [TPM08] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10) :1713–1727, 2008.
- [TS92] D. Terzopoulos and R. Szeliski. Tracking with Kalman snakes. *Active vision*, pages 3–20, 1992.
- [TSG⁺14] A. Tavanai, M. Sridhar, F. Gu, A. G. Cohn, and D. C. Hogg. Context Aware Detection and Tracking. In *IEEE International Conference on Pattern Recognition*, pages 2197–2202, Aug 2014.
- [TTL09] J. Y. Tou, Y. H. Tay, and P. Y. Lau. Gabor filters as feature images for covariance matrix on texture classification problem. In *Advances in Neuro-Information Processing*, pages 745–751. Springer, 2009.
- [Vap98] V. N. Vapnik. *Statistical learning theory*. Wiley, September 1998.
- [VBvdWV11] E. Vazquez, R. Baldrich, J. van de Weijer, and M. Vanrell. Describing Reflectances for Color Segmentation Robust to Shadows, Highlights, and Textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5) :917–930, 2011.
- [vdSGS10] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Trans. on PAMI*, 32(9) :1582–1596, 2010.
- [VJ04] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2) :137–154, 2004.
- [vK05] J. A. von Kries. *Handbuch der Physiologie der Menschen, volume 3, chapter III*. Vieweg, Brunswick, 1905.
- [VM11] T. Vojtík and J. Matas. Robustifying the flock of trackers. In *16th Computer Vision Winter Workshop*, pages 91–97. Graz Univ. of Tech., 2011.
- [WKSL11] H. Wang, A. Kläser, C. Schmid, and C. L. Liu. Action recognition by dense trajectories. In *CVPR 2011*, pages 3169–3176, June 2011.

- [WRHS13] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid. DeepFlow : Large displacement optical flow with deep matching. In *International Conference on Computer Vision*, Sydney, Australia, December 2013.
- [WS13] H. Wang and C. Schmid. Action recognition with improved trajectories. In *2013 IEEE International Conference on Computer Vision*, pages 3551–3558, Dec 2013.
- [XJM10] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1293–1300, June 2010.
- [YCHC10] H-H. Yeh, J-Y. Chen, C-R. Huang, and C-S. Chen. An adaptive approach for overlapping people tracking based on foreground silhouettes. In *IEEE International Conference on Image Processing (ICIP)*, pages 3489–3492. IEEE, 2010.
- [YJS06] A. Yilmaz, O. Javed, and M. Shah. Object tracking : A survey. *ACM Computing Surveys (CSUR)*, 38(4) :13, 2006.
- [YL15] J. Yang and H. Li. Dense, accurate optical flow estimation with piecewise parametric model. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1019–1027, June 2015.
- [YO11] J. Yao and J-M. Odobez. Fast human detection from joint appearance and foreground feature subset covariances. *Computer Vision and Image Understanding*, 115(10) :1414–1426, 2011.
- [YPC⁺16] J. Yang, B. Price, S. Cohen, H. Lee, and M. H. Yang. Object Contour Detection with a Fully Convolutional Encoder-Decoder Network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–202, June 2016.
- [YRMFF16] S. Yeung, O. Russakovsky, G. Mori, and L. Fei-Fei. End-to-end learning of action detection from frame glimpses in videos. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2678–2687, June 2016.
- [yTLZG13] A. Romero Mier y Terán, L. Lacassagne, A. H. Zahraee, and M. Gouiffès. Real-time covariance tracking algorithm for embedded systems. In *2013 Conference on Design and Architectures for Signal and Image Processing*, pages 104–111, Oct 2013.
- [YYL11] Y. Cong, J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. In *IEEE conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011, pages 3449–3456, 2011.
- [ZBC13] C. Zhu, C-E. Bichot, and L. Chen. Image Region Description Using Orthogonal Combination of Local Binary Patterns Enhanced with Color Information. *Pattern Recognition*, 46(7) :1949–1963, July 2013.
- [Zen86] S. Di Zenzo. A Note on the Gradient of Multi-Image. *Computer Vision, Graphics, and Image Processing*, 33 :116–125, 1986.
- [ZGL⁺16] T. Zhang, B. Ghanem, S. Liu, C. Xu, and N. Ahuja. Robust visual tracking via exclusive context modeling. *IEEE Transactions on Cybernetics*, 46(1) :51–63, Jan 2016.
- [ZL11] Y. Zhang and S. Li. Gabor-LBP Based Region Covariance Descriptor for Person Re-identification. In *Image and Graphics (ICIG)*, 2011 Sixth International Conference on, pages 368–371, 2011.
- [Zt09] Q. Zhao and H. tao. A motion observable representation using color correlogram and its application to tracking. *Computer Vision and Image Understanding*, 113 :273–290, 2009.

Annexe A

Curriculum Vitae

Née le 20 janvier 1979 à Quimper (29)
michele.gouiffes@limsi.fr
01 69 85 81 13

LIMSI-CNRS
rue John Von Neuman
Campus d'Orsay, bât. 508
91403 ORSAY CEDEX

A.1 Parcours et diplômes

2018 **Habilitation à Diriger des Recherches**(Université Paris-Sud

2006 -... **Maître de Conférences Université Paris-Sud**

- Enseignement au dépt. EES (Électronique, Énergie, Systèmes) de PoPS (Polytech'Paris Sud)
- **Depuis 2013** : Recherche au LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur) au département CHM (Communication Homme-Machine), groupe AMI (Architectures et Modèles pour l'Interaction)
- **2006-13** : Recherche au laboratoire IEF (Institut d'Électronique Fondamentale)

2005-06 **Attachée Temporaire à l'Enseignement et à la Recherche (ATER) à 50%**

- Enseignement au dépt. SFA (Sciences Fondamentales et Appliquées) de l'Université de Poitiers
- Recherche effectuée au laboratoire SIC (Signal Image Communications) désormais XLIM.

2005 **Diplôme de doctorat**, Université de Poitiers, mention Très Honorable.

Apports de la couleur et des modèles de réflexion pour l'extraction et le suivi de primitives visuelles.

Travaux menés au Cemagref de Rennes (maintenant IRSTEA), en collaboration avec le SIC (XLIM) Univ. Poitiers, et le LIGIV (laboratoire H. Curien), Univ. Saint-Etienne.

Composition du jury :

Michel Dhome	(Rapporteur) DR CNRS, LASMEA, Univ. Blaise Pascal
Bernard Péroche	(Rapporteur) PU, Univ. Claude Bernard Lyon 1
Christophe Collewet	(Examinateur) CR Cemagref
Sylvie Philip Foliguet	(Examinatrice) PU ENSEA Cergy Pontoise
Christine Fernandez	(Directrice de thèse)PU Univ. de Poitiers
Alain Trémeau	(Directeur de thèse) PU Univ. de Saint-Etienne
Theo Gevers	(Invité) Professeur, Univ.of Amsterdam

2002 **Diplôme d'Études Approfondies** STIR (Signal, Télécommunications, Image, Radar)
Université de Rennes I, mention Bien.

Diplôme d'ingénieur de l'INSA de Rennes

Département EII (Électronique et Informatique Industrielle), mention Bien.

A.2 Activités de recherche

A.2.1 Publications et communications

La liste complète des publications et communications est fournie dans l'annexe B. La plupart des publications sont disponibles sur ResearchGate.

Journaux internationaux publiés/acceptés	12
Journaux francophones	2
Conférences internationales	37
Conférences francophones	9
Chapitre de livre	1
Rapport de recherche	1
Présentations diverses	5
Base de données	1

A.2.2 Participation significative à des projets de recherche

• Projets de collaboration externe au laboratoire

2017-18 Membre projet Farman CAT (Context-Aware Tracking)

Porteur : Sergio Rodriguez (SATIE, ENS Cachan).

Contributions : Co-encadrement d'un stage de M2 sur la prise en compte du contexte dans le suivi multi-objet.

2017 -18 Porteur projet Carnot Tremplin Cognition LIMSI-IRIT : Analyse automatique de vidéos de communication visuo-gestuelle

Partenaires : Annelies Braffort (LIMSI), Christophe Collet (IRIT, Toulouse)

Contributions : Co-encadrement d'un stage de M2 sur l'analyse des primitives visuelles pour la caractérisation de la langue des signes.

2016 -20 Membre projet ANR ETAE (Écoulement Tournant et Actionneur Électroactifs)

Partenaires : Laurent Martin-Witowski (porteur, LIMSI), Yves Bernard (Geeps).

L'objectif est, à partir d'excitations induites par des actionneurs électro-actifs et parfaitement contrôlées en fréquence et en amplitude, de manipuler l'écoulement afin de mieux comprendre les mécanismes qui, en présence de bruit, conduisent aux motifs d'instabilité à grande échelle.

Contributions : calcul de déformation de la surface libre, en temps-réel à l'aide d'une caméra et de lumière structurée (profilométrie par transformée de Fourier). À partir de cette carte de profondeur, des caractéristiques seront définies et extraites pour mesurer le niveau de turbulence et transmettre ces données au système de contrôle du disque tournant. Cette proposition fait suite au projet d'AI *Vision par ordinateur pour la profilométrie de surfaces dynamiques* obtenu en 2013-2014.

2015- 18 Membre du projet industriel 3D Reno

Porteur : Mehdi Ammi (groupe AMI, LIMSI).

Contributions : co-encadrement de la thèse d'Adrien Arnaud sur la reconstruction 3D sur tablette pour l'évaluation des performances énergétiques des bâtiments.

-
- 2014-16** **Porteur du projet Digiteo MAPOCA bourse post-doctorale (Modélisation des Activités par Processus pOnctuels et Covariance de cAractéristiques)**
Partenaire : Antoine Manzanera (ENSTA ParisTech).
Contributions : co-encadrement d'un stage post-doctoral Fabio Martinez, 18 mois sur la reconnaissance d'activités et de gestes.
- 2011-12** **Membre du projet d'action incitative LIMSI-IEF : Recherche de surfaces de projection optimales et compensation de la distorsion géométrique pour la réalité augmentée**
Partenaires : Franck Bimbard, Samia Bouchafa, Emmanuelle Frenoux, Michèle Gouiffès, Christian Jacquemin
Cette collaboration a permis de définir de nouveaux axes de recherche communs en vue d'une intégration au sein du LIMSI (réalisée en 2013).
- 2011-13** **Membre du projet Européen ITEA2 SPY (Surveillance imProved sYstem)**
Porteur : Cassidian (Airbus Defence and Space). Partenaires industriels : Cassidian, Arc Informatique, Aselsan (Turquie), Eolane Les Ulis, CogVis (Autriche), C²Tech, KiloSoft, PPSL, Roger-GPS (Finlande). Laboratoires : IEF, ENSTA-Paritech, Telecom Sud Paris, VTT (Finlande).
Contributions : coordinatrice locale (au sein de l'IEF), gestion du budget (environ 300 K€), recrutement, encadrement, développement d'algorithmes d'analyse vidéo, rédaction des rapports d'avancement et des livrables, participation aux réunions.
- 2006-09** **Membre du projet systém@tic LOVe (Localisation et Observation de Vulnérables)**
Porteur : Laurent Trassoudaine (LASMEA, Clermont-Ferrand)
Contributions : coordinatrice locale (au sein de l'IEF) du sous-projet 2 (traitement d'images) : recrutement et encadrement de 2 ingénieurs (Laurent Schenider en 2008 puis Antoine Patri en 2009), développement d'algorithmes de détection d'obstacles par stéréovision et de suivi temporel de piétons, rédaction des livrables.

• **Projets de collaboration interne au LIMSI : les actions incitatives (AI)**

Les projets d'actions incitatives visent à initier de nouvelles collaborations de recherche entre deux groupes ou deux thèmes différents du LIMSI.

2015-16 Descripteurs de mouvement pour l'annotation semi-automatique de vidéos de langue des signes

Collaboration avec Annelies Braffort. Co-encadrement d'un stage de M2.

Contributions : étude bibliographique, utilisation des matrices de covariance de primitives pour segmenter les signes en signes lexicaux et signes à visée illustrative.

2013-14 Fusion multi-sources pour la localisation d'objets dans l'ambient (iRoom)

Collaboration avec Yacine Bellik. Co-encadrement de 2 stages de M1.

Contributions : fusion des données issues d'une caméra avec celles fournies par des émetteurs radio-fréquence. Mise en place d'un algorithme de suivi d'objets utilisant l'image et la profondeur, fusion probabiliste et expériences *in-situ* dans la salle d'interaction ambiante iRoom.

2013-14 Vision par ordinateur pour la profilométrie de surfaces dynamiques

Collaboration avec Laurent Martin-Witowski. Co-encadrement d'un stage de M2.

Contributions : mise en place d'un algorithme de profilométrie par transformée de Fourier. Des motifs sinusoïdaux sont projetés sur la surface. L'image de ces motifs déformés permet de déduire la déformation engendrée.

• **Projets art-science soutenus financièrement par la Diagonale Paris-Saclay**

2017 Porteur du projet *Vera Icona (Les Invisibles)*

Partenaires : Véronique Caye (Laboratoire Victor Vérité) et Panagiotis-Alexandros Bokaris. Bourse Coup de Pouce obtenue en mai 2017.

Contributions : encadrement d'un stage. Acquisition d'images, détection et transformations de visages.

2017 Membre du projet Phare *L'œil de Mars*

Partenaires : Mathieu Vincendon, (IAS, CNRS, Université-Paris-Sud); Ikse Maître, Le sas, La métonymie.

Contributions : recalage de 16 caméras IP, détection de mouvement.

2015-16 Membre du projet *Pixel's Pixels*

Partenaires : Christian Jacquemin (LIMSI) et Bertrand Planes (artiste).

Contributions : encadrement d'un ingénieur, fusion multi-image et simulation d'un zoom infini à l'aide d'ondelettes.

2015-16 Porteur du projet *Gardien du Temple*

Partenaires : Véronique Caye (Laboratoire Victor Vérité), Christian Jacquemin (LIMSI) et Panagiotis-Alexandros Bokaris, Jean-Marc Chomaz (LadHyX, école Polytechnique).

Contributions : supervision technique : encadrement de deux stages de développement informatique, mise en place de l'installation. Installation au festival CURIOSITAS 2015.

• **Autres participations à des montages de projets**

Depuis 2010, j'ai participé au montage de plusieurs projets qui n'ont pas été retenus. En voici les principaux.

-
- 2017** Porteur ANR FlyBIM. Génération de maquettes numériques d'intérieurs de bâtiments par acquisition automatique de données 3D avec drone autonome : application à la rénovation. Partenaires : LIMSI et entreprise RPE.
- 2016** Membre ANR Délim 3.0. Degrés de liberté en imagerie médicale 3.0. Le cas de l'élastographie par résonance magnétique. Partenaires : IR4M (porteur X. Maître), U1149-CRI, IRPHIL, CAPHRI, Lab-STICC.
- 2013,14** Membre ANR ContexTrack. Commutation de Modalités et de Modèles pour le Suivi visuel. Partenaires : ISIR (UPMC) et IBISC (UEVE).
Membre ANR INESM. Interfaces numériques en situation de mobilité : dispositifs cognitifs, artistiques, ludiques et fictionnels. Partenaires : ACTE (Panthéon-Sorbonne), EHESS, ESADHAR, L2C2, XiLabs Company.
- 2014** Membre PACOME (DGA RAPID 2014) Système de veille PANoramique COmpacte Multi-Eléments. Partenaires : entreprise HGH et le LRI (UPSUD).
- 2012-14** Porteur ANR Beamshare, collaboration bilatérale Franco-Allemande (DFG-ANR). *Mixed reality information space for interactive social experience sharing*. Partenaires : HU (Humboldt Universität zu Berlin), TUC (Clausthal University of Technology), IBISC (UEVE).
- 2011** ANR Virtual Palimpsest. Partenaires : LIMSI (porteur C. Jacquemin), MAP (AMU), LIG (UGA), IEF (UPSUD), entreprise Immersion.
- 2010** ANR MetaScale *Meta-programming for the Petascale*. Partenaires : LRI (porteur J. Falcou), IEF.

A.2.3 Encadrement

Thèses en cours	4	T-K. Mai (depuis 2015), A. Arnaud (depuis 2016), H. Chaaban et V. Belissen (début en octobre 2017)
Thèses soutenues	5	N. Lertchuwongsa (2011), F. Laguzet (2013), A. Romero (2013), A. Setkov (2015), P-A. Bokaris (2016)
Post-doctorants	6	F. Martinez Carillo (18 mois en 2015-16), A. Chamseddine Ben Abdallah (11 mois en 2013), A. Hassan Zahraee (11 mois en 2013), Y. Almehio (10 mois en 2012), H. Tabia (10 mois en 2012), H-V Lee (1 an en 2007)
Stages M2	6	K. Benseghir (2016), E. Guerci (2014), J. Gao (2014), H. Itoudj (2012), S. Benma (2012), F. Ouchène (2009)
Autres	17	14 stages M1 et 3 stages de Bac+2 (cycle préparatoire)

• Encadrement doctoral

- 2017-...** **Valentin Belissen.** *Analyse de vidéos de communication gestuelle. Application à la Langue des Signes Française.* Encadrement à 50 % avec A. Braffort (LIMSI). Contrat doctoral ED STIC Univ. Paris Saclay.
- 2017-...** **Hussein Chaaban.** *Reconnaissance d'activités humaines dans les vidéos.* Encadrement à 80 % avec A. Braffort (LIMSI). Contrat doctoral ED STIC Univ. Paris Saclay.
- 2016-...** **Adrien Arnaud.** *Reconstruction et édition de scènes 3D sur dispositifs mobiles.* Encadrement à 60 % avec M. Ammi (LIMSI). Contrat CIFRE avec RPE (Rénovation Plaisir Énergie).
- 2015- ...** **Tan Khoa Mai.** *Vers un système de vision opportuniste pour l'analyse de scènes complexes.* Encadrement à 50 % avec S. Bouchafa (IBISC UEVE). Contrat doctoral ED STIC Univ. Paris Saclay.
- 2013-16** **Panagiotis-Alexandros Bokaris.** *Réalité augmentée vidéo-projetée : Compensation photométrique pour l'effacement statique et dynamique.* Encadrement à 50% avec C. Jacquemin (LIMSI). Contrat doctoral ED STIC Univ. Paris Saclay.
Situation actuelle : Ingénieur R&D chez L'Oréal France depuis 2017.
- 2012-15** **Aleksandr Setkov.** *Image et Vision par ordinateur pour la réalité augmentée : Invariance couleur et mise en correspondance pour la définition d'un système caméra-vidéoprojecteur.* Encadrement à 50% avec C. Jacquemin (LIMSI). Contrat doctoral ED STIC Université Paris Saclay. Participation au programme EIT ICT Labs.
Situation actuelle : CDD Ingénieur R&D chez CEA LIST depuis septembre 2016.
- 2010-13** **Andres Romero Mier y Teran.** *Real time multi-target tracking : a study on color texture covariance matrices and descriptor/operator switching.* Encadrement à 50% avec L. Lacassagne (UPMC LIP6). Bourse Conacyt du Mexique.
Situation actuelle : Ingénieur R&D chez Apple, San José, USA depuis 2015.
- 2009-13** **Florence Laguzet.** *Contribution en suivi d'objets couleur. Parallélisation des algorithmes pour systèmes embarqués.* Encadrement à 40%, avec L. Lacassagne(UPMC LIP6) à 40% et D. Etiemble (LRI) à 20%. Contrat doctoral EDIPS UPSud.
Situation actuelle : ingénieur en vision par ordinateur chez ASL à Brighton (UK) depuis 2014.
- 2008-11** **Noppon Lertchuwongsa.** *Mise en correspondances de primitives couleur robustes pour la stéréovision.* Encadrement à 50 % avec B. Zavidovique (IEF). Bourse du gouvernement thaïlandais.
Situation actuelle : enseignant-chercheur à la Prince of Songkla University, Thaïlande depuis 2012.

- **Encadrement post-doctoral**

- 2014-16** **Fabio Martinez Carrillo.** *Reconnaissance d'activités et reconnaissance de la langue signée.*
Durée 18 mois, projet Digiteo MAPOCA, encadrement à 50% avec A. Manzanera.
Situation actuelle : enseignant-chercheur en Colombie, Escuela de Ingeniería de Sistemas e Informática, Universidad Industrial de Santander.
- 2013** **Ahmed Chamseddine Ben Abdallah.** *Détection d'anomalies dans les vidéos.*
Durée 11 mois, projet SPY, encadrement à 80% avec L. Lacassagne.
Situation actuelle : Ingénieur en métrologie dans l'entreprise ASML, Pays-Bas (depuis 2014).
- 2013** **Ali Hassan Zahraee** *Adéquation algorithme-architecture : parallélisation d'algorithmes de traitement d'image et intégration sur la plateforme EOLCAM de l'entreprise Eolane-Les-Ulis.*
Durée 11 mois, projet SPY, encadrement à 20% avec L. Lacassagne.
Situation actuelle : chercheur en robotique à l'Université de Sheffield.
- 2012-13** **Yasser Almechio** *Mise en correspondance de lignes de niveaux.*
Durée 10 mois, projet SPY, encadrement à 50% avec B. Zavidovique.
Situation actuelle : Ingénieur R&D chez Thales.
- 2012** **Hedi Tabia** *Détection d'anomalies dans les vidéos. Reconnaissance d'activités.*
Durée 10 mois, projet SPY, encadrement à 80% avec L. Lacassagne.
Situation actuelle : Maître de Conférences à l'ENSEA/ETIS, Cergy-Pontoise.
- 2007** **Ha-Vu Le** *Estimation de l'amplitude et la direction de vibrations sinusoïdales appliquées sur les composants MEMS.*
Durée 1 an, projet Nanores, encadrement à 40%.

- **Étudiants de niveau M2**

- 2017** **Hussein Chaaban** *Analyse automatique de vidéos de communication visuo-gestuelle. Application à la langue des signes française.*
M2R SETI (UPSUD). Encadrement à 80% avec A. Braffort (LIMSI).
- 2016** **Kenza Benseghir** *Descripteurs de mouvement pour l'annotation semi-automatique de vidéos de langue des signes*
M2R SETI (UPSUD). Encadrement à 80% avec A. Braffort (LIMSI).
- 2014** **Erika Guerci** *Vision par ordinateur pour la profilométrie de surfaces dynamiques*
M2 Computer engineering, Computer science and Multimedia, Univ. de Pavie, Italie, 2014. Encadrement à 80% avec L. Martin-Witkowski (LIMSI, dpt MECA).
- 2014** **Jie Gao** *Analyse d'images de cristallographie grâce à l'intelligence artificielle, pour l'aide au diagnostic*
M2R ATSI (UPSUD). Encadrement à 20% avec E. Frenoux (LIMSI).
- 2012** **Hamid Itoudj** *Automatic detection of planar surfaces for an ideal videoprojection indoor or outdoor. Compensation of the geometrical distortions for augmented reality on multi-planar surfaces.*
M2R Systèmes Avancés et Robotique, Université UPMC, 2012. Encadrement à 40% avec S. Bouchafa de l'UEVE/IBISC.
- Sondes Benma** *Mise en œuvre d'un système de reconnaissance adaptatif pour la définition d'un système d'assistance aux forces de sécurité publique.*
M2R Systèmes Avancés et Robotique, Université UPMC. Encadrement à 80% avec Andres Romero Mier y Téran (20%).
- 2009** **Faiza Ouchène** *Suivi d'objets par Mean-Shift*
M2R SETI. Encadrement à 100 %.

• Etudiants ingénieur Bac +4

La plupart des étudiant-e-s que j'ai encadré-e-s viennent de la formation EES4 de Polytech'Paris Sud. J'ai supervisé Jason Lhies, un étudiant de Supélec Metz. Dans le cadre d'un partenariat d'échange entre la BME Budapest et Polytech'Paris Sud, j'ai accueilli 5 étudiants Hongrois de niveau M1 pendant 3 mois. Enfin, j'ai accueilli 3 étudiants indiens de l'Indian Institute of Technology (IIT) de niveau M1 pendant 3 mois.

2017	Toky Rakotomalala <i>Développement logiciel pour le projet art-science Vera Iona.</i>	EES4
2015	Kevin Heems et Xi Wang <i>Développement logiciel pour le projet art-science Gardien du Temple : interface graphique et réalité augmentée.</i>	EES4
2014	Jie GAO <i>Localisation et suivi d'objets dans l'ambient par traitement d'image.</i>	EES4
	Jason Lhies <i>Fusion de données image et RF pour la localisation d'objets dans l'ambient.</i>	Supélec
	Peter Neukirchner <i>Localization of Pololu robots by image processing.</i>	BME
2013	Balazs Opra <i>Automatic resistors recognition. Implementation on Android Platform.</i>	BME
	Tamas Garacszi <i>Object tracking and recognition in videos. A comparison between SURF and Covariance matching.</i>	BME
2012	Barnabas Szirmay-Kalos <i>Using Scale Invariant Image Feature Points for homography detection between images</i>	BME
	Gábor Weszelovszki <i>SIFT keypoints for compensating projections</i>	BME
2011	Manas Chaudhari et Siddharth Sarangdhar <i>Color covariance tracking and color space switching</i>	IIT
	Krishna Kishore <i>Structure from motion</i>	IIT
2008	Xuefeng Peng <i>Segmentation d'images par Watershed appliquée à la détection de piétons.</i>	EES4

A.2.4 Participation à des comités

• Examinatrice à des jurys de thèse

2015	Julien Gosseaume. <i>Développement d'un algorithme de génération de trame de camouflage optique et infrarouge.</i> Encadrement : K. Kpalma, J. Ronsin, IETR, INSA Rennes.
2014	Weizhi Lu. <i>Dimension reduction techniques and related applications.</i> Encadrement : K. Kpalma, IETR, INSA Rennes.
	Hasan C. Faridul. <i>Analysis of 3D color matches for the creation and consumption of video content.</i> Encadrement : A. Trémeau, Univ. de Saint-Etienne, J. Stauder, Technicolor R&D, France.
2011	Xiaohu Song. <i>Descripteurs couleur locaux invariants aux conditions d'acquisition.</i> Encadrement : D. Muselet et A. Trémeau, Univ. de Saint-Etienne.

• Participation à des comités de suivi de thèse

2017	Florian Lemaître UPMC-LIP6 et CERN
	Shuo Zheng. Telecom PariTech et L2S-CentraleSupélec, UPSay.

• Comités de relecture

Relectures régulières pour des journaux ($\simeq 10$ relectures par an).

IEEE Signal Processing Letters , IEEE Trans. on Image Processing, Integrated Computer-Aided Engineering (ICAE), International Journal of Neural Systems (IJNS), Journal of Visual Communication and Image Representation (JVCI), Journal of Real-Time Image Processing (JRTIP), Computers & Electrical Engineering, IET Computer Vision, IET Image Processing, Leonardo MIT Press, Journal on Multimodal User Interfaces (JMUI), Electronic Letters, IEEE Transactions on Robotics (T-RO), International Journal of Reconfigurable Computing, Ain Shams Engineering Journal.

Relecture pour les conférences

Conférence Interaction Homme-Machine (IHM 2017), Int. Conf. on Intelligent Human Computer Interaction (IHCI 2017), IEEE Int. Conf. on Intelligent Transportation Systems (ITSC depuis 2013), Int. Joint Conference on Computational Intelligence (IJCCI depuis 2011), IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2016), IEEE Int. Conf. on Automation Science and Engineering (CASE 2012), IEEE Int. Conf. on Pattern Recognition (ICPR 2012), IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man 2012), IEEE Int. Conf. on Networking, Sensing and Control (ICNSC 2013), DASIP (depuis 2013), Int. Conf. on Image Processing Theory, Tools and Applications (IPTA, depuis 2014), Colloque GRETSI 2013, IEEE Int. Conf. on Networking, Sensing and Control (ICNSC), Int. Conf. on Telecommunications (ICT 2012), IEEE Int. Conf. on Robotics and Automation (ICRA 2011)

Participation à l'évaluation de projets

Projets exploratoires premier soutien, Université de Lille / CNRS 2016; expertise régulière de projets Art-Science et Sciences et Société de la Diagonale Paris-Saclay depuis 2013 ; projets émergents du département STIC Pars-Saclay (2017).

A.3 Responsabilités administratives

A.3.1 Responsabilités pédagogiques

- 2016 - ...** Co-responsable des stages en EES4 PoPS (30 étudiants) : validation des sujets, suivi, lecture rapports, planning, soutenances.
- 2015 - ...** Coordinatrice spécialité EES4 PoPS (30 étudiants) : réflexions pédagogiques, mise en place de l'emploi du temps, coordination, dialogue avec les étudiants.
- 2011 -15** Responsable des stages spécialité EES5 PoPS.

A.3.2 Participation aux instances locales

- 2014 - ...** Membre CCSU sections 60-61-62 UPSud
Membre CCSU section 27 UPSud
- 2010-14** Vice-Président B CCSU section 60-61-62 UPSud.
Membre de la commission recherche du Département de Physique de l'UPSud
- 2007-10** Membre élu au Conseil de Perfectionnement de PoPS

A.3.3 Participation à des comités de recrutement

Depuis 2008, participation à :

- 14 comités de recrutement de MCF (à l'UFR des Sciences UPSud ou à PoPS, à l'IUT de Cachan, à l'INSA de Rennes, l'Université de Lille, à l'UVSQ et un échange UPMC/Polytech'Marseille),

- 2 comités de recrutement de professeurs agrégés (IUT de Cachan),
- 1 comité de recrutement d'IE CNRS au Laboratoire Hubert Curien (Univ. Saint-Etienne).

A.4 Activités d'enseignement

J'ai été recrutée en 2006 à Polytech'Paris Sud sur un poste dont le profil d'enseignement couvre l'informatique industrielle et le traitement d'images et des signaux. Après deux années où j'ai pu bénéficier d'une décharge de 42 heures par an, mon service affiche de 200 à 260h équivalent TD par an depuis 2008, qui se répartissent en cours magistraux, TP, TD, proposition et encadrement de projets, suivi de stages et décharges liées aux responsabilités (voir section A.3). J'interviens devant des publics variés, comme le montre le récapitulatif des filières dans le tableau A.1. En effet, j'enseigne à différents niveaux de Polytech'Paris Sud, allant de la classe préparatoire à la 5^e année principalement dans le département EES de Polytech'Paris Sud, en apprentissage et en filière classique. Par ailleurs j'interviens ou je suis intervenue ponctuellement dans d'autres formations. Les tableaux A.2, A.3 et A.4 font une synthèse des différents cours magistraux, travaux dirigés et travaux pratiques.

TABLE A.1 – Filières dans lesquelles je suis intervenue.

Acronyme	Intitulé de la filière
PeiP2	2 ^e année du cycle préparatoire Parcours Elève Ingénieur Polytech
IUT	2 ^e année IUT de Cachan, département GEII (Génie Électronique et Informatique Industrielle)
APP3	4 ^e année filière par apprentissage spécialités EES et PSO, Polytech'Paris-Sud
APP4	4 ^e année filière par apprentissage spécialités EES et PSO, Polytech'Paris-Sud
EES4	4 ^e année filière étudiante spécialité EES, Polytech'Paris-Sud
EES5	5 ^e année filière étudiante spécialité EES, Polytech'Paris-Sud
M1 Ergo	Master 1 Ergonomie, Université Paris-Sud, UFR STAPS
M2 SETI	Master 2 Recherche Systèmes Embarqués et Traitement de l'Information, Université Paris-Sud, UFR Sciences
PoT	Polytech'Tours
INSA	Institut National des Sciences Appliquées de Rennes, département Génie Électronique et Informatique Industrielle, département GMA (Génie Mécanique et Automatique)

A.4.1 Postes occupés

2006 – .. Maître de Conférences UPSud affectée à PoPS, dpt EES

Service : 230 et 260h eq. TD/an répartis entre CM ou C-TD (27% d'h. eq. TD), TP ou projets (51%), suivi de stages, responsabilités (voir section A.3) et recrutement (22%), en traitement des images et des signaux, informatique et systèmes embarqués

Composantes : PoPS (Polytech'Paris Sud), IUT de Cachan, UFR Sciences (2007-2013)

Publics : DUT2, Peip2 (cycle préparatoire PoPS 2^eme année), EES4 et EES5 (PoPS), APP3 et APP4 (apprentissage PoPS, dpt. EES et PSO). Ponctuellement : M1 IST (Information, Systèmes et Technologie) et M2 SETI (Systèmes Embarqués et Traitement de l'Information), M1 Ergonomie.

2005-06 ATER, Univ. Poitiers

81 h eq. TD en Programmation Langage C, Transmissions numériques, Réseaux Industriels.

2002-05 Vacataire à l'INSA de Rennes

111.5h eq. TD en Traitement du Signal, Logique Séquentielle, Programmation Langage C, Systèmes d'exploitation, Électronique, Traitement d'images couleur.

A.4.2 Principaux enseignements dispensés depuis 2006

Informatique et systèmes embarqués ($\simeq 30\%$ du service en 2016)

- Depuis 2015 : responsable d'un cours sur les systèmes embarqués (*Info Elec*) en Peip2 : notions d'architecture des microprocesseurs, programmation. Mise en place de 7h de cours et 16h de TP de programmation sur la plateforme *mbed*. Le défi de cet enseignement est d'intéresser à la fois des étudiants qui se destinent au département EES ou Informatique mais aussi des étudiants qui n'ont pas d'attrance initiale pour ces matières. Dans l'objectif de proposer des TPs à la fois formateurs et ludiques, les étudiants doivent manipuler les API de la carte pour programmer entre autre un chenillard à partir de leds et de boutons poussoirs, un instrument de musique contrôlé par l'accéléromètre, un jeu (soit Snake, Tetris) contrôlé par joystick.
- Depuis 2010 : Informatique industrielle en EES5 (7,5 h cours bus de communication, 24h TP C++ sur robots Pololus).
- 2008-14 : Architecture des ordinateurs. Cours de logique combinatoire et séquentielle en Peip2.
- 2007-14 : Informatique industrielle. Architecture des Microcontrôleurs en EES4.
- 2007-08 : Remise à niveau en C en ET3.

Traitement d'images et de vidéos ($\simeq 24\%$ du service en 2016)

- Prévu en 2018 : cours et TPs à l'ENS Cachan pour la préparation au concours d'agrégation
- Depuis 2012 : responsable de l'option vidéo en EES5. Mise en place de 16h de cours-TP (détection de mouvement, flot optique, suivi d'objets, stéréovision) en C++ et openCV
- Depuis 2008 : responsable d'un cours-TP d'initiation en DUT2 (4h CM et 4h TP)
- Depuis 2007 : cours, TPs et projets en traitement d'image en EES4 et EES5 (6h CM et 32h TP)
- 2007-10 : remplacement de congés (maternité ou maladie) dans le M2 SETI : cours et TP. Logiciels Khoros, Visiquest, Matlab, C++.

Projets tutorés en EES4 et EES5 ($\simeq 12\%$ du service en 2016)

J'ai proposé et encadré plusieurs projets tutorés en EES5 et EES4. Les étudiants ont environ 100 h de travail à réaliser par groupes de 3 à 6 étudiants. Les trois principaux projets sont décrits ici.

- **Virtual Drums** (2012, 2013, 2014, 2017). Jeu de batterie virtuelle (12 étudiants) : capteur Kinect, C++, OpenGL, openAL, openCV.

L'objectif du projet est de développer un jeu de batterie virtuelle utilisant un capteur Kinect. La batterie comporte 4 toms, chacun d'entre eux est symbolisé à l'écran par quatre rectangles de couleurs différentes. Lorsque le joueur frappe l'un des rectangles, le jeu envoie un son associé à cet objet et signale à l'écran le tom effectivement actionné.

- **Shazam** (2014). Reconnaissance de morceaux musicaux (4 étudiants) C++, multithreading, interface homme-machine, Traitement des signaux.

L'objectif de ce projet est de réaliser un programme de reconnaissance de musique à l'aide de leur représentation fréquentielle et temporelle (spectrogramme, Short Time Fourier Transform). Des points clés sont extraits dans celui-ci et comparés avec ceux de la base de données.

- **PololuMaps** (2014, 2016). Localisation de robots par traitement d'images (7 étudiants). C++, robots m3pi Pololu, traitement d'image, communication sans fil ZigBee.

Les robots m3pi Pololu ne disposent pas de capteurs d'odométrie permettant de leur fournir des informations de localisation. Ils ne disposent que de capteurs au sol permettant de détecter et suivre une ligne. L'objectif du projet PololuMaps est de localiser les robots à l'aide d'une caméra, de les cartographier afin de pouvoir les contrôler dynamiquement et permettre des interactions entre ces robots : suivi de robots, chorégraphie, évitement.

TABLE A.2 – Synthèse des cours magistraux (heures équivalent TD par an).

Filière	Intitulé du module	h.eq.TD	Année(s)
PeiP 2	Architecture des ordinateurs. Cours de logique combinatoire et séquentielle	6	2008-2014
Peip 2	Électronique Informatique. Cours systèmes embarqués	12	depuis 2015
IUT	Traitement d'images	6	depuis 2008
APP3	Traitement numérique du signal	24	depuis 2007
APP4	Traitement numérique du signal	36	2007-2014
EES4	Informatique industrielle. Cours sur les microcontrôleurs	11,25	2007-2014
EES4	Traitement d'image	8	2006-2012
EES5	Informatique industrielle. Cours sur les bus de communication	10	depuis 2011
EES5	Traitement d'image	9	depuis 2006
EES5	Traitement des vidéos	12	depuis 2006
M1 Ergo	Éclairagisme et colorimétrie	36	2013
M2 SETI	Traitement d'image	4-8	2007-2011
PoT	Vision par ordinateur	4	2005

TABLE A.3 – Synthèse des Travaux dirigés (heures équivalent TD par an).

Filière	Intitulé du module	h.eq.TD	Année(s)
PeiP 2	Architecture des ordinateurs. TD de logique combinatoire et séquentielle	4,5	2008-2014
ET3	Remise à niveau en langage C	12	2007-2010

Traitement numérique du signal ($\approx 10\%$ du service en 2016)

- **Responsable des modules de TNS** en APP3 et APP4 (2007-16).

Échantillonnage, transformée de Fourier, quantification, systèmes numériques, transformée en Z, filtrage numérique, corrélation.

Mise en place de 24h de cours -TD , rédaction d'un polycopié pour chaque année.

Mise en place de 16h de TP : une première séance se fait sous Matlab et permet d'illustrer les concepts vus en cours. En APP4, j'ai mis en place avec Marc Bachelier un TP de 4 h sur DSP avec Simulink dont l'objectif était d'illustrer le filtrage numérique en comparant la qualité de filtrage passe bande d'un RIF et d'un RII. Les étudiants devaient extraire les signaux vocaux dans un fichier sonore extrêmement bruité.

A.4.3 Participation à la vie collective de Polytech'Paris Sud

- Participation aux Assises de réseau Polytech en avril 2014
- Déplacement à Budapest avec les Relations Internationales en 2012
- Mise en place en 2009 d'une page internet sur le département (appelé à l'époque ESE)
- Participation régulière aux journées portes ouvertes de l'école et aux entretiens de recrutement post-bac ou Bac+3, en formation initiale et par apprentissage.

TABLE A.4 – Synthèse des Travaux pratiques (heures équivalent TD par an).

Filière	Intitulé du module	h.eq.TD	Année(s)
PeiP 1	Véhicule du Futur	8	2007
PeiP 2	Architecture des ordinateurs. TP de logique combinatoire et séquentielle. TP d'architecture des ordinateurs.	36	2006-2014
PeiP 2	Électronique Informatique. TP de programmation sur systèmes embarqués.	36	depuis 2015
IUT GEIII1	Traitement d'images	4	depuis 2008
ET3	Remise à niveau en langage C	4.5	2007-2010
APP3	Traitement numérique du signal	12	depuis 2006
APP4	Traitement numérique du signal	16	2006-2014
EES4	Informatique industrielle	16	2006-2014
EES4	Traitement d'image	≈ 12	depuis 2006
EES5	Informatique industrielle	de 20 à 32	depuis 2011
EES5	Traitement d'image	≈ 14	depuis 2006
EES5	Traitement des vidéos	≈ 8	depuis 2006
M1 Ergo	Éclairagisme et colorimétrie	36	2013
M1 IST	Traitement d'images	24	2007
M2 SETI	Traitement d'image	32	2006
INSA	Électronique analogique		2002-2005
INSA	Langage C		2002-2005

A.5 Diffusion de la culture scientifique et technique

Mon activité de diffusion de la culture scientifique passe principalement par la participation à des installations art-sciences dans le cadre de l'axe VIDA au LIMSI et au sein du sas (Sciences Arts Sociétés), localisé à l'Université Paris-Sud et géré par l'association La Métonymie et Xavier Maître (CR CNRS IR4M). Mes contributions passent par la mise en place de l'interaction et du traitement d'images sur les installations, par l'aide à leur réalisation technique et l'évaluation par les expériences. Par ailleurs, j'ai participé à l'encadrement de plusieurs missions doctorales pour la mise en œuvre des installations et la médiation scientifique : Alexandra Roatis, Alexandr Setkov, Florian Folcone. Voici les principales réalisations dans lesquelles je suis impliquée :

L'œil de Mars (Astrophysique et interaction homme-machine)

Partenaires : X. Maître et T. Boucneau (IR4M), M. Courgeon (Lab-STICC), F. Poulet (IAS)

Contributions : Recalage d'images et détection de présence

Expositions : IAS, Fête de la Science (Orsay, 2016), Nuit Blanche (Paris, 7 octobre 2017)

Vois-moi à Travers Toi (Astrophysique et interaction homme-machine)

Partenaires : M. Courgeon, T. Giraud, et X. Maître.

Contributions : Suivi temporel des trajectoires des visiteurs pour l'interaction par capteur RGB + D

Expositions : L'Odyssée de la Lumière, Cité des Sciences et de l'Industrie (Paris, Janvier-Décembre 2015)

La lumière ne s'arrête pas là (Astrophysique, relativité espace-temps, Interaction homme-machine, Réalité augmentée)

Partenaires : La Métonymie, Holorapt et des Vues de l'Esprit, de X. Maître, l'artiste C. Vachez, F. Couchot (directeur de recherche au LAL).

Contributions : mise en place de l'interaction par analyse de l'image de profondeur (capteur RGB+D)

Expositions : Journée du développement des logiciels JDEV, à l'École Polytechnique en 2013, anniversaire des 100 ans de la naissance de Nicolas Schöffer, précurseur de l'art cybernétique, dans la cour de la Villa des Arts (Paris, XVIIe) en 2013.

High Time Range Imaging (Photographie et traitement d'images)

Partenaires : B. Planes et C. Jacquemin

Contributions : algorithme de traitement de séquences d'images d'une même scène, saillance visuelle

Exposition : *Printemps de la culture*, Univ. Paris-Sud en mai 2012

Autoportrait, Intimité de l'être, Turbulent (Interaction et imagerie IRM)

Partenaires : des Vues de l'Esprit, X. Maître, T. Giraud, M. Courgeon

Contributions : Mise en place de l'interaction. Encadrement de missions doctorales (A. Roatis, A. Setkov, P-A. Bokaris) sur la mise en œuvre des installations.

Expositions : les journées arts-sciences au *Printemps de la culture* en mai 2012 ; *La vie en transparence*, exposition CNRS au musée des Arts et Métiers (janvier 2012-janvier 2013) ; *La science de l'art* en Essonne en novembre 2011 ; Musée Scientifique du Pass à Mons en Belgique depuis 2013 ; Inauguration du Google Cultural Institute, Paris, Décembre 2013 ; Festival South by South West SXSW à Austin (Texas) en 2014

Gardiens du Temple (Réalité augmentée, traitement d'image, spectacle vivant)

Partenaires : avec V. Caye, C. Jacquemin, P-A. Bokaris, J-M. Chomaz

Contributions : Supervision technique, encadrement de stages, mise en place de l'installation

Expositions : festival CURIOSITAS 2015

Annexe B

Liste de publications

Articles de journaux internationaux publiés ou acceptés pour publication	12
Article soumis dans un journal international	1
Articles parus dans des journaux francophones	2
Conférences internationales	37
Conférences francophones	9
Chapitre de livre	1
Rapport de recherche	1

ARTICLE SOUMIS DANS UN JOURNAUX INTERNATIONAUX

- AIA** **3D Reconstruction and Mock-Up Generation : A Mobile Platform for Indoor Building Renovation**
Adrien Arnaud, Julien Christophe, **Michèle Gouiffès** et Mehdi Ammi
soumis en septembre 2017 à Elsevier Automation in Construction

ARTICLES DE JOURNAUX INTERNATIONAUX ACCEPTÉS POUR PUBLICATION

- Leonardo** **Gardien du Temple : concealment and revelation. An interactive installation involving poetry, performance and spatial augmented reality**
Panagiotis-Alexandros Bokaris, **Michèle Gouiffès**, Véronique Caye, Christian Jacquemin, Jean-Marc Chomaz
MIT Press Journals, accepté en avril 2017

ARTICLES PUBLIÉS DANS DES JOURNAUX INTERNATIONAUX

- JVCI17** **Color Enhanced Local Binary Patterns in Covariance Matrices Descriptors (ELBCM)**
Michèle Gouiffès, Andres Romero Mier y Terán, Lionel Lacassagne
Elsevier Journal of Visual Communication and Image Representation, 49 : 447-458 (2017)
- MVA16** **A modular system for global and local abnormal event detection and categorization in videos**
Ahmed Chamseddine Ben Abdallah, **Michèle Gouiffès**, Lionel Lacassagne
Machine Vision and Applications, 1-19. (2016).
- JVCI16** **Evaluation of color descriptors for projector-camera systems**
Aleksandr Setkov, **Michèle Gouiffès**, Christian Jacquemin

Journal of Visual Communication and Image Representation 36 : 11-27 (2016)

- Eurasip14** **Covariance tracking : architecture optimizations for embedded systems**
Andres Romero Mier y Terán, Lionel Lacassagne, **Michèle Gouiffès** and Ali Hassan Zahraee
EURASIP Journal on Advances in Signal Processing, 2014(1), 1-15.
- JVCI13** **HTRI : High Time Range Imaging**
Michèle Gouiffès, Bertrand Planes and Christian Jacquemin
Journal of Visual Communication and Image Representation, Elsevier (2013)
- JRTIP13** **Color tracking for contextual switching : Real-time implementation on CPU**
Florence Laguzet, Andres Romero Mier y Terán, **Michèle Gouiffès**, Lionel Lacassagne, Daniel Etiemble
Journal of Real-Time Image Processing -Special Issue on Real-Time Color Image Processing. Octobre 2013.
- ICAE12** **Enhancing A Disparity Map by Color Segmentation**
Noppon Lertchuwongsa, **Michèle Gouiffès** and Bertrand Zavidovique
Journal of Integrated Computer-Aided Engineering (2012)
- CVIU12** **A study on local photometric models and their application to robust tracking**
Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne, Alain Trémeau
Computer Vision and Image Understanding (2012). 166(8). pp 896-907.
- ICAE11** **Mixed color/level lines and their stereo-matching with a modified Hausdorff distance**
Noppon Lertchuwongsa, **Michèle Gouiffès** and Bertrand Zavidovique
Journal of Integrated Computer-Aided Engineering (2011).
- JVCI10** **Body color sets : a compact and reliable representation of images**
Michèle Gouiffès and Bertrand Zavidovique
Journal of Visual Communication and Image Representation, Elsevier (2010)
- IVP 08** **A color topographic map based on the dichromatic reflectance model**
Michèle Gouiffès and Bertrand Zavidovique
Eurasip Journal on Image and Video Processing (2008)

JOURNAUX FRANCOPHONES

- TS 10** **Parallélisation d'opérateurs de TI : multi-cœurs, Cell ou GPU ?**
Antoine Pédrón, Florence Laguzet, Tarik Saidani, Pierre Courbin, Lionel Lacassagne, **Michèle Gouiffès**
Revue Traitement du Signal. 27(2), 2010, pp.161-187
- TS04** **Invariants couleur dédié à la transparence de l'encre. Application à la segmentation de marquages**
Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne, Alain Trémeau et Philippe Loisel
Revue Traitement du Signal. 21(5), Tome 1. 2004

- ICINCO17** **Exploiting optical flow field properties for 3D structure identification**
Tan Khoa MAI, **Michèle Gouiffès**, Samia Bouchafa
International Conference on Informatics in Control, Automation and Robotics, 24-28 July 2017, Madrid, Spain
- Visapp17** **Optical Flow refinement using reliable flow propagation**
Tan Khoa MAI, **Michèle Gouiffès**, Samia Bouchafa
12th international Conference on Computer Vision Theory and Applications, pages 451-458, 2017, Porto, Portugal
- VRST16** **3D Reconstruction of Indoor Building Environments with New Generation of Tablets**
Adrien Arnaud, Julien Christophe, **Michèle Gouiffès** et Mehdi Ammi
22nd ACM Symposium on Virtual Reality Software and Technology (VRST), 02-04 November 2016, Munich, Germany
- ICMI2016** **Exploration of Virtual Environments on Tablet : Comparison Between Tactile and Tangible Interaction Techniques**
Adrien Arnaud, Jean-Baptiste Corrége, Céline Clavel, **Michèle Gouiffès** et Mehdi Ammi
18th ACM International Conference on Multimodal Interaction, Tokyo, Japon, 12-16 Novembre 2016
- MOCO16** **Motions and Experiences in a Multiuser Interactive Installation : Towards an Embodied Account of Visitors' Journey**
Tom Giraud, Matthieu Courgeon, **Michèle Gouiffès**, Michel Bertier, Aymard de Mengin, Xavier Maitre
Proceedings of the 3rd International Symposium on Movement and Computing, 5-6 July 2016, Thessaloniki, Greece
- Visapp16** **Action-centric polar representation of motion trajectories for online action recognition**
Fabio Martínez, Antoine Manzanera, **Michèle Gouiffès**, Phuong Nguyen
12th International Conference on Computer Vision Theory and Applications, 2016. Italy
- ISVC15a** **DACImPro : A Novel Database of Acquired Image Projections and Its Application to Object Recognition**
Aleksandr Setkov, Fabio Martínez Carillo, **Michèle Gouiffès**, Christian Jacquemin, Maria Vanrell, Ramon Baldrich
International Symposium on Vision Computing (2) 2015 : 463-473. Las Vegas - USA
- ISVC15b** **A Gaussian mixture representation of gesture kinematics for on-line Sign Language video annotation**
Fabio Martínez, Antoine Manzanera, **Michèle Gouiffès**, Annelies Braffort
International Symposium on Vision Computing. Las Vegas - USA
- IEEE ICIP15** **One-frame delay for dynamic photometric compensation in a projector-camera system**
Panagiotis-Alexandros Bokaris, **Michèle Gouiffès**, Christian Jacquemin, Jean-Marc Chomaz, Alain Trémeau
IEEE International Conference on Image Processing, 27-30 September 2015, Quebec, Canada.
- ECCV14** **Photometric Compensation to Dynamic Surfaces in a Projector-Camera System**

Panagiotis-Alexandros Bokaris, **Michèle Gouiffès**, Christian Jacquemin, Jean-Marc Chomaz
Computer Vision-ECCV 2014 Workshops (pp. 283-296). Springer International Publishing.

- Mirage 13a** **Color Invariant Feature Matching For Image Geometric Correction**
Aleksandr Setkov, **Michèle Gouiffès** and Christian Jacquemin
6th ACM Mirage International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications 2013.
- Mirage 13b** **Enhanced Local Binary Covariance Matrices (ELBCM) for texture analysis and object tracking.**
Andres Romero Mier y Terán, **Michèle Gouiffès** and Lionel Lacassagne
6th ACM Mirage International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications 2013.
- IEEE ICIP 13** **Total Bregman Divergence for Multiple Object Tracking**
Andres Romero Mier y Terán, **Michèle Gouiffès**, Lionel Lacassagne
IEEE International Conference on Image Processing. Melbourne, 15-18 sept 2013
- DASIP 13** **Covariance tracking architecture optimizations for embedded systems**
Andres Romero Mier y Terán, Lionel Lacassagne, Ali Hassan Zahraee and **Michèle Gouiffès**
Conference on Design and Architectures for Signal and Image Processing (DASIP 2013). Cagliari, Italy 2013, 8-10 oct 2013
- ACCV12** **Covariance Descriptor for Multiple Object Tracking and Re-Identification with Colorspace Evaluation**
Andres Romero Mier y Teran, **Michèle Gouiffès**, Lionel Lacassagne
ACCV (Asian Conference on Computer Vision) Workshop on Detection and Tracking in Challenging Environments. November 5-9, 2012, Korea
- ICPR12** **Motion histogram quantification for human action recognition**
Hedi Tabia, **Michèle Gouiffès** and Lionel Lacassagne
International Conference on Pattern Recognition. November 11-15, 2012, Tsukuba Science City, Japan
- ISIVC12** **Motion modeling for abnormal event detection in crowd scenes**
Hedi Tabia, **Michèle Gouiffès** and Lionel Lacassagne
International Symposium on signal, Images, Video and Communications, 4-6 July 2012 at University of Valenciennes, France. 2012
- WIAMIS 11** **Adaptive clustering by analysis of the Connectedness Degree. Application to skin segmentation**
Michèle Gouiffès
International Workshop on Image Analysis for Multimedia Interactive Services. Delft, The Netherlands, April 2011
- IEEE ICSIPA 11** **Adaptive Spatio-Color Clustering Based on the Connectedness in a Chrominance-Luminance Space**
Michèle Gouiffès et Bertrand Zavidovique IEEE International Conference on Signal and Image Processing Applications. Malaysia, November 2011
- IEEE ICSIPA 11** **Feature Points Tracking Adaptive to Saturation**
Andres Romero Mier y Teran, **Michèle Gouiffès** and Lionel Lacassagne
IEEE International Conference on Signal and Image Processing Applications 2011. Malaysia, November 2011

-
- IEEE ICSPA 11 Automatic Color Space Switching for Robust Tracking**
Florence Laguzet, **Michèle Gouiffès** and Lionel Lacassagne
IEEE International Conference on Signal and Image Processing Applications. Malaysia, November 2011
- IEEE ICASSP 10 Body sets and lines : a reliable representation of images**
Michèle Gouiffès, Bertrand Zavidovique
International Conference on Acoustics, Speech and Signal Processing. Dallas, March 2010
- IEEE ICIP 10 Projection Histogram For Mean-Shift Tracking**
Michèle Gouiffès, Florence Laguzet, Lionel Lacassagne
IEEE International Conference on Image Processing 2010. Hong Kong, September 2010
- ICPR 10 Color Connectedness Degree For Mean-Shift Tracking**
Michèle Gouiffès, Florence Laguzet, Lionel Lacassagne
International Conference on Pattern Recognition 2010. Istanbul, August 2010
- SPIE 10 Robust obstacles detection and tracking using disparity for car driving assistance**
Michèle Gouiffès, Antoine Patri, Marius Vasiliu
IS&T SPIE Electronic Imaging 2010. San Jose, USA, January 2010
- ICINCO10 Mixed color/level lines and their stereo-matching with a modified Hausdorff distance**
Noppon Lertchuwongsa, **Michèle Gouiffès**, Bertrand Zavidovique
International Conference on Informatics in Control, Automation and Robotics. Funchal, Madeira, June 2010
- Nvidia 10 Complex Automative Applications**
Marius Vasiliu and **Michèle Gouiffès**
GPU technology conference 2010
- ICINCO 09 Segments of color lines : a comparison through a tracking procedure**
Michèle Gouiffès, Samia Bouchafa et Bertrand Zavidovique
International Conference on Informatics in Control, Automation and Robotics 2009
- IEEE ICIP 08 Dichromatic Level Lines**
Michèle Gouiffès and Bertrand Zavidovique
IEEE International Conference on Image processing 2008
- CGIV 08 A topographic map based on the dichromatic model**
Michèle Gouiffès et Bertrand Zavidovique
European Conference on Color in Graphics, Imaging and Vision, 2008
- CGIV 08 Adaptive spatio-colorimetric classification**
Michèle Gouiffès
European Conference on Color in Graphics, Imaging and Vision, 2008
+ Poster [ClasSpec 08] Project ClasSpec08, Université Lille, 10/15/2008
- IEEE ICIP 07 Tracking by combining photometric normalization and color invariants according to their relevance**
-

Michèle Gouiffès

IEEE International Conference on Image processing, 2007

- SPIE 07** **Image blur analysis for the subpixel-level measurement of in-plane vibration parameters of MEMS resonators**
Ha Vu Lee, Michèle Gouiffès, Fabien Parrain, Alain Bossebœuf, Bertrand Zavidovique
SPIE Applications of Digital Image Processing, 2007
- IEEE ICIP 06** **A photometric model for specular highlights and lighting changes. Application to feature points tracking**
Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne et Alain Trémeau
IEEE International Conference on Image Processing, 2006
- ECCV06** **Feature points tracking : robustness to specular highlights and lighting changes**
Michèle Gouiffès, Christophe Collewet, and Christine Fernandez-Maloigne et Alain Trémeau
European Conference on Computer Vision, 2006
- CGIV06** **Feature points tracking by using photometric model and colorimetric invariants**
Michèle Gouiffès, Christophe Collewet, and Christine Fernandez-Maloigne et Alain Trémeau
Colour in Graphics, Imaging and Vision, 2006
- IEEE ICIP 04** **Color segmentation of inked characters. Application to meat traceability control**
Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne et Alain Trémeau
IEEE International Conference on Image processing. Singapour, Octobre 2004
- CGIV04** **Reflectance modeling of an anisotropic surface : the ham**
Michèle Gouiffès, Pascale Marty-Mahé, Christine Fernandez-Maloigne et Alain Trémeau
2nd European Conference Color in Graphics, Imaging and Vision. Aachen, Allemagne, Avril 2004

CONFÉRENCES FRANCOPHONES

- IHM 15** **Miroir 3D augmenté par imagerie médicale : la perception de soi en question**
Mathieu Courgeon, Tom Giraud, Marion Tardieu, Alexandra Roatis, Michèle Gouiffès, Xavier Maître
27ème conférence francophone sur l'Interaction Homme-Machine, Octobre 2015.
- IHM 15** **Exploration d'environnement virtuel sur tablette : Comparaison entre des modalités tactiles et tangibles**
Adrien Arnaud, Jean-Baptiste Corrège, Céline Clavel, Michèle Gouiffès, Mehdi Ammi
27ème conférence francophone sur l'Interaction Homme-Machine. (p. w18) ACM, Octobre 2015.
- Coresa12** **Reconnaissance des activités humaines à partir de vecteurs de mouvements quantifiés**
Hedi Tabia, Michèle Gouiffès and Lionel Lacassagne
COMpression et REprésentation des Signaux Audiovisuels, 2012

- Gretsi 11** **Choix automatique de l'espace couleur pour le suivi robuste d'objets**
 Florence Laguzet, **Michèle Gouiffès**, Lionel Lacassagne, Daniel Etiemble
 Congrès GRETSI, Traitement du Signal et des Images, Bordeaux 2011
- Gretsi 11** **Classification par analyse des degrés de connexité d'ensembles couleur**
Michèle Gouiffès, Bertrand Zavidovique
 Congrès GRETSI, Traitement du Signal et des Images, Bordeaux 2011
- Gretsi 07** **Suivi dans des séquences d'images par coopération luminance/couleur**
Michèle Gouiffès
 Congrès GRETSI, Traitement du Signal et des Images, 2007
- Gretsi 04** **Suivi de points robuste aux spéularités et aux changements d'éclairage**
Michèle Gouiffès, Christophe Collewet, Christine Fernandez-Maloigne et Alain Trémeau
 Colloque GRETSI sur le Traitement du Signal et des Images, Louvain-la-Neuve, Belgique, Septembre 2005
- JJCR04** **Suivi de points : extension aux objets non-lambertiens**
Michèle Gouiffès, Christophe Collewet, and Christine Fernandez-Maloigne et Alain Trémeau
 Journée des Jeunes Chercheurs en Robotique, Douai, France, Septembre 2004
- RFIA 04** **Comparaison de deux méthodes de segmentation couleur appliquées à traçabilité de produits carnés**
Michèle Gouiffès, Pascale Marty-Mahé Christine Fernandez-Maloigne et Alain Trémeau
 14ieme Congrès Francophone AFRIF-RFIA de Reconnaissance des Formes et Intelligence Artificielle, Toulouse, Janvier 2004

CHAPITRE DE LIVRE

- 2016** **Some Technical and Scientific Issues in Mobile : Video-Projection and Possible Applications**
 Christian Jacquemin, Panagiotis.-Alexandros Bokaris, **Michèle Gouiffès**, Aleksandr Setkov
 In Digital Interfaces in Situations of Mobility : Cognitive, Artistic, and Game Devices, Bernard Guelton, Ed. Common Ground Publishing

RAPPORT DE RECHERCHE

- 2007** **Modeling of photometric variations for matching correspondences, its validity and its application to tracking**
Michèle Gouiffès, Christophe Collewet
<https://hal.archives-ouvertes.fr/hal-00430799v1>, 58 pages. 2007

AUTRES

Logiciel 2014 : Dépôt auprès de l'APP (Agence pour la Protection des Programmes) du programme CovTrack, un ensemble de bibliothèques dédiées au suivi d'objets et à la reconnaissance, développées au cours de la thèse d'Andres Romero Mier y Téran (LRI) en co-encadrement avec Lionel Lacassagne (LRI).

Base de données DAcImPro : The Database of ACquired IMage PROjections

Ensemble d'images de vidéoprojections réalisées sur des surfaces uniformes ou non, planes ou non, sous différents éclairages. Cette base de données a été construite pour évaluer les algorithmes définis pour les systèmes caméra-projecteurs. Elle est accessible sur <https://dacimpro.limsi.fr/>

Présentations diverses

- Réunion du GDR ISIS *Atelier scientifique sur les interactions Homme/Machine (2/12/2015)* Online action recognition by modeling local kinematic primitive, Fabio Martínez, Michèle Gouiffès, Antoine Manzanera.
- Christian Jacquemin, Panagiotis-Alexandros Bokaris, Michèle Gouiffès, Aleksandr Setkov, Quelques enjeux technoscientifiques et applications possibles de la vidéo-projection mobile. Interfaces Numériques en Situation de Mobilité : Dispositifs artistiques et ludiques, Colloque et Tables rondes, 2014
- Présentation à la journée *Regards sur la Lumière 2015*. Vision par ordinateur et vidéo-projection : contributions et applications
- Présentation à Technicolor, juillet 2014. *Image processing at LIMSI : matching robust descriptors for video-surveillance and spatial augmented reality*.
- Réunion du GDR ISIS. *Suivi temps-réel (16/11/2007)* : Apports de la couleur et des modèles de réflexion pour le suivi de points.