



HAL
open science

Fusion de données multimodales par combinaison de l'incertain et de modèles de perception

Hanan Ait Errani Anzid

► **To cite this version:**

Hanan Ait Errani Anzid. Fusion de données multimodales par combinaison de l'incertain et de modèles de perception. Ingénierie assistée par ordinateur. Université Bourgogne Franche-Comté; Université Ibn Zohr (Agadir, Maroc). Faculté des sciences, 2019. Français. NNT : 2019UBFCK046 . tel-02936328

HAL Id: tel-02936328

<https://theses.hal.science/tel-02936328>

Submitted on 11 Sep 2020

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.



كلية العلوم
+ⴰⴳⴷⴰⴳⴰⴳⴰ | +ⴷⴰⴳⴰⴳⴰⴳⴰ
FACULTÉ DES SCIENCES



THESE COTUTELLE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE
DOCTORAT NATIONAL MAROC
PREPAREE A L'UNIVERSITE DE FRANCHE-COMTE
FACULTE DES SCIENCES IBN ZOHR

Ecole doctorale n°37

SPIM - Sciences pour l'Ingénieur et Microtechniques

Doctorat en Informatique
Spécialité : Imagerie et vision Artificielle

Par

ANZID HANAN

Fusion de données multimodales par combinaison de l'incertain et de modèles de perception

Thèse présentée et soutenue à Agadir, Le 30 Octobre 2019

Composition du Jury :

Pr Abdelaziz, El Fazziki	Professeur, Université Cadi Ayyad de Marrakech, Maroc	Président
Pr Abderrahim ,El Moataz billah	Professeur, Université de Caen Normandie, France	Rapporteur
Pr Mohammed, SADGAL	Professeur, Université Cadi Ayyad de Marrakech, Maroc	Rapporteur
Pr Driss, MAMMASS	Professeur, Université Ibn Zohr, Agadir, Maroc	Directeur de Thèse
Pr Alamin , MANSOURI	Professeur, Université de Bourgogne-Franche-Comté, France	Directeur de Thèse
Pr Gaetan, Le Goic	MC, Université de Bourgogne-Franche-Comté, France	Co-encadrant

À la mémoire de mon père,

À ma mère

À mon mari et mon fils Lokmane

À ma sœur Noura

À mon frère Zakaria

À mes nièces Tasnime et Sarah

Remerciement

Cette thèse s'inscrit dans le cadre d'une cotutelle entre le laboratoire ImVia, Dijon France et du laboratoire IRF-SIC de la faculté des sciences d'AGADIR, Maroc. Avant tout, je tiens à exprimer ma sincère gratitude à ces deux laboratoires de m'avoir accueillie, et de m'avoir permis de poursuivre mes études dans le domaine du traitement d'images.

Je tiens à remercier M. abderrahim el moataz billah, professeur à l'université de Caen, à M. Mohammed Sagdal, professeur à la faculté des sciences Semlalia université de Cadi Ayyad, à M. Soufiane Idbraim, professeur à l'université des sciences Agadir université IBN ZOHR d'avoir accepté de rapporter sur mes travaux de thèse. J'exprime ma gratitude à M. Abdelaziz El fazziki, professeur à la faculté des sciences Semlalia université Cadi Ayyad d'avoir été examinateurs et membre de mon jury.

Je remercie mes directeurs de thèse M. Driss Mammass et M. Alamin Mansouri, tout d'abord pour leur confiance qui m'ont accordée en acceptant de m'encadrer, pour leurs directives et conseils, et d'avoir mis à ma disposition tous les moyens matériels pour la réussite de ce travail.

Mes remerciements vont également à mon encadrant M. Gaëtan Le Goïc pour ses conseils et ses encouragements précieux, je remercie également M. Aissam Bekkari pour sa disponibilité, ses déplacements, pour ses encouragements dans les grands moments de doute, pour ses qualités humaines et sa bonne humeur.

Je suis très reconnaissante envers M. Ali Idarrou, M. Youssef Essady et Azeddine Elhassouny pour leur soutien, leur expérience et pour leurs conseils et leur qualité humaines.

Je tiens à remercier vivement l'ensemble des membres des laboratoires ImVia et IRF-SIC, en particulière à Loubna, Fatima Zahra, Meina, Hasna, Rabia, Salim, Yousef, Ismail, Dominica, Dalicia, Soumia et Salma. Je tiens à remercier profondément Aicha, Hanan, Leila, Jihad, Khadija et Yassine pour leur amitié et leur soutien.

Mes remerciements vont à ma famille.

À mon papa que j'aurais souhaité être présent en ce moment pour partager ma joie. Que son âme repose en paix. À ma maman pour son amour, ses sacrifices, ses prières. À mon mari pour son soutien inconditionnel durant toutes ces longues années d'études et pour être toujours à mes côtés et m'épauler. À ma sœur Noura et mon frère Zakaria de m'avoir soutenu toute au long de ma vie. À mon fils Lokmane d'avoir apporté toute cette joie et ce bonheur à ma vie, et d'avoir été patient durant mes multiples déplacements. À mes chères cousines Zahra, Aicha et Fatima pour leurs encouragements et pour être toujours à mes côtés.

Table des matières

RESUME	10
ABSTRACT	12
1. INTRODUCTION GENERALE	13
1.1. Contexte scientifique et objectifs	13
1.2. Contributions.....	15
1.3. Organisation du manuscrit	18
1.4. Cadre de thèse	19
PARTIE 1	20
ÉTUDE BIBLIOGRAPHIQUE	20
CHAPITRE 1 IMAGES MULTIMODALES ET FUSION	21
1.1. Introduction.....	21
1.2. Les images multimodales : Acquisition et caractérisation	21
1.2.1. Contexte général d'acquisition des données utilisées	21
1.2.2. Le processus d'acquisition et de caractérisation des données.....	23
1.3.1. Définition.....	25
1.3.2. Les types d'imperfections	26
1.3.3. Les niveaux de fusion	28
1.3.4. Les étapes de la fusion	30
1.4. Problématique générale et méthodologie pour la fusion	30
1.4.1. Problématique générale	30
1.4.2. Le processus de fusion.....	32
1.5. Conclusion	33
CHAPITRE 2 LE RECALAGE ET L'EXTRACTION DES CARACTERISTIQUES	34
2.1. Introduction.....	34
2.2. Le prétraitement : le recalage	35
2.2.1. Définition du recalage.....	35
2.2.1. Extraction des caractéristiques.....	36
2.2.2. Le modèle de transformation	46
2.3. L'extraction des caractéristiques.....	47
2.3.1. L'information Spectrale et Spatiale	47
2.3.1.1. Les descripteurs spectraux.....	47

2.3.1.1.1.L'espace RVB.....	47
2.3.1.1.2.L'espace CIELab	48
2.3.1.2. Les descripteurs spatiaux.....	49
2.3.2. L'information perceptuelle & la saillance	52
2.4. Conclusion	56
CHAPITRE 3 LA FUSION DES IMAGES BASEE SUR LES THEORIES DST ET DSMT	57
3.2. FORMALISATION DU PROBLEME DE FUSION	58
3.3. Théorie de Dempster-Shafer (DST).....	59
3.4. Théorie de Dezert-Smarandache (DSmT)	61
PARTIE 2.....	78
CONTRIBUTIONS	78
CHAPITRE 4 RECALAGE DES IMAGES MULTIMODALES PAR FILTRAGE AUTOMATIQUE ET ADAPTATIF DE CARACTERISTIQUES SURF	79
4.1. Introduction.....	79
4.2. La problématique de correspondance rencontrée	79
4.3. Le filtrage des points de correspondance basé sur la distribution gaussienne :... 80	80
4.4. Le filtrage des points de correspondance basé le critère de la distance et de l'orientation	90
4.5. Conclusion	100
CHAPITRE 5 CLASSIFICATION DES IMAGES MULTIMODALES BASEE SUR LA FUSION DES CARACTERISTIQUES SPECTRALES ET DENSE SURF	101
5.1. Introduction.....	101
5.2. Le classifieur SVM.....	101
5.3. L'établissement des échantillons d'apprentissage	102
5.4. La classification des images multimodales basée sur Dense SURF et l'information spectrale	103
5.5. Résultat et discussion	106
5.6. Conclusion	119
CHAPITRE 6 CLASSIFICATION/FUSION DES IMAGES MULTIMODALES BASEE SUR DSMT-SAILLANCE	120
6.1. Introduction.....	120
6.2. La fusion des images multimodales basée sur la DSmT en intégrant le modèle perceptuel.....	121
6.3. Résultat et discussion.....	130
6.4. Conclusion	138
7. CONCLUSION ET PERSPECTIVES.....	140
ANNEXE 1 LA CLASSIFICATION SVM.....	142
Introduction.....	142

Généralité sur la classification SVM	142
SVM linéaire.....	143
• Données linéairement séparables	144
• Données non-linéairement séparables.....	147
SVM non-linéaire	151
• Formalisation de problème	151
• La notion des noyaux	153
SVM multiclasse	154
• Un-contre-tous	154
• Un-contre-un.....	155
Conclusion.....	157

Liste des figures

Figure 1 – Vue Générale du palais de Germolles	22
Figure 2 - Plan du palais de Germolles.....	22
Figure 3 - Exemple de restauration effectuée dans le dressing de Margurite.....	23
Figure 4 : photographie macro-technique.....	24
Figure 5: Exemple de quatre modalités acquises sur la lettre 'P' sur le mur sud du dressing de la Comtesse.....	25
Figure 6 Les niveaux de fusion	29
Figure 7 – la problématique générale de la fusion des données	32
Figure 8 – le processus de la fusion des images.....	33
Figure 9 Exemple du problème du recalage.....	36
Figure 10 les relations de l'intensité entres l'image modèle et l'image requête[47]	37
Figure 11 Exemples de l'intensité globale des images multimodales correspondant à la même scène ((Figure5)	38
Figure 12 - Exemples de l'intensité locale des images multimodales extraites au long de la même ligne pour les quatre modalités	39
Figure 13 les différentes octaves de l'image dans la pyramide de Gauss [64]	41
Figure 14 l'assignement de l'orientation au descripteur SIFT[64]	41
Figure 15 Principe de l'image intégrale [66]	43
Figure 16 Box-Filtre [66].....	43
Figure 17 Analyse d'échelle avec une taille constante de l'image [66].....	44
Figure 18 Exemple de la fenêtre coulissante d'orientation de taille $\pi 3$ [66].....	44
Figure 19 le Cube RVB (source Wikipédia)	48
Figure 20 : Représentation de l'espace CIELab (source Wikipédia).....	48
Figure 21: Codification de Venn pour $\Theta = 3$ modèle libre de DSMT [34].....	65
Figure 22 Diagramme de Venn pour le modèle hybride $\mathcal{M}'(\Theta)$ [34]	67
Figure 23 : les fonctions de croyance généralisées [19]	69
Figure 24 Organigramme du premier algorithme proposé. Deux filtres : le premier basé sur la distance F1 (D) et le deuxième sur l'orientation F2 (O).....	83

Figure 25 Distributions d'orientation sur la modalité IRR (à droite) avant le filtrage (à gauche) après le filtrage (région zoomée).....	84
Figure 26 Distributions d'orientation sur la modalité UVR (à droite) avant le filtrage (à gauche) après le filtrage (région zoomée).....	84
Figure 27 Distributions d'orientation sur la modalité UVF (à droite) avant le filtrage (à gauche) après le filtrage (région zoomée).....	85
Figure 28 les quatre modalités pour la première scène sur le mur ouest de la chambre de la comtesse Nevers	86
Figure 29 Application du recalage entre IRR et VIS de la première scène - (a) Méthodologie SURF (b) Notre méthodologie.....	86
Figure 30 Application du recalage entre UVR et VIS de la première scène - (a) Méthodologie SURF (b) Notre méthodologie.....	86
Figure 31 Application du recalage entre UVF et VIS de la première scène - (a) Méthodologie SURF (b) Notre méthodologie.....	87
Figure 32 les quatre modalités pour la deuxième scène sur le mur nord du dressing de Marguerite de Flandre	87
Figure 33 Application du recalage entre IRR et VIS de la deuxième scène - (a) Méthodologie SURF (b) Notre méthodologie.....	88
Figure 34 Application du recalage entre UVR et VIS de la deuxième scène - (a) Méthodologie SURF (b) Notre méthodologie.....	88
Figure 35 Application du recalage entre UVF et VIS de la deuxième scène - (a) Méthodologie SURF (b) Notre méthodologie.....	89
Figure 36 Schéma représentant l'approche proposée.....	92
Figure 37 : Exemple de distances calculées et d'orientation pour un point inlier ..	93
Figure 38 : Distributions des orientations d'IRR (à droite) avant filtrage (à gauche) après filtrage	93
Figure 39 Distributions des orientations d'UVR (à droite) avant filtrage (à gauche) après filtrage	94
Figure 40 Distributions des orientations d'UVF (à droite) avant filtrage (à gauche) après filtrage	94
Figure 41 les quatre modalités pour la deuxième scène sur le mur ouest du Duc de Bourgogne	96
Figure 42 Application du recalage entre IRR et VIS de la première scène pour la deuxième contribution- (a) Méthodologie SURF (b) Notre méthodologie	97

Figure 43	Application du recalage entre UVR et VIS de la première scène pour la deuxième contribution- (a) Méthodologie SURF (b) Notre méthodologie	97
Figure 44	Application du recalage entre UVF et VIS de la première scène pour la deuxième contribution- (a) Méthodologie SURF (b) Notre méthodologie	98
Figure 45	les quatre modalités pour la deuxième scène sur le mur nord du dressing de la comtesse de Nevers	98
Figure 46	Application du recalage entre IRR et VIS de la deuxième scène pour la deuxième contribution- (a) Méthodologie SURF (b) Notre méthodologie	99
Figure 47	Application du recalage entre IRR et VIS de la deuxième scène pour la deuxième contribution- (a) Méthodologie SURF (b) Notre méthodologie	99
Figure 48	Application du recalage entre IRR et VIS de la première scène pour la deuxième contribution- (a) Méthodologie SURF (b) Notre méthodologie	100
Figure 49	Architecture générale de l'approche proposée	104
Figure 50	Exemple de la décoration murale des chambres du palais de Germolles	106
Figure 51	les quatre modalités pour la scène M sur le mur nord du dressing de duc de Bourgogne	107
Figure 52	Cartes de classification avec le classifieur SVM sur la modalité Visible du premier ensemble des images (a), utilisant l'information spectrale seulement (b) et la combinaison de SURF dense avec l'information spectrale (c)	109
Figure 53	Cartes de classification avec le classifieur SVM sur la modalité IRR du premier ensemble des images (a), utilisant l'information spectrale seulement (b) et la combinaison de SURF dense avec l'information spectrale (c)	110
Figure 54	Cartes de la classification avec classifieur SVM sur la modalité UVF du premier ensemble des images (a), utilisant l'information spectrale seulement (b) et la combinaison de SURF dense avec l'information spectrale (c)	111
Figure 55	Cartes de classification avec le classifieur SVM sur la modalité UVR du premier ensemble des images (a), utilisant l'information spectrale seulement (b) et la combinaison de SURF dense avec l'information spectrale (c)	112
Figure 56	les quatre modalités pour la scène P sur le mur nord du dressing de duc de Bourgogne	113
Figure 57	Cartes de la classification avec classifieur SVM sur la modalité Visible du deuxième ensemble des images (a), utilisant Dense SURF (b), Information spectrale (c) et la combinaison de SURF dense avec l'information spectrale (d)	114

Figure 58 Cartes de la classification avec classifieur SVM sur la modalité IRR du deuxième ensemble des images (a), utilisant Dense SURF (b), Information spectrale (c) et la combinaison de SURF dense avec l'information spectrale (d)	115
Figure 59 Cartes de la classification avec classifieur SVM sur la modalité UVF du deuxième ensemble des images (a), Information spectrale (b) et la combinaison de SURF dense avec l'information spectrale (c)	116
Figure 60 Cartes de la classification avec classifieur SVM sur la modalité UVR du deuxième ensemble des images (a), Information spectrale (b) et la combinaison de SURF dense avec l'information spectrale (c)	117
Figure 61 le system global de la fusion des images multimodales basé sur la DSMT intégrant la saillance.....	121
Figure 62 Comparaisons de cartes de saillance dans la base de données d'images multimodales: (a) l'image originale, (b) Carte de fixation, (c) [101],(d)[111], (e) [112], (f) [113], (g) [80], (h) [110], (i) [100], (j) [105].	124
Figure 63 le calcul de la carte de la saillance selon la méthode de Rahtu <i>et al.</i>	127
Figure 64 Cas d'application 1 - scène sur le mur sud du dressing de la comtesse du Château de Germolles	130
Figure 65 Cas d'application 2 – scène sur le mur sud du dressing de la comtesse du Château de Germolles	131
Figure 66 : les cartes de classification SVM de la première scène, (a) Classification SVM de la modalité UVF, (b) Classification SVM de la modalité UVR, (c) Classification SVM de la modalité IRR.....	132
Figure 67 : Les cartes de la saillance correspondant à la première scène, (a) la carte de la saillance de la modalité VIS, (b) la carte de la saillance de la modalité UVF, (c) la carte de la saillance de la modalité UVR, (d) la carte de la saillance de la modalité IRR.....	133
Figure 68 : la classification SVM de la modalité VIS (a), Classification DSMT des images multimodales (b), Classification en se basant sur DSMT et la saillance des images multimodales (c)	134
Figure 69 : les cartes de classification SVM de la deuxième scène, la carte de la classification SVM de la modalité UVF (a), la carte de la classification SVM de la modalité UVR (b), la carte de la classification SVM de la modalité IRR (c).	135
Figure 70 : Les cartes de la saillance correspondant à la première scène, (a) la carte de la saillance de la modalité VIS, (b) la carte de la saillance de la modalité	

UVF, (c) la carte de la saillance de la modalité UVR, (d) la carte de la saillance de la modalité IRR.....	136
Figure 71 : la carte classification SVM de la modalité VIS (a), la carte de la classification DSMT des images multimodales (b), la carte de classification en se basant sur DSMT et la saillance des images multimodales (c).....	137
Figure 72 Classification linéaire : Données linéairement séparables [187].....	145
Figure 73 Cas des données non-linéairement séparables [187].....	148
Figure 74 Cas de SVM non linéaire	151
Figure 75 : Exemple de séparation linéaire par hyperplan optimale dans un nouvel espace de description	152
Figure 76 Cas de SVM multiclass	154

Liste des tableaux

Tableau 1 Séquence de Dedekind.....	63
Tableau 2 Taille de la mémoire requise pour D.....	64
Tableau 3 Évaluation des performances du recalage sur la totalité des images multimodales de notre base de données.....	89
Tableau 4 Évaluation des performances du recalage sur les images satellitaires .	90
Tableau 5 Évaluation des performances du recalage avec la deuxième méthode sur la totalité des images multimodal de note base de données	96
Tableau 6 Précision globale de la classification Images multimodales basées uniquement sur les informations spectrales et la combinaison des informations spectrales et de la densité SURF.....	118
Tableau 7 Métrique NSS des méthodes comparées	125
Tableau 8 Précision et erreurs de classification résultant de différentes méthodes	138

Résumé

Le travail que nous présentons dans cette thèse s'inscrit dans une problématique générale de la fusion des images multimodales. Bien que les images multimodales aient apparues depuis plusieurs décennies, leurs utilisations dans divers domaines ne sont devenues effectives qu'avec l'évolution des domaines d'applications d'imagerie notamment dans le domaine médical, le domaine satellitaire, et très récemment dans le domaine de l'héritage et patrimoine dans lequel s'inscrivent les travaux de cette thèse. Cette technique présente de nombreux avantages et son utilisation présente un réel bénéfice dans plusieurs domaines d'application.

La tendance actuelle consiste à intégrer plusieurs modalités d'imagerie dans un et même système. Une autre pratique consiste à combiner les données de différentes modalités, concernant un même objet pour tirer profit de la redondance et de la complémentarité des informations délivrées par celles-ci.

Cette thèse est préparée en cotutelle entre l'équipe IRF du laboratoire d'Image et Reconnaissance de Formes — Systèmes Intelligents et Communicants (IRF-SIC) de l'université Ibn Zohr au Maroc et l'équipe CORES du Laboratoire Imagerie et vision artificielle de l'Université de Bourgogne-Franche-comté en France. Ce projet de thèse a bénéficié d'un soutien financier à la mobilité dans le cadre du projet PHC TOUBKAL TBK-16-31.

Dans cette thèse, nous proposons et développons des méthodes spectrales-spatiales-perceptuelles pour la classification des données multimodales. Ces méthodes permettront la combinaison de l'information spectrale et spatiale ainsi que l'intégration de l'information perceptuelle.

Dans les deux premières contributions, nous traitons l'alignement (appelé aussi recalage) des images que nous cherchons à fusionner en se basant sur la méthode géométrique SURF. Le recalage de ces différentes images soulève le problème de la présence des points aberrants, vu la nature et les conditions d'acquisition différentes entre les modalités, ce qui nous a poussés à réfléchir à une méthode permettant de filtrer les points d'appariement détectés.

La première contribution intègre deux filtres qui se basent sur la distribution gaussienne des données. La première opération de filtrage est basée sur un calcul de distance entre descripteurs de points caractéristiques précédemment extraits. La seconde implémente un filtrage basé sur l'orientation qui est situé dans une plage automatiquement définie par la distribution gaussienne des données.

La deuxième contribution est basée dans un premier temps sur un filtrage automatique utilisant la distribution gaussienne des distances calculées entre les points

putatifs détectés sur les deux images (modèle et requête). Ensuite, et en se basant sur le résultat de ce premier filtrage, nous fixons deux points comme références sur chacune des images (modèle et requête) et puis nous calculons les distances entre ces deux points et le point en question ainsi que leurs orientations. Par la suite, un critère strict a été mis en place en utilisant les distances calculées et les orientations afin de supprimer tout point aberrant.

Dans la troisième contribution, une classification SVM a été réalisée en utilisant une combinaison de caractéristiques multi-spectrales et les caractéristiques Dense SURF. La méthode proposée extrait les caractéristiques SURF pour chaque pixel dans chacune des images appelées Dense SURF, en exploitant uniquement la partie de description du descripteur SURF. Puis, nous nous sommes concentrés sur l'utilisation conjointe de l'information Dense SURF avec l'information spectrale et les incorporer dans le processus de la classification.

La quatrième contribution aborde la fusion des images multimodales. Dans le but d'améliorer la tâche de la classification dans des images multimodales, nous proposons une méthodologie permettant, en premier lieu, de fusionner, à l'aide de DSMT, les caractéristiques spectrales et les caractéristiques denses SURF extraites de chaque modalité, combinées, et pré-classées à l'aide du classifieur SVM. Ensuite, nous intégrons le modèle de perception visuelle dans le processus de fusion.

Toutes les approches proposées dans cette thèse ont permis une amélioration significative des performances de la classification et une flexibilité pour la répartition équilibrée entre les informations spatiale et spectrale ainsi que l'information de perceptuelle dans le processus de la fusion.

Mots clés : Fusion des images, Images multimodales, Approche statistique, Théorie de l'incertain, saillance.

Abstract

Although multimodal images have appeared for several decades, their use in various fields has only become effective with the evolution of imaging applications fields, particularly in the medical field, the satellite field, and very recently in the field of heritage in which the work of this thesis is based. This technique has many advantages and its use has real benefits in several areas of application.

The current trend is to integrate several imaging modalities into one system. Another practice is to combine data from different modalities, concerning the same object, to take advantage of the redundancy and complementarity of the information provided by them.

In this thesis, we propose and develop spectral-spatial-perceptual methods for the classification of multimodal data. These methods will allow the combination of spectral and spatial information as well as the integration of perceptual information.

In the first two contributions, we discuss the alignment of the images we seek to merge based on the SURF geometric method. The misalignment of these different images raises the problem of the presence of outliers, given the different nature and acquisition conditions between the modalities, which prompted us to consider a method to filter the detected matching points.

In the third contribution, an SVM classification was performed using a combination of multispectral characteristics and Dense SURF characteristics. The proposed method extracts the SURF characteristics for each pixel in each of the images called Dense SURF, using only the description part of the SURF descriptor. Then, we focused on the joint use of Dense SURF information with spectral information and incorporating them into the classification process.

The fourth contribution deals with the fusion of multimodal images. In order to improve the task of classification in multimodal images, we propose a methodology that allows, first, to merge, using DSMT, the spectral characteristics and dense SURF characteristics extracted from each modality, combined, and pre-classified using the SVM classification. Then we integrate the visual perception model into the fusion process.

All the proposed approaches were tested on imagery scenes collected as part of this project are multimodal images of the historical wall paintings of Château de Germolles located in Burgundy, and have allowed a significant improvement in classification performance and flexibility for balanced distribution between spatial and spectral information as well as perceptual information in the fusion process.

Keywords: Image fusion, Multimodal Images, Statistical Approach, Uncertainty Theory, Saliency.

1. Introduction générale

1.1. Contexte scientifique et objectifs

Ce travail de thèse s'inscrit dans une problématique générale de traitement d'image et s'intéresse particulièrement à la classification et à la fusion des images multimodales, d'objets complexes, issues de sources différentes.

Les systèmes d'imagerie et les outils de traitement et d'analyses subséquents connaissent un développement rapide et sont devenus incontournables dans différents domaines. Ce développement est dû au caractère non invasif et non destructif, mais aussi à la fiabilité, la répétabilité et à la précision qu'offrent ces systèmes. Ces derniers exploitent divers phénomènes physiques et/ou biologiques, et en génèrent des mesures locales ou des cartographies sous forme d'images.

Cependant, une unique image ne permet souvent pas d'obtenir une information complète sur un aspect donné, tel que par exemple l'aspect morphologique de la surface imagée. Les différentes modalités d'imagerie permettent d'accéder à un aspect particulier de l'objet acquis. Les aspects métrologiques tels que l'incertitude, la répétabilité et la reproductibilité sont aussi à prendre en compte. Ainsi, afin de pouvoir caractériser correctement des objets complexes, et notamment les objets naturels qui sont généralement par nature multidimensionnels (physique/biologique et/ou perceptuelle), il est souvent judicieux de combiner des acquisitions de différents types, en faisant varier par exemple le type de source d'illumination de l'objet imagé. Ces acquisitions sont appelées aussi acquisitions multimodales.

Ainsi, La tendance actuelle consiste à intégrer plusieurs modalités d'imagerie dans un et même système. Une autre pratique consiste à combiner les données de différentes modalités, concernant un même objet pour tirer profit de la redondance et de la complémentarité des informations délivrées par celles-ci. Ces pratiques sont courantes pour le traitement des images dans différentes disciplines, notamment, en médecine et biologie [1][2], en agriculture [3][4][5], en aérospatiale, en robotique, en intelligence artificielle, ou encore dans le domaine du patrimoine et de la conservation [6].

C'est dans ce dernier champ d'application que s'inscrivent mes travaux de thèse. En effet, dans ce domaine en particulier, les images multimodales sont utilisées pour une analyse rapide, avec un faible coût et sans causer de dommage aux surfaces étudiées. Toutefois, l'exploitation des données issues des acquisitions multimodales soulève plusieurs problématiques. Cette étape de la chaîne d'analyse, de l'acquisition à la visualisation, est essentielle et critique. S'il est incontestable que les données multimodales sont riches et complémentaires - et aussi parfois redondantes - il n'en reste pas moins qu'elles ne sont généralement pas exploitables directement notamment de leur

différente typologie, résolution, position d'acquisition. Ceci génère des axes de recherche à la fois pluriels et complémentaires.

Le premier aspect concerne l'alignement et la mise en correspondance des données acquises. Il s'agit de faire en sorte que chaque pixel, issu potentiellement différentes modalités, représente le même point physique de l'objet après recalage. L'appariement des acquisitions multimodales soulève aussi le problème de la présence des points aberrants, qui peut perturber les opérations de recalage, ce qui nous a poussés à réfléchir à une méthode permettant de filtrer les points d'appariement afin d'améliorer la performance de cette opération sur des données multimodales d'imagerie.

Le deuxième aspect concerne la représentation et le format. Il s'agit ici de préciser quelles représentations et quels formats de données adopter ? À ce niveau, on est amené à penser à une représentation unifiée permettant l'analyse globale des données. La représentation de ces derniers est un élément clé dans le choix des outils mathématiques à utiliser et à adapter au traitement des images multimodales.

Le troisième aspect concerne l'exploitation conjointe des données multimodales, ce qui nous mène aux problématiques liées à la fusion des données, Le modèle de fusion est conditionné par la nature des données, par l'objectif et par les connaissances disponibles. Ainsi, l'enjeu consiste à combiner les informations liées aux acquisitions, de façon intelligente, en prenant en compte les aspects liés :

(1) à l'imprécision potentielle des données acquises qui caractérise le manque d'exactitude de la connaissance en terme quantitatif ;

(2) à l'incertitude des données qui décrit le degré de conformité à la réalité d'une information ;

(3) aux conflits potentiels entre les informations à fusionner : celles-ci peuvent être en contradictions à cause par exemple de la non-fiabilité des sources, ou du fait que ces sources observent des phénomènes différents ;

(4) à l'incomplétude qui caractérise l'absence de l'information apportée par la source et le manque de données, ou encore l'ambiguïté lorsqu'il est difficile de distinguer entre deux situations, d'où plusieurs interprétations qui pourraient être envisagées.

Ceci n'empêche pas les données d'offrir une quantité d'information importante sur les objets/surfaces imagés, d'où la nécessité d'une exploitation adéquate à travers la mise en œuvre d'approches mathématiques combinées numériquement, à travers un formalisme adapté.

Néanmoins, une revue de la littérature montre que la majorité des méthodes actuelles ne sont pas adaptées aux données multimodales, et ne prennent pas en compte ni les informations qualitatives, ni les informations intrinsèques au système visuel humain et sa perception. Nous visons à travers ces travaux de thèse à apporter des réponses à cette

problématique, ce qui constitue l'originalité de la recherche que nous visons à travers la combinaison d'approches statistiques et d'attributs perceptifs pour créer des cartes de fusion répondant au besoin de l'analyse de l'utilisateur. Le deuxième aspect novateur réside dans l'élaboration d'un cadre de représentation et de traitement unique, afin d'uniformiser les traitements et les outils d'analyse.

Il est à noter que les méthodologies et instrumentations que nous avons développées restent génériques et ne sont pas limitées au domaine du patrimoine. Elles offrent la possibilité d'application et d'interaction dans différents domaines, et notamment en agriculture (pour la reconnaissance d'adventices, la rationalisation des traitements, la planification des plans d'occupation des sols), en biomédecine et dermatologie où l'imagerie multimodale a ouvert beaucoup de perspectives en termes d'aide au diagnostic et décision.

1.2. Contributions

Comme nous l'avons présenté ci-dessus, le travail que nous proposons s'articule donc autour du traitement des images multimodales avec une application sur les images issues du domaine du patrimoine.

Nous avons aussi souligné que ce domaine de recherche présente plusieurs problématiques liées notamment au recalage de ces images, à la représentation et à l'exploitation conjointe des données multimodales. Ces trois axes de recherche ont été traités dans le cadre de ce travail de thèse.

Afin d'aborder ces problématiques, nous avons jugé nécessaire d'étudier les travaux de recherche sur le recalage, l'extraction de caractéristiques, la classification, la fusion des images, etc... Ces outils sont présentés dans ce document.

Le recalage des images, même s'il ne fait pas partie en tant que tel de la fusion, est un prétraitement primordial et nécessaire dans le processus. Sa finalité réside dans la représentation des informations issues de différentes modalités dans un même référentiel. Dans notre cas, nous avons utilisé les méthodes d'appariement géométrique pour effectuer l'alignement des images multimodales. Ces méthodes ont l'avantage d'être rapides et efficaces avec ce type de données. Cependant, elles présentent l'inconvénient de générer un grand nombre de points de correspondance, qui peut contenir des points aberrants, et ainsi nuire à la performance du recalage effectué.

L'extraction des caractéristiques sur chaque modalité et la projection dans un espace de représentation commun a pour finalité de faciliter leur comparaison, et de servir à traiter le problème de pré-classification des images multimodales. Le classificateur utilisé dans ce travail est le SVM(Support Vector Machine). Ce choix a été fait en raison de son efficacité à traiter les données de grandes dimensions extraites combinant informations

spatiale et spectrale. Le choix des caractéristiques à extraire est essentiel dans le processus de classification et il doit être indépendant de la modalité traitée.

La fusion des images multimodales est très utilisée pour les opérations de classification[7]. Elle permet de s'affranchir des limites inhérentes aux problèmes d'utilisation des images uni-modales. Par ailleurs, cette technique permet de bénéficier de la redondance d'informations pour diminuer les incertitudes et les imprécisions, et de profiter de la complémentarité des informations lorsque les différentes modalités utilisées donnent des informations sur des aspects différents du problème étudié, afin d'avoir une version plus complète et plus générale. Bien que cette classification donne des résultats acceptables, elle n'atteint pas toujours le niveau du modèle de perception visuelle qui a une grande capacité à classifier facilement les scènes observées grâce au mécanisme puissant du cerveau humain.

Nous allons revenir sur ces points dans la partie état de l'art (Chapitre 3). Ces réflexions vérifiées dans la littérature nous ont amenés à formaliser les questions de recherche suivantes :

Q1 : comment peut-on filtrer les points aberrants en gardant un nombre suffisant de données correspondances valides pour effectuer le recalage ?

Q2 : Quelle sont les caractéristiques à utiliser comme étant des descripteurs spatiaux ? Comment les combiner avec les descripteurs spectraux dans le processus de la classification ?

Q3 : Comment choisir le formalisme adéquat ? Comment incorporer l'information perceptuelle dans le processus de fusion ?

Nous répondons, dans cette thèse, à ces trois questions par trois contributions :

Première contribution. Elle traite les limites des approches géométriques de recalage citées précédemment. Dans le but de fournir une meilleure confiance dans le recalage des images multimodales, la méthode proposée vise à détecter et filtrer les points aberrants identifiés. Une stratégie de filtrage automatique est proposée pour permettre une opération de recalage plus efficace et plus robuste. Cette stratégie intègre deux filtres qui se basent sur la distribution gaussienne des données. La première opération de filtrage est basée sur un calcul de distance entre descripteurs de points de caractéristiques précédemment extraits. La seconde implémente un filtrage basé sur l'orientation qui est situé dans une plage automatiquement définie par la distribution gaussienne des données. Cette méthode a montré son efficacité dans les cas de déformation, de translation et de changement d'échelle, mais elle a montré ses limites dans certains cas de rotation d'amplitude angulaire importante entre les données

Cette première contribution a fait l'objet d'une communication intitulée *'Improving point matching on multimodal images using distance and orientation automatic*

filtering’, Hanan Anzid, Gaëtan Le Goïc, Aissam Bekkari, Alamin Mansouri, Driss Mammass à AICCSA’2016[8].

Deuxième contribution. Cette deuxième contribution vise à surmonter la limite rencontrée dans la première contribution, qui se produit lorsque nous avons d’importantes déformations de rotation. Cette version améliorée se base dans une première étape sur un filtrage automatique basé sur la distribution gaussienne des distances calculées entre les points putatifs détectés sur les deux images (modèle et requête). Ensuite et en se basant sur le résultat de ce premier filtrage, deux points de références sont fixés (choisis) sur chacune des images (modèle et requête) puis les distances entre ces deux points sont calculées, ainsi que les descripteurs d’orientation. Enfin, un critère strict a été mis en place en utilisant les distances calculées et les orientations afin de supprimer tout point aberrant, et permettre ainsi d’améliorer la performance et la robustesse du recalage réalisé sur les données multimodales.

Cette deuxième contribution est intitulée ‘*A SURF-based methodology for robust registration of multi-modal images data*’.

Troisième contribution. Cette contribution concerne particulièrement l’extraction des caractéristiques et la pré-classification des images multimodales. Les descripteurs de la méthode SURF ont été choisis comme descripteur spatiaux en raison de la forte utilisation de la méthode, de son efficacité et sa capacité en termes d’invariance au changement d’échelle et à la rotation, et de sa rapidité. La méthode proposée est basée sur la présentation de toutes les modalités utilisées dans un espace commun en extrayant les caractéristiques SURF pour chaque pixel de l’image, appelées ‘Dense SURF’. Cette méthode exploite uniquement la partie description de la méthode SURF sans passer par sa partie de détection. Le ‘Dense SURF’ est ensuite utilisé conjointement avec les caractéristiques spectrales de l’espace CIELab. Les caractéristiques fusionnées sont ensuite fournies au classifieur SVM pour l’étape de classification.

Cette troisième contribution a fait l’objet d’une publication intitulée « *Multimodal images classification using dense-SURF* , Spectral information and Support Vector Machine”, dans *Procedia Computer Science Journal* (ISSN: 1877-0509)[9].

Quatrième contribution. Elle aborde la fusion des images multimodales et a pour but d’améliorer la tâche de classification des images multimodales. Nous proposons ainsi une méthodologie permettant de fusionner, à l’aide de DSMT (le formalisme de la théorie du raisonnement plausible et paradoxal de Dezert-Smarandache) les caractéristiques spectrales et les caractéristiques dense-SURF extraites de chaque modalité, combinées et pré-classées à l’aide du classifieur SVM. Ensuite, un modèle de perception visuelle est intégré dans le processus de fusion.

Cette quatrième contribution a fait l’objet d’une publication intitulée ‘*Improvement of Multimodal Images Classification Based on DSMT Using Visual Saliency Model Fusion*

With SVM', Hanan Anzid, Gaetan Le Goic, Aissam Bekkari, Alamin Mansouri, Driss Mammass à International Journal of computer & Technology[10].

1.3. Organisation du manuscrit

Ce manuscrit est organisé en une introduction, sept chapitres et une conclusion et perspectives.

Le chapitre 1 porte sur les images multimodales et le prétraitement. Nous présentons brièvement les étapes de l'acquisition de différentes modalités et ensuite les fondamentaux de la fusion des données ainsi que la problématique générale de ce travail.

Le chapitre 2 aborde, dans un premier temps, la problématique liée au prétraitement nécessaire concernant le recalage pour la fusion des données. Deux approches principales sont présentées : l'approche géométrique et l'approche iconique. Un accent particulier sera mis sur les algorithmes géométriques utilisés dans notre travail. Nous aborderons dans un deuxième lieu une revue de la littérature sur les modèles linéaires et non linéaires de transformation.

Puis, nous abordons les problèmes de l'extraction des caractéristiques et de la classification. Nous présentons dans un premier temps, les descripteurs spectraux dans les deux espaces RVB et CIELab en soulignant l'avantage d'utilisation de l'espace CIELab dans le cas de données multimodales. Nous décrivons ensuite les méthodes d'extraction des caractéristiques spatiales. Enfin, nous présentons une étude comparative de primitives perceptuelles sur notre base de données.

Dans le chapitre 3, nous présentons les formalismes DST et DSmT. Ensuite, nous citons les applications liées à la classification des images utilisant le processus de la fusion avec le formalisme d'évidence et celui de Dezert-Smarandache.

Dans le chapitre 4, nous présentons de façon détaillée notre première contribution, qui intègre un filtrage automatique, basé sur la distance et l'orientation des points d'intérêt extraits au niveau de la mise en correspondance avec des résultats. Une version améliorée est ensuite présentée en détail (deuxième contribution).

Le chapitre 5 est consacré à notre troisième contribution qui intègre l'information spectrale et spatiale dans la classification SVM. Les résultats obtenus sont présentés.

Dans le chapitre 6, nous présentons notre 4^{ème} contribution basée sur l'intégration de l'information perceptuelle\saillance dans le processus de la fusion et les formalismes de l'incertain.

1.4. Cadre de thèse

Cette thèse est préparée en cotutelle entre l'équipe IRF du laboratoire d'Image et Reconnaissance de Formes — Systèmes Intelligents et Communicants (IRF-SIC) de l'université Ibn Zohr au Maroc et l'équipe CORES du Laboratoire ImViA (Imagerie et Vision Artificielle) de l'Université de Bourgogne Franche Comté en France.

Ce projet de thèse a bénéficié d'un soutien financier à la mobilité dans le cadre du projet PHC TOUBKAL TBK-16-31.

Les travaux de l'équipe IRF du laboratoire IRF-SIC, visent principalement le traitement de l'image, la reconnaissance de formes, le watermarking, l'analyse de documents, reconnaissance de l'écriture, du caractère et de la signature manuscrite, l'imagerie satellitaire, l'analyse vidéo et le suivi du mouvement, la classification et la fusion des images. Nos travaux s'appuient sur l'expérience de l'équipe sur ce dernier volet.

Les travaux de l'équipe CORES sont centrés sur les différents axes touchant au domaine de la vision artificielle en utilisant l'imagerie multimodale non conventionnelle pour diverses applications (patrimoine, biomédical, industrie, etc..). Ils regroupent des thématiques de recherche liées notamment : (1) à la conception de systèmes d'imagerie et vision temps réel (2) aux outils de mesures et de diagnostics basés sur la vision (3) à l'imagerie multimodale pour la numérisation et la modélisation fonctionnelle de l'apparence des surfaces complexes, etc. Notre travail s'inscrit dans le cadre de cette dernière thématique.

Partie 1

Étude bibliographique

Chapitre 1

Images multimodales et fusion

1.1. Introduction

Bien que les images/données multimodales aient apparues depuis plusieurs décennies, leurs utilisations dans divers domaines ne sont devenues effectives qu'avec l'évolution des domaines d'applications d'imagerie notamment dans le domaine médical, le domaine satellitaire, et plus récemment dans le domaine de l'héritage et patrimoine dans lequel s'inscrivent les travaux de cette thèse, où les besoins des acteurs sont en forte croissance. Cette technique présente de nombreux avantages et son utilisation peut apporter un réel bénéfice dans de nombreux domaines d'application.

Dans ce chapitre, nous présentons dans un premier temps le contexte général de notre projet. Nous nous intéresserons dans un deuxième temps à la notion d'images multimodales et décrirons leurs caractéristiques. Dans un troisième temps, les généralités sur la fusion des données sont présentées, afin de préciser le cadre de l'étude, ses types, ses niveaux ainsi que ses étapes. Ensuite, nous abordons la problématique générale qui nous a amenés à utiliser un système de fusion et le processus adopté dans ce travail. Enfin, une conclusion sera donnée.

1.2. Les images multimodales : Acquisition et caractérisation

1.2.1. Contexte général d'acquisition des données utilisées

Dans cette section, nous décrivons la base de données utilisée dans ce travail pour évaluer l'efficacité et la robustesse des approches proposées. Comme nous l'avons mentionné, le contexte général de ce travail s'inscrit dans le domaine du patrimoine dans le sens historique du terme.

Les images acquises et utilisées dans le cadre de ce projet sont des images multimodales des peintures murales historiques du château de Germolles situées en Bourgogne (figure 1). Situé à 10kms de Chalon-sur-Saône, ce château est une propriété

privée, protégée par la loi en raison de son importance historique ; il reste néanmoins est ouvert au public toute l'année.



Figure 1 – Vue Générale du palais de Germolles

En revenant sur les aspects historiques de ce château, celui-ci a été offert par le duc de Bourgogne Philippe le Flandre à son épouse, Marguerite de Flandre en 1380. Il représentait à l'époque un petit domaine fortifié, qui a été transformé en un palais luxueux par les meilleurs artistes de l'école bourguignonne dont Claus Sluter et Jean de Beumetz. Il s'agit du dernier château des ducs de Bourgogne à avoir été si bien conservé.

Les peintures murales des chambres du palais se situent dans le premier étage de ce dernier (Figure 2) et datent de 1382 à 1392 et ont été réalisées par l'atelier de Jean de Beumetz. Ces murales ont été partiellement ou entièrement restaurées entre 1989 et 1991. À l'exception des aspects visibles du travail de conservation (comme le remplissage des trous et la réinsertion visuelle), l'ampleur de l'intervention réalisée en 1989 est totalement inconnue. La Figure 3 présente la restauration effectuée dans le dressing de Marguerite.

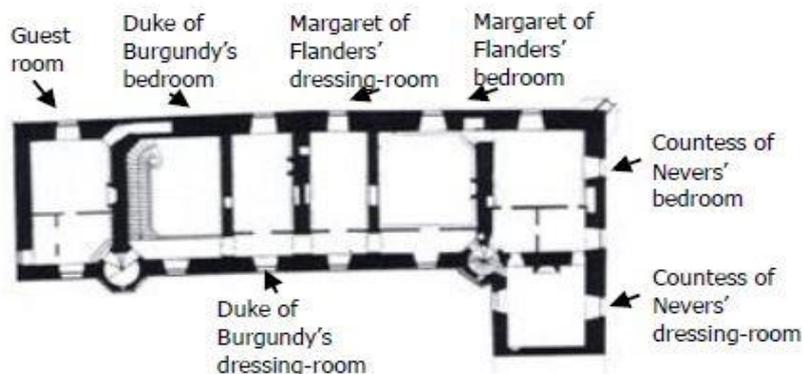


Figure 2 - Plan du palais de Germolles



Figure 3 - Exemple de restauration effectuée dans le dressing de Marguerite

Le manque de correspondance entre les archives historiques de l'époque concernant notamment la liste des matériaux, et les analyses effectuées lors de la campagne de restauration de 1989-1991 soulèvent des questions essentielles. Une des difficultés majeures que l'on peut recenser est la possibilité de différenciation entre les zones de peintures originales et les zones restaurées. En effet, des traces de la décoration d'origine subsistent. Il est donc très important de les identifier afin de déterminer et de limiter tout risque éventuel de contact direct et externe tel que l'humidité, la pollution et les dommages accidentels.

Afin de détecter les parties d'origine dans la zone restaurée, le restaurateur de Germolles a utilisé un ensemble de solutions d'imagerie multimodales, ce qui a permis d'apporter une solution rapide et relativement peu coûteuse pour l'examen de grandes surfaces de peintures murales sans causer de dommages aux surfaces étudiées.

1.2.2. Le processus d'acquisition et de caractérisation des données

L'acquisition des données a été effectuée en 2013 dans le cadre du projet européen *COST* (www.cosch.info). Ce projet consistait en un examen visuel et une documentation avec différentes modalités d'images, de portions de peintures murales du palais de Germolles (fin du XIV^e siècle). L'objectif était d'identifier les zones avec une peinture originale et d'utiliser des techniques de photographie sur des sections significatives et représentatives pour documenter ces décorations.

La technique de la photographie utilisée consistait à prendre un ensemble d'images de la même zone/scène avec un appareil photographique numérique de type *CANON EOS 5D Mark II*, avec capteur *EOS* plein cadre (36 mm x 24 mm), modifié par des professionnels afin d'utiliser pleinement la bande spectrale du capteur en supprimant le

bloqueur IR (infrarouge) standard et en le remplaçant par *DAS/AR* filtre clair pour permettre la mise au point (Figure 4). De cette manière, il est possible d'acquérir des images de la lumière visible réfléchi (Vis), de la lumière infrarouge réfléchi (IRr) et de la lumière ultraviolette réfléchi (UVr) ainsi que UV fluorescence.



Figure 4 : photographie macro-technique

Pour l'acquisition de la modalité VIS (visible light), la lumière Halogène lowel V (500 W) avec le filtre *Idas* ont été utilisés, la même lumière est aussi exploitée avec le filtre *rg830* pour obtenir la modalité infrarouge (Reflected infrared light IRr). L'utilisation de la Lampe ultraviolette à main de haute intensité mobile, *UVAHAND 250 GS HI/BL de Honle* technologie UV a permis d'acquérir avec le filtre *kv418* la modalité UV fluorescence (UV-induced fluorescence UVF), tandis que leur utilisation avec le filtre *dug11* a permis d'obtenir la modalité ultraviolet (reflected ultraviolet light), dont les modalités UVR et IRR sont composés d'une seule bande. La Figure 5 illustre un exemple de quatre modalités acquises sur la lettre 'P' sur le mur sud du dressing de la comtesse.

Chaque modalité fournit des informations sur le comportement optique de la surface lorsque les différents types de lumière l'atteignent. L'image IRR pourrait indiquer les parties refaites sur une surface verte de la scène. La modalité de fluorescence induite par les UV (UVF) montre une fluorescence relativement forte correspondant aux restes d'une couche de peinture ancienne sur le blanc de la lettre 'P' tout comme l'image UVR qui permet d'identifier l'original repeint sur le blanc du 'P'.

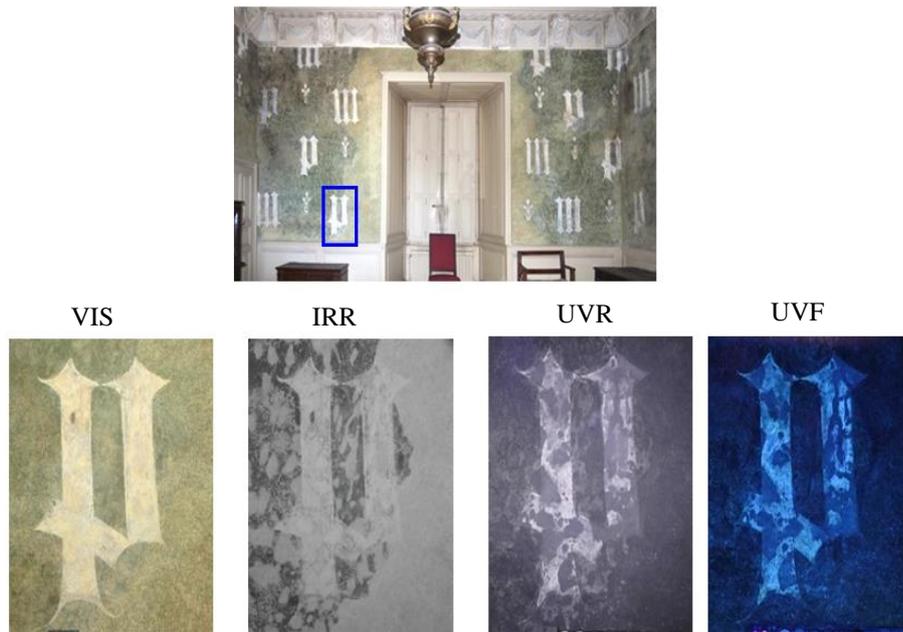


Figure 5: Exemple de quatre modalités acquises sur la lettre 'P' sur le mur sud du dressing de la Comtesse

1.3. Généralité sur la fusion

1.3.1. Définition

La fusion des données avec des méthodes numériques et automatisées est apparue récemment [11]. Sa première définition a été proposée par le regroupement JDL (Joint Directors of Laboratories) [12] : *'A process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats, and their significance. The process is characterized by continuous refinements of its estimates and assessments, and the evaluation of the need for additional sources, or modification of the process itself, to achieve improved results.'*

Cette notion a été introduite dans les domaines qui disposent d'information multiples[13] tel que le domaine militaire [14] dont plusieurs informations de différentes sources (radar, caméra visible, caméra infrarouge ...) peuvent être fournies pour des opérations de surveillance par exemple. De même, le domaine médical et celui du patrimoine sont aussi concernés par le besoin de traiter conjointement les informations multiples. Dans la fusion des images, plusieurs définitions existent dans la littérature et

nous proposons d'adopter celle de *Bloch* [15] : ‘‘la fusion d'images est l'utilisation conjointe d'images hétérogènes pour l'aide à la prise de décision’’, qui été améliorée par *I. Bloch* juste après pour être plus spécifique [16] : ‘‘ la fusion d'informations consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision’’. Toujours en optant pour la définition présentée par *Isabelle Bloch*, une définition plus spécifique au domaine d'imagerie est donnée par *I. Bloch* et al. [17]: ‘‘la fusion d'images est l'utilisation conjointe d'images hétérogènes pour l'aide à prise de décision’’.

Un processus de fusion doit satisfaire les conditions suivantes [13]:

- L'hétérogénéité : les informations sont issues soit de différents capteurs soit d'un même capteur, mais alors utilisé dans des conditions différentes.
- L'utilisation conjointe des informations : chaque source peut fournir une décision séparée, qui peut être combinée pour prendre une décision globale.
- L'aide à décision : En les combinant, les informations fournies doivent aider à la prise de la décision ou à son amélioration.

1.3.2. Les types d'imperfections

Les informations issues de différentes sources/images sont entachées d'imperfections. Les causes sont nombreuses et multifactorielles. Elles peuvent être liées directement à la technologie des capteurs, c.à.d. être liée à leur mécanisme d'interaction physique avec la surface imagée, ou bien être liées aux conditions d'acquisitions. Les informations entachées de ces imperfections peuvent fournir un aspect même partiel de la réalité qui doit être modélisée. D'où l'intérêt du processus de la fusion qui réside dans l'amélioration de l'information en utilisant tout type de donnée.

Il existe plusieurs types d'imperfections, dont les principales sont [18] [19] :

- **L'incertitude** [20] : Elle exprime le degré de conformité de l'information à la réalité. Elle est ainsi liée d'une façon subjective à l'opinion de l'expert sur la réalité. Une information jugée certaine présenterait une connaissance totale de la réalité[18]. Elle est souvent présente dans les images [17], et correspond

généralement à une ignorance soit totale ou partielle de la connaissance. Ce type d'imperfection est bien modélisé par les approches probabilistes.

- **L'imprécision** [21]: Elle exprime le manque d'exactitude de la connaissance en terme quantitatif. Deux types se distinguent, l'imprécision sans erreur, qui est relative à l'opinion de l'expert (ce qui la rend difficile à estimer) ; et l'impression avec erreur, qui traduit un défaut quantitatif mesurable. Ce type d'imperfection a été mis en évidence par les théories des sous-ensembles flous et la théorie de croyance.

L'incertitude et l'imprécision[22] sont duales et nous ne pouvons que favoriser l'un au détriment de l'autre [13]. Ces deux types d'imperfections traduisent de manière rigoureuse l'ignorance totale ou partielle, l'imprécision étant relative au contenu de l'information tandis que l'incertitude est relative à sa vérité.

- **Le conflit** [23] [18] [22] : Deux informations sont conflictuelles si elles sont contradictoires. Autrement dit, le conflit se présente lorsque nous avons plusieurs informations qui conduisent à des interprétations paradoxales. Ce type d'imperfection peut être provoqué par plusieurs raisons :
 - Le manque de fiabilité des sources utilisées (lorsque chaque source fournit une information différente de l'autre sur la même situation donnée)
 - Les informations sont acquises dans différentes conditions, par exemple, Jour, Nuit
 - La non-exhaustivité du cadre de discernement utilisé.
 - Les caractères inconsistants des informations fournies par les différentes sources

La modélisation de conflit est une étape contraignante dans la majorité des formalismes dédiés à effectuer la fusion. Cependant, l'approche de *Dezert-Smarandache (DSMT)* [24] permet de modéliser ce type d'imperfection assez aisément.

- **L'incomplétude** [25] [18] [22] : traduit l'absence d'information fournie par une source sur une situation donnée. Elle peut être provoquée par l'incertitude et/ou par l'imprécision. Ceci peut être mesuré si nous disposons d'une connaissance

générale et explicite sur la quantité de l'information qui doit être fournie par la source, la mesure est alors obtenue en calculant la différence entre la quantité de l'information réellement fournie et celle qui doit être normalement fournie.

Cependant, la quantification de l'information est une tâche difficile et la mesure la plus utilisée est celle de *Shanon* [26] connue sous le nom d'entropie de *Shanon*.

- **Redondance et la complémentarité** [127] [129] [22] : l'information est dite redondante si elle est fournie par plusieurs sources sur une situation donnée . Dans le cas de la fusion des informations cela peut permettre de confirmer une information et ainsi de réduire le degré d'incertitude/d'imprécision. A l'inverse, la complémentarité est réalisée lorsque les sources fournissent des caractéristiques différentes ce qui conduit à une vision plus complète sur la même situation donnée.
- **Ambiguïté** [16] [18]: l'information est dite ambiguë si elle supporte au moins deux interprétations. Nous ne pouvons alors plus choisir entre les décisions. Elle peut être provoquée par d'autres imperfections telles que l'incertitude, l'imprécision, ou encore le conflit.

1.3.3. Les niveaux de fusion

La fusion des données consiste à combiner les données dans le but d'améliorer une situation donnée. Selon *Dasarathy* [27], cette combinaison peut se faire en trois niveaux à savoir (Figure 6):

- **La fusion au niveau pixel** : Considérée comme étant une fusion bas niveau, elle consiste à fusionner les données brutes[28] [29] ou très proches des paramètres physiques mesurés. Dans les images multimodales, l'établissement d'une correspondance directe entre les pixels de différentes modalités est appliquée [30]. La fusion bas niveau mène à une amélioration de la résolution spatiale [31] (détails de la structure), tout en assurant la fidélité des données spectrales originales [32]. Dans ce cas, l'étape de recalage est très importante en prétraitement de l'étape de fusion, car chaque pixel doit avoir la même position physique, d'une manière très précise, que son correspondant dans les autres images.

- **La fusion au niveau objet/caractéristiques** [33]: la fusion à ce niveau est effectuée après l'extraction des caractéristiques, des objets et /ou des descripteurs pertinents [34]. Les caractéristiques extraites peuvent être des régions, les coins des bords, des contours, des lignes, des structures de textures ou les composantes fréquentielles [28] [35]. La fusion à ce niveau consiste à utiliser les primitives extraites au lieu des données d'origine ce qui entraîne une perte d'information importante. Cela permet en retour l'utilisation d'un référentiel commun pour évaluer les attributs [19]. Le recalage n'est pas forcément contraignant à ce niveau en se basant sur la mise en correspondance uniquement de l'objet en question avec le référentiel commun connu auparavant.
- **La fusion au niveau décision** : elle consiste à combiner les décisions issues de chaque source/image fournies par un expert ou un système. Selon *Martin* [18], la fusion à ce niveau permet de combiner les décisions des classificateurs préliminaires. Dans ce cas, chaque source ou image est traitées individuellement, ses caractéristiques sont extraites séparément, puis classifiées. Enfin , les décisions de chaque classifieur sont combinées et fusionnées avec les autres en introduisant la modélisation des imperfections rencontrées, dans le but d'obtenir une décision finale [36].

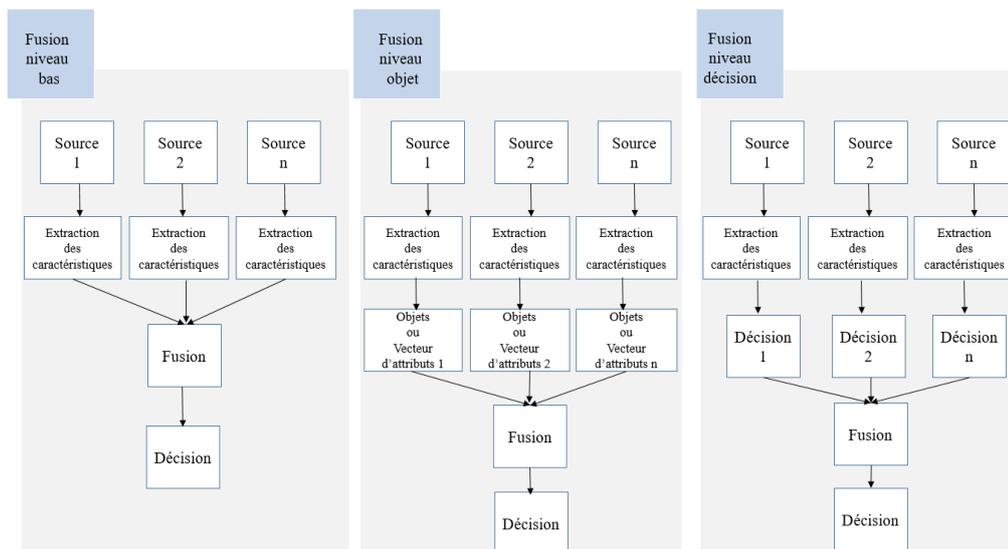


Figure 6 Les niveaux de fusion

1.3.4. Les étapes de la fusion

Le processus de la fusion des images selon les théories de l'incertain consiste en quatre étapes [13] [19] [33]:

- **Modélisation** : la modélisation des données est la première étape dans le processus de la fusion. C'est à cette phase que s'effectue le choix du formalisme en se basant sur les théories de probabilité, les théories des ensembles flous et de possibilités, la théorie de croyance ou la théorie de *Dezert-Smarandache (DSmT)* (choix adopté dans ces travaux). Nous détaillerons ces formalismes dans la section suivante.
- **Estimation** : A ce stade, chaque élément d'information est représenté par un nombre ou une masse. Selon le formalisme choisi, cette représentation peut prendre la forme d'une probabilité, d'un degré de possibilité ou d'une fonction de masse.
- **Combinaison** : cette étape est considérée comme étant le cœur de la fusion car à ce niveau les informations sont regroupées en utilisant une opération de fusion ou une règle de combinaison en fonction du formalisme choisi ainsi que des types d'imperfections modélisées.
- **Décision** : c'est la dernière étape qui consiste à prendre une décision optimale sur un évènement donné en utilisant les règles dédiées, toujours selon la théorie appliquée.

1.4. Problématique générale et méthodologie pour la fusion

1.4.1. Problématique générale

L'utilisation de plusieurs sources d'information, sous forme de plusieurs images de différentes modalités, doit permettre de fournir une classification plus fiable. En effet, la classification fondée sur une seule image/source d'information peut être associée à un risque d'erreur important car une image ne donne qu'un aperçu partiel et/ou déformé de la réalité.

Chaque image (ou modalité) permet de fournir des informations différentes, qui peuvent parfois être incomplètes, redondantes, complémentaires ou même conflictuelle/paradoxe par rapport aux informations des autres modalités. Les méthodes de classification classiques cherchent à éliminer ces informations via les prétraitements tels que par exemple la suppression du bruit, et à superposer les images sans prendre en

compte l'hétérogénéité des données, d'où l'intérêt d'adapter un système de fusion basé sur les théories qui modélisent ces informations multimodales et qui offrent les outils permettant de les combiner pour aider à prendre une décision.

Dans ce travail, nous cherchons à proposer une solution pour la classification des peintures murales de château de Germolles (cas d'application), qui ont fait l'objet d'une campagne d'acquisition par imagerie multimodale récemment [37]. Ces peintures ont subies une restauration entre 1989 et 1991, dont nous ne disposons que de peu d'information.

Chaque modalité d'imagerie permet de bien visualiser ces deux parties soit d'une manière redondante et/ou complémentaire ou bien conflictuelle. Ainsi, la question posée par cette application est comment avoir une carte de classification finale qui combine/fusionne ces informations.

Pour répondre à cette problématique, nous avons besoin d'un système de fusion de l'information flexible pour gérer toutes les quantités très importantes de données, robuste et tolérant aux imperfections qui peuvent être présentes dans ces données. Ceci nous a amenés au choix d'utilisation des théories de l'incertain, afin de combiner des vues partielles et d'obtenir une vue complète en profitant de la totalité des informations hétérogènes disponibles, tout en assurant la gestion des imperfections. La figure 7 décrit le problème rencontré et la nécessité de l'utilisation de la fusion des données pour ce type de problématique.

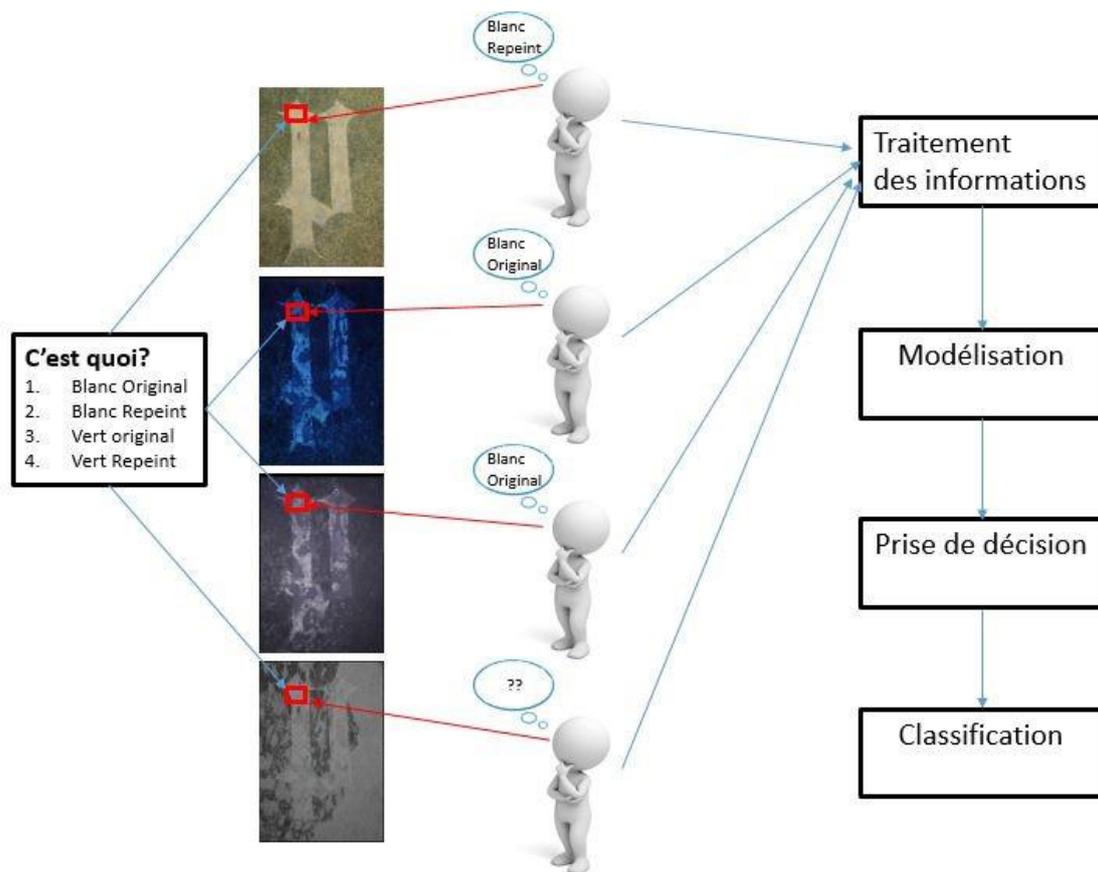


Figure 7 – la problématique générale de la fusion des données

1.4.2. Le processus de fusion

Le processus de la fusion des images commence généralement par un prétraitement consistant à recalcr les images/données multimodales entre elles. Dans cette étape nous avons rencontré notre premier problème causé par la grande différence entre les différentes modalités à apparier. Ensuite, la deuxième étape de ce processus est l'étape de l'extraction de caractéristiques et enfin la classification. Le résultat issu de chaque pré-classification de chaque modalité est ensuite modélisé et combiné afin de prendre une décision finale. Ce processus est décrit en Figure 8.

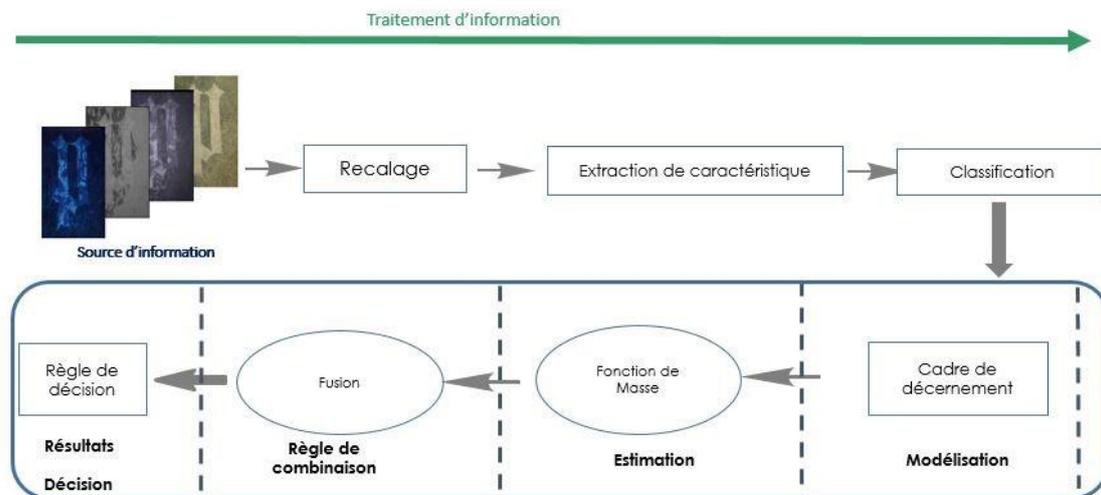


Figure 8 – le processus de la fusion des images

1.5. Conclusion

Dans la première partie de ce chapitre, nous avons présenté une description des données multimodales utilisées, leurs acquisitions, leurs caractéristiques et la problématique globale de leurs utilisations conjointement. Ensuite, nous avons présenté les généralités sur la fusion des données, les définitions de la fusion des images, les différentes imperfections qui peuvent être présentes dans l'information acquise ainsi que les niveaux et les étapes du processus de fusion des données.

Le processus de la fusion adopté est réalisé au niveau pixel (niveau bas) ce qui nécessite d'effectuer un recalage performant et une pré-classification de chaque pixel. Ces deux axes sont détaillés dans le chapitre suivant.

Chapitre 2

Le recalage et l'extraction des caractéristiques

2.1. Introduction

Lors d'une acquisition multimodale, chaque image contient généralement des informations différentes qui peuvent être parfois des informations redondantes et complémentaires pour une autre modalité. Leur utilisation conjointe peut présenter un avantage significatif moyennant l'établissement de correspondance entre les pixels des différentes modalités. En effet, le processus de fusion exige que chaque pixel présente le même emplacement physique sur toutes les modalités (étape de prétraitement d'alignement ou du recalage). Cette étape du recalage est souvent complexe à réaliser dans le cas de données multimodales à cause du changement non linéaire d'intensités entre les différentes modalités.

L'étape de recalage est suivie par l'étape du choix des caractéristiques à extraire. L'extraction des différents types de caractéristiques permet de tenir compte de la diversité des données, et amène aussi à la nécessité de projeter cet ensemble d'images dans un espace de représentation commun.

Dans ce chapitre, nous abordons la problématique de recalage dans sa globalité puis dans le cas spécifique de données multimodales. Nous expliciterons par la suite les problèmes rencontrés et les solutions envisagées et mises en œuvre. Puis, nous traitons l'extraction des caractéristiques spectrales dans divers espaces de représentation. Ensuite nous présentons le cas de l'extraction de caractéristiques spatiales en nous appuyant sur les caractéristiques *SURF*. Enfin, nous présentons les primitives perceptuelles en comparant plusieurs méthodes sur notre base de données.

2.2. Le prétraitement : le recalage

2.2.1. Définition du recalage

Le recalage est une tâche importante dans de nombreuses applications de vision par ordinateur telles que la stéréovision [38], le mosaïquage [39], la détection de changement [40] ou la fusion de données [41].

De nombreuses définitions associées à la notion de recalage ont été proposées dans la littérature, relativement hétérogènes. Dans le cadre de ces travaux nous proposons de retenir celle de L. Brown qui définit le recalage de la façon suivante [42] :

“ The registration problem is the task involved in finding the optimal spatial and intensity transformations so that the images are matched with regard to the misregistration source. The intensity transformation is frequently not necessary, except, for example, in cases where there is a change in sensor type or where a simple look up table determined by sensor calibration techniques is sufficient. “

Autrement dit, le recalage vise à établir une correspondance entre deux ou plusieurs images, issues d'une scène obtenue par un ou plusieurs capteurs, relativement à des positions et à des échelles spatiales différentes.

D'un point de vue du formalisme mathématique, Si **I1** est l'image de référence et **I2** l'image requête (l'image à recalier- exemple de la Figure 9), le recalage peut être exprimé mathématiquement par l'équation 1.1 :

$$I1(x, y) = k(I2(x', y')) \quad 2.1$$

Où (x, y) et (x', y') représentent respectivement les coordonnées spatiales sur les images **I1** et **I2** (les points en rouges sur la Figure 9), et k symbolise la fonction radiométrique.

L'objectif du recalage consiste à trouver une transformation spatiale optimale et radiométrique entre les deux images, noté $T(\cdot)$, vérifiant l'équation 1.2 :

$$(u, v) = T(x', y') \quad 2.1$$

Avec u et v représentant les coordonnées spatiales transformées en utilisant T , de telle sorte que les pixels dans les deux images représentent les mêmes points physiques après l'application de la transformation.

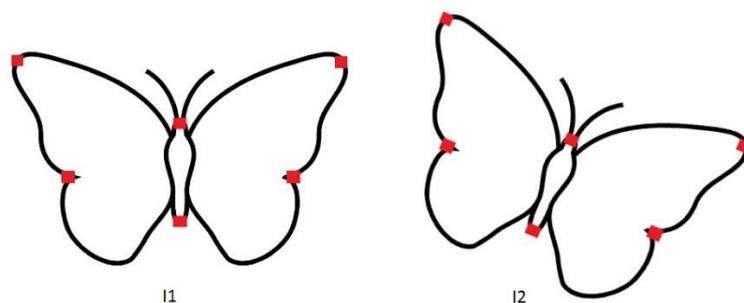


Figure 9 Exemple du problème du recalage

La problématique du recalage réside dans l'identification de ces paires de points correspondants, sur lesquels la transformation entre l'image référence et l'image requête sera calculée. Ainsi, le processus de recalage peut être résumé en trois étapes [19] [43]:

1. Détection et appariement des primitives communes aux images à aligner,
2. Identification de la transformation, en se basant sur les caractéristiques extraites,
3. La mise en place de la transformation pour réaliser l'alignement de l'image requête.

2.2.1. Extraction des caractéristiques

L'extraction des caractéristiques est une étape essentielle pour le succès du recalage. Elle permet de définir les paramètres de la transformation appropriée. Il existe deux principales catégories [44] : les approches iconiques et les approches géométriques.

2.2.1.1. Approches iconiques

Les approches iconiques sont des algorithmes itératifs qui se basent sur l'intégralité des pixels des images [45]. Cette approche utilise les comparaisons locales de l'intensité, d'une manière automatique, pour déterminer les points de correspondances sur les deux images, en optimisant la métrique de similarité [46].

Ces méthodes ne nécessitent pas une réduction préliminaire des données, par segmentation comme exemple. A contrario, elles utilisent l'information dense contenue dans chaque pixel de l'image, ce qui est avantageux dans certaines applications. Cependant, cette méthode présente aussi un inconvénient majeur en ce qui concerne le coût en temps de calcul dans le cas des données de grande dimension. En effet, la relation entre l'intensité de l'image requête et celle de l'image modèle est influencée par la relation entre les deux images, la Figure 10 illustre les différentes relations possibles entre les images. Cette relation est souvent difficile à déterminer quand il s'agit d'un changement non linéaire de l'intensité, ce qui relève un autre inconvénient des méthodes iconique, et rend cette approche peu appropriée au cas de données multimodales.

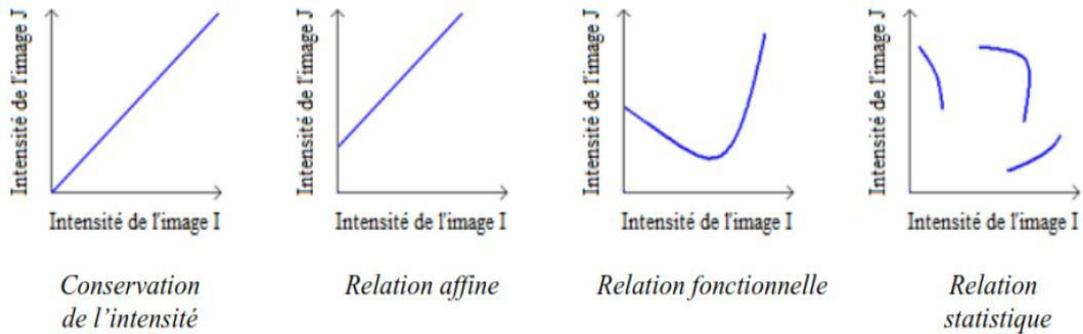


Figure 10 les relations de l'intensité entre l'image modèle et l'image requête[47]

On peut distinguer deux types de recalage dans cette catégorie, le recalage monomodal où les deux images (l'image modèle et l'image requête) sont issues de mêmes capteurs/source, et (2) le recalage multimodal, où les images sont acquises de différentes sources. Diverses métriques de similarité peuvent être implémentées selon le type du recalage [48].

(1) Cas du recalage monomodal

Les images utilisées dans le recalage monomodal se caractérisent par une relation d'intensité similaire entre les mêmes structures imagées [49], vu l'utilisation du même capteur pour l'acquisition des deux images.

Le bruit d'intensité sur les images détermine le choix entre la métrique SSD (*Sum Square Difference*) [50] [51] et SAD (*Sum Absolute Difference*) pour mesurer la similarité entre les deux images. Quand il s'agit d'une relation linéaire entre les intensités de deux images les métriques optimales sont alors la corrélation croisée (*Cross Correlation CCor*) et le coefficient de corrélation (*Correlation Coefficient CCoef*)[42] [52].

(2) Cas du recalage multimodal

Dans le cas du recalage de données multimodales, les images utilisées sont généralement très différentes au niveau de la distribution de l'intensité [53][54], ce qui rend le choix de la métrique de similarité difficile. Dans ce cas, deux méthodes ont été proposées [55] :

- Les méthodes de la théorie de l'information et plus spécialement l'entropie et l'information mutuelle (MI) [56].
- Les méthodes de statistique globales telles que le rapport de corrélation [57].

Pour répondre aux problèmes liés à la complexité du cas de multimodalité, une solution alternative a été adoptée. Elle consiste à ramener ce problème au cas du recalage monomodal, à travers deux approches :

- La transformation d'une modalité en une autre de telle sorte que les deux images présentent la même modalité. Cette approche nécessite une connaissance préalable des propriétés physiques du capteur utilisé.
- La réalisation d'un mapping entre les deux modalités dans un domaine de représentation commun.

La transformation qui permet de maximiser la métrique de similarité entre l'image référence et l'image requête produira un recalage plus efficace. Par conséquent, le choix de la métrique influe fortement sur les performances du processus du recalage.

Dans ce travail, nous utilisons des images multimodales de grandes tailles, présentant des imperfections et du bruit, et qui se caractérisent par une relation d'intensité non linéaire entre elles. Ceci soulève le problème du coût de calcul d'une part, et la complexité à déterminer la métrique de similarité d'autre part.

Ces aspects couplés au fait que nous nous basons sur l'information d'intensité, qui fournit une description incomplète sur un pixel et son voisinage, mènent souvent à une correspondance ambiguë [58], et nous ont amenés à utiliser les méthodes géométriques, détaillées dans la section suivante.

Les Figures 11 et 12 montrent le changement non linéaire de l'intensité sur les différentes modalités utilisées de la scène présentée sur la Figure 5.

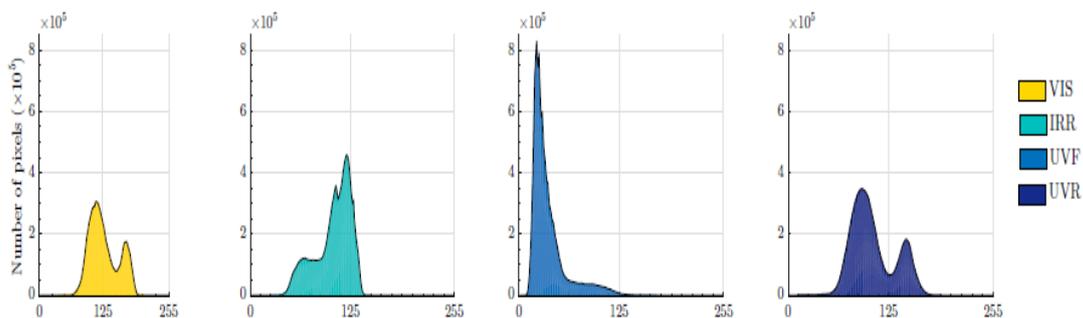


Figure 11 Exemples de l'intensité globale des images multimodales correspondant à la même scène ((Figure5)

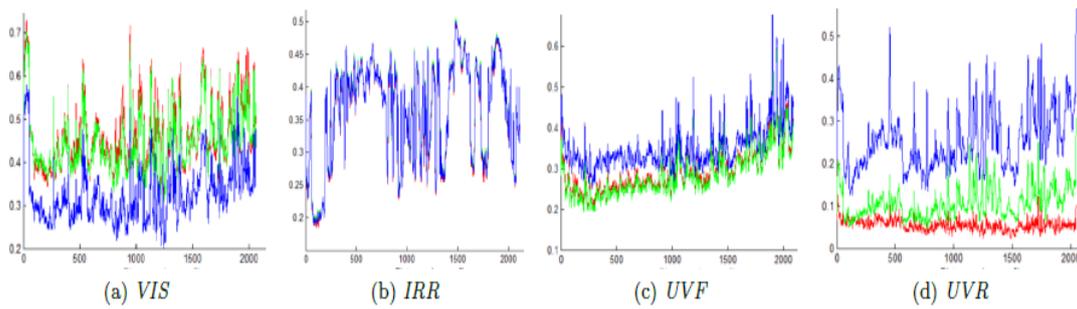


Figure 12 - Exemples de l'intensité locale des images multimodales extraites au long de la même ligne pour les quatre modalités

2.2.1.2. Approches géométriques

Les approches géométriques [48] permettent de répondre à certaines limites des méthodes iconiques. Elles reposent sur un ensemble de caractéristiques extraites localement, telles que les points d'intérêts, les bordures, les contours, l'orientation, la forme, etc. La mise en correspondance de ces caractéristiques est déterminée à partir d'indicateurs de robustesse [59].

Dans ce cas, le processus particulier du recalage peut être défini en quatre étapes [60] [61] :

2. *Extraction des caractéristiques* : cette étape se base sur la détection des primitives fines [48] afin de garantir une bonne robustesse de la méthode, puis elle réalise ensuite leur extraction sous forme de descripteurs.
3. *La mise en correspondance des caractéristiques* : Cette étape consiste à chercher la meilleure correspondance entre l'ensemble des caractéristiques détectées sur l'image requête (c'est-à-dire l'image à recaler) et l'ensemble des caractéristiques détectées sur l'image modèle (ou image de référence). La meilleure correspondance est réalisée en comparant chaque caractéristique de l'image requête et l'ensemble des caractéristiques du modèle. Cette comparaison est effectuée en fonction de leur relation spatiale et des descripteurs invariants calculés à l'étape précédente. Les paires de points correspondants sont ainsi identifiées et les meilleures correspondances sont sélectionnées pour le calcul de la transformation T.
4. *L'estimation de la transformation* : Dans cette étape, la transformation géométrique à appliquer sur l'image requête est déterminée à partir des meilleurs points correspondants identifiés dans l'étape 2.
5. *Application de la transformation et ré-échantillonnage* : L'image requête est transformée par la fonction de transformation établie à l'étape 3. Ensuite, si

nécessaire, un ré-échantillonnage/interpolation est appliqué pour fournir l'image recalée.

La détection et l'extraction des points de correspondances fiables présentent une problématique ouverte et un domaine de recherche actif [62] dans cette catégorie et plusieurs méthodes ont été proposées dans ce cadre. Nous citons, dans ce qui suit, les plus connues et les plus utilisées d'entre elles, en détaillant le principe et les étapes des méthodes *SURF* et *SIFT* que nous utilisons dans ces travaux.

- **Iterative Closest Point ICP :**

La méthode du plus proche voisin [63] est très répandue pour le recalage des données *3D*. Elle consiste, tout d'abord, à extraire des caractéristiques sur les deux images ou les deux nuages de points puis à mettre en correspondance les primitives extraites. La transformation est ensuite calculée d'une manière itérative en minimisant la distance entre les points appariés (ou les deux nuages de points) jusqu'à la convergence de l'algorithme (c'est-à-dire, lorsque l'erreur résiduelle de la distance entre ces points sera inférieure à un seuil donné).

- **Scale-Invariant Feature Transform SIFT :**

Scale-Invariant Features est un descripteur local invariant à l'intensité proposé par Lowe[64], qui est conçu pour la détection et la description des caractéristiques locales dans les images. L'algorithme SIFT est divisé en trois étapes principales :

- ✓ *Construction d'un espace d'échelle* : Cette construction est d'abord effectuée en appliquant une séquence de filtre gaussien sur l'image en augmentant itérativement la valeur du paramètre σ . Chaque valeur de σ correspond à une échelle ce qui construit ainsi un ensemble d'échelles par octave, puis un échantillonnage est appliqué sur l'image pour avoir l'octave suivante. Ensuite la différence de Gauss (DoG) est calculée entre chaque image successivement afin de déterminer les points extrêmes sur les images filtrées (Figure 13).

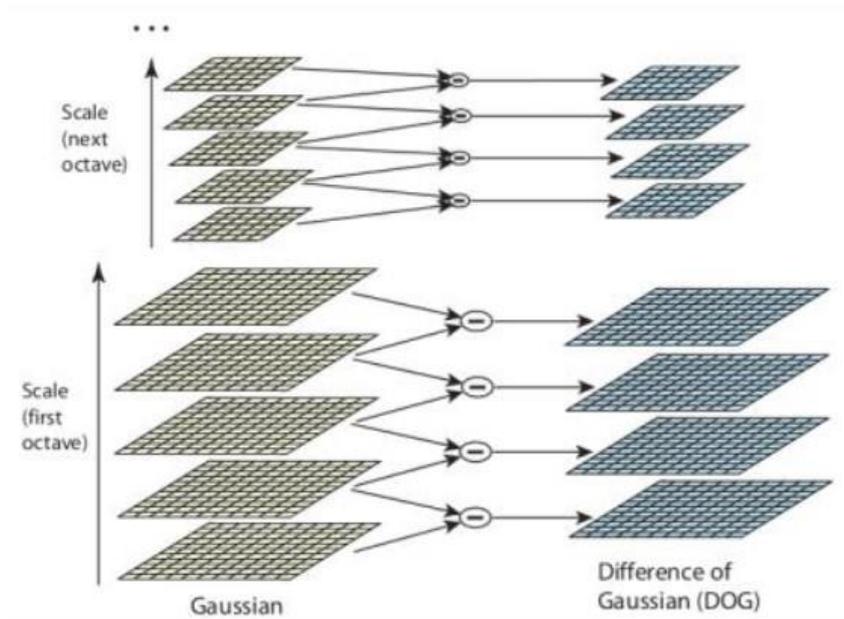


Figure 13 les différentes octaves de l'image dans la pyramide de Gauss [64]

Les points d'intérêt sont alors identifiés en comparant chaque point de l'échantillon avec ses 26 voisins (la taille du voisinage est définie à $3 \times 3 \times 3$). Un point qui a une valeur supérieure à celle de son voisinage est considéré comme un point d'intérêt.

- ✓ *Attribution de l'orientation* : dans cette étape, un calcul de la magnitude et de la direction de l'image filtrée sur l'échelle du point d'intérêt est effectué. Puis, un histogramme de 36 bins chacun de 10 degrés (couvrant 360 angles) est créé. En utilisant les gradients d'image, chaque bin représente le nombre de pixels ayant un degré dans cette plage. L'orientation du descripteur du point intérêt est ensuite associée au bin avec la plus grande valeur (Figure 14).

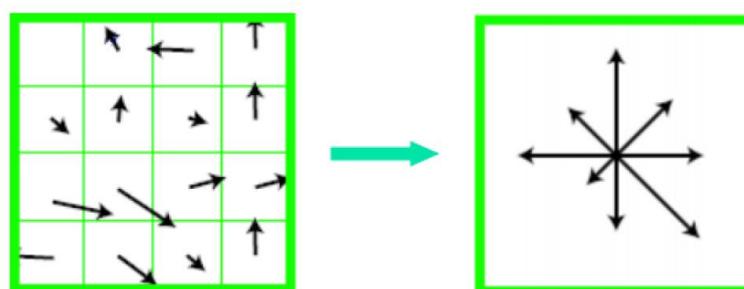


Figure 14 l'assignement de l'orientation au descripteur SIFT[64]

- ✓ *Calcul de descripteur SIFT* : cette étape consiste à dériver le descripteur en calculant l'orientation relative et la magnitude dans un voisinage 16×16 . Le voisinage est divisé par 4 sous-blocs, ensuite un histogramme de 8 bins est calculé pour chaque bloc. Les histogrammes sont concaténés pour former la dimension de descripteur $4 \times 4 \times 8 = 128$.

Plusieurs approches de recalage basées sur SIFT ont été proposées. *Firmenichy* et al. [65] ont proposé une détection de coins sur les images multispectrales basées sur SIFT afin d'aligner une image infrarouge sur une image en lumière visible et d'améliorer les performances d'extraction des points d'intérêts. Le descripteur Symmetric-SIFT est une extension de SIFT, qui est symétrique par rapport au contraste, développé par *Jian* et al. [48] pour surmonter le problème lié à la recherche sur l'espace de transformation globale en fonction du désalignement des images. Mais l'invariance à la multimodalité se fait au prix de la perte d'informations importantes. De plus, Symmetric-SIFT souffre encore de la génération d'un grand nombre de points aberrants (outliers), ce qui rend non robuste le calcul, et donc peut nuire à la qualité du recalage. Pour remédier à cet inconvénient, *Tanvir Hossain* et al. [48] ont proposé une amélioration de Symmetric-SIFT [48] dans laquelle les différences d'orientation globale entre l'image modèle et l'image requête sont estimées en se basant sur l'analyse de l'appariement initial. La nouvelle valeur estimée est utilisée dans la normalisation et la reconstruction des descripteurs au lieu d'utiliser l'orientation dominante. Un autre travail basé sur SIFT est proposé par *Xia* et al. [44], qui définit un algorithme consistant à construire une représentation de congruence de phase PCR (Phase Congruency Representation) de l'image requête, puis à obtenir des points caractéristiques importants en appliquant le détecteur SIFT. La correspondance est effectuée à l'aide de l'intégration de la description SIFT dans une méthode cohérente de déviation de point (*Coherent point drift CPD*) pour garder les meilleures correspondances en combinant l'apparence avec les propriétés de distance de la correspondance.

Dès que la méthode SIFT a été proposée, plusieurs versions et extensions ont été formulées [48]. Pour s'affranchir de la limite de coût de calcul, la méthode SURF, a été proposée dans le but de réduire la dimension du vecteur de descripteur qui décrit le point d'intérêt de 128 à 64 éléments et rend la méthode moins distinctive et plus rapide lors de la phase de la mise en correspondance. La méthode de recalage utilisée dans ce travail est basée sur cette méthode SURF, qui permet d'extraire rapidement des points d'intérêt robustes et invariants au changement d'échelle, d'illumination et de rotation, ce qui est appropriée dans le cas de données multimodales, tout en assurant la détection d'un nombre suffisant des points d'intérêts partagés à différentes positions pour permettre une étape de correspondance rapide et répétable. Le détail de ce processus est décrit ci-dessous (paragraphe SURF).

- **Speeded Up Robust Feature SURF** : Cet algorithme proposé par *H. Bay* [66] est une version améliorée de SIFT en termes de vitesse. Il est caractérisé par sa robustesse et son invariance relative vis-à-vis des changements d'échelle, de la rotation, de la translation et le changement d'illumination [19]. La technique SURF se base dans les calculs sur ce qui est appelé l'image intégrale $I_{\Sigma}(x)$, qui représente la somme des pixels de l'image d'entrée située au-dessus et à gauche du pixel à la position $(x ; y)$, comme décrit dans l'équation (1.3) et illustré dans la Figure 15 :

$$I_{\Sigma}(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j < y} I(i, j) \quad 2.3$$

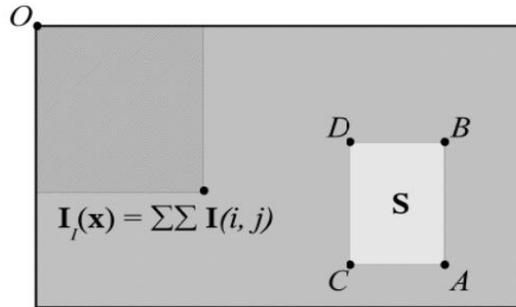


Figure 15 Principe de l'image intégrale [66]

Cette technique aide à incrémenter et décrémenter le temps de calcul [61]. De plus, elle présente un avantage en termes de rapidité et de précision [67]. Une fois l'image intégrale calculée, les principales étapes de cette approche sont les suivantes :

- ✓ *Détection et localisation de points d'intérêt* : Une fois que l'image intégrale est calculée, une approximation de la matrice de Hessian est utilisée (équation 1.4) pour détecter les points d'intérêt, notés IP dans ce document. $L_{xx}(X, \sigma)$ représente la convolution de la dérivée gaussienne du second ordre sur l'axe x avec chaque point X dans l'image (de même pour $L_{xy}(X, \sigma)$ et $L_{yy}(X, \sigma)$).

$$H(X, \sigma) = \begin{pmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{pmatrix} \quad 2.4$$

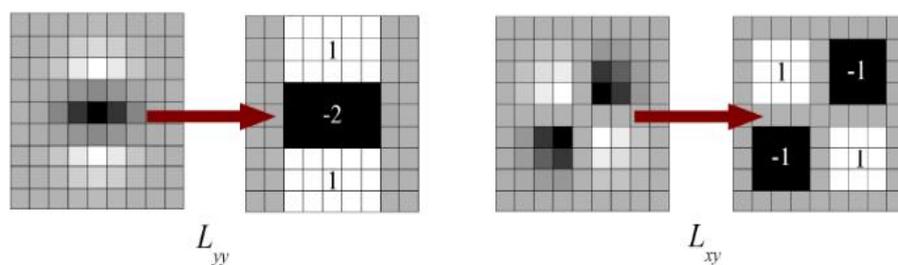


Figure 16 Box-Filtre [66]

Une approximation simple des dérivées gaussiennes du second ordre est ensuite appliquée, (Figure 16). Celle-ci est utilisée sur l'image intégrale pour calculer le déterminant de Hessian, afin de rendre le détecteur invariant aux changements d'échelle. Le point doit être détecté à différentes échelles, puis une "carte de réponse blob" à différentes échelles est construite en augmentant la taille du filtre (Figure 17). Les points d'intérêts où le déterminant de la matrice

Hessian qui sont positifs et qui sont des maxima locaux dans un voisinage de $3 \times 3 \times 3$ sont conservés.

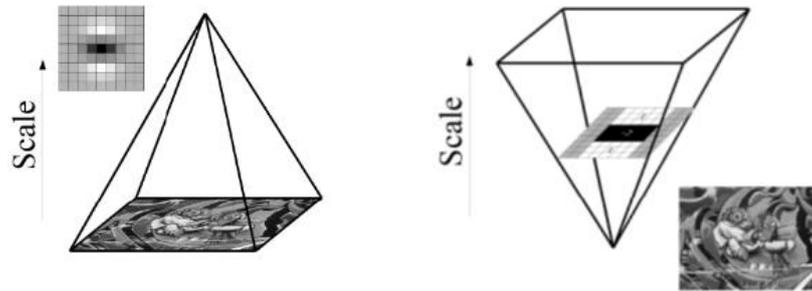


Figure 17 Analyse d'échelle avec une taille constante de l'image [66]

- ✓ *L'attribution de l'orientation* : Cette étape consiste à déterminer l'orientation dominante dans la région autour des points d'intérêt précédemment identifiés, afin de rendre la méthode invariante à la rotation d'image. Ceci est effectué en calculant les réponses en ondelettes de Haar. Les réponses en ondelettes de Haar utilisent l'image intégrale et le BOX-Filter pour chaque point d'intérêt de x et y dans un voisinage circulaire de rayon $6s$ (où s est l'échelle dans laquelle le point d'intérêt a été identifié) tout en tenant compte de la nature discrète de l'image. Les résultats des réponses sont ensuite pondérés par une fenêtre gaussienne centrée sur le point intérêt considéré, puis représentée sous forme de vecteurs. La somme de toutes les réponses situées dans une fenêtre de taille $\frac{\pi}{3}$ tournant autour du centre de la région d'intérêt est alors calculée (Figure 18). Enfin, la direction du vecteur le plus long est attribuée en tant qu'orientation du point d'intérêt.

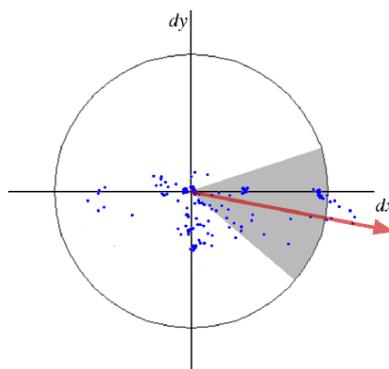


Figure 18 Exemple de la fenêtre coulissante d'orientation de taille $\frac{\pi}{3}$ [66]

- ✓ *Description du point d'intérêt* : Pour construire les descripteurs SURF sur chaque point d'intérêt, une zone rectangulaire est formée autour du point d'intérêt orienté dans la direction principale sélectionnée à l'étape précédente.

La zone est ensuite divisée en sous-régions carrées plus petites de 4×4 . Dans chaque sous-région, les réponses aux ondelettes de Haar, dx et dy , sont calculées à partir de 5×5 points d'échantillonnage, où dx représente les réponses aux ondelettes de Haar dans le sens horizontal et dy dans le sens vertical. dx et dy sont pondérés avec un noyau gaussien. Ensuite, les réponses et leurs valeurs absolues sur chaque sous-région pour dx et dy ($\sum dx, \sum dy, \sum |dx|, \sum |dy|$) sont résumées et un vecteur composé de $4 \times 4 \times 4$ valeurs est construit, formant le descripteur SURF de 64 dimensions. Enfin, le descripteur SURF est normalisé en un vecteur unitaire pour être invariant au facteur d'échelle.

- ✓ *La mise en correspondances des points d'intérêt* : Une mesure de similarité est estimée à partir des descripteurs SURF de deux images, en calculant une distance entre l'ensemble des descripteurs de la requête et les points d'intérêt de l'image modèle, afin de déterminer la meilleure correspondance. Le voisin le plus proche avec la distance euclidienne minimale pour le vecteur descripteur invariant est qualifié comme point d'intérêt correspondant.

Pour le recalage multimodal, de nombreuses recherches se sont basées sur l'utilisation de SURF, telles que MM-SURF [68], qui est une méthode robuste qui tente d'aligner des images multimodales en lumière visible et infrarouge, en introduisant dans SURF un concept de gradient similaire permettant de générer un grand nombre de points d'intérêt robustes. SURF-PIIFD est une autre approche qui a été introduite par Wang et al. [69] pour recalibrer les images multimodales rétinienne complexes, en utilisant la méthode SURF pour détecter les points d'intérêts plus répétables et invariants aux changements d'échelle et le descripteur PIIFD. Dans [70], les auteurs ont utilisé la méthode SURF pour représenter d'une manière compacte une séquence d'images et les signatures spectrales audio suivies par l'utilisation d'une fenêtre glissante (DTW) afin de calculer les alignements temporels, puis l'utilisation à nouveau de l'algorithme SURF pour générer ces alignements. Une utilisation conjointe d'Histogramme de gradient orienté (HOG) avec le descripteur SURF a été mis en place par Patel et al. [71] dans le but de surmonter le problème de correspondance incorrecte causée par les variations d'illumination dans les images satellitaires multimodales, multicateur ainsi que multi-temporel. Lukashevich et al. [72] ont présenté une approche basée sur le détecteur SURF pour extraire les caractéristiques sur les images CT, dans différentes séries d'étude d'un patient, afin d'effectuer un recalage efficace. Su et al. [73] ont proposé une combinaison de la méthode SURF avec le flot optique pour améliorer le recalage des images multimodales cérébrales, en commençant par l'utilisation de la spécification d'histogramme dans le but de transformer la modalité d'images avant le recalage. Puis SURF a été utilisé pour extraire les caractéristiques afin d'obtenir un recalage préliminaire de l'image. L'algorithme de flot optique est ensuite appliqué pour avoir des résultats plus précis.

L'utilisation de différentes modalités d'acquisition produit une différence significative d'intensité acquise par le capteur [74]. Ceci est confirmé dans notre cas d'acquisition par les histogrammes d'intensité globale (Figure 11) et les intensités locales (extractions le long du même profil linéaire - Figure 12). De plus, les variations d'intensité entre les quatre modalités d'acquisition sont clairement non linéaires. Cela constitue un problème à la fois pour les étapes de mise en correspondance et de recalage. En effet, les descripteurs construits à partir de l'étape d'extraction des caractéristiques sont d'une importance primordiale pour l'opération d'appariement, et l'extraction des caractéristiques est elle-même profondément affectée par les changements d'intensité de voisinage du point de calcul, qui diffèrent clairement d'une modalité à une autre. Par conséquent, on peut s'attendre à ce que les mêmes points clés soient décrits par des descripteurs très différents et génèrent des valeurs incohérentes qui altéreraient la performance du recalage.

2.2.2. Le modèle de transformation

L'extraction des caractéristiques est une phase déterminante pour avoir une information sur la nature de la transformation à appliquer sur l'image requête. Généralement, deux catégories de modèles de transformation sont distingués[55] :

2.2.2.1. Modèle paramétrique

Le modèle paramétrique nécessite une formulation explicite de la déformation et à son tour, il peut être divisé en modèle global ou semi-local, dont chaque paramètre influence la totalité de l'image. De plus, les deux modèles peuvent être classifiés en rigide ou non rigide.

Le modèle rigide est largement utilisé dans les applications d'imagerie médicale. Il consiste à prendre en compte les déformations de la rotation et de translation, tout en conservant les angles, les distances et le parallélisme. Tandis que le modèle affine corrige le positionnement en prenant en considération les déformations de la rotation, de translation, de cisaillement ainsi que de l'étirement, en conservant en retour le parallélisme, ce qui le rend plus approprié dans le cas de données multimodales. Nous avons adopté ce dernier modèle dans ces travaux. Le modèle semi-local, dit aussi non linéaire et non rigide permet la modification de la forme globale des structures en autorisant leur variation géométrique.

2.2.2.2. Modèle non paramétrique

Inversement, le modèle non paramétrique est considéré comme étant une solution d'un problème d'optimisation, dont chaque paramètre agit sur une région restreinte de l'image. Le modèle élastique est parmi les modèles de cette catégorie, dans lequel le problème du recalage est considéré comme étant une déformation élastique [55] [75], ce

modèle utilise une modélisation par un polynôme de degré supérieur à 2, tout en assurant la régularité et la conservation de la topologie. L'estimation de grande déformation reste la limite majeure de ce modèle. Ceci a motivé l'utilisation du modèle fluide par [76] [77] et qui consiste à considérer l'image comme un fluide dans un cadre eulérien, ce qui soulève le problème du coût de calcul. Le modèle de diffusion est un des modèles non paramétriques existants, il peut être utilisé comme une grille de déformation dont les contours de l'image sont orientés de l'intérieur vers l'extérieur de l'objet [55].

2.3. L'extraction des caractéristiques

2.3.1. L'information Spectrale et Spatiale

2.3.1.1. Les descripteurs spectraux

Se baser uniquement sur des descripteurs spectraux seulement pour la classification peut permettre d'obtenir des résultats satisfaisants, notamment dans le cas d'images satellitaires et d'images multimodales. Les images multimodales sont caractérisées par la haute résolution obtenue sur une bande étroite des canaux qui les composent et dont chaque thème/classe a une signature spectrale qui peut permettre une bonne classification. En outre, chaque pixel est considéré comme étant un vecteur d'attributs qui seront directement utilisés comme une entrée du classificateur. L'information couleur est le descripteur spectral utilisé dans ces travaux. Cette information peut être représentée dans des modèles mathématiques en tant qu'attributs de couleur. Sa représentation dans l'espace RVB et CIELab est présentée dans cette section.

2.3.1.1.1. L'espace RVB

Les informations spectrales couleur généralement acquises par des appareils photographiques numériques sont enregistrées en combinant les attributs additifs rouge, vert et bleu dans un espace tridimensionnel. Cet espace, appelé espace RVB, est utilisé dans la plupart des applications qui utilisent les images couleur [78]. Le principe de cette espace [79] est que la combinaison de ces trois attributs primaires de base définit un trièdre qui correspond à une chromaticité spécifiée, la Figure 19 présente le cube de l'espace RVB. Cependant, cet espace présente plusieurs inconvénients tels que la corrélation entre les primaires, le métamérisme [80], la non représentation des longueurs d'onde non visibles, ainsi que le problème de la non-conformité à la perception humaine.

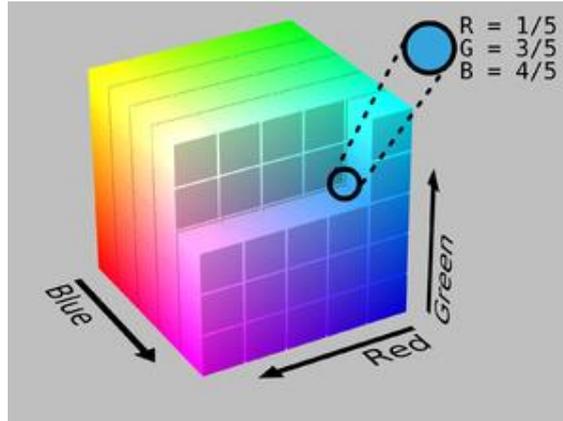


Figure 19 le Cube RVB (source Wikipédia)

2.3.1.1.2. L'espace CIELab

Dans l'espace CIELab (L^* , a^* , b^*), la luminance et la chrominance sont représentées séparément : un attribut représente la luminance et les deux autres représentent la chrominance. Cet espace défini par le CIE (Commission Internationale de l'Eclairage) fournit une meilleure corrélation avec la perception humaine. Le modèle CIELab [81] utilise la luminosité (L) d'une couleur comme une fonction ($L=0$ représente le noir et $L=100$ le blanc). La transformation des informations de couleur de l'espace RVB à cet espace (a ; b) peut être effectuée par les équations (2.1) et (2.2). Une illustration de cet espace est donnée dans la Figure 20.

$$a^* = \frac{Red}{Green} \quad 2.5$$

$$b^* = \frac{Yellow}{Blue} \quad 2.6$$

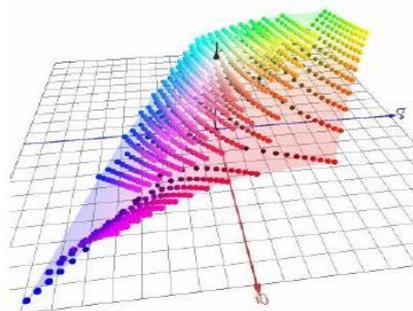


Figure 20 : Représentation de l'espace CIELab (source Wikipédia)

L'utilisation de cette espace de représentation a montré sa pertinence dans diverses applications. Cependant, l'information spatiale contenue dans les différentes structures de l'image n'est pas prise en compte.

2.3.1.2. Les descripteurs spatiaux

Les descripteurs spatiaux sont associés aux intensités des pixels pris par blocs. Ils visent à tenir compte des relations spatiales entre les pixels dans un voisinages [80]. Ces descripteurs spatiaux peuvent être classés en descripteurs locaux et descripteurs globaux.

- Les descripteurs globaux cherchent à décrire l'image dans son ensemble. Par exemple :
 - **Le descripteur de la texture** : une définition formelle de la texture [82] n'existe pas dans la littérature, cependant ce descripteur permet en utilisant des méthodes structurelles statistiques et spectrales de fournir des mesures des propriétés de la texture [82]. Les descripteurs de Haralick et GLCM sont les plus utilisées :
 - Les descripteurs GLCM permettent de calculer quatre caractéristiques afin d'estimer le mouvement des images[83] à travers
 - *Le deuxième moment angulaire* connu aussi sous le nom de *l'énergie* et qui mesure l'homogénéité de l'image,
 - *Le moment de différence inverse* (MDI) qui mesure l'homogénéité locale,
 - *L'entropie* qui est une mesure de l'information et de sa structuration dans l'image,
 - La corrélation qui mesure la dépendance linéaire des niveaux de gris des pixels voisins.
 - Les descripteurs de Haralick permettent de mesurer l'information sur les motifs de la texture en calculant les matrices de co-occurrence. Leur estimation est coûteuse en termes de temps de calcul. Cette matrice permet de calculer 13 caractéristiques (dites de Haralick) [84] qui sont l'énergie, la corrélation, l'inertie, l'entropie, le moment de

différence inverse, la somme moyenne, la somme variance, la moyenne de différence, la différence variance, l'entropie par différence, la mesure d'information de la corrélation 1 et la mesure d'information de la corrélation 2.

- **Les descripteurs de la forme** : ce sont des descripteurs très utiles pour la correspondance de la similarité. Ils sont invariants au changement d'échelle, aux changements de rotation ainsi qu'aux translation [85]. Cette catégorie contient quatre types de descripteurs qui sont :
 - le descripteur de forme-spectre (histogramme de l'indice de forme)
 - les descripteurs basés sur les moments invariants tels que de Zernike et de Hu, pour extraire les caractéristiques invariantes
 - les descripteurs de forme basée sur le contour dont les caractéristiques peuvent être décrites en utilisant *MPEG7*
 - les descripteurs basés sur les contours se basant sur les représentations de contours par échelle de courbure (CCS), etc.
- Les descripteurs locaux représentent les zones intéressantes dans l'image telles que les bords d'objets. Les descripteurs utilisés dans cette catégorie sont SIFT et SURF (présentés en détail dans la première partie), mais aussi MSER, BRISK et FREAK
 - MSER : ce descripteur permet d'extraire les blobs/régions de covariants dans l'image. En outre, ils représentent une composante stable et connectée de certains ensembles de niveaux de gris de l'image. Ce descripteur a montré sa performance dans les cas de changements de point de vue et d'éclairage. Ce descripteur n'est pas conçu pour les applications en temps-réel.
 - BRISK [86] : ce descripteur permet aussi de détecter et d'extraire les points d'intérêts dans les images dont aucune information préalable sur les scènes n'est disponible. Il est basé sur le calcul d'une gaussienne pondérée sur un motif en utilisant des configurations simples et faciles d'échantillonnage circulaire autour d'un point. Il est largement utilisé dans les applications en temps-réel.

- FREAK [87] : Ce descripteur permet d'extraire des points d'intérêts en s'inspirant de la rétine. Il est basé sur les comparaisons d'intensité par paires. Les descripteurs extraits sont sous forme d'une chaîne binaire, dont les cercles concentriques se chevauchent en provoquant la redondance.

Dans ce travail, nous utilisons les descripteurs locaux car, d'une part, nous n'avons pas d'informations de texture ou de forme sur la base de données utilisée. Et d'autre part, les descripteurs locaux sont préférables dans le cas de classification car ils permettent de représenter les caractéristiques dans une zone réduite, ce qui peut permettre une description plus robuste de l'information qu'en se basant sur l'image entière. À cet effet, nous avons utilisé le descripteur local SURF d'une manière dense pour représenter toutes les modalités utilisées dans un espace de représentation commun. Notre motivation pour le choix du descripteur SURF est basée sur les avantages de stabilité du descripteur, ainsi que pour son efficacité en termes de calcul et ses performances [86].

Plusieurs travaux ont utilisé la méthode SURF pour extraire les caractéristiques dans le but d'effectuer la classification. Nous commençons par *Kim* et al. [88] qui ont utilisé le descripteur local SURF dans le but de générer les vecteurs de caractéristiques avec un classifieur SVM pour effectuer la classification dans un premier temps entre les deux classes face et background, et ensuite pour effectuer un étiquetage entre les composantes yeux, bouche et nez. La méthode SURF a été employée par *Schmitt* et al. [89] qui ont présenté une technique basée sur SURF pour avoir une classification en extrayant les caractéristiques et en les associant à une base de données prétraitée. Les auteurs utilisent une structure pyramidale pour la localisation d'objets.

La combinaison de descripteur SURF et de classifieur de mots visuels (Visual Words) a montré son efficacité dans l'étude menée par *Alfanindya* et al. [90], qui ont proposé l'utilisation du descripteur SURF avec le sac des mots visuels (BOVW) dans le but d'effectuer la classification. En ce sens, *Slavomir* et al. [91] ont effectué une classification SVM basée sur une combinaison des méthodes SURF et SIFT comme un détecteur hybride. Ce détecteur hybride a été combiné avec le classifieur de mots visuels (BOW) dans le but d'améliorer la classification et la reconnaissance des objets animaux.

Dans le domaine de reconnaissance de forme 3D, *Knopp* et al. [92] ont proposé une méthode de classification et de reconnaissance des classes de formes dans les données 3D basées sur les descripteurs locaux au lieu des descripteurs globaux, qui sont dominantes dans la littérature sur la classification/reconnaissance des formes 3D. Leur méthode est basée sur l'utilisation de SURF dans le contexte des formes 3D combiné avec les méthodes probabilistes basées sur la transformation de Hough.

2.3.2. L'information perceptuelle & la saillance

Les images multimodales constituent un enjeu conséquent (avec de nombreux défis scientifiques) dans le traitement des images pour plusieurs raisons que nous avons expliquées précédemment. Cependant, ce domaine intéresse de nombreux chercheurs en raison de l'avantage significatif apporté dans certains domaines par la multimodalité de l'information. Notamment, l'acquisition, le traitement et l'analyse de données multimodales présente un intérêt fort pour les problèmes de perception, et particulièrement de détection et de saillance visuelle.

Plusieurs définitions de la saillance visuelle existent dans la littérature. Il s'agit de prédire ce qui est susceptible d'attirer l'attention de l'homme [93] [65] [95]. Cette information peut être un bénéfice pour une grande variété d'applications du monde réel, en segmentation d'images [96], en reconnaissance d'objet [97] [98] [64], en suivi visuel [99], en compression image/vidéo [100], ou encore en détection d'objets [97].

L'extraction de la saillance est ainsi un outil précieux pour de nombreuses applications. Généralement, les travaux existants répondent exclusivement à une application spécifique et aux jeux de données disponibles. Nous nous contentons dans cette section de présenter une analyse comparative de la détection de la saillance pour un ensemble de données d'images multimodales.

Les méthodes de détection de la saillance peuvent être divisées en deux types : le problème de détection de saillance Bottom-Up et le problème de détection de saillance Top-Down

- Les méthodes de détection de la saillance Bottom-Up ne nécessitent aucune connaissance préalable de l'image et sont basées sur une attention particulière aux caractéristiques visuelles, telles que l'intensité, la couleur, l'orientation des

bords et la direction du mouvement. Les méthodes présentées dans [101] [94] sont des exemples de mise en œuvre de ce modèle.

- Les méthodes de détection de la saillance Top-Down exigent des connaissances a priori sur le contexte pour identifier les régions saillantes, en s'appuyant sur des tâches visuelles de haut niveau, telles que la recherche d'un objet spécifique qui capte intentionnellement l'objectif et l'attention des observateurs, tels que des visages, des humains, ou encore des voitures. Un exemple de la littérature appartenant à cette catégorie est le travail de *Torralba* et al. [102][103] [104].

Globalement, les méthodes de détection de la saillance Bottom-Up sont regroupées en quatre catégories principales [95] :

- **Modèle basé sur l'espace** : il est basé sur la sélection de caractéristiques de motivation biologique, suivie du calcul des différences entourant le centre des primitives sélectionnées (couleur, bord, intensité). Le modèle d'Itti [101][94] est l'exemple séminal.
- **Modèle basé sur la fréquence** : ce modèle utilise un domaine spectral au lieu d'une image véritable. Le modèle bien connu de cette catégorie est le modèle SR [105], qui repose sur une notion de résidu spectral. Il utilise le spectre logarithmique de l'image dérivée du spectre de Fourier moyen.
- **Modèle basé sur les objets** : ce modèle repose sur la théorie voulant que l'attention soit attribuée à l'objet perceptuel créé par les règles de la Gestalt, telles que la proximité, la similarité, la fermeture, la continuation, etc. Le modèle de Bruce et al présente un travail pionnier de ce type de modèles qui utilise le principe de maximisation de l'information [106].
- **Modèle basé sur des graphes** : ce modèle est basé sur le fait que l'attention visuelle sélective sélectionne différentes caractéristiques de l'objet, ainsi que la notion de l'information est plus puissante que l'échantillonnage attentif. GBVS présente le modèle classique [107].

Les méthodes de la catégorie Top-Down comprennent deux étapes principales : l'apprentissage par dictionnaire et le calcul de la saillance [108]. Cependant, cette

catégorie retient toujours moins d'attention des chercheurs en raison de la complexité d'imiter un processus cognitif d'attention visuelle complexe [109].

Dans ce travail, étant donné la nature des images dont nous disposons, nous avons choisi d'utiliser les méthodes de type Bottom-Up. Nous détaillons ci-après les méthodes utilisées dans l'étude comparative que nous avons effectuée.

Dans [94], *Itti* propose un modèle antérieur et classique reflétant ce qui attire probablement l'attention de l'œil humain. Ce modèle est largement utilisé pour calculer la zone de saillance dans une image donnée et est considéré comme un guide pour plusieurs travaux en tant qu'outil de comparaison. Les auteurs ont utilisé les caractéristiques de couleur, d'intensité et d'orientation comme information primitive. Les caractéristiques extraites sont traitées de manière parallèle sur plusieurs échelles spatiales après un filtrage linéaire. Ensuite, les cartes de visibilité sont calculées par le centre entouré. La valeur moyenne de ces cartes est calculée pour obtenir une carte de saillance finale.

R. *Bharath* [110] présente un cadre évolutif qui repose sur une méthode basée sur les objets pour la classification des scènes à l'aide d'arbres de décision. Il a proposé une détection automatique de ROI afin de fournir une découverte d'objets saillants guidée, utilisée comme marqueur de premier plan pour la segmentation d'objet par découpage de graphe. L'étape de classification des objets est ensuite réalisée à l'aide d'un ensemble de caractéristiques contenant les descripteurs SIFT et ceux issus de la SVM. Dans cette étape, L'auteur intègre une correspondance pyramidale spéciale pour améliorer la précision des objets. Une classification des scènes est ensuite réalisée à l'aide d'un arbre de décision construit à l'aide d'images d'apprentissage étiquetées.

Dans [111], *Hou* et al. suggèrent un algorithme pour séparer le premier plan de la partie arrière-plan de l'image. La ségrégation est effectuée à l'aide de la fonction de signe de la transformée en cosinus discrète (DCT) d'une image. Ensuite, la carte de saillance est générée en calculant l'inverse de la transformée en cosinus discrète (IDCT) pour extraire la partie au premier plan de l'image qui est généralement considérée comme la plus importante. L'approche est testée pour les canaux de couleur RVB et CIELab et ce dernier produit de meilleurs résultats.

Zhang et al. présentent dans [100] un nouveau modèle de saillance de la carte booléenne. Le modèle proposé s'appuie sur l'indication de voisinage comme principe de Gestalt, qui tient compte de la relation topologique entre la figure et le fond de l'image, afin d'appliquer une séparation invariante à la transformation. Dans ce modèle, une image est caractérisée par un ensemble de cartes booléennes obtenues à partir de canaux de caractéristiques sélectionnés. Sur ces cartes, une technique d'image binaire est appliquée pour mettre en évidence le contour des régions, afin de calculer une carte d'attention. Ensuite, une combinaison linéaire de la carte d'attention est appliquée pour obtenir une carte d'attention moyenne qui peut être utilisée avec certains post-traitements pour fournir la carte de saillance.

Dans [112], les auteurs proposent un nouveau type de saillance par la détection de points de fixation ou d'objets dominants basée sur des informations environnantes locales et globales. L'algorithme repose sur quatre principes issus de la littérature psychologique :

- des considérations locales de bas niveau telles que le contraste et la couleur
- des considérations globales qui visent à conserver les caractéristiques qui dérogent à la norme et donnent à un niveau inférieur un contexte immédiat
- des règles organisationnelles visuelles regroupant les pixels saillants au lieu de les répartir dans l'image
- des facteurs de haut niveau tels que les objets reconnus ou la détection de visage appliqués en post-traitement.

Le Moan et al. proposent dans [80] une nouvelle méthode de détection de la saillance pour les images multi/hyperspectrales inspirée du modèle de *Itti*, à des fins de comparaison, de visualisation et d'interprétation de ces images. Ce modèle est réalisé en trois étapes. La première étape consiste en des comparaisons centre-surround dans lesquelles les centres et ses voisinages sont définis, puis comparés à l'aide de quatre mesures de comparaison : distance euclidienne (norme L2), valeurs trichromatiques CIELab, angle entre pixels et image filtrée par Gabor du premier composant principal. La réflectance, les composantes principales et les pixels filtrés par Gabor sont les caractéristiques utilisées par ces mesures, qui sont ensuite comparées en utilisant la

distance ou l'angle euclidien. La deuxième étape est la normalisation qui vise à concentrer la luminosité dans un emplacement moins important sur une carte fusionnée. Et la dernière étape est des images en basse dimension.

Dans [105] *Hou* et al. proposent un nouveau modèle qui utilise le domaine spectral (au sens fréquentiel) de l'image au lieu de l'image véritable. Le principe de l'algorithme consiste à utiliser le résidu spectral généré par la transformation logarithmique par différence et la forme généralisée de l'image, afin de supprimer les informations redondantes et de ne mettre l'accent que sur les nouvelles informations. La carte de saillance est générée en appliquant une transformation de Fourier inverse sur le résidu spectral et les valeurs sont au carré et lissées par une Gaussienne.

Rahtu et al. [113] proposent une nouvelle méthode de segmentation basée sur un objet saillant pour une image et une séquence vidéo. Cette méthode utilise des contrastes de luminance locaux pour les primitives locales, de couleur et de motivation contrôlée par le paramètre de lissage avec un cadre statistique afin de fournir une mesure cartographique de la saillance. Une mesure de saillance régularisée est ensuite introduite pour la rendre plus robuste au bruit puis cette mesure est combinée à un champ aléatoire conditionnel (CRF) afin d'atteindre le but de la segmentation.

2.4. Conclusion

Dans la première partie de ce chapitre, nous avons décrit le principe de fonctionnement du recalage des images, suivi par une étude bibliographique de ces approches, en s'appuyant sur les approches géométriques avec lesquelles nous avons effectué le recalage dans nos travaux. Les différents modèles de transformation sont ensuite présentés.

La deuxième partie est partagée en 3 axes principaux : nous avons commencé par les méthodes d'extraction des caractéristiques spectrales avec la mise en évidence de l'avantage d'utiliser l'espace CIELab dans le cas des images multimodales objets de ce travail. Ensuite nous avons abordé les caractéristiques spatiales tout en citant quelques travaux qui utilisent les descripteurs SURF pour la classification. Enfin, nous avons traité les informations perceptuelles accompagnées d'une description comparative de quelques méthodes de saillance.

Chapitre 3

La fusion des images basée sur les théories

DST et DSMT

3.1. Introduction

Lorsque l'on ne dispose que d'une seule modalité d'acquisition, la qualité de la classification, par exemple, peut être rapidement altérée par des phénomènes de bruits, des informations incomplètes ou encore des informations déformées. Certaines de ces limitations peuvent parfois être surmontées en couplant plusieurs modalités d'imagerie. Chaque pixel de la surface est alors représenté par plusieurs valeurs, associées à chacune des modalités d'acquisition, ce qui permet de disposer d'informations différentes, parfois redondantes ou au contraire complémentaires du fait de la diversité des technologies des capteurs et de leurs mécanismes d'interaction physique. L'utilisation combinée de cet ensemble d'information présente ainsi un avantage significatif dans de nombreux cas d'applications, et peut notamment permettre d'améliorer la qualité de la classification [114].

Cependant, comme indiqué précédemment, les données présentent généralement des imperfections (conflit, ignorance, incertitude, imprécision, ou encore incomplétude) qui doivent être traitées et prises en compte par un formalisme dédié. Ainsi, plusieurs théories existent pour combiner ces informations numériquement.

Nous présentons dans ces travaux le cadre théorique de la fusion, en se basant sur les formalismes DST et DSMT. Notre choix d'utiliser le formalisme DSMT ainsi que les travaux basés sur DST et DSMT en vue d'application dans la classification est ensuite motivé.

3.2. Formalisation du problème de fusion

Depuis l'apparition de la fusion des données, les théories qui traitent ces données et modélisent les imperfections continuent à évoluer dans le but de pallier aux problèmes rencontrés en fonction des domaines d'application. On peut citer de façon non exhaustive :

- Les théories de probabilités[115] sont applicables dans divers domaines. Elles sont performantes pour la modélisation de l'incertitude,
- les théories des ensembles flous [116] et de possibilités [20],
- la théorie de l'évidence [117] qui permet de bien représenter l'incertitude et l'imprécision en trouvant un compromis entre les deux,
- La plus récente théorie de Dezert-Smarandache[118] [24] qui permet de représenter des données conflictuelles.

Dans cette section nous développerons de façon plus détaillée le formalisme de la théorie de Dezert-Smarandache choisi dans le cadre de ces travaux.

Dans ce travail, en termes de vocabulaire, chaque modalité est considérée comme étant une source fournissant de l'information. Ainsi, le problème de la fusion de ces sources peut être exprimé de la manière suivante :

Soit m sources $S_j (j = 1, \dots, m)$ qui fournissent une information sur un élément x (qui peut être soit un pixel ou un objet). Chaque source S_j prend une décision d_i dans un ensemble de n décision d_1, \dots, d_n , pour que l'élément x appartienne à une classe C_i de l'ensemble des n classes présentées dans le cadre de discernement $\Theta = \{C_1, C_2, \dots, C_n\}$. La décision obtenue d_i est représentée par $M_{ji}(x)$, elle relie l'élément x à une classe C_i .

Dans le cas d'une fusion globale, aucune information n'est négligeable, car la décision est prise à partir de la matrice des $M_{ji}(x)$ pour chaque source et pour l'ensemble des décisions C_i .

3.3. Théorie de Dempster-Shafer (DST)

La théorie de l'évidence a été introduite par *Dempster* [119] sur les probabilités supérieure et inférieure, puis utilisée et développée par *Shafer* [117] en montrant son efficacité au niveau de la représentation des informations incertaines [120] ainsi que l'ignorance. Cette méthode repose sur la distribution des jeux de masses pour affecter un degré de confiance à une hypothèse simple ou bien la réunion d'un ensemble d'hypothèses dans le cas d'ignorance.

- **Cadre de discernement**

Le cadre de discernement présente l'ensemble des réponses possibles pour un problème donné noté $\Theta = \{H_1, H_2, \dots, H_n\}$. Ce cadre est composé de n hypothèses exhaustives de telle sorte qu'une seule solution H_i du cadre de discernement est nécessaire et exclusive en exigeant l'unicité de la solution avec

$$H_i \cap H_j = \emptyset, \forall i \neq j.$$

- **Modélisation**

La théorie d'évidence permet la représentation de l'ignorance de l'information et l'affectation des degrés de confiance à la réunion des hypothèses dans le cas où la connaissance disponible présente un ensemble sans plus de précision.

L'ensemble $2^\Theta \in [0,1]$ est déduit du cadre de discernement et représente l'ensemble de toutes les disjonctions possibles. Il contient tous les sous-ensembles A_k qui peuvent constituer une hypothèse unique ou bien une union de plusieurs hypothèses, où les fonctions de masses sont définies. La modélisation dans cette théorie s'effectue en assignant les masses notées m_j aux éléments de 2^Θ appelé aussi 'Basic belief assignment BBA' en satisfaisant les conditions suivantes :

$$\sum_{A_k \subset 2^\Theta} m_{A_k}^{S_j}(x) = 1 \quad 3.1$$

$$m(\emptyset) = 0 \quad 3.2$$

- **Estimation**

Il n'existe pas une méthode générique reconnue dédiée à l'estimation des jeux de masses et cela dépend de l'application en question. La masse m qu'un élément x appartienne à une classe A_k selon la source S_j est définie par :

$$M_{A_k}^{S_j}(x) = m_{A_k}^{S_j}(x) \quad 3.3$$

À partir d'une fonction de masse normalisée, on peut alors notamment calculer les fonctions suivantes :

- La fonction de crédibilité : elle modélise l'incertitude sur un ensemble A_k et peut être interprétée [121] comme étant le degré d'évidence que nous avons sur A_k :

$$Cr_{A_k}^{S_j}(x) = \sum_{B \subset A_k} m_B^{S_j}(x) \quad 3.4$$

- La Fonction de plausibilité : elle modélise la probabilité supérieure qui traduit le degré pour lequel l'évidence n'est pas contradictoire avec A_k .

$$Pl_{A_k}^{S_j}(x) = \sum_{B \cap A_k \neq \emptyset} m_B^{S_j}(x) \quad 3.5$$

Ces deux dernières grandeurs présentent des bords inférieurs et supérieurs pour qu'un degré d'évidence soit affecté à un sous-ensemble de cadre de discernement, ce qui est exprimé par l'équation 3.6 :

$$Pl_{A_k}^{S_j}(x) = 1 - Cr_{A_k}^{S_j}(\bar{x}) \quad 3.6$$

- **Combinaison**

L'étape de la combinaison consiste à fusionner les masses estimées de sources multiples en utilisant une règle de combinaison. Pour ce faire, il existe plusieurs opérateurs dont la règle de base de *Dempster* [117] qui est une sommation orthogonale [13] basée sur une règle conjonctive [122]. Dans le cas de deux sources, cette règle s'écrit de la manière suivante :

$$m(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i)m_2(B_j)}{1 - K} \quad 3.7$$

Avec $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)$ qui représente le coefficient de conflit entre les sources m_1 et m_2 .

- **Décision**

La décision peut être prise en se basant sur [13] :

- Le maximum de crédibilité, la décision est alors dite pessimiste.
- Le maximum de plausibilité (choix dit optimiste).
- Un intervalle de confiance $[Cr, Pl]$. Celui-ci exprime le mieux la connaissance que nous possédons sur une proposition donnée par rapport aux décisions prises séparément. Ce choix permet de chercher la proposition la moins incertaine ; Cependant, celle-ci ne peut être qu'une proposition simple et non pas une proposition composée.
- La probabilité pignistique: elle est considérée comme étant un compromis entre le maximum de crédibilité et celui de la plausibilité[18].

Le choix de la règle de décision à adopter dépend de l'application étudiée et du résultat désiré ainsi que les attentes des utilisateurs, s'ils s'intéressent aux hypothèses simples pour avoir une décision plus précise ou composées pour une décision certaine.

3.4. Théorie de Dezert-Smarandache (DSmT)

- **Cadre de raisonnement et la modélisation**

Pour remédier au problème lié au cadre de raisonnement 2^Θ de la théorie de l'évidence, qui comprend seulement les singletons et leurs unions, du cadre de discernement qui est discret, exclusif et exhaustif, Dezert et Smarandache ont conçu conjointement un cadre de raisonnement étendu à l'ensemble d'hyper-power-set D^Θ .

Le cadre de discernement est composé de n hypothèses $\Theta = \{H_1, H_2, \dots, H_n\}$ exhaustives et non pas forcément exclusifs avec H_i qui peut supporter des informations provenant d'une ou plusieurs sources [19]. Le cadre de raisonnement

D^Θ contient les singletons et leurs unions ainsi que leurs intersections en vérifiant les 3 règles suivantes :

1. $\emptyset, H_1, H_2, \dots, H_n \in D^\Theta$.
2. Si $X, Y \in D^\Theta$, alors $X \cup Y \in D^\Theta$ et $X \cap Y \in D^\Theta$.
3. Seuls les éléments issus de 1 et 2 appartiennent à D^Θ .

Pour bien comprendre comment construire ce cadre de raisonnement, nous donnons un exemple pour la dimension ≤ 3 .

Exemple :

1. Pour $n = 0$, $\Theta = \{\}$, alors $D^\Theta = \{\emptyset\}$ et $|D^\Theta| = 1$.
2. Pour $\Theta = \{H_1\}$, $D^\Theta = \{\alpha_0 \triangleq \emptyset, \alpha_1 \triangleq H_1\}$ et $|D^\Theta| = 2$.
3. Pour $\Theta = \{H_1, H_2\}$, $D^\Theta = \{\alpha_0 \triangleq \emptyset, \alpha_1 \triangleq H_1 \cap H_2, \alpha_2 \triangleq H_1, \alpha_3 \triangleq H_2, \alpha_4 \triangleq H_1 \cup H_2\}$ et $|D^\Theta| = 5$.
4. Pour $\Theta = \{H_1, H_2, H_3\}$, $D^\Theta = \{\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{18}\}$ et $|D^\Theta| = 19$, ou D^Θ

$\alpha_0 \triangleq \emptyset$	$\alpha_9 \triangleq H_1$
$\alpha_1 \triangleq H_1 \cap H_2 \cap H_3$	$\alpha_{10} \triangleq H_2$
$\alpha_2 \triangleq H_1 \cap H_2$	$\alpha_{11} \triangleq H_3$
$\alpha_3 \triangleq H_1 \cap H_3$	$\alpha_{12} \triangleq (H_1 \cap H_2) \cup H_3$
$\alpha_4 \triangleq H_2 \cap H_3$	$\alpha_{13} \triangleq (H_1 \cap H_3) \cup H_2$
$\alpha_5 \triangleq (H_1 \cup H_2) \cap H_3$	$\alpha_{14} \triangleq (H_2 \cap H_3) \cup H_1$
$\alpha_6 \triangleq (H_1 \cup H_3) \cap H_2$	$\alpha_{15} \triangleq H_1 \cup H_2$
$\alpha_7 \triangleq (H_2 \cup H_3) \cap H_1$	$\alpha_{16} \triangleq H_1 \cup H_3$
$\alpha_8 \triangleq (H_1 \cap H_2) \cup (H_1 \cap H_3) \cup (H_2 \cap H_3)$	$\alpha_{17} \triangleq H_2 \cup H_3$
$\alpha_{18} \triangleq H_1 \cup H_2 \cup H_3$	

La cardinalité de l'ensemble hyper-power-set D^Θ s'incrémente en fonction de l'incrémentation du cadre de discernement Θ sur lequel elle est basée. Cette cardinalité est calculée par 2^{2^N} avec N qui représente la cardinalité de Θ .

Le problème est de déterminer si cette cardinalité est similaire à la séquence des numéros *Dedekind* [123] [124]. Cette expression est alors utilisée pour calculer la séquence de cardinalité de D^Θ comme indiqué dans le Tableau 1.

Cardinal de Θ	0	1	2	3	4	5	6	7	8
Cardinal de D^Θ	1	2	5	19	167	7580	7828353	241×10^{10}	561×10^{20}

Tableau 1 Séquence de Dedekind

Ensuite chaque source S_j distribue les masses m entre 0 et 1 ($D^\Theta \rightarrow [0,1]$) qui indiquent le degré de croyance sur un ensemble d'évènements ou d'hypothèses $A_i \in D^\Theta$ [118] :

$$0 \leq m(A_i) \leq 1 \quad 3.8$$

La modélisation se fait de la même manière que dans la théorie d'évidence, mais cette fois le jeu de masse généralisé de croyance (GBBA) est défini par :

$$\sum_{A_i \in D^\Theta} m(A_i) = 1 \quad 3.9$$

$$m(\emptyset) = 0 \quad 3.10$$

où $m(A_i)$ représente la masse élémentaire de croyance généralisée en A_i et tout sous-ensembles $A_i \in D^\Theta$ ayant une masse non nulle est un **élément focal**.

L'ensemble A de tous les éléments focaux constituent le **noyau** de la façon suivante :

$$N_\Theta = \{A \in D^\Theta \mid m(A) > 0\} \quad 3.11$$

Contrairement au jeux de masse classiques présentés dans la théorie DST (*Dempster-Shafer*), les jeux de masses dans DSMT permettent la pondération des informations

fortement conflictuelles, dites paradoxales, à travers la formalisation de l'ensemble vide \emptyset par $m(\emptyset) \neq 0$ [18][34].

- **Les modèles de fusion**

La théorie de DSMT propose trois modèles de fusion en fonction de l'application considérée et de la nature intrinsèque des événements du problème de fusion étudié[34]. Initialement, la DSMT considère tous les éléments du cadre de discernement Θ non précisément définis et non obligatoirement exclusifs [118][34]. Cependant, pour l'adapter aux cas réels où nous avons des contraintes d'intégrités, le cadre peut être raffiné selon le modèle adopté à savoir le modèle libre, le modèle de Shafer ou le modèle hybride.

- **Modèle libre \mathcal{M}^f :**

Ce modèle prend l'ensemble d'hyper-power-set au complet sans aucune contrainte. Il prend en compte toutes les possibilités de combinaison entre les singletons soit par les unions ou bien par les intersections ce qui est une limite majeure de ce modèle à cause de la complexité de sa mise en place en raison de la taille de mémoire nécessaire au stockage des éléments de D^Θ lorsque la cardinalité du cadre de discernement est supérieure à 6. Le tableau 2 illustre la taille de la mémoire en fonction de cardinalité de Θ .

$ \Theta $	2	3	4	5	6	7	8
Taille de θ_i	1 bytes	1 bytes	2 bytes	4 bytes	8 bytes	16 bytes	32 bytes
Card(D^Θ)	4 bytes	19 bytes	166	7569	7828352	$\approx 2.4 \times 10^{12}$	$\approx 5.6 \times 10^{12}$
Mémoire de stockage (D^Θ)	4 bytes	18 bytes	0.32 Kb	≈ 30 Kb	≈ 59 Mb	$\approx 3.6 \times 10^4$ Gb	$\approx 1.7 \times 10^{15}$ GB

Tableau 2 Taille de la mémoire requise pour D

Les éléments de l'ensemble hyper-power-set D^Θ peuvent être présentés par le diagramme de Venn. Cet outil graphique facilite la codification et l'implémentation de tous les éléments de D^Θ [118] et permet de présenter la nature pure des éléments simples

par les singletons, la nature conflictuelle par les éléments d'intersections et la nature d'ignorance par les éléments d'unions [34] [19].

Exemple :

Nous reprenons l'exemple de $\Theta = 3$, Smarandache[118] a proposé une codification qui permet de codifier 2^n-1 parties distinctes du diagramme de Venn de la manière suivante :

- Chaque $\langle i \rangle$ représente les parties qui appartiennent à θ_i uniquement, dites aussi les parties qui ne se chevauchent pas avec les autres parties de θ_j avec $i \neq j$.
- Chaque $\langle ij \rangle$ représente les parties d'intersection entre la partie $\langle i \rangle$ θ_i et $\langle j \rangle$.
- Chaque $\langle ijk \rangle$ représente les parties d'intersection entre la partie $\langle i \rangle$, $\langle j \rangle$ et $\langle k \rangle$.

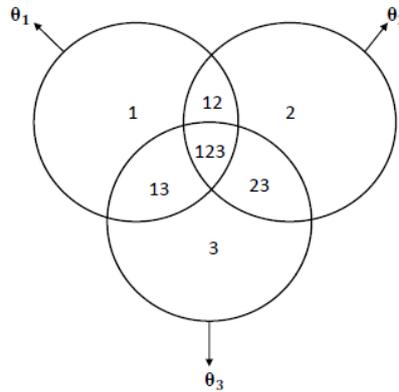


Figure 21: Codification de Venn pour $\Theta = 3$ modèle libre de DSmT [34]

D'après le diagramme de Venn présenté dans la Figure 21, la classe θ_1 est représentée par l'union des éléments raffinés $\{\langle 1 \rangle, \langle 12 \rangle, \langle 13 \rangle, \langle 123 \rangle\}$. $\theta_1 \cap \theta_2$ est représenté par $\{\langle 12 \rangle, \langle 123 \rangle\}$.

Le vecteur d_n de tous les éléments de D^Θ est obtenu en résolvant l'équation suivante :

$$d_N = D_N \cdot u_N \tag{3.12}$$

Avec D_N qui représente la matrice binaire de Dedekind et u_N la base de codification choisie. Le vecteur des éléments de diagramme de Venn est défini par l'équation suivante :

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \\ \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \\ \alpha_{15} \\ \alpha_{16} \\ \alpha_{17} \\ \alpha_{18} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \langle 1 \rangle \\ \langle 2 \rangle \\ \langle 12 \rangle \\ \langle 3 \rangle \\ \langle 13 \rangle \\ \langle 23 \rangle \\ \langle 123 \rangle \end{bmatrix} \quad 3.13$$

○ **Modèle de Shafer \mathcal{M}^0 :**

Ce modèle correspond à l'ensemble de puissance 2^Θ de la théorie de l'évidence, où tous les singletons doivent être exclusifs et distincts et seules les unions de ces singletons sont considérées.

○ **Modèle Hybride \mathcal{M}' :**

Ce modèle noté $\mathcal{M}'(\Theta)$ permet d'inclure les contraintes d'intégrités qui sont impossibles dans les cas où l'on dispose des connaissances du problème de fusion étudié.

Trois types de contraintes d'intégrité peuvent être cités :

- Les contraintes d'exclusivité : lorsque nous disposons de certaines intersections de Θ qui sont vides.
- Les contraintes non existentielles : lorsque nous disposons de certaines unions de Θ qui sont vides.

- Les contraintes hybrides/mixtes : lorsque nous disposons à la fois de contraintes exclusives et non existentielles.

La Figure 22 illustre le diagramme de Venn pour le modèle hybride en considérant la contrainte d'exclusivité $\theta_1 \cap \theta_2 = \emptyset$, ce qui implique que $\theta_1 \cap \theta_2 \cap \theta_3 = \emptyset$.

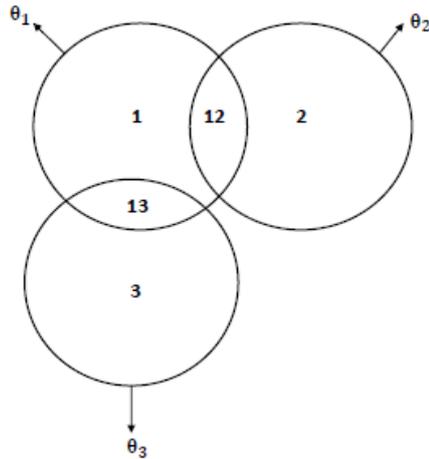


Figure 22 Diagramme de Venn pour le modèle hybride $\mathcal{M}'(\Theta)$ [34]

Le vecteur des éléments de diagramme de Venn est défini par l'équation suivante :

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \\ \alpha_{11} \\ \alpha_{12} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \langle 1 \rangle \\ \langle 2 \rangle \\ \langle 3 \rangle \\ \langle 13 \rangle \\ \langle 23 \rangle \end{bmatrix} \quad 3.14$$

- **Estimation**

L'étape de l'estimation des masses est une étape difficile car il n'existe à ce jour pas de solution universelle [34]. L'estimation est effectuée généralement en utilisant

les modèles de transfert qui servent à estimer les fonctions de masses à partir de la densité des probabilités bayésiennes pour une source donnée. Parmi les modèles utilisés, nous citons [34], le modèle de probabilité inférieure et supérieure de Dempster[119] et le modèle de transfert *Appriou* [34].

Les fonctions de croyance généralisées telles que la fonction de crédibilité, la fonction de plausibilité et celle de la transformation pignistique sont aussi estimées à partir de la fonction de masse généralisée m :

○ La fonction de crédibilité généralisée

Cette fonction définit le degré de croyance sur un événement A , que ce soit une masse placée exactement sur A ou bien la somme de toutes les masses des propositions B incluses dans A par.

$$Cr(A_i) = \sum_{B|B \subseteq A_i} m(B) \quad 3.15$$

Avec $Cr: D^\Theta \rightarrow [0,1]$ $Cr(\emptyset) = 0$ et $Cr(\Theta) = 1$.

○ La fonction de plausibilité généralisée

Cette fonction quantifie le degré de croyance que nous ne doutons pas de A , et est définie par la somme de toutes les masses des propositions B qui présente une intersection non vide avec A .

$$Pl(A_i) = \sum_{B|B \cap A_i \neq \emptyset} m(B) \quad 3.16$$

Dont $Pl(\emptyset) = 0$ et $Pl(\Theta) = 1$.

La différence entre la fonction de crédibilité et celle de plausibilité sur une proposition donnée A_i mesure l'incertitude sur la proposition A_i comme illustré sur la Figure 23 et elle est définie par:

$$Cr(A_i) - Pl(A_i) \quad 3.17$$

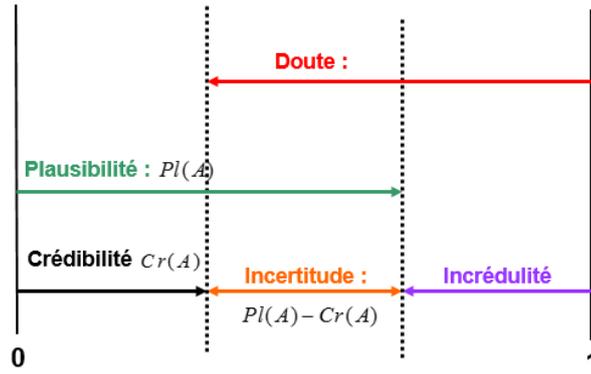


Figure 23 : les fonctions de croyance généralisées [19]

○ La transformation pignistique généralisée

Cette notion dépend de la cardinalité de DSMT de chaque élément $A_i \in D^\Theta$ noté $C_M(A_i)$ qui représente le nombre de parties non vide de A_i dans le diagramme de Venn du modèle utilisé.

La transformation pignistique dans DSMT est définie sur l'ensemble hyper-power-set D^Θ comme suit :

$$\forall A \in D^\Theta, \text{Bet}P(A) = \sum_{X \in D^\Theta} \frac{C_M(X \cap A)}{C_M(X)} m(X) \quad 3.18$$

Avec $C_M(X)$ qui représente la cardinalité de la proposition X pour le modèle M de DSMT.

La transformation DSMP

La transformation DSMP est une version sophistiquée de la transformation pignistique [118] [125] qui présente une nouvelle version de la transformation de masse pour chaque ignorance.

• **Combinaison**

Le formalisme de DSMT fournit plusieurs règles de combinaison telles que de façon non exhaustive la règle de *Dempster-Shafer* normalisée, la règle *Smets* [126] [127], la règle DSMT, règle DSMT, la règle de *Dubois* et *Prade*[128], la règle de *Yager* [129] [130], la règle PCR et toutes ses version[131] ou encore la règle de *Martin* et *Osswald*. Dans cette section, nous détaillons les deux règles les plus utilisées à savoir la règle classique DSMT [118], la règle hybride DSMT[132].

○ La Règle de combinaison classique (DSmC)

Cette règle correspond à la règle de combinaison de la théorie d'évidence, mais sur un ensemble étendu que l'ensemble 2^Θ et qui est l'hyper-power-set D^Θ . C'est une règle associative et commutative qui est définie dans le cas deux sources $n=2$ comme suit :

$$m(C) = \sum_{A \cap B = C} m_1(A) m_2(B), A, B \in D^\Theta, \forall C \in D^\Theta \quad 3.19$$

○ La règle de combinaison hybride DSmH

Cette règle est une alternative à la règle de combinaison classique qui permet de prendre en compte les divers types de contraintes ce qui n'est pas possible dans la règle classique. Elle est définie pour deux sources comme suit :

$$m_{M(\Theta)}(A) = \phi(A) [S_1(A) + S_2(A) + S_3(A)], \forall A \in D^\Theta \quad 3.20$$

$$\text{Avec } \phi(A) = \begin{cases} 0 & \text{si les ensemble } A \text{ sont vides} \\ 1 & \text{Sinon} \end{cases}$$

$$S_1(A) = \sum_{X_1 \cap X_2 = A} m_1(X_1) m_2(X_2), \forall X_1, X_2 \in D^\Theta \quad 3.21$$

$$S_2(A) = \sum_{[(\mu(X_1) \cap \mu(X_2) = A) \vee ((\mu(X_1) \cap \mu(X_2) \in \phi) \wedge (A = I_t))]} m_1(X_1) m_2(X_2), \forall X_1, X_2 \in \phi \quad 3.22$$

$$S_3(A) = \sum_{X_1 \cup X_2 = A} m_1(X_1) m_2(X_2), \forall X_1, X_2 \in D^\Theta, \forall X_1, X_2 \in \phi \quad 3.23$$

Avec $\mu(X)$ qui représente l'union de toutes les parties de l'ensemble X.

Dans l'équation 3.20, l'intersection est définie comme étant une contrainte d'intégrité, la masse est affectée à l'union des ensembles X_1 et X_2 , tandis que dans l'équation 3.19, si l'union qui est définie comme étant la contrainte, la masse est alors attribuée soit à l'union de tous les singletons qui forment l'ensemble ou bien à l'ignorance totale I_t .

○ La règle de combinaison PCR5 et PCR6

La famille de règles de combinaison à Redistribution Proportionnelle du Conflit (PCR) consiste à répartir proportionnellement le conflit partiel aux éléments impliqués dans sa génération [118]. La version PCR5 est une alternative qui a montré son efficacité dans plusieurs applications et elle consiste à redistribuer la masse conflictuelle (partielle ou totale) après un calcul des jeux de masses par une règle conjonctive. Pour deux sources, la règle PCR5 est définie par :

$$\left\{ \begin{array}{l} m_{PCR5}(\emptyset) = 0 \\ m_{PCR5}(X) = \sum_{\substack{Y \in D^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \quad \forall X \in D^\Theta \setminus \{\emptyset\} \end{array} \right. \quad 3.24$$

La règle PCR6 présente une généralisation de PCR5 et montre son efficacité lorsque le nombre des sources est supérieur à 2.

La règle PCR6 est utilisée dans ces travaux car elle permet la gestion des données conflictuelles, ce qui correspond à notre cas d'application.

• **Décision**

La prise de décision dépend de l'application étudiée et du résultat désiré. Dans la DSMT, tout comme dans la théorie de l'évidence, la décision peut être prise en se basant sur le maximum de crédibilité, le maximum de plausibilité, le maximum de DSMP, le maximum de transformation pignistique ou bien définie sur l'intervalle de confiance [Cr, Pl] [133] [134] [135].

3.5. Le choix du formalisme DSMT et la motivation

L'application étudiée et la nature des données à fusionner impliquent de choisir un formalisme spécifié. Pour ce faire, nous avons comparé les différents formalismes existants en commençant par la théorie de probabilité. Celle-ci est en effet simple à mettre en œuvre et, elle permet de modéliser de façon performante l'incertain grâce à la représentation des connaissances a priori sur chaque classe de cadre de discernement. Cependant, cette théorie confond les notions d'incertitude et d'imprécision, et ne permet pas la prise en compte de l'ignorance, qui s'exprime

généralement par une égalité des probabilités a priori [33][18]. Ceci peut provoquer des conflits et des incohérences au niveau de la modélisation. En outre, ce formalisme est basé sur un espace fermé porté sur seulement les singletons sur lesquelles les décisions prises sont exhaustives et exclusives.

Ensuite, nous avons étudié les théories des ensembles flous et des possibilités qui ont été introduites pour surmonter les limites de la théorie de probabilité. Elles ont l'avantage d'exprimer la notion d'imprécision d'une manière explicite ? De plus, contrairement à la théorie de probabilité, il existe une grande flexibilité dans les règles de combinaison [22] [13]. Malgré ces avantages, ce formalisme n'est pas approprié à nos données car il est encore peu développé au niveau de la modélisation du conflit entre les sources et le choix entre ses règles est très complexe à cause des distributions des possibilités issues de chaque source. En outre, l'utilisation des degrés de nécessités et de possibilités restreint l'utilisation de toutes les masses en tenant en compte juste du maximum ou du minimum de ces degrés, ce qui provoque une perte d'information [13] [18].

Puis nous avons examiné la théorie de l'évidence, qui surmonte les limites de deux formalismes précédents par l'introduction des grandeurs de la crédibilité et de la plausibilité. Ce formalisme permet de modéliser et d'exprimer le conflit entre les sources qui remonte lors de calcul des masses de confiance. Mais, ce conflit entre les sources présente ainsi la limite majeure de cette théorie quand il devient très grand. De plus, la taille du cadre de discernement est une limite importante. En effet, plus ce cadre est grand plus la complexité augmente. Ce cadre de discernement n'est pas souhaitable pour plusieurs applications, car il implique que l'ensemble de propositions, issues de sources indépendantes et distinctes, soit discret, exhaustif et exclusif, ce qui est une contrainte très forte souvent non vérifiée dans les cas réels.

Le dernier formalisme est la DS_mT. Il a été introduit par Jean Dezert et Florentin Smarandache [118]. Il s'agit d'une généralisation de la théorie d'évidence au niveau de la gestion des conflits importants entre les sources, réalisée en introduisant de nouvelles règles de combinaison dédiées à la gestion des conflits. Elle rend aussi le cadre de discernement plus souple par la possibilité d'introduction des contraintes

d'intégrité selon l'application étudiée, et permet de travailler avec n'importe quel type de modèle.

Dans ces travaux, nous cherchons à fusionner des données provenant de plusieurs modalités qui présentent des imperfections de type incertitude, imprécision, l'ignorance et parfois des conflits importants. Afin de prendre en compte ces imperfections de manière efficace et performante, nous avons choisi le formalisme de la théorie de *Dezert-Smarandache* (DSmT), connue également sous le nom de la théorie de raisonnement plausible et paradoxal. Ce formalisme a par ailleurs montré au niveau expérimental qu'il est approprié pour l'application de ces travaux.

3.6.L'application de la fusion dans la classification des images

La fusion de données issues de modalités d'imagerie est apparue en premier temps avec l'utilisation de l'imagerie aérienne et satellitaire dans le domaine militaire par le groupe de travail sur la fusion des données JDL (*Joint Directors of Laboratories*) [16]. La classification basée sur la théorie de l'évidence DST est largement exploitée dans de nombreux travaux dans ce domaine.

Dans le domaine satellitaire, les données traitées sont souvent incertaines, incomplètes et imprécises, et demandent à être qualifiés et quantifiés au niveau des résultats de classification. *Lee* et al. [136] ont par exemple utilisé la théorie de DST avec une approche probabiliste pour modéliser l'incertitude des données, et ensuite pour combiner le contenu de l'information des images satellitaires issues de sources multiples avec des données spatiales en donnant plus d'importance à une source et en tenant en compte la qualité du capteur utilisé. La comparaison entre les deux approches a montré la performance de la théorie de DST.

La modélisation des données incertaines et imprécises était aussi l'objet du travail du *Zahzah* et al. [137] qui ont traité la problématique de la classification des images satellitaires en utilisant la règle orthogonale de Dempster de la théorie de DST, pour la combinaison des données multi sources présentant incertitude et imprécision. Le traitement numérique des images a été effectué en utilisant les réseaux de neurones artificiels pour affecter chaque point à la classe recherchée. Les réseaux neuro-flous sont ensuite utilisés pour produire une carte de décision à partir d'une connaissance symbolique.

La notion d'incertitude a été traitée avec succès par le travail de *Kim* et al. [138] où ils ont proposé une nouvelle méthode pour la classification multi-sources basé sur la DST, dont chaque source de données fournit un ensemble de masses avec un certain degré de croyance, pour l'occupation du sol des données du spectromètre d'imagerie satellitaire à

haute résolution. Les masses de l'évidence issues de chaque source sont exprimées par des probabilités évaluées par intervalle dans le but d'incorporer la notion de l'incertitude des données.

Pour la détection des mines en utilisant l'imagerie satellitaire, *Milisavjevic* et al. [139] ont utilisé la théorie de *Dempster-Shafer* DST. L'extraction des caractéristiques est effectuée dans un premier temps, suivi par la modélisation. Ces deux étapes sont exécutées pour chaque source/capteur. Ensuite les masses issues de chaque source/capteur sont combinées par la règle de *Dempster-Shafer* en fournissant la première intuition de la présence des mines. Les auteurs ont proposé une nouvelle règle de décision pour simplifier l'étape de prise de décision.

La classification par la théorie de DST a été utilisée aussi pour les images radar ou SAR (Synthetic Aperture Radar) dans [140], et encore dans [141] où les auteurs ont proposé une nouvelle méthode basée sur l'estimation des noyaux, en utilisant les fonctions de densité de probabilité non paramétrique avec le contexte de Markov. Ce qui permet des pré-classifications qui seront ensuite combinées par la théorie Dempster-Shafer d'une façon non-supervisée suivie d'une régularisation. Cette méthode ne demande aucune distribution d'observation ce qui présente un avantage dans les cas des observations compliquées et arbitraires. DST est utilisée plus récemment par [142] pour des fins de cartographie des terres, aussi dans [143] [144] et pour la cartographie forestière dans [145].

Toujours dans l'imagerie satellitaire, l'introduction du contexte sémantique de l'information a été utilisé récemment par [146] pour effectuer une classification multi-label en se basant sur la théorie de DST. La méthode proposée permet de faire face au manque de probabilités antérieures et des données incertaines fournies par les primitives de bas niveau. L'information spatiale été introduite par [147] pour l'amélioration de la qualité des cartes d'occupation de sol issue de *Landsat 7 ETM+* en utilisant la théorie de *Dempster-Shafer* et celle de *Dezert-Smarandache*.

La théorie de l'évidence DST a également été utilisée avec succès dans diverses applications de l'imagerie médicale. Dans le but de la sélection des caractéristiques pertinentes, La DST a été utilisée dans [148] pour chercher un ensemble de caractéristiques qui augmentent les performances de prévision et diminuent l'imprécision des données d'une façon simultanée. Et dans [149] pour la sélection des caractéristiques extraites d'images FDG-PET pour fournir un système de prévision des résultats du traitement du cancer. Et très récemment, *Razi* et al. [150] ont proposé une nouvelle méthode basée sur la théorie de *Dempster Shafer* pour fusionner les résultats de classifieurs binaires constitutifs issus des modèles spatiaux communs (CSP).

La théorie de DST a été utilisée dans plusieurs applications médicales des images multimodales. *Bloch* [151] a utilisé la DST pour la fusion des images médicales dans le but d'avoir une classification des tissus cérébraux dans les images multimodales RM pathologiques à double écho. *Zhu et al.* [152] se sont basés sur DST et l'information du voisinage spatial pour la segmentation des images multimodales du cerveau en utilisant la logique floue pour l'estimation des fonctions de masse. *Wafa et al.* [153] ont proposé une approche basée sur la multi-caractéristiques FCM et la théorie de l'évidence, appliquée sur les images multimodales IRM pour la classification et la segmentation des tumeurs cérébrales.

Malgré le fait que la théorie de la DST ait été largement utilisée dans plusieurs applications, sa performance est décevante quand le conflit entre les sources est important ainsi que sa capacité à fusionner des données multi-sources pour la classification est à vérifier.

La théorie de *Dezert-Smarandache* DS m T est plus récente. Elle a été appliquée en imagerie satellitaire pour la fusion des données multi-temporelles pour la prédiction de la couverture hivernale des sols [154] et sur des données multi-dates par [155]. Toujours dans l'imagerie satellitaire et pour des fins de fusion et de classification des images multi-sources et multi-temporelles, le formalisme DS m T a été utilisé pour l'amélioration de la carte d'occupation du sol et l'établissement des cartes de changements [156], ainsi que pour une classification supervisée en intégrant des informations contextuelles obtenues du classifieur ICM avec des informations temporelles en tenant en compte de l'aspect parcellaire des classes thématiques [157]. Les auteurs ont proposé une nouvelle règle de décision adaptative (ADR) basée sur le modèle hybride de DS m T. Une nouvelle règle de décision a aussi été proposée par les mêmes auteurs dans [158], basée sur la fonction de croyance généralisée et la transformation pignistique (DS m P) pour pallier les limites inhérentes aux règles de décision utilisant le maximum de fonctions de masses généralisées.

La DS m T a été utilisée très récemment par *Okaingni et al.* [159] pour traiter l'aspect paradoxal des données qui est exprimé par l'intersection entre les classes dans les images satellitaires. Ils ont développé un modèle de classifications des pixels en agrégeant l'information en utilisant la théorie DS m T basée sur la règle PCR5 et l'utilisation des indices spectraux obtenus à partir des images satellitaires ASTER.

À des fins de cartographie des terres, *Abbass et al.* [133] ont utilisé la théorie de DS m T sur les images satellitaires multi-sources, multi-échelles et multi-temporelles pour mettre en évidence les changements survenus dans une région de l'espace terrestre. Les auteurs ont appliqué le modèle *Appriou* généralisé pour les données multi-sources et multi-temporelles pour estimer les fonctions de masses et développer une estimation

particulière pour les données multi-échelles. Bien qu'une comparaison de classification et de fusion issue de la théorie de DST et de maximum de vraisemblance ait montré la performance de la DST pour fournir une discrimination thématique meilleure, cette dernière a montré des limites lorsque le conflit entre les classes spectrales devient important. Pour cette raison, les auteurs ont opté pour utilisation de la théorie de DSMT qui a donné des résultats encourageants et prometteurs. *Pannetier et al.* [160] ont proposé une nouvelle méthode basée sur la fusion des attributs obtenus à partir des indicateurs de cibles mobiles au sol et des capteurs d'imagerie à l'aide de DSMT pour le suivi et la classification.

Le formalisme DSMT est utilisé aussi dans les images médicales pour fusionner les caractéristiques hétérogènes de plusieurs capteurs qui seront inclus dans les systèmes CBR [161]. *Jousselme et al.* [162] ont présenté une nouvelle approche basée sur la théorie du raisonnement plausible et paradoxale (DSMT) pour combiner les informations issues de différents capteurs composés de détecteurs/classificateur afin de fournir une réponse de la décision complète.

Le domaine de la classification de manuscrit fait aussi l'objet de travaux où les auteurs ont appliqué la DSMT pour la reconnaissance de chiffres manuscrits en combinant deux implémentations SVM OAA (One-Against-All) à l'aide de la règle de combinaison PCR6 [163]. Les fonctions des masses sont calculées par la technique de sigmoïde de Platt à partir des probabilités issues de sorties de classificateur SVM.

Toutes ces études montrent que la performance de la classification basée sur la DST est moins efficace que la classification fondée sur la DSMT lorsque le conflit entre les sources utilisées est élevé. Les performances de la DSMT pour la modélisation des données incertaines et imprécises (peu ou pas de conflit) sont comparables à celle de la méthode DST. La méthode DSMT présente cependant toujours un avantage comparatif en termes de souplesse et de flexibilité du cadre de discernement (par rapport à celui dans la DST), grâce à la possibilité d'introduction des contraintes d'intégrité selon l'application étudiée fournie par les modèles de la méthode DSMT [158], [163], [164].

La plupart des méthodes que nous avons étudiées agissent au niveau pixel puisqu'elles affectent chaque pixel à l'une des classes en analysant ses propriétés spectrales. Et parce que les pixels voisins sont corrélés dans une image réelle, l'introduction de la contrainte du contexte spatial dans le processus de la fusion pour des fins de la classification par les théories DST et DSMT a été la question discutée dans plusieurs applications. Cependant, l'intégration de l'information perceptuelle (visuellement saillante) n'a pas été abordée à notre connaissance. Par conséquent, afin d'améliorer les résultats de classification, il nous est apparu pertinent d'incorporer l'information de la perception visuelle qui a une grande capacité à classifier facilement

les scènes observées grâce au mécanisme puissant du cerveau humain. Les travaux présentés dans la partie 2 vont dans ce sens.

3.7. Conclusion

Dans ce chapitre, nous avons donné une description des formalismes de la DST et la DSMT en respectant les étapes de la fusion pour chaque théorie. Nous avons motivé le choix de la théorie de *Dezert-Smarandache* pour ces travaux, puis nous avons présenté une revue de la littérature sur les applications de la théorie de *Dempster-Shafer* (DST) et la théorie de *Dezert-Smarandache* (DSMT) dans le domaine de classification d'images. Ces deux formalismes ont prouvé leur efficacité pour la classification dans de nombreuses applications. Cependant, l'introduction de l'information perceptuelle dans le processus de la fusion basé sur la DST et la DSMT n'a pas encore été faite. A notre sens, la prise en compte du caractère saillant de l'information, de manière analogue au fonctionnement du capteur humain, serait pertinente dans le cadre du processus de classification.

Partie 2

Contributions

Chapitre 4

Recalage des images multimodales par filtrage automatique et adaptatif de caractéristiques SURF

4.1. Introduction

Dans ce chapitre, la problématique du recalage est abordée dans le cas général, puis nous détaillons les spécificités et les limites de cette opération de prétraitement nécessaire à la fusion dans le cas de données issues de différentes modalités d'acquisition. Deux nouvelles approches sont ensuite proposées afin de répondre à ces limitations. Elles sont basées sur l'utilisation de la méthode SURF pour la détection, la description et la mise en correspondance des points d'intérêts. Nos deux contributions interviennent spécifiquement au niveau de l'étape de mise en correspondance. Elles mettent en œuvre un ensemble de de filtres afin de supprimer les points aberrants et d'augmenter ainsi la performance et la robustesse du recalage réalisé.

4.2. La problématique de correspondance rencontrée

La méthode de recalage proposée est basée sur la méthode SURF. En effet, cette méthode est pertinente car elle permet une extraction rapide de points d'intérêt robustes et assure la détection d'un nombre suffisant de ces points à différentes positions pour effectuer un efficace appariement en termes de répétabilité et de rapidité du temps de calcul. Le principal problème avec l'augmentation du nombre de modalités pour une scène/objet est que le même pixel est représenté par différentes intensités dans chaque modalité. Ainsi, les descripteurs extraits peuvent différer de manière significative, voire même être conflictuels, ce qui engendre des difficultés lorsque de la mise en correspondance réalisée pour déterminer la transformation géométrique permettant d'aligner deux paires d'images [165]. Il en résulte ainsi souvent un

nombre important de correspondances incorrectes ou non conformes, qui altèrent la qualité du recalage.

Il existe peu de méthodes pour résoudre les problèmes liés à la mise en correspondance des points caractéristiques. L'approche la plus répandue est basée sur l'algorithme RANSAC (*Random Sampling Consensus*), qui permet d'identifier et si nécessaire de supprimer les points aberrants [166], et permet alors de ne pas faire correspondre tout l'ensemble des points détectés mais seulement ceux qui ont été identifiés comme non aberrants. L'inconvénient de cette méthode provient d'une part de la nécessité de régler de façon empirique ses différents paramètres [167]. D'autre part, dans notre cas d'application, cet algorithme ne couvre qu'une proportion inférieure à 50 % des correspondances correctes.

La méthodologie proposée dans ce travail [8] vise à rendre la technique plus robuste en détectant les correspondances aberrantes, et en effectuant le recalage en se basant uniquement sur les informations les plus pertinentes, c'est-à-dire sans faire correspondre systématiquement toutes les caractéristiques. Deux versions de cette méthode ont été proposées, elles sont développées en détail dans cette section.

4.3. Le filtrage des points de correspondance basé sur la distribution gaussienne :

Dans le but de fournir une meilleure confiance au recalage des images multimodales, la méthode proposée vise à détecter et filtrer les points aberrants identifiés. Une stratégie de filtrage automatique est proposée pour permettre une opération de recalage plus efficace et plus robuste. Cette stratégie intègre deux filtres identifiés respectivement par F1 (D) et F2 (O) :

- F1(D) : La première opération de filtrage est basée sur un calcul de distance entre les descripteurs des points des caractéristiques précédemment extraits.
- F2 (D) : La seconde implémente un filtrage basé sur l'orientation.

Cette stratégie est décrite dans l'algorithme 1 et dans l'organigramme présenté sur la Figure 24 et s'applique comme suit :

A1.1 L'algorithme SURF est d'abord appliqué pour extraire les points caractéristiques notés $X1$ et $X2$ de l'image requête et de l'image modèle respectivement, dont des descripteurs SURF de dimension 64 sont associés à chaque point d'intérêt.

A1.2 Pour effectuer le recalage, l'étape suivante consiste à faire correspondre les points caractéristiques $X1$ et $X2$. Cet appariement est mis en œuvre en calculant une distance euclidienne notée D , entre chaque potentiel pair $X1$ et $X2$. La distance D est calculée à partir des descripteurs associés aux points caractéristiques. La (meilleure) correspondance est considérée comme étant la correspondance de points caractéristiques qui minimise la valeur D . À cette étape, des paires de points de caractéristiques appariées sont obtenues et triées dans un ordre croissant de valeur D . Une distribution typique des distances triées est présentée en **A1.2** sur la Figure 24. À ce stade, nous obtenons généralement un grand nombre de paires de points caractéristiques.

L'étape suivante consiste à réduire ce nombre. La stratégie proposée consiste à n'utiliser que les paires de points caractéristiques qui minimisent la valeur D , c'est-à-dire pour lesquelles la confiance correspondante est plus élevée.

A1.3 Afin de garantir un nombre adapté de points caractéristiques permettant d'effectuer le recalage sans utiliser un seuil fixé arbitrairement ou manuellement, un seuil λ est automatiquement calculé à l'aide de caractéristiques de distribution de données. Ce seuil est exprimé par l'équation (4.1). Cet ensemble de points peut éventuellement encore contenir des valeurs aberrantes. Une illustration de cette opération est présentée dans la Figure 24 (**A1.3**).

$$\lambda = \mu - \frac{k}{2} \times \sigma(D) \text{ avec } \begin{cases} \mu & \text{la moyenne de la distribution des données} \\ k = 6 & \Leftrightarrow \pm 3\sigma \text{ (qui couvre 95\% de données)} \end{cases} \quad 4.1$$

Une matrice de transformation notée T est ensuite calculée en se basant sur quelques correspondances dans différentes positions où la confiance de correspondance est la plus élevée (au moins quatre points sont recommandés). Pratiquement, 20 points de paires sont utilisés pour estimer de manière fiable la transformation.

Ensuite, la transformation calculée est appliquée à l'ensemble des caractéristiques de l'image requête $X1$. Un ensemble de caractéristiques transformées est obtenu, identifié par $X1_{filt}$. Afin de supprimer tous les points aberrants restants, l'étape suivante consiste à effectuer un autre filtre sur un ensemble filtré de caractéristiques obtenues à partir de l'étape précédente.

A1.4 Les vecteurs entre les ensembles des caractéristiques de l'image requête et leur correspondance dans l'image modèle sont construits et identifiés par \overline{VM} . Ces vecteurs sont ensuite normalisés et leurs orientations O_{VM} sont calculées.

A1.5 Un deuxième filtre est appliqué consistant à retenir tous les points compris dans la plage $\left[\frac{-k\sigma}{2}, \frac{+k\sigma}{2}\right]$ ($k = 6$ est utilisé). Une illustration de cette opération sur la scène de la Figure 5 est présentée dans les Figures 25, 26 et 27, et montre que seules les orientations comprises dans la plage déterminée sont conservées. Ensuite, une vérification itérative est implémentée pour couvrir le changement d'écart-type de la distribution de données une fois qu'un point aberrant est identifié et supprimé.

A1.6 Les points identifiés comme aberrants sont supprimés, puis nous calculons la transformation T avec des correspondances correctes (inliers).

Algorithme 1 Filtrage des points aberrants

Entrée :

D : la distance euclidienne entre des paires points.

K : coefficient de l'écart type.

X1 : points caractéristiques de l'image requête.

X2 : points caractéristiques de l'image modèle.

Stdvar : pourcentage de variation de l'écart type (σ) lorsque nous supprimons les points aberrants.

Sortie :

- 1 : 1ere filtre sur la distance : Calcule le premier ensemble d'inliers où

$$D_{\text{filtered}} \leftarrow D (D \leq \text{mean}(D) - \frac{k}{2} \times \sigma)$$
- 2 : $X1 \leftarrow X2_{\text{filtered}}$
- 3 : Sélectionnez 20 paires de points
- 4 : Calcul de la transformation affine T
- 5 : Application de T sur X1
- 6 : Pour chaque paire de points faire
- 7 : Construire les vecteurs de correspondances \overline{VM}
- 8 : Normaliser les vecteurs de correspondances
- 9 : $\|\overline{VM}\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$
- 10 : $(VM)^* = \frac{\overline{VM}}{\|\overline{VM}\|}$
- 11 : Calculer les orientations des vecteurs \overline{VM}
- 12 : $O_{VM} = \arctan\left(\frac{Y_{\overline{VM}^*}}{X_{\overline{VM}^*}}\right)$
- 13 : Le deuxième filtre sur orientation
- 15 : Tant que delta > stdvar faire
- 16 : $OVM \leftarrow OVM(\text{mean}(OVM) - \frac{k}{2} \times \sigma (OVM)) < OVM < \text{mean}(OVM) + \frac{k}{2} \times \sigma (OVM)$
- 17 : Fin tant que
- 18 : Fin pour

- 19 : inliers ← Garder ensemble de points où OVM
- 20 : Recalculer la transformation affine T sur tous les inliers

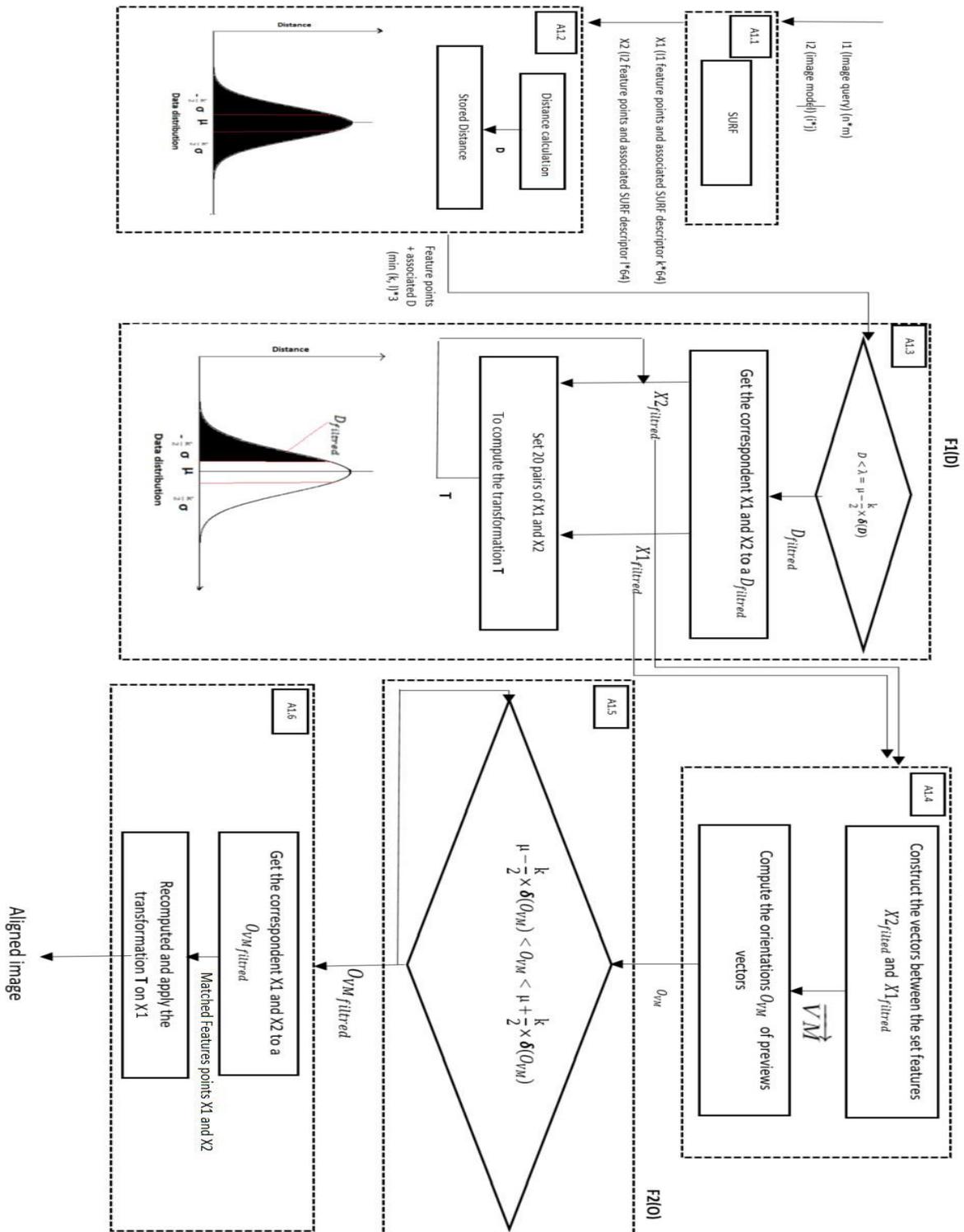


Figure 24 Organigramme du premier algorithme proposé. Deux filtres : le premier basé sur la distance F1 (D) et le deuxième sur l'orientation F2 (O)

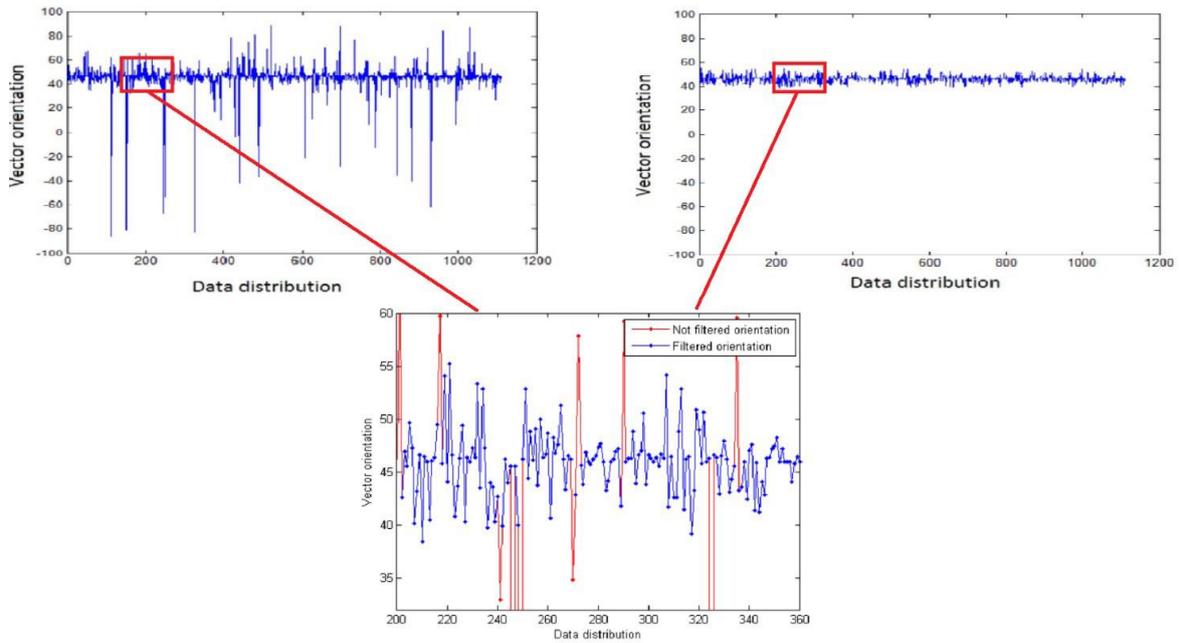


Figure 25 Distributions d'orientation sur la modalité IRR (à droite) avant le filtrage (à gauche) après le filtrage (région zoomée)

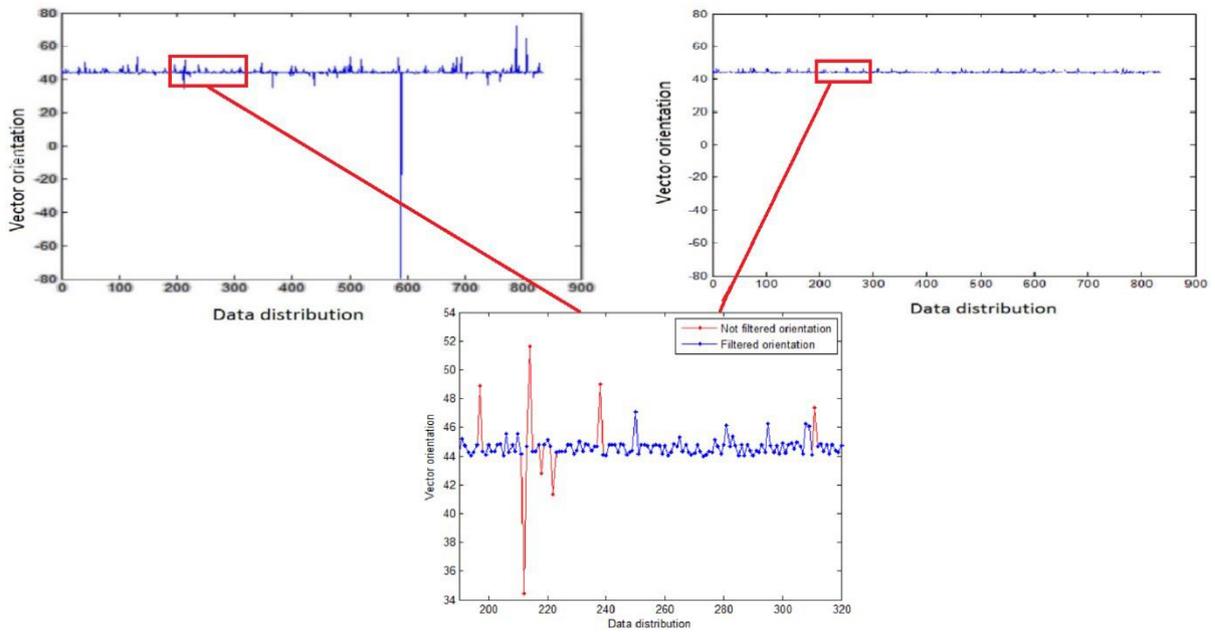


Figure 26 Distributions d'orientation sur la modalité UVR (à droite) avant le filtrage (à gauche) après le filtrage (région zoomée)

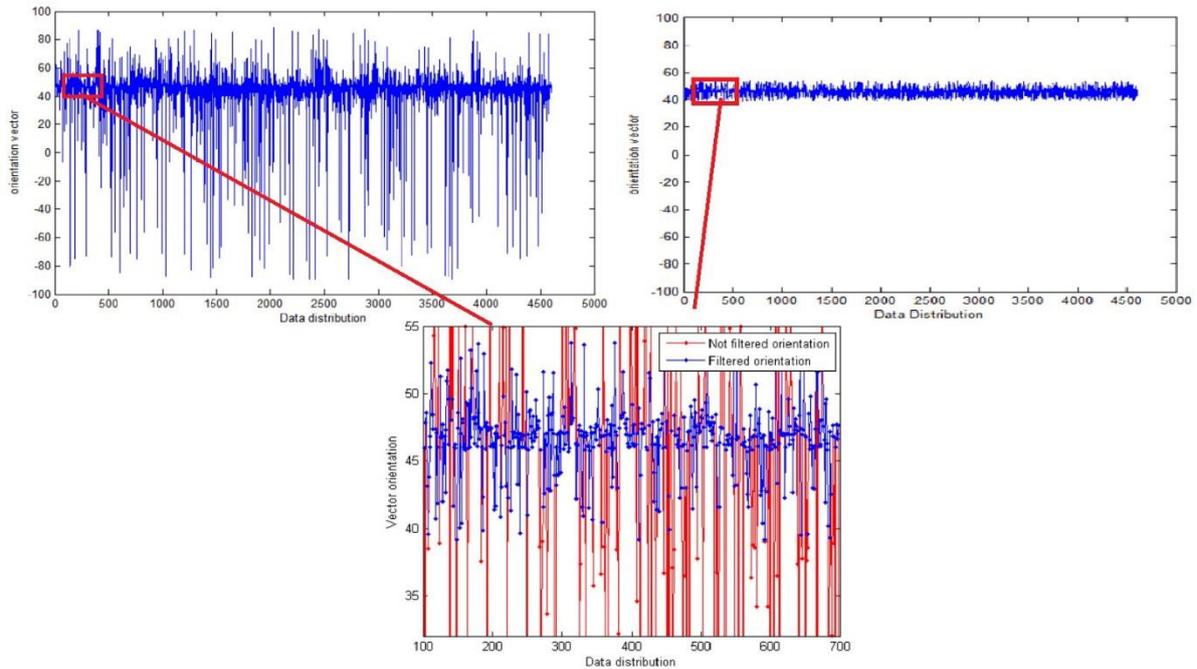


Figure 27 Distributions d'orientation sur la modalité UVF (à droite) avant le filtrage (à gauche) après le filtrage (région zoomée)

Une comparaison avec les autres méthodes de recalage conventionnelles basées sur SIFT/SURF est effectuée afin d'évaluer les performances de la méthode proposée. Les résultats obtenus ont été réalisés sur une base de données de 100 images multimodales fournies par les gestionnaires du château de Germolles. Dans ce cas d'application, l'objectif final est la fusion des données d'imagerie multimodales afin de comprendre l'évolution des attributs d'apparence des zones inspectées, et en particulier de comparer ces attributs pour les zones originales et celles ayant fait l'objet d'une restauration. La performance de l'utilisation du descripteur local SURF, sans et avec notre conditionnement de données, est évaluée en calculant le taux de rappel R_{rec} (équation 4.2) :

$$R_{rec} = \frac{N_{cm}}{N_c} \text{ avec } \begin{cases} N_{cm} & \text{nombre de points d'intérêt bien appariés} \\ N_c & \text{nombre de correspondances existantes} \end{cases} \quad 4.2$$

Pour l'illustration, nous avons choisi deux ensembles des images multimodales de différentes scènes du château de Germolles à savoir :

- Le premier ensemble utilisé présente une scène sur le mur ouest de la chambre de la comtesse de Nevers, dont des exemples d'acquisitions sont présentés dans la Figure 28.

Les Figures 29, 30 et 31 représentent respectivement le recalage entre les images IRR, UVR et UVF (l'image requête), avec l'image modèle VIS, où (a) montre les correspondances

de points putatives et (b) indique les correspondances de points après des points aberrants supprimés avec notre méthode proposée.

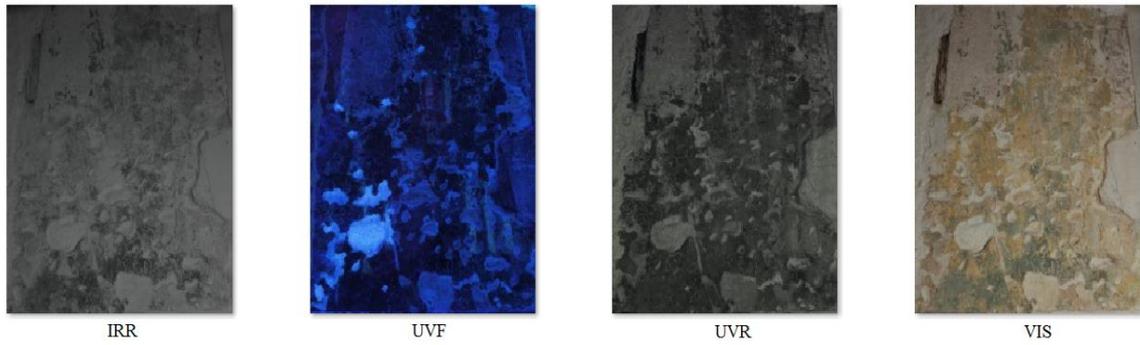


Figure 28 les quatre modalités pour la première scène sur le mur ouest de la chambre de la comtesse Nevers

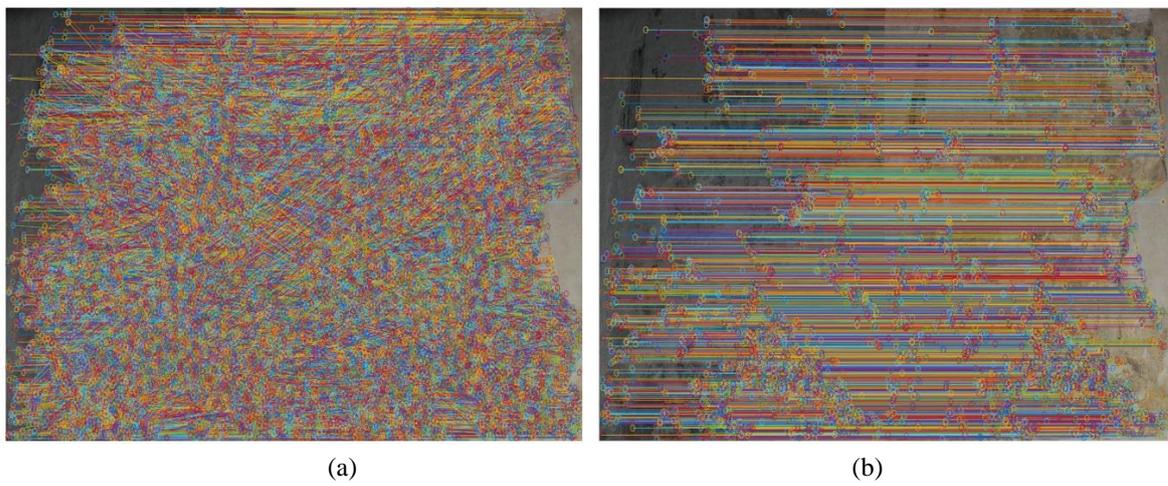


Figure 29 Application du recalage entre IRR et VIS de la première scène - (a) Méthodologie SURF (b) Notre méthodologie

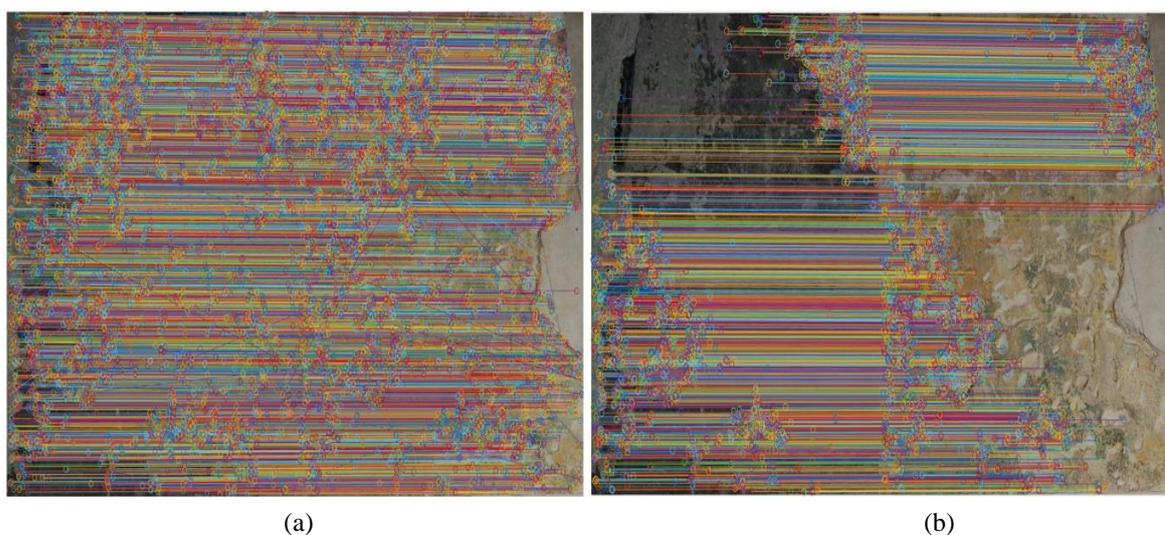
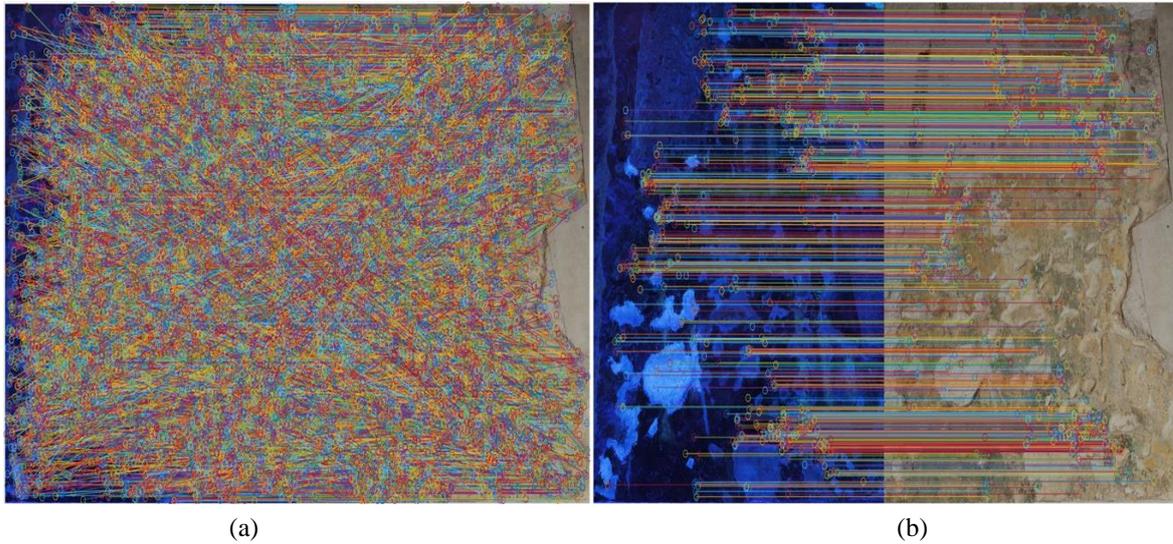


Figure 30 Application du recalage entre UVR et VIS de la première scène - (a) Méthodologie SURF (b) Notre méthodologie



**Figure 31 Application du recalage entre UVF et VIS de la première scène - (a) Méthodologie SURF
(b) Notre méthodologie**

- La seconde série d'images utilisées pour l'illustration présente une scène du mur ouest du vestiaire de Margaret de Flandres (Figure 32).

Les Figures 32, 33 et 34 représentent respectivement le recalage entre les images IRR, UVR et UVF (l'image), avec l'image modèle VIS, où (a) montre les correspondances de points putatives et (b) indique les correspondances de points après des points aberrants supprimés avec notre contribution.

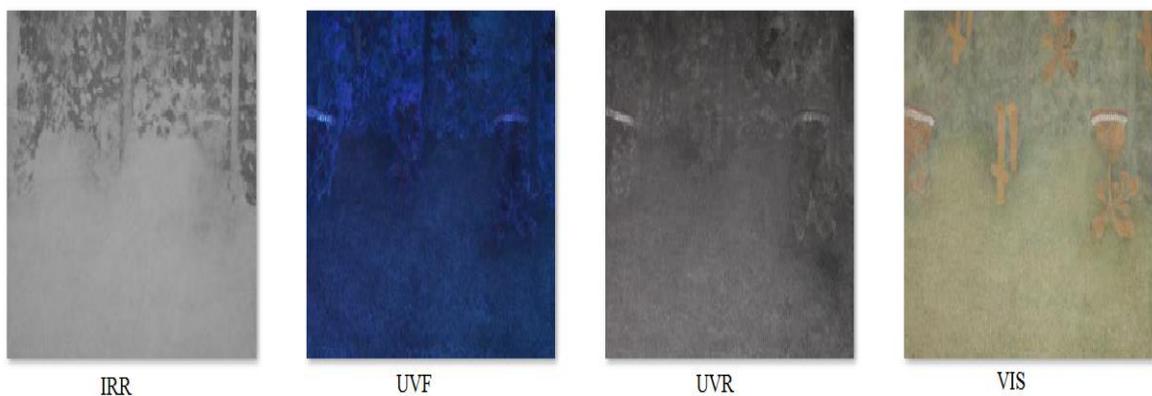


Figure 32 les quatre modalités pour la deuxième scène sur le mur nord du dressing de Marguerite de Flandre

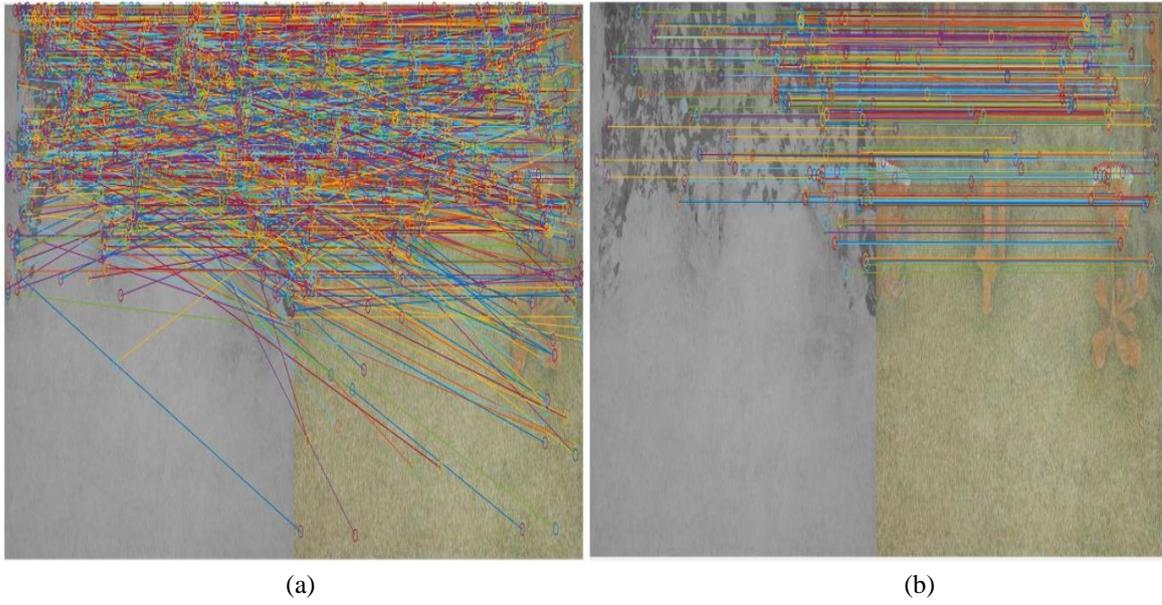


Figure 33 Application du recalage entre IRR et VIS de la deuxième scène - (a) Méthodologie SURF
 (b) Notre méthodologie

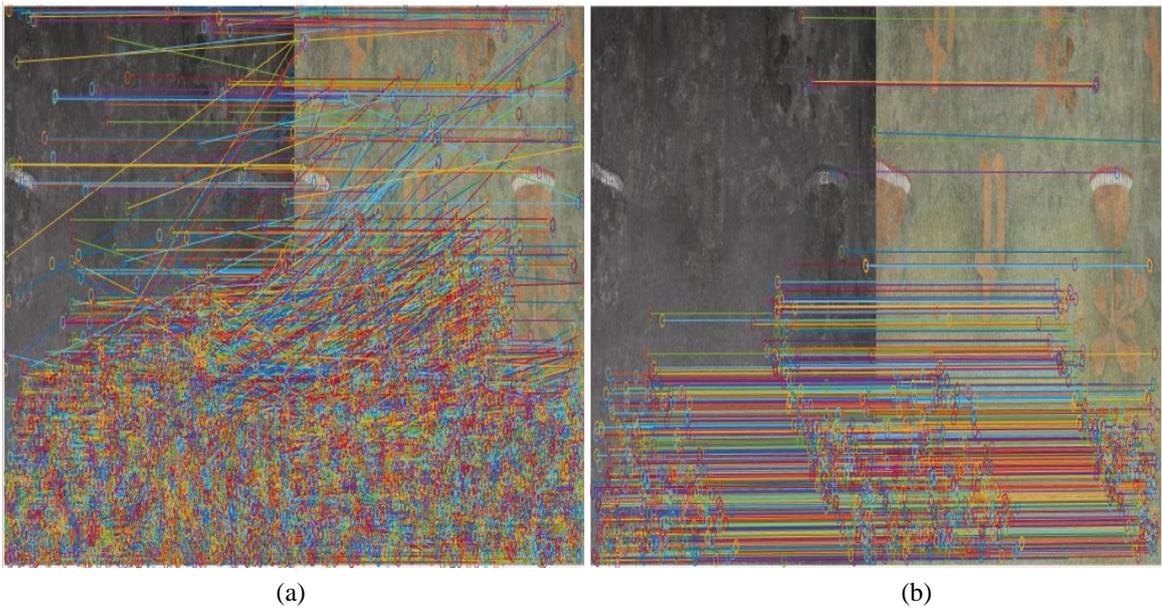


Figure 34 Application du recalage entre UVR et VIS de la deuxième scène - (a) Méthodologie SURF
 (b) Notre méthodologie

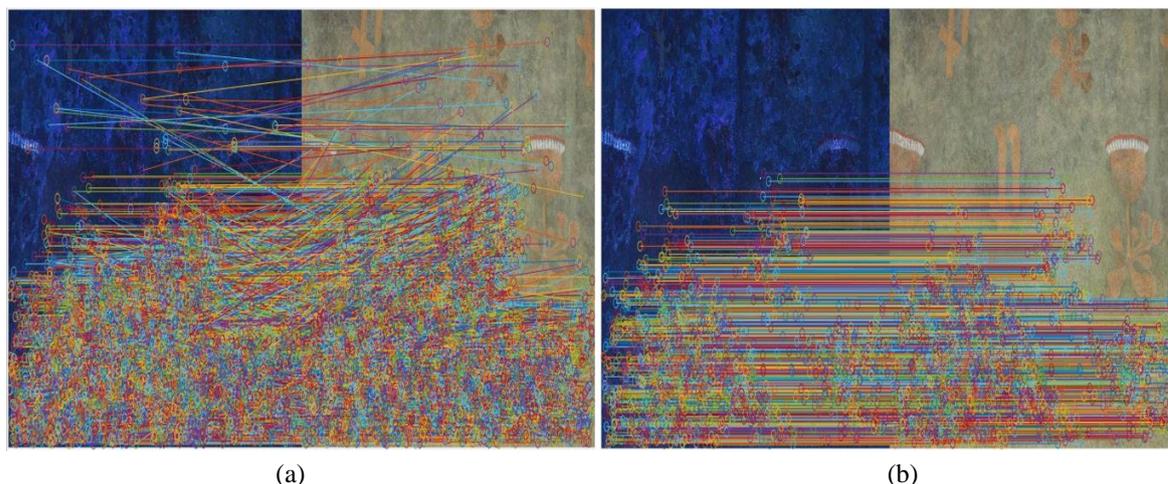


Figure 35 Application du recalage entre UVF et VIS de la deuxième scène - (a) Méthodologie SURF
(b) Notre méthodologie

Notre algorithme est rapide pour les jeux de correspondances et supprime tous les points aberrants (le cas des grandes rotations sera traité dans le deuxième algorithme) sans aucune intervention manuelle grâce au seuil adaptatif avec tous les types de modalités contrairement à l'algorithme RANSAC qui exige de définir manuellement les paramètres qui sont difficiles à déterminer dans certains cas. Les expérimentations ont montré que SURF avec notre méthode donne de bons résultats par rapport au SURF et SIFT avec RANSAC qui ont montré de faibles performances en termes de rappel. Les résultats expérimentaux présentés dans le Tableau 1 illustrent l'efficacité de la méthode proposée.

On constate particulièrement que même un petit nombre de points aberrants peut altérer le processus de recalage des images en produisant des images recalées certes, mais déformées. Pour vérifier si le recalage a un résultat optimal et pour mesurer la similarité structurelle entre les deux images, une mesure de similarité "mean SSIM" (MSSIM) entre l'image requête (après le recalage) et l'image modèle est implémentée.

Une comparaison de la méthode proposée avec d'autres algorithmes (SURF classique, SURF avec RANSAC et SIFT avec RANSAC) est proposée et la moyenne des valeurs MSSIM sur les tests globaux de la base de données de 100 images est présentée dans le Tableau 3.

Nous observons que les meilleures moyennes de SSIM (MSSIM) sont obtenues lorsque SURF est combiné avec notre méthode. L'identification des points aberrants des descripteurs SURF et la sélection des valeurs les plus fiables pour l'estimation de la transformation permettent au recalage d'être plus efficace et robuste.

Modalité	MSSIM (en %)			R_{rec} ratio		
	IRR	UVR	UVF	IRR	UVR	UVF
SURF-Classique	65,30	61,65	60,15	-	-	-
SURF + RANSAC	78,08	72,28	70,06	0,42	0,46	0,48
SIFT + RANSAC	74,54	75,01	74,07	0,67	0,63	0,83
SURF + Notre méthode	95,12	93,05	92,80	0,69	0,87	0,93

Tableau 3 Évaluation des performances du recalage sur la totalité des images multimodales de notre base de données

Pour vérifier la généralité et la généricité de notre approche, nous l'avons évalué aussi en utilisant des images satellitaires de haute résolution représentant les scènes des zones [168].

Les images utilisées proviennent du satellite QUICKBIRD haute résolution, avec une résolution spatiale élevée de 2,4M/pixels, ainsi que des images satellitaires IKONOS haute résolution, avec une résolution spatiale élevée de 4,1M pixels, qui représentent les zones d'une scène urbaine.

Comme pratiqué dans le domaine de télédétection, nous avons évalué le recalage à l'aide de trois métriques de similarité pour mesurer la correspondance entre images [29]. Nous avons utilisé la mesure de similarité "Erreur quadratique moyenne" (MSE), l'"Informations mutuelles" (MI) et la "Corrélation croisée normalisée" (NCC) entre l'image de requête après le recalage, avec notre méthode, et l'image modèle, et ce en comparant les résultats avec les méthodes classiques SURF et SURF + RANSAC.

Notre méthode montre clairement de meilleurs scores en termes de MSE, MI et NCC. Ces résultats, présentés dans le tableau 4, montrent que notre méthode est plus distinctive et plus performante que l'algorithme de RANSAC.

Modalité	IKONOS			Quickbird		
	MI	MSE	NCC	MI	MSE	NCC
SURF-Classique	0.0821	1.1234e+03	0.7741	0.4984	813.9660	0.8129
SURF + RANSAC	0.7710	380.959	0.9389	0.5583	731.5788	0.8522
SURF + Notre méthode	0.8204	376.1443	0.9500	1.0492	260,3233	0,9466

Tableau 4 Évaluation des performances du recalage sur les images satellitaires

4.4. Le filtrage des points de correspondance basé le critère de la distance et de l'orientation

Bien que la méthode proposée dans la section précédente ait montré de bonnes performances comparées aux autres méthodes, dans le cas de déformations de translations et de changement d'échelle, et même sur les images satellitaires, une limite a été relevée lorsque la déformation de rotation est importante. Ceci nous a amenés à réfléchir à une amélioration pour surmonter ce problème. Cette amélioration de la méthodologie, présentée dans cette section, est basée sur une modification du deuxième filtre de telle sorte que les déformations de rotation soient prises en compte. Cette nouvelle approche, décrite par l'algorithme 2 et dans l'organigramme de la Figure 36, s'applique comme suit :

Nous avons gardé le premier filtre basé sur la distance comme indiqué sur **A1.1**, **A1.2**, **A1.3** sur l'organigramme de la Figure 24, et déjà détaillé dans la section précédente.

Après l'application du premier filtre, un ensemble de caractéristiques filtrées est obtenu et identifié par $X1_{filt}$ et $X2_{filt}$ à partir de l'image modèle Im_{Is} et de l'image de requête Im_{Ir} . Ensuite, deux paires de points caractéristiques sont définies comme des points de référence et identifiés respectivement par $X1_{ref1}$ et $X1_{ref2}$ dans Im_{Is} et $X2_{ref1}$ et $X2_{ref2}$ dans Im_{Ir} . Ces paires de points choisies sont celles qui minimisent la distance en termes de descripteurs.

Afin de supprimer tous les points aberrants restants et de garantir l'invariance à la rotation, l'étape suivante consiste à effectuer un autre filtre sur l'ensemble de caractéristiques obtenues comme suit :

A1.4 Dans cette étape, plusieurs opérations sont effectuées pour obtenir toutes les mesures utilisées dans le deuxième filtre.

- Les deux distances, entre $X1_{ref1}$ et $X1_{ref2}$ notée $dist_{Is}$ et entre $X2_{ref1}$ et $X2_{ref2}$ notée $dist_{Ir}$, sont calculées. Le rapport $\frac{dist_{Ir}}{dist_{Is}}$ entre ces deux distances est ensuite calculé, il représente l'échelle entre les deux images et est noté k_{scale} .
- Les distances $dist_{ref1_X1}$ et $dist_{ref1_X2}$ sont obtenues à partir du calcul de la distance euclidienne ($X1_{ref1}, X1_{filt}$) et ($X2_{ref1}, X2_{filt}$) à partir de Im_{Is} et Im_{Ir} . Ensuite, la mesure k_{dist_ref1} est acquise en calculant $\frac{dist_{ref1_X2}}{dist_{ref1_X1}}$.
- La troisième mesure k_{dist_ref2} est obtenue à partir du calcul de la distance ($X1_{ref2}, X1_{filt}$) et ($X2_{ref2}, X2_{filt}$) de Im_{Is} et Im_{Ir} , respectivement, comme suit : $k_{dist_ref2} = \frac{dist_{ref2_X2}}{dist_{ref2_X1}}$.
- Les vecteurs entre (le premier point de référence, le deuxième point de référence) et (le deuxième point de référence, les caractéristiques filtrées définies) dans les deux images sont construits, puis normalisés et leurs orientations sont notées $Orientation_{X1}$ et $Orientation_{X2}$ dans Im_{Is} et Im_{Ir} respectivement. Ensuite, la mesure $\frac{Orientation_{X2}}{Orientation_{X1}}$ est calculée et identifiée par $k_{orientation}$. Une illustration est présentée à la Figure 37.

A1.5 Le deuxième filtre appliqué consiste à prendre toutes les caractéristiques satisfaisant la condition suivante :

$$(k_{dist_ref1} = k_{scale}) \cap (k_{dist_ref2} = k_{scale}) \cap (k_{orientation} = 1).$$

La Figure 2 montre un exemple des distances et de l'orientation calculées pour un point inlier.

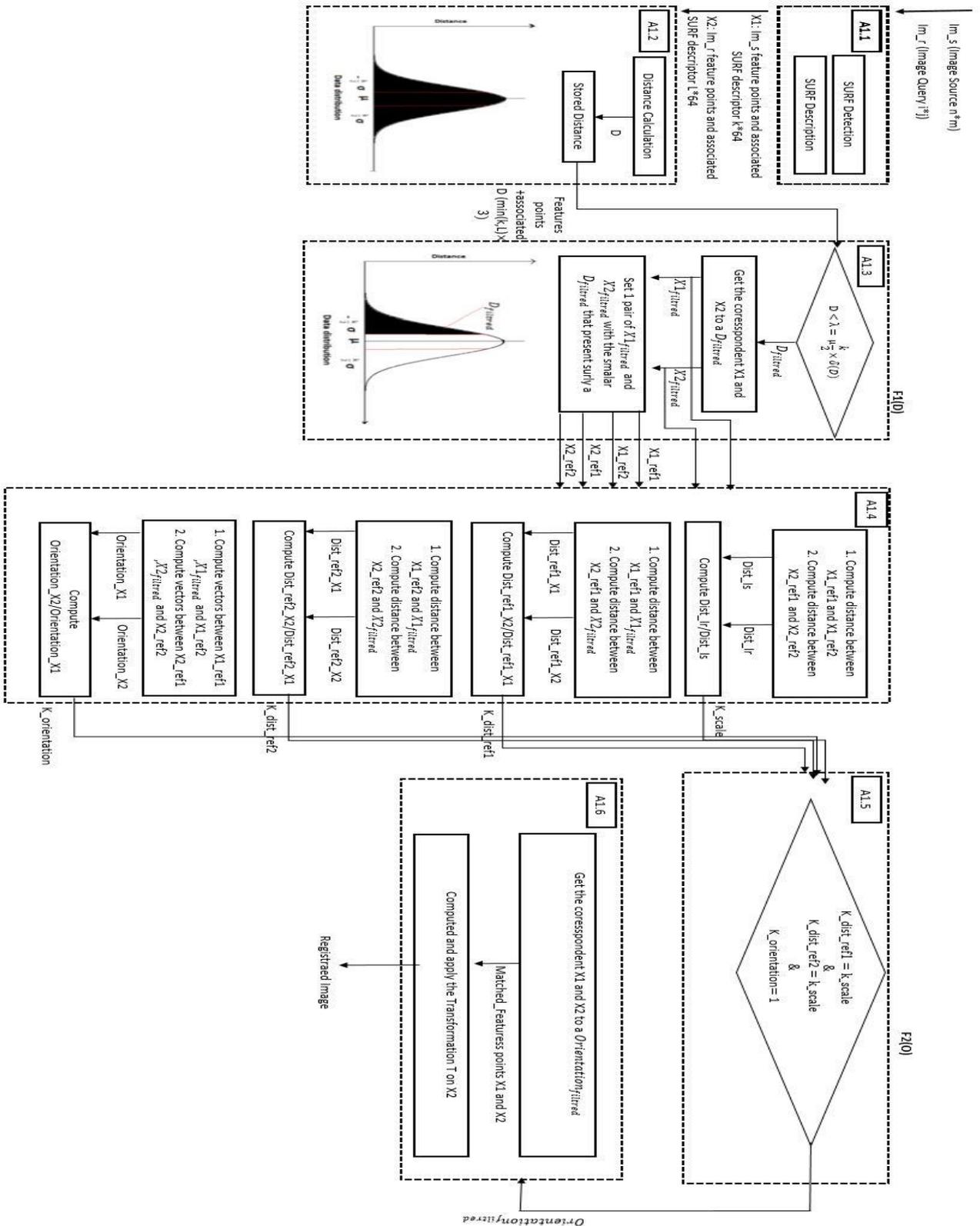


Figure 36 Schéma représentant l'approche proposée.

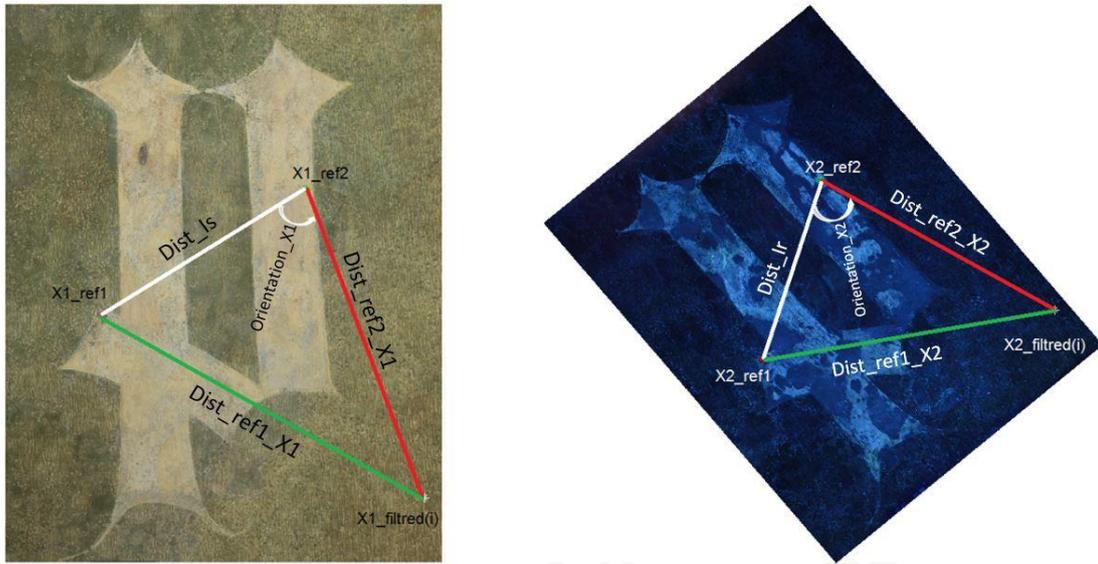


Figure 37 : Exemple de distances calculées et d'orientation pour un point inlier

Les résultats de cette opération de filtrage présentés dans les Figures 38, 39, 40, qui illustrent le fait que seules les paires de points ayant $Orientation_{X1} = Orientation_{X2}$ de Im_{Is} et Im_{Ir} respectivement, sont conservées.

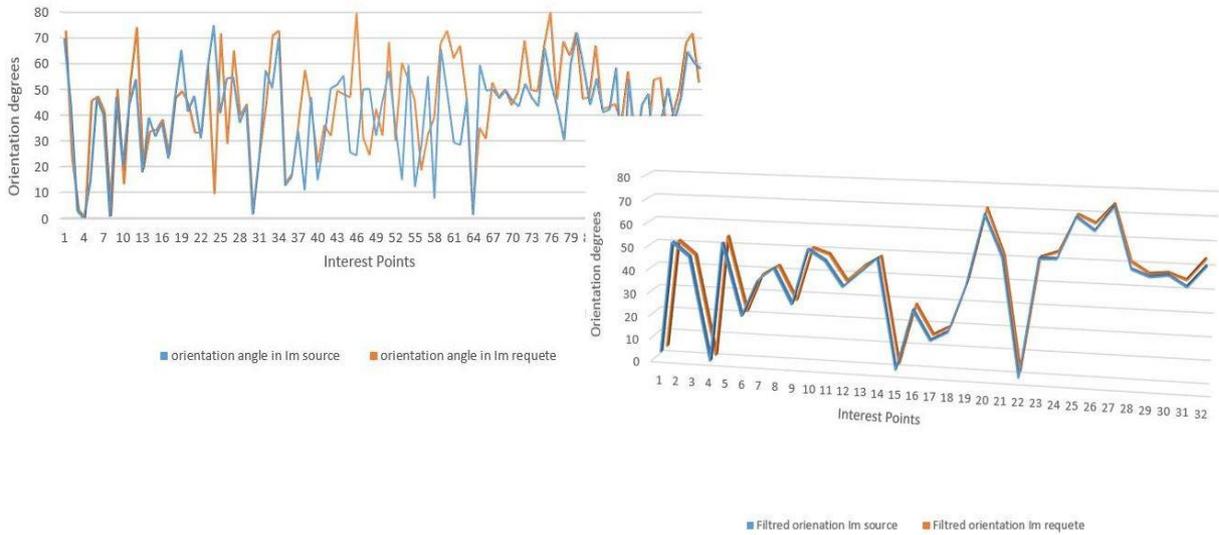


Figure 38 : Distributions des orientations d'IRR (à droite) avant filtrage (à gauche) après filtrage

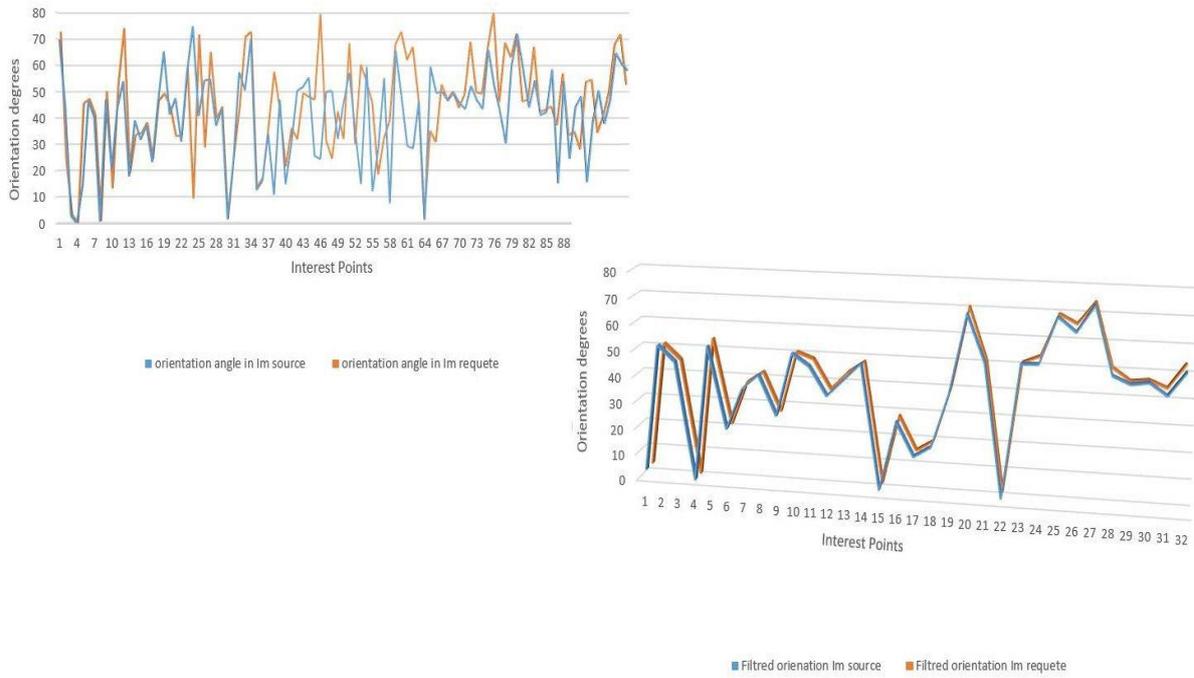


Figure 39 Distributions des orientations d’UVR (à droite) avant filtrage (à gauche) après filtrage

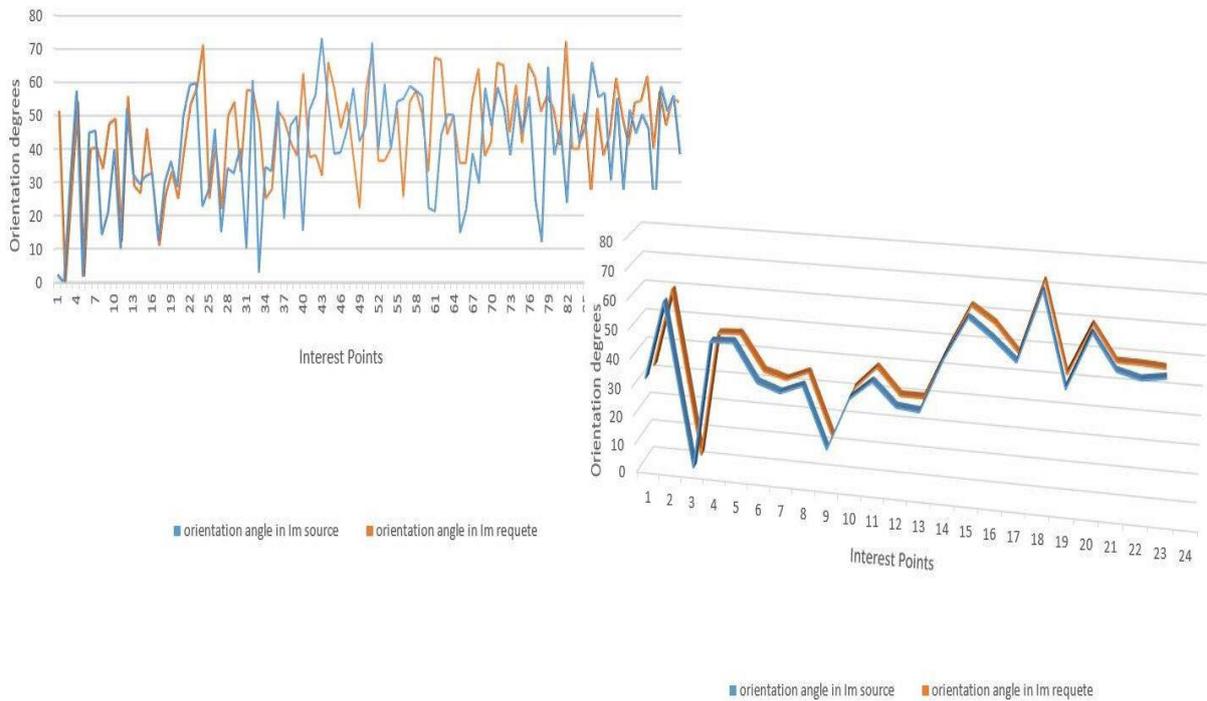


Figure 40 Distributions des orientations d’UVF (à droite) avant filtrage (à gauche) après filtrage

Algorithme 2

Entrée :

$centroid_{Is}$: le centre de l'image modèle

$centroid_{Ir}$: le centre de l'image requête

D : la distance euclidienne entre les paires caractéristiques.

K : coefficient de l'écart type.

X1 : les points caractéristiques de l'image requête.

X2 : les points caractéristiques de l'image modèle.

Sortie :

1 : 1ere filtre sur la distance: Calcule le premier ensemble d'inliers

2 : $D_{\text{filtred}} \leftarrow D (D \leq \text{mean}(D) - \frac{k}{2} \times (\sigma))$

3 : Triez D_{filtred}

4 : $X1_{\text{filt}} \leftarrow X1(D_{\text{filtred}})$, $X2_{\text{filt}} \leftarrow X2(D_{\text{filtred}})$

5 : Sélectionnez la première paire d'éléments filtrés $X1_{\text{filt}}$ et $X2_{\text{filt}}$

6 : $X1_{\text{ref}} \leftarrow X1_{\text{filt}}(1)$ $X2_{\text{ref}} \leftarrow X2_{\text{filt}}(1)$

7 : $\text{dist}_{Is} = \text{dist}(\text{centroid}_{Is}, X1_{\text{ref}})$

8 : $\text{dist}_{Ir} = \text{dist}(\text{centroid}_{Ir}, X2_{\text{ref}})$

9 : $k_{\text{scale}} = \frac{\text{dist}_{Ir}}{\text{dist}_{Is}}$

10 : $\text{dist}_{\text{centre_X1}} = \text{dist}(\text{centroid}_{Is}, X1_{\text{filt}})$

11 : $\text{dist}_{\text{centre_X2}} = \text{dist}(\text{centroid}_{Ir}, X2_{\text{filt}})$

12 : $K_{\text{dist_centre}} = \frac{\text{dist}_{\text{centre_X2}}}{\text{dist}_{\text{centre_X1}}}$

13 : $\text{dist}_{\text{ref_X1}} = \text{dist}(X1_{\text{ref}}, X1_{\text{filt}})$

14 : $\text{dist}_{\text{ref_X2}} = \text{dist}(X2_{\text{ref}}, X2_{\text{filt}})$

15 : $K_{\text{dist_ref}} = \frac{\text{dist}_{\text{ref_X2}}}{\text{dist}_{\text{ref_X1}}}$

16 : Construire les vecteurs de correspondances $\overrightarrow{V1_{Is}}$, $\overrightarrow{V2_{Is}}$, $\overrightarrow{V1_{Ir}}$, $\overrightarrow{V2_{Ir}}$

17 : Normaliser les vecteurs $\text{norm}V1_{Is}$, $\text{norm}V2_{Is}$, $\text{norm}V1_{Ir}$, $\text{norm}V2_{Ir}$

18 : Calculer les orientations des vecteurs

19 : $\text{Orientation_X1} = \arctan\left(\frac{\text{norm}V2_{Is}}{\text{norm}V1_{Is}}\right)$

20 : $\text{Orientation_X2} = \arctan\left(\frac{\text{norm}V2_{Ir}}{\text{norm}V1_{Ir}}\right)$

21 : calculer $K_{\text{orientation}} = \frac{\text{Orientation_X2}}{\text{Orientation_X1}}$

22 : le deuxième filtre sur l'orientation

23 : Inliers \leftarrow Garder les points où $K_{\text{dist_centre}} = K_{\text{scale}} \cap K_{\text{dist_ref}} = K_{\text{scale}} \cap K_{\text{orientation}} = 1$

24 : Calculer la transformation affine T surtout les Inliers

Les performances de l'utilisation du descripteur SURF, sans et avec notre conditionnement de données, sont évaluées en calculant l'erreur moyenne RMSE. Cet indicateur est utilisé comme une mesure de la qualité, les valeurs proches de zéro étant les meilleures.

Les expériences ont montré que SURF avec notre méthode donne de bons résultats par comparaison avec SURF et SIFT avec RANSAC. SURF et SIFT avec RANSAC ont enregistré de faibles performances en termes de RMSE et MSSIM. Les résultats expérimentaux sur la totalité des images de notre base de données sont présentés dans le tableau 5.

Modalité	MSSIM (en %)			RMSE		
	IRR	UVR	UVF	IRR	UVR	UVF
SURF-Classique	69,02	63,73	65,73	-	-	-
SURF + RANSAC	77,78	73,3	72,8	13.2697	15.7216	15.6296
SIFT + RANSAC	76,62	76,61	76,9	15.9687	15.97	15.9480
SURF + Notre méthode	98.90	98.31	98,26	2.3606	0.1346	0.8240

Tableau 5 Évaluation des performances du recalage avec la deuxième méthode sur la totalité des images multimodal de notre base de données

- La première série d'images utilisées pour l'illustration présente une scène d'une portion du mur ouest du vestiaire du Duc de Bourgogne, la première scène est présentée dans la Figure 41.

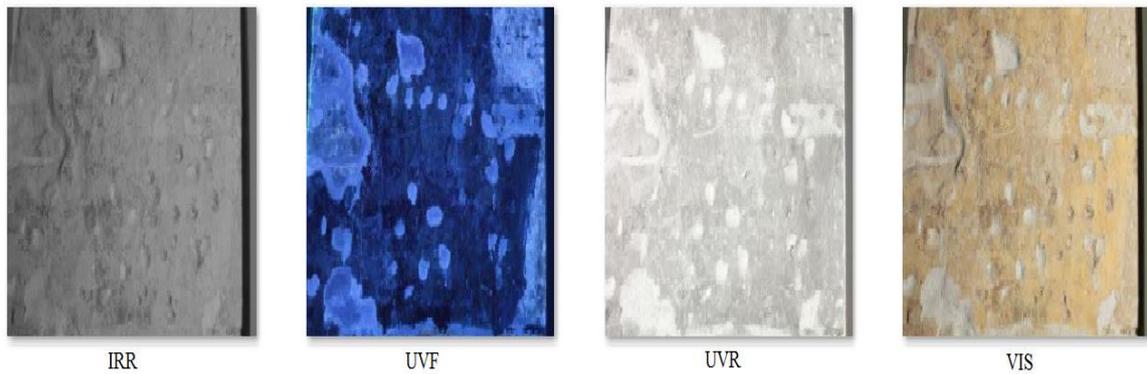
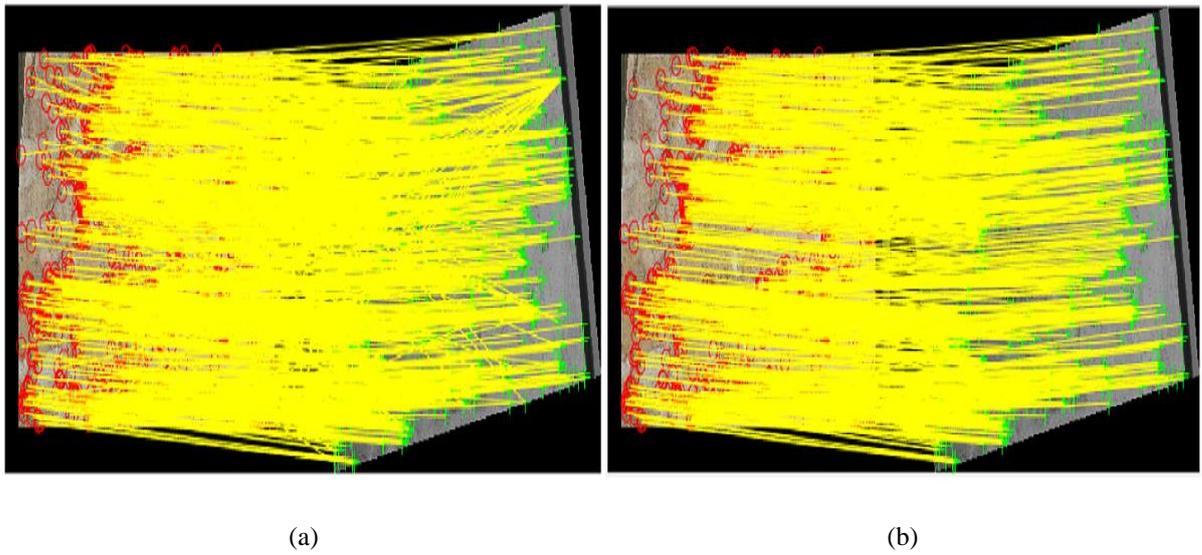


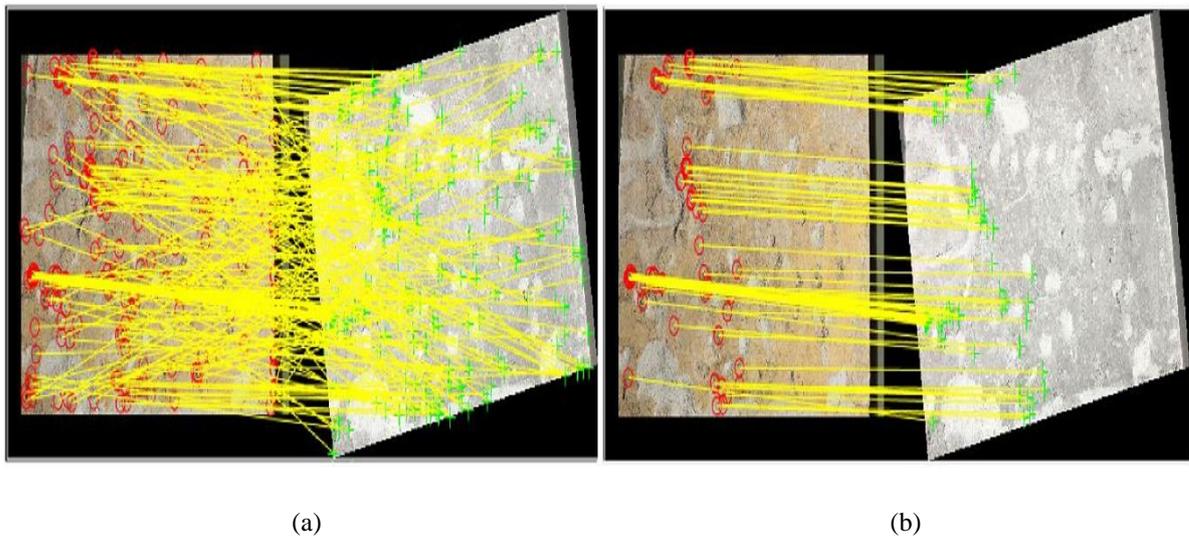
Figure 41 les quatre modalités pour la deuxième scène sur le mur ouest du Duc de Bourgogne

Les Figures 42, 43 et 44 représentent respectivement le recalage entre les images IRR, UVR et UVF (les images requêtes), et l'image modèle VIS de la scène illustrée dans la Figure 41, où (a) montre les correspondances de points putatives et (b) indique les points de correspondances après suppression des « outliers » en utilisant notre deuxième méthode.



◊ Les points correspondants sur l'image source I_s
 + Les points correspondants sur l'image requête I_r

Figure 42 Application du recalage entre IRR et VIS de la première scène pour la deuxième contribution-
 (a) Méthodologie SURF (b) Notre méthodologie



◊ Les points correspondants sur l'image source I_s
 + Les points correspondants sur l'image requête I_r

Figure 43 Application du recalage entre UVR et VIS de la première scène pour la deuxième contribution-
 (a) Méthodologie SURF (b) Notre méthodologie

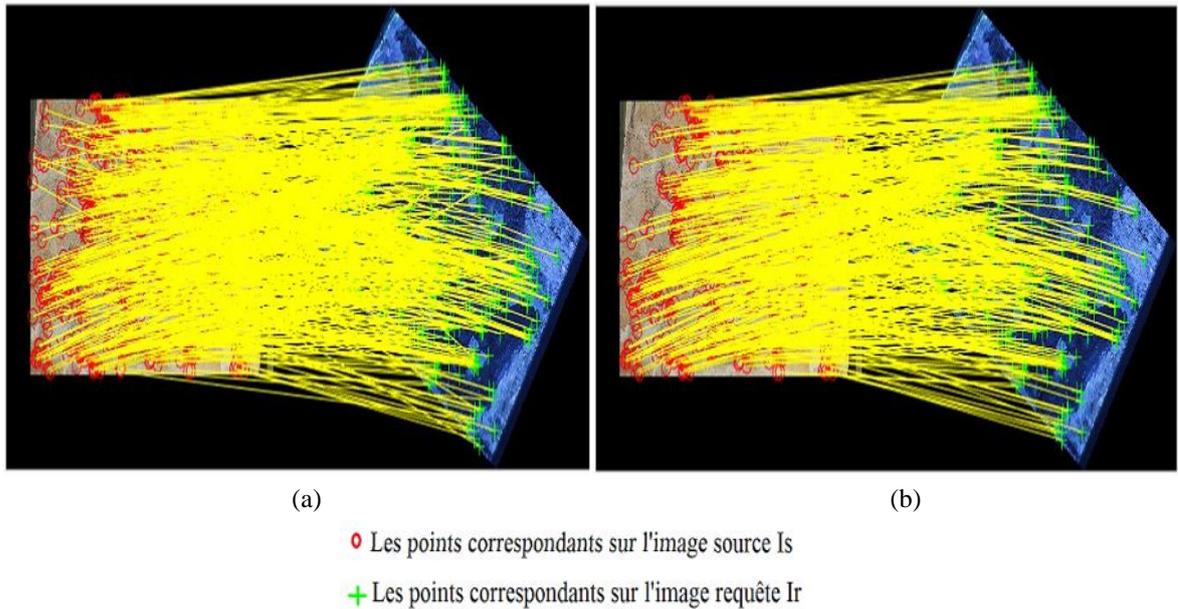


Figure 44 Application du recalage entre UVF et VIS de la première scène pour la deuxième contribution-
 (a) Méthodologie SURF (b) Notre méthodologie

- La deuxième série des images utilisées pour l'illustration présente l'une des moulures des peintures murales sur le mur nord du dressing de la comtesse de Nevers (Figure 45).

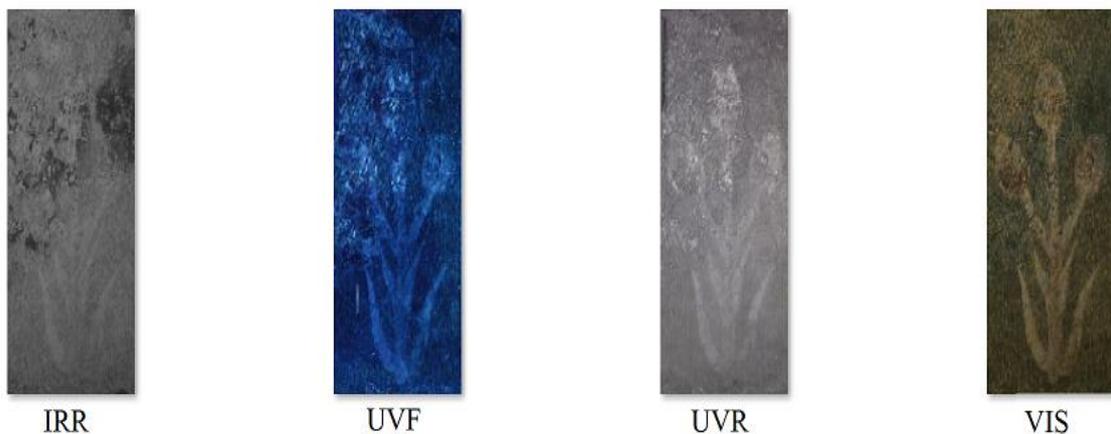


Figure 45 les quatre modalités pour la deuxième scène sur le mur nord du dressing de la comtesse de Nevers

Les Figures 46, 47 et 48 présentent respectivement le recalage entre les images IRR, UVR et UVF (les images requêtes), et l'image modèle VIS de la scène illustrée dans la Figure 45, où (a) montre les correspondances de points putatives et (b) indique les points de correspondances après enlevés les « outliers » en utilisant la deuxième méthode proposée.

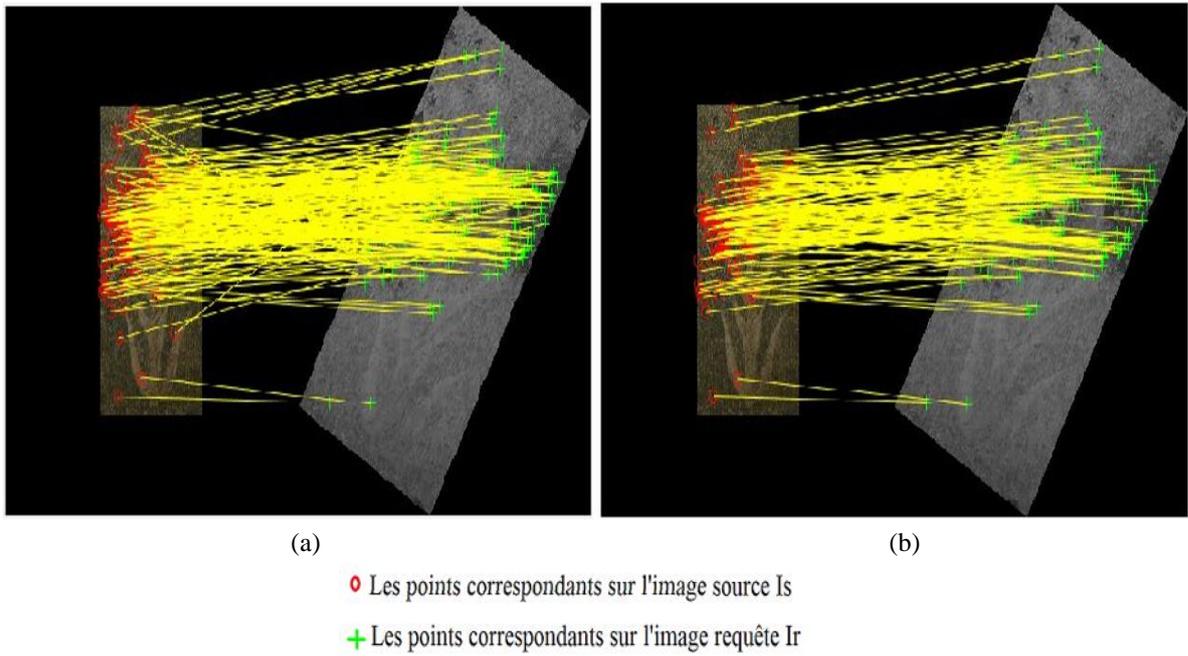


Figure 46 Application du recalage entre IRR et VIS de la deuxième scène pour la deuxième contribution-
 (a) Méthodologie SURF (b) Notre méthodologie

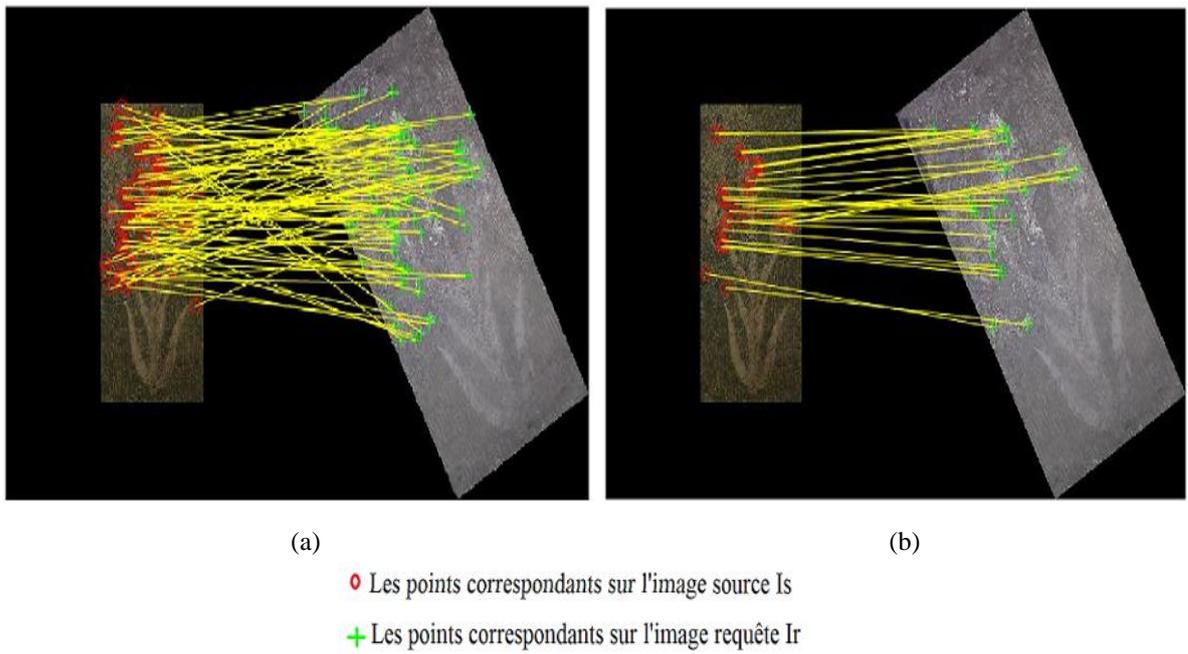


Figure 47 Application du recalage entre IRR et VIS de la deuxième scène pour la deuxième contribution-
 (a) Méthodologie SURF (b) Notre méthodologie

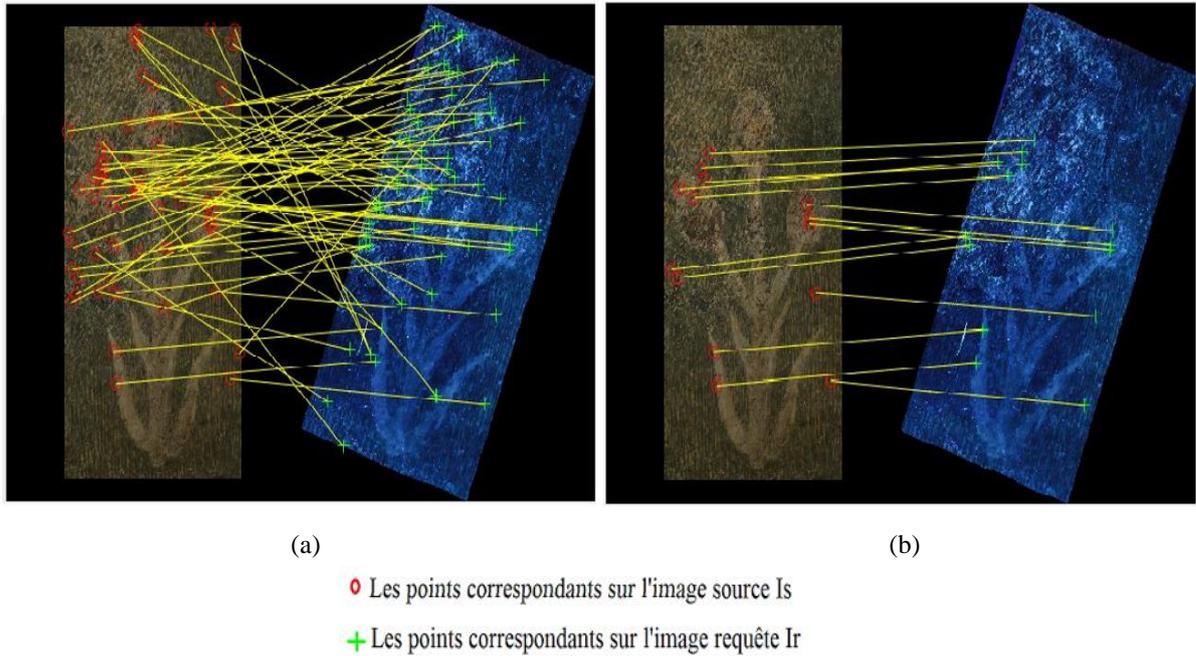


Figure 48 Application du recalage entre IRR et VIS de la première scène pour la deuxième contribution- (a) Méthodologie SURF (b) Notre méthodologie

4.5. Conclusion

Dans ce chapitre, nous avons commencé par présenter la problématique liée au recalage des images multimodales basé sur l'utilisation des méthodes géométriques, qui consiste à appairer les points d'intérêts détectés et extraits sur l'image modèle et requête par SURF. Puis, une mise en correspondance des points détectés a été effectuée pour déterminer le recalage à effectuer.

Ensuite, nous avons décrit notre première approche visant à détecter les correspondances correctes à partir d'un ensemble de correspondances candidates. Nous avons adapté des filtres existants de détection des points aberrants, en nous basant sur la distribution statistique des données des distances et des orientations calculées entre chaque paire de points de correspondance dans les cas de déformation de translation et du changement d'échelle. Puis, nous avons présenté une amélioration de cette première méthode pour couvrir les déformations liées à de grandes rotations. Un deuxième filtre basé sur l'orientation calculée entre les paires de correspondance a été utilisé. Il consiste à combiner la distance et l'orientation comme une condition à vérifier pour que la paire de points soit qualifiée pour le calcul du recalage. Les résultats expérimentaux obtenus sont présentés. La combinaison de la méthode SURF et des algorithmes mis en place dans cette section semble constituer un moyen performant et robuste de recherche des correspondances cohérentes entre deux ensembles de caractéristiques en présence de variations d'intensité non linéaires élevées.

Dans le chapitre suivant, nous présentons l'étape qui suit le recalage dans le processus de fusion, à savoir l'extraction caractéristique et la pré-classification.

Chapitre 5

Classification des images multimodales basée sur la fusion des caractéristiques spectrales et Dense SURF

5.1. Introduction

Ce chapitre présente la troisième contribution de notre travail de thèse. Elle consiste dans un premier temps à projeter les quatre modalités d'imagerie utilisées dans ce travail dans un espace de représentation commun afin de faciliter leur comparaison. Les images projetées vont servir à traiter le problème de pré-classification des images multimodales pour fournir les matrices de probabilités pour la phase de la fusion.

L'approche proposée est basée sur le principe Dense SURF que nous introduisons et qui consiste en l'extraction des caractéristiques spatiales SURF pour chaque pixel pour chacune des modalités. Ainsi pour la tâche de classification, il ne s'agit pas uniquement d'utiliser l'information spectrale comme unique caractéristique, mais de la combiner avec les caractéristiques spatiales et de les utiliser conjointement comme entrée du classificateur SVM.

Le présent chapitre est organisé comme suit : les deux premières sections présentent brièvement le background du classifieur SVM. Nous décrirons dans la troisième section l'étape de préparation des échantillons d'apprentissage. La méthode proposée sera décrite en détail dans la quatrième section puis nous présenterons et discuterons les résultats obtenus et nous donnerons des illustrations avec quelques scènes de notre base de données.

5.2. Le classifieur SVM

Le travail que nous présentons dans cette thèse se base sur la classification SVM. Ce choix est motivé par les résultats d'une étude comparative que nous avons menée entre le classifieur Arbre de décision (Decision Tree DT) et le classifieur KNN (K Nearest Neighbors). Les meilleurs résultats obtenus par la classification SVM s'expliquent à notre sens par le fait que d'une part, l'approche SVM permet au niveau apprentissage de faire une analyse statistiquement significative même avec peu d'échantillon. D'autre part, au niveau de la classification, il a la capacité de traiter des données de haute dimensionnalité, ce qui est le cas de nos données.

Le classifieur SVM a été initialement utilisé pour les classifications binaires. Son principe [169] [170] [171] consiste à trouver une séparation des classes en maximisant la marge séparatrice entre celles-ci. Il comporte deux modèles : le modèle linéaire et le modèle non linéaire (détaillé en annexe). La relation non linéaire entre les données extraites de chaque modalité a été l'une des fortes motivations qui nous ont poussés à utiliser le SVM non linéaire. La nature de la base de données utilisée exige en effet une séparatrice de décision plus complexe qu'un simple hyperplan linéaire. Pour cette raison, une généralisation des SVMs linéaires était introduite en utilisant la notion de noyaux.

Afin d'y parvenir, les données non linéaires sont mappées dans un nouvel espace de représentation intermédiaire H de grande dimension :

$$\Phi: \mathbb{R}^n \rightarrow H \quad 5.1$$

La recherche de l'hyperplan séparateur optimal dans l'espace H est ensuite appliquée : une application Φ de \mathbb{R}^n est utilisée muni d'un produit scalaire dans H , dont une fonction noyau $K: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ est définie en vérifiant :

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle, \quad 5.2$$

L'application visée dans ce travail est une application classification multi-classes, ce qui implique l'utilisation de la version multi-classes du classifieur SVM et cette extension peut être par deux approches :

Un contre le reste (One against all (OAA)) : k classificateurs binaires sont appliqués sur chaque classe contre les autres. Chaque échantillon est affecté à la classe avec la sortie maximale.

Un contre un (One against one (OAO)) : $k(k - 1)/2$ des classificateurs binaires sont appliqués à chaque paire de classes. Chaque échantillon est attribué à la classe qui obtient le plus grand nombre de votes. Un vote pour une classe donnée est défini comme un classifieur assignant le motif à cette classe.

5.3. L'établissement des échantillons d'apprentissage

Les stratégies d'échantillonnage qui existent dans la littérature sont nombreuses. On peut citer ici de façon non exhaustive la méthode aléatoire simple, la méthode aléatoire stratifiée, la méthode systématique, la méthode stratifiée systématique non-alignée ou encore la méthode stratification par grappes [172]. Toutefois, il est à noter qu'il n'existe pas une règle générale pour définir la meilleure stratégie.

Selon la comparaison effectuée par *Congalton* et al. [172], les méthodes aléatoires, stratifiées ou non, donnent des résultats acceptables dans tous les cas. Cependant, la difficulté de garantir un nombre suffisant et représentatif des échantillons pour chaque classe est un inconvénient majeur des méthodes non stratifiées. L'échantillonnage stratifié systématique

non-aligné a de plus l'avantage d'utiliser une maille systématique évitant ainsi le problème de la périodicité. La méthode stratification par grappes permet l'adaptation de la stratégie utilisée à chaque strate.

La stratégie utilisée dans ce travail est l'échantillonnage stratifié systématique non-aligné car cette approche combine à notre sens les avantages des méthodes stratifiées et aléatoires.

L'unité d'échantillonnage dépend de l'unité spatiale d'analyse et de l'application visée [173]. L'unité adoptée dans ce travail est le pixel sans aucune généralisation. Ainsi, pour chaque modalité de la même scène, nous avons créé deux fichiers, le premier contient les pixels d'apprentissage et le deuxième contient les pixels de la classification. Les pixels utilisés sont divisés en classes selon la modalité de la manière suivante :

- La modalité VIS : les pixels sont divisés en quatre classes qui sont : le blanc original (*Original White (WO)*), le vert original (*Original Green (GO)*), le blanc repeint (*Repainted White (WR)*), le vert repeint (*Repainted Green (GR)*).
- La modalité IRR : les pixels sont divisés en trois classes qui sont : le vert original (*Original Green (RO)*), le vert repeint (*Repainted Green (GR)*) et une classe appelée Otherwise qui contient les pixels dont on ignore l'appartenance au blanc original (*Original White (WO)*) ou bien au blanc repeint (*Repainted White (WR)*).
- La modalité UVR : les pixels sont divisés en deux classes qui sont : le blanc original (*Original White (WO)*) et une classe Otherwise qui contient les pixels appartenant au vert original (*GO*), le vert repeint (*Repainted Green (GR)*) et blanc repeint (*Repainted White (WR)*).
- La modalité UVF : La stratégie est identique à celle utilisée avec la modalité UVR, les pixels sont divisés en deux classes qui sont : le blanc original (*Original White (WO)*) et une classe Otherwise qui contient tous les pixels appartenant au vert original (*GO*), le vert repeint (*Repainted Green (GR)*) et blanc repeint (*Repainted White (WR)*).

5.4. La classification des images multimodales basée sur Dense SURF et l'information spectrale

Afin d'assurer une bonne classification des images multimodales, notre contribution est basée sur deux axes :

- (1) le premier axe consiste à étendre l'extraction des descripteur SURF à tous les pixels de l'images (non seulement aux points d'intérêt) dans le but de pouvoir projeter toutes les modalités utilisées dans un espace de représentation homogène.
- (2) Le deuxième consiste à combiner les informations spatiales (Dense SURF) et les informations spectrales pour obtenir une meilleure classification.

Le processus de classification peut être divisé en trois étapes principales comme suit :

- (1) l'extraction des caractéristiques
- (2) l'étape d'apprentissage
- (3) l'étape de test ou de classification.

L'utilisation conjointe des informations spatiales et spectrales améliore la capacité de représentation des primitives extraites [174]. Ainsi, lors de l'étape d'extraction des caractéristiques, un descripteur « dense », c'est-à-dire combiné avec les caractéristiques spectrales est appliqué à chaque pixel des différentes images/modalités.

Dans la phase d'apprentissage, les caractéristiques combinées sont apprises par SVM pour fournir le modèle de classification. Ce modèle est utilisé dans l'étape de classification pour attribuer à chaque pixel une étiquette de classe. Une architecture simplifiée du processus adopté est illustrée dans la Figure 49.

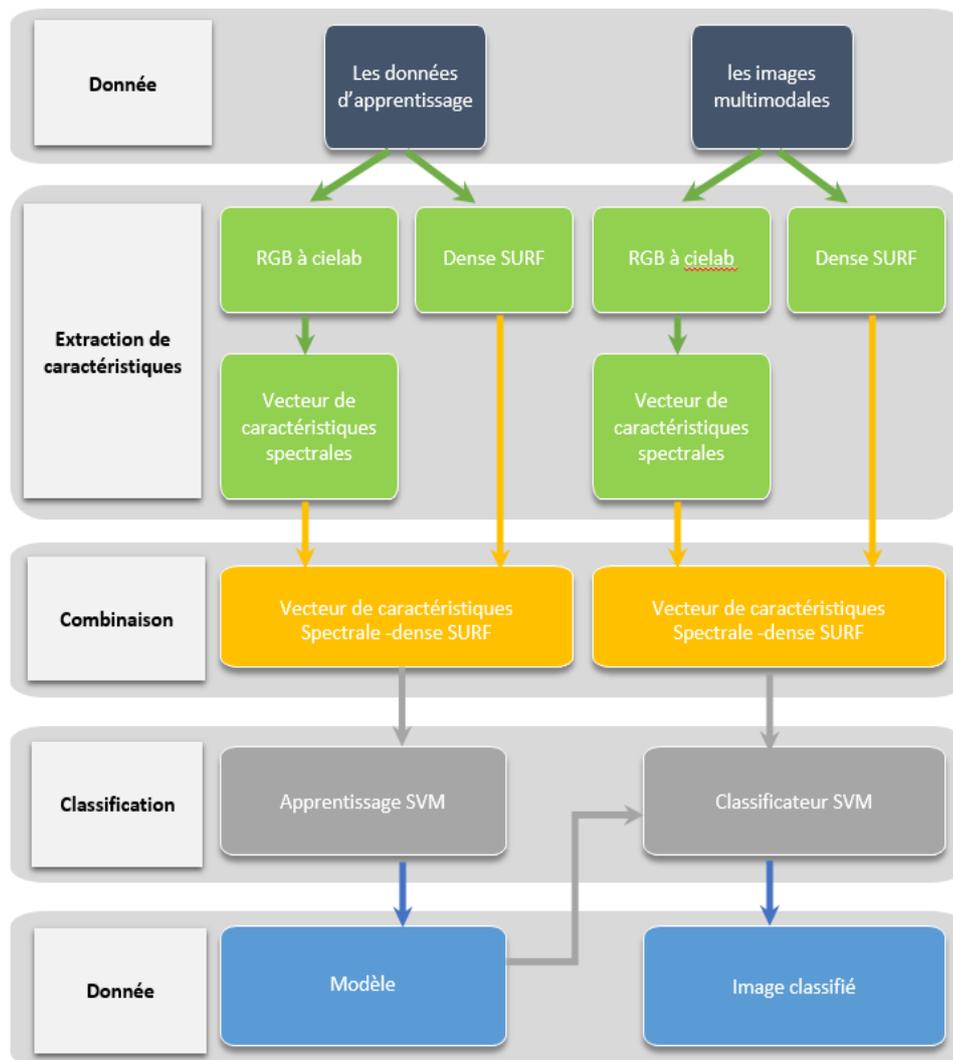


Figure 49 Architecture générale de l'approche proposée

Les procédures de la classification des images multimodales que nous proposons peuvent être décrites comme suit :

➤ Phase d'apprentissage :

Nous commençons par l'extraction des caractéristiques spectrales pour chaque échantillon d'apprentissage, présentées initialement dans l'espace RVB puis exprimées dans l'espace CIELab. Cet espace fournit en effet une meilleure corrélation/approximation de la couleur avec la perception humaine, et de meilleurs résultats expérimentaux. Les valeurs des pixels extraites dans l'espace CIELab sont considérées directement comme données d'entrée du classifieur SVM. Nous extrayons ensuite les caractéristiques spatiales pour chaque pixel des échantillons d'apprentissage. L'idée proposée pour effectuer cette étape est d'utiliser les caractéristiques (SURF) proposées par *Herbert Bay* [175].

L'originalité de notre approche réside dans l'utilisation des caractéristiques Dense au lieu des caractéristiques éparsees fournies par SURF. L'objectif recherché est de projeter toutes les modalités dans un espace de représentation commun. Pour ce faire, nous avons ignoré la phase de détection et nous avons effectué une description pour chaque pixel de l'image. Ceci est fait en assignant à chaque pixel l'orientation dominante calculée en combinant les résultats d'ondelettes de Haar dans un voisinage circulaire autour de chaque pixel, ensuite nous divisons simplement l'image en créant des sous-régions de 4x4 superposées autour du pixel sur chaque point. Les réponses en ondelettes de Haar sont calculées dans une grille de points d'échantillonnage de 5×5 , où les distances s entre les points sont égales à un pixel $s = 1$. Dans chaque sous-région, nous additionnons ces réponses en ondelettes de Haar exprimées en pixels pour former un descripteur de 64 éléments.

Une fois les deux types de caractéristiques extraites, celles-ci sont fusionnées en concaténant les informations spectrales transformées à l'espace CIELab avec le descripteur SURF pour chaque pixel de l'image en formant un vecteur de dimension 67. Cette combinaison sera utilisée comme entrée pour le classifieur SVM pour générer un jeu de vecteurs supports, que nous appelons **le modèle d'apprentissage**.

➤ Dans la phase de classification :

Durant cette phase, nous appliquons les mêmes étapes décrites pour la phase d'apprentissage afin de calculer les descripteurs denses SURF, c'est-à-dire l'extraction des descripteurs spectraux, puis la combinaison de ces derniers d'une manière identique dans un même vecteur avec les descripteurs SURF.

C'est à l'étape de la prise de décision, qu'on se sert du modèle SVM obtenu à l'étape d'apprentissage. La prise de décision, concernant l'appartenance d'un pixel à une classe précise, s'effectue directement en utilisant la fonction de décision des SVMs.

5.5. Résultat et discussion

L'approche proposée a été testée sur des images multimodales fournies par les gestionnaires du château de Germolles, décrite dans la section 1.2 du chapitre 1. Généralement, les décorations murales dans les chambres portent sur de grandes lettres blanches P (pour Philip) et M (pour Marguerite) séparés par des moulures blanches (marque de fidélité) recouvrent les murs peints en vert comme présenté sur la Figure 50. Ces motifs caractérisaient l'esprit d'amour courtois très en vogue parmi les cours princières à la fin du XIVe siècle.



Figure 50 Exemple de la décoration murale des chambres du palais de Germolles

Les données que nous traitons sont des données de très grande dimension et non linéairement séparables, ce qui nous a conduit à l'utilisation du classifieur SVM non linéaire. Dans ce cas, le choix du noyau et le réglage des paramètres associés ont des effets sur le résultat final de la classification. Plusieurs types de noyaux ont été proposés dans la littérature, nous en avons retenu deux en particulier, à savoir :

1. Le noyau polynomial qui cherche une séparation des données par une fonction polynomiale de degré au plus de r ; exprimé comme suit :

$$K(x, y) = (x' y + c)^r, \text{ avec } c \in \mathbb{R}_+ \text{ et } r \in \mathbb{N} \quad 5.3$$

2. Le noyau gaussien qui cherche une séparation des données en utilisant les frontières radiales basées sur la gaussienne ; exprimé comme suit :

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

Etant donné la nature des données d'application de ces travaux, le choix de l'un ou l'autre des noyaux linéaires susmentionné n'a pas d'effet direct sur les résultats obtenus. En nous basant sur les tests effectués sur notre base de données, nous avons choisi de travailler avec le noyau RBF, qui nous a semblé avoir la capacité de représenter des relations de plus en plus complexes en fonction de la taille des données. Ces caractéristiques manquent au noyau polynomial, qui semble moins approprié lorsque les données dépassent une certaine taille.

L'utilisation du noyau RBF gaussien exige la définition de deux paramètres à savoir :

- le paramètre de régularisation (C) aussi appelé facteur de pénalisation, qui consiste à pondérer les erreurs de classification pour optimiser la fonction de séparation finale.
- La largeur du noyau (σ), qui permet de déterminer la similarité d'un exemple par rapport à ceux qui l'entourent, et par conséquent de contrôler la capacité du classifieur.

Pour déterminer la paire qui va donner le meilleur résultat, nous avons utilisé la méthode de grille de recherche couplée avec la méthode de validation croisée. Cette technique teste les différentes paires de paramètres et qualifie celle qui donne la validation croisée la plus élevée.

Afin d'illustrer nos propos, nous avons choisi deux ensembles d'images multimodales de différentes chambres du château de Germolles que nous allons décrire dans ce qui suit.

1. Le premier ensemble utilisé présente une scène sur le mur nord du dressing du Duc de Bourgogne. Cette scène porte sur la lettre M (Figure 51).

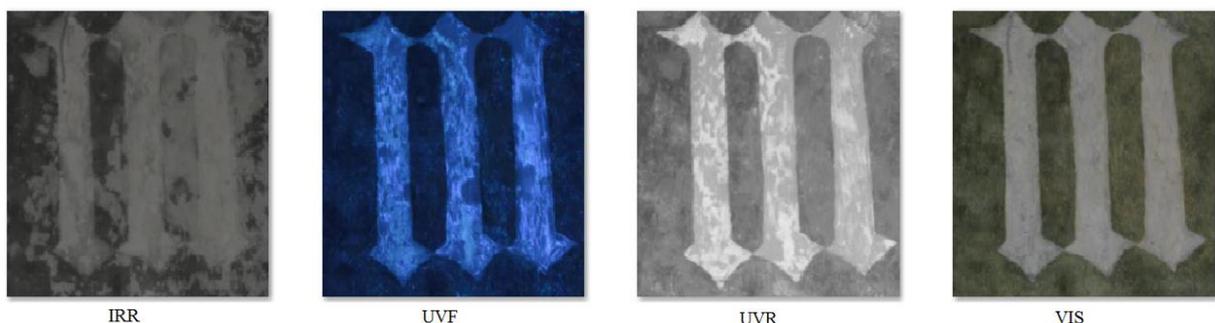


Figure 51 les quatre modalités pour la scène M sur le mur nord du dressing de duc de Bourgogne

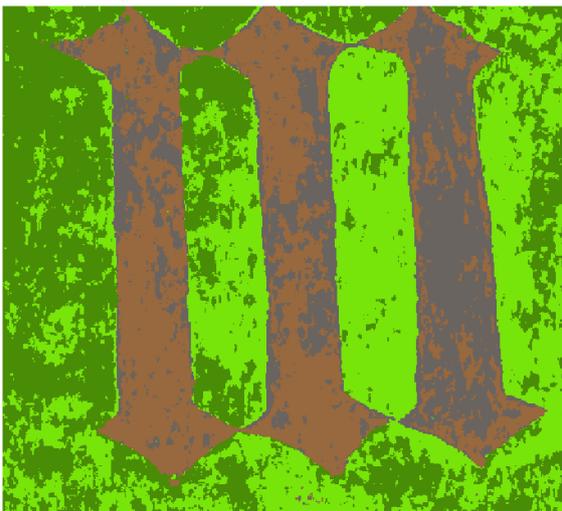
Pour chaque modalité de cette scène, nous construisons deux fichiers. Le premier fichier contient uniquement les informations spectrales sur lesquelles nous appliquons une classification SVM avec le noyau RBF gaussien basé sur la distance euclidienne minimale. Le

deuxième fichier contient la concaténation entre les informations spectrales et les informations spatiales obtenues en utilisant Dense SURF sur lesquelles nous appliquons la classification SVM basée sur la distance euclidienne minimale.

Les cartes de classification représentées par (b) dans les Figures 52, 53, 54, 55 sont obtenues en utilisant le premier fichier contenant des informations spectrales uniquement et elles donnent un aperçu sur le résultat d'une classification SVM basé sur l'information spectrale seule. Celles représentées par (c) dans ces figures sont obtenues en utilisant le deuxième fichier généré et donnent un aperçu sur la classification SVM basée sur l'utilisation conjointe de l'information spectrale et spatiale.



(a)

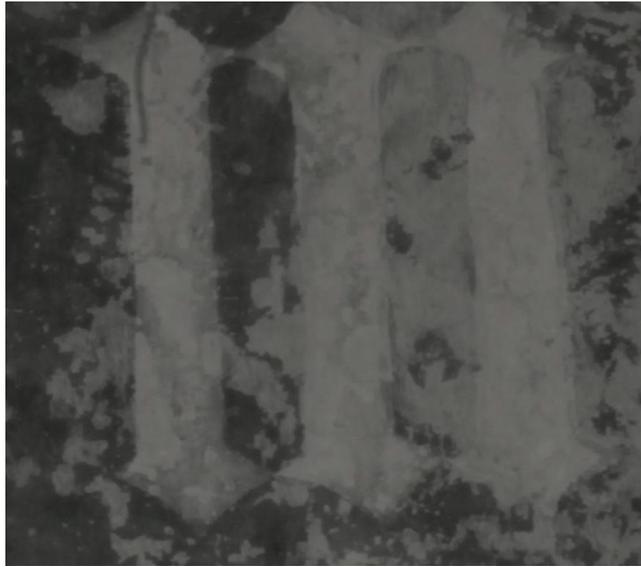


(b)



(c)

Figure 52 Cartes de classification avec le classifieur SVM sur la modalité Visible du premier ensemble des images (a), utilisant l'information spectrale seulement (b) et la combinaison de SURF dense avec l'information spectrale (c)



(a)

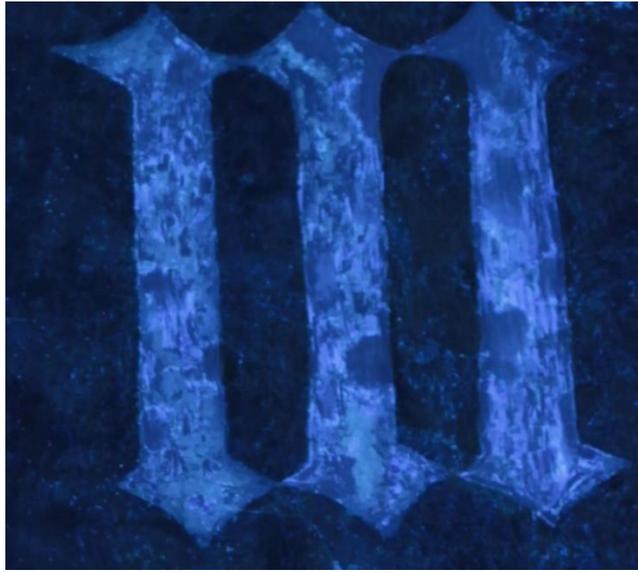


(b)



(c)

Figure 53 Cartes de classification avec le classifieur SVM sur la modalité IRR du premier ensemble des images (a), utilisant l'information spectrale seulement (b) et la combinaison de SURF dense avec l'information spectrale (c)



(a)

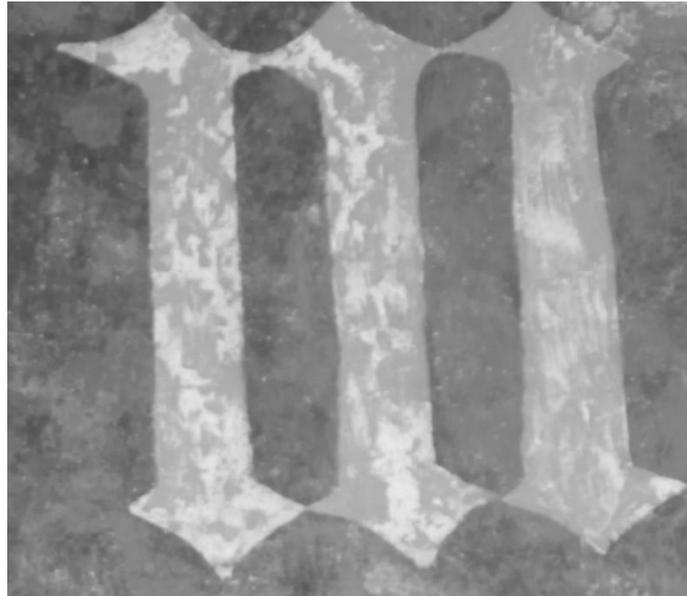


(b)



(c)

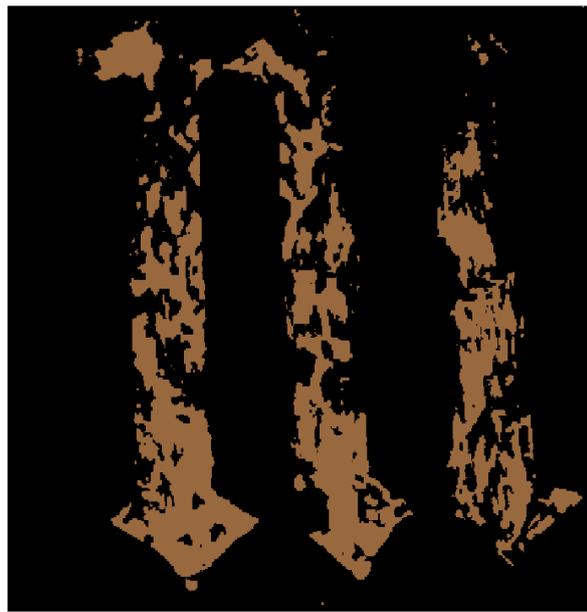
Figure 54 Cartes de la classification avec classifieur SVM sur la modalité UVF du premier ensemble des images (a), utilisant l'information spectrale seulement (b) et la combinaison de SURF dense avec l'information spectrale (c)



(a)



(b)



(c)

Figure 55 Cartes de classification avec le classifieur SVM sur la modalité UVR du premier ensemble des images (a), utilisant l'information spectrale seulement (b) et la combinaison de SURF dense avec l'information spectrale (c)

2. Le deuxième ensemble utilisé présente une autre scène sur le mur nord du dressing du duc de Bourgogne (Figure 56).

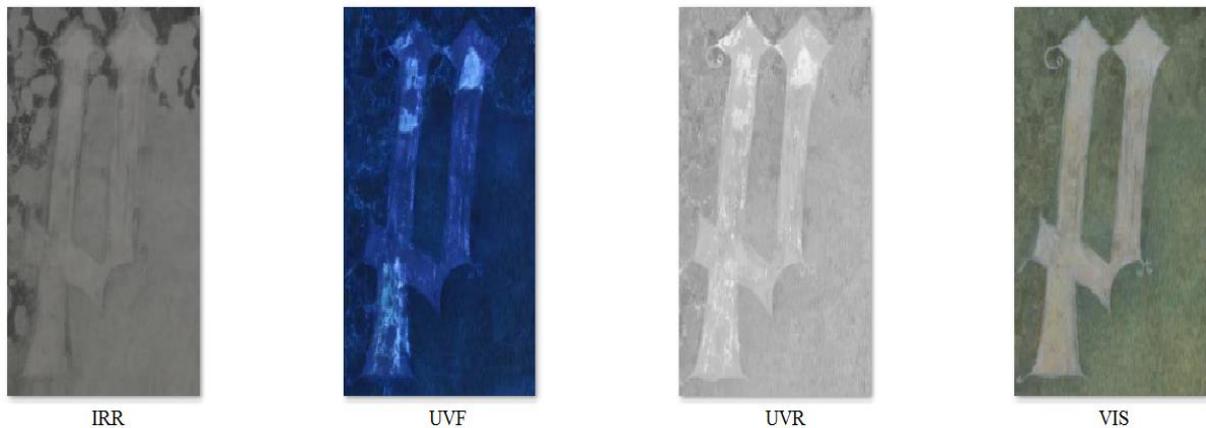


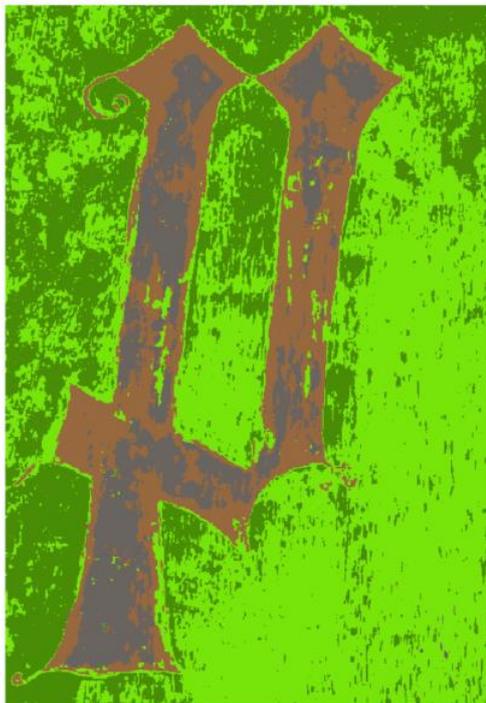
Figure 56 les quatre modalités pour la scène P sur le mur nord du dressing de duc de Bourgogne

Tout comme dans la scène précédente, nous utilisons deux fichiers, le premier contient uniquement les informations spectrales et le deuxième contient la concaténation entre les informations spectrales et les informations spatiales obtenues en utilisant Dense SURF. Nous appliquons ensuite à partir de ces deux fichiers de descripteurs la classification SVM basée sur la distance euclidienne minimale.

Les cartes de classification représentées par (b) dans les Figures 57, 58, 59, 60 sont obtenues en utilisant le premier fichier des informations spectrales uniquement, elles donnent un aperçu sur le résultat d'une classification SVM basé sur l'information spectrale seule. Celles représentées par (c) sont obtenues en utilisant le deuxième fichier généré, elles donnent un aperçu sur la classification SVM basée sur l'utilisation conjointe de l'information spectrale et spatiale.



(a)

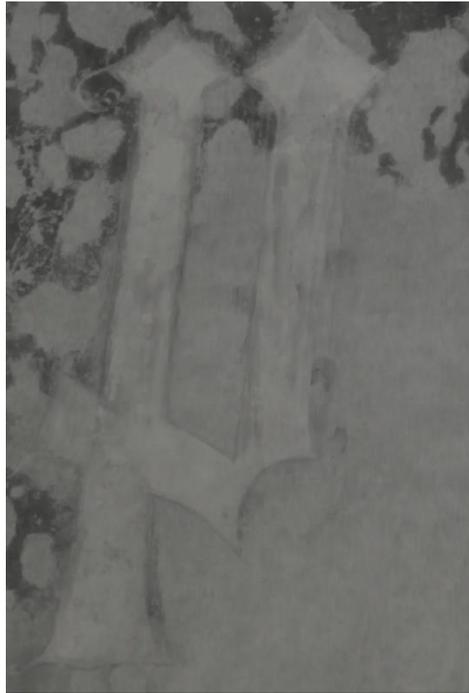


(b)



(c)

Figure 57 Cartes de la classification avec classifieur SVM sur la modalité Visible du deuxième ensemble des images (a), utilisant Dense SURF (b), Information spectrale (c) et la combinaison de SURF dense avec l'information spectrale (d)



(a)

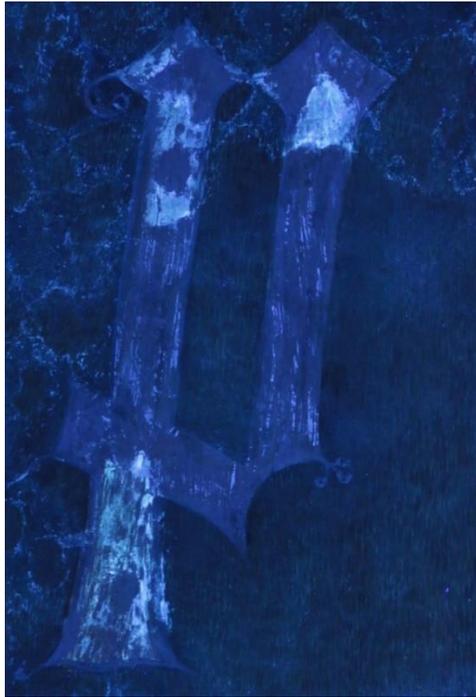


(b)



(c)

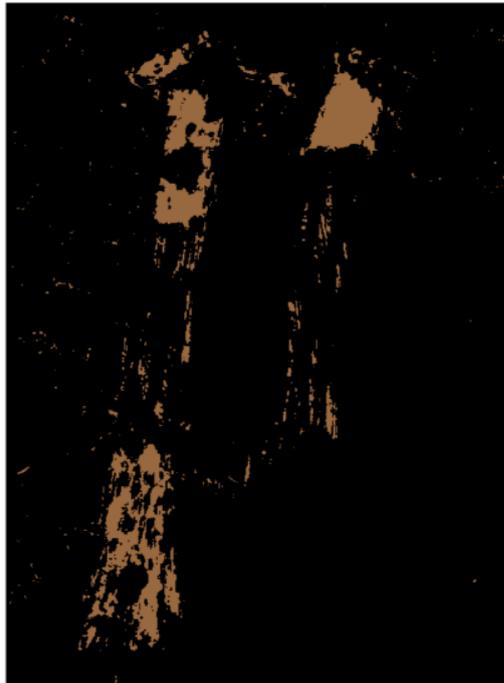
Figure 58 Cartes de la classification avec classifieur SVM sur la modalité IRR du deuxième ensemble des images (a), utilisant Dense SURF (b), Information spectrale (c) et la combinaison de SURF dense avec l'information spectrale (d)



(a)



(b)



(c)

Figure 59 Cartes de la classification avec classifieur SVM sur la modalité UVF du deuxième ensemble des images (a), Information spectrale (b) et la combinaison de SURF dense avec l'information spectrale (c)



(a)



(b)



(c)

Figure 60 Cartes de la classification avec classifieur SVM sur la modalité UVR du deuxième ensemble des images (a), Information spectrale (b) et la combinaison de SURF dense avec l'information spectrale (c)

On observe que Les cartes de classification illustrées sur les Figures 52b, 53b, 54b, 55b, 57b, 58b, 59b et 60b sont obtenues en utilisant uniquement des informations spectrales, ce qui donne également une classification erronée comme étant bruitées. Les cartes de classification illustrées sur les Figures 52c, 53c, 54c, 55c, 57c, 58c, 59c et 60c résultent de la fusion des caractéristiques spectrale et dense SURF. On observe que ces cartes de classification sont moins bruitées et que les performances de classification progressent globalement. L'analyse visuelle des cartes montre une bonne classification avec l'utilisation combinée des caractéristiques de dense SURF et des informations spectrales en termes de précision et homogénéité. Cette approche semble donc prometteuse pour la classification des images multimodales.

L'analyse comparative des cartes de classification montre que les cartes basées sur l'utilisation de l'information spectrale uniquement fournissent parfois une classification erronée. Les cartes de la classification issues de l'utilisation conjointe de l'information spectrale avec l'information spatiale (Dense SURF) sont moins bruitées et les performances de classification progressent globalement en termes de précision et homogénéité et elle semble très prometteuse pour la classification d'images multimodales.

Afin de valider ces résultats, la précision globale est la métrique d'évaluation qui a été utilisée (équation 5.5). Cet indicateur exprime le pourcentage des pixels correctement classés de la classe en surbrillance présentés dans chaque modalité :

$$\text{Précision globale} = \frac{\text{pixels correctement classés}}{\text{nombre total des pixels de test}} \quad 5.5$$

Le tableau 6 décrit la moyenne de la précision globale obtenue sur la totalité des tests, en utilisant uniquement des informations spectrales, puis en combinant ces informations avec combinaison avec des caractéristiques SURF.

Il apparaît que la méthode proposée montre un meilleur résultat en termes de précision globale par rapport à la précision obtenue en utilisant uniquement des caractéristiques spectrales, dont les résultats de classification sont par ailleurs dégradés.

Les méthodes	Visible	UV fluorescence	UVR	IRR
L'information Spectrale	80.489 %	88,77 %	85,31%	82,43 %
Dense SURF + l'information Spectrale	85,506%	91,27 %	91,99%	85,78 %

Tableau 6 Précision globale de la classification Images multimodales basées uniquement sur les informations spectrales et la combinaison des informations spectrales et de la densité SURF.

5.6. Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode de classification des images multimodales à haute résolution. Cette méthode est basée sur la représentation de toutes les modalités utilisées dans un espace commun en extrayant des caractéristiques SURF pour chaque pixel, appelée SURF Dense. La méthodologie consiste à combiner les descripteurs SURF avec des caractéristiques spectrales de l'espace CIE Lab. Les caractéristiques fusionnées sont ensuite utilisées comme données d'entrée du classifieur SVM pour l'étape de classification.

La méthode proposée exploitant les modalités séparément donne des résultats satisfaisants en termes de précision ce qui est facile à observer par l'analyse visuelle des cartes de classification obtenues. Néanmoins, nous pensons que la combinaison des différentes modalités pourrait améliorer davantage la classification. Pour atteindre cet objectif, les cartes de pré-classification obtenues par la méthode proposée sur chaque modalité seront fusionnées à l'aide des méthodes de l'incertain. Ceci est l'objet du chapitre suivant.

Chapitre 6

Classification/fusion des images multimodales basée sur

DSmT-Saillance

6.1. Introduction

L'imagerie multimodale permet d'accéder à différentes typologies d'informations sur les surfaces et/ou les objets ainsi numérisés. Comme indiqué précédemment, par leur complémentarité à décrire les surfaces, ces informations multiples sont souvent très riches et utiles dans de nombreux cas d'application, bien qu'il puisse aussi y avoir des redondances. Ce type de données peut aussi permettre d'améliorer la robustesse certaines opérations fréquentes en traitement des images et du signal, comme la classification, à condition de modéliser et/ou de combiner ces informations multiples d'une façon adaptée. C'est ce que nous avons commencé à effectuer dans la méthode présentée préalablement (Chapitre 5).

Cependant, bien que cette approche améliore la performance et la robustesse du processus de classification, on constate encore dans certains cas d'application des défaillances de la méthodologie. Plus particulièrement, on constate que la qualité de la classification est souvent évaluée au regard de la perception visuelle humaine, c'est-à-dire que de façon consciente ou non celle-ci sert souvent de référence. Celle-ci fait appel à un processus mêlant des mécanismes complexes et multi échelle, qui opère (via le cerveau humain) en un temps très court. Nous proposons donc dans ce chapitre une méthode permettant d'intégrer dans notre approche de classification d'images multimodales des aspects liés à la perception, permettant de se 'rapprocher' du système de visuel humain. Nous décrivons ensuite (deuxième section) le système global du processus de fusion des images proposé, de la phase de prétraitement jusqu'à la prise de décision. Les résultats (de la fusion) sont ensuite présentés et discutés. Enfin une conclusion est donnée dans la quatrième section.

6.2. La fusion des images multimodales basée sur la DSMT en intégrant le modèle perceptuel

L'approche proposée se base sur l'utilisation d'un processus de fusion fondé sur le modèle hybride de DSMT suivi de la règle de décision PCR6, afin de permettre la prise en compte du conflit élevé entre les différentes sources/modalités de notre cas d'application. Afin d'intégrer des aspects perceptuels à ce processus de fusion, et dans l'objectif d'améliorer la performance et la robustesse de la classification, nous avons choisi d'intégrer l'information de saillance visuelle, qui est essentielle dans le processus humain de perception et de classification visuelle. L'approche proposée est représentée à travers le flow-chart de la Figure 61 ci-dessous.

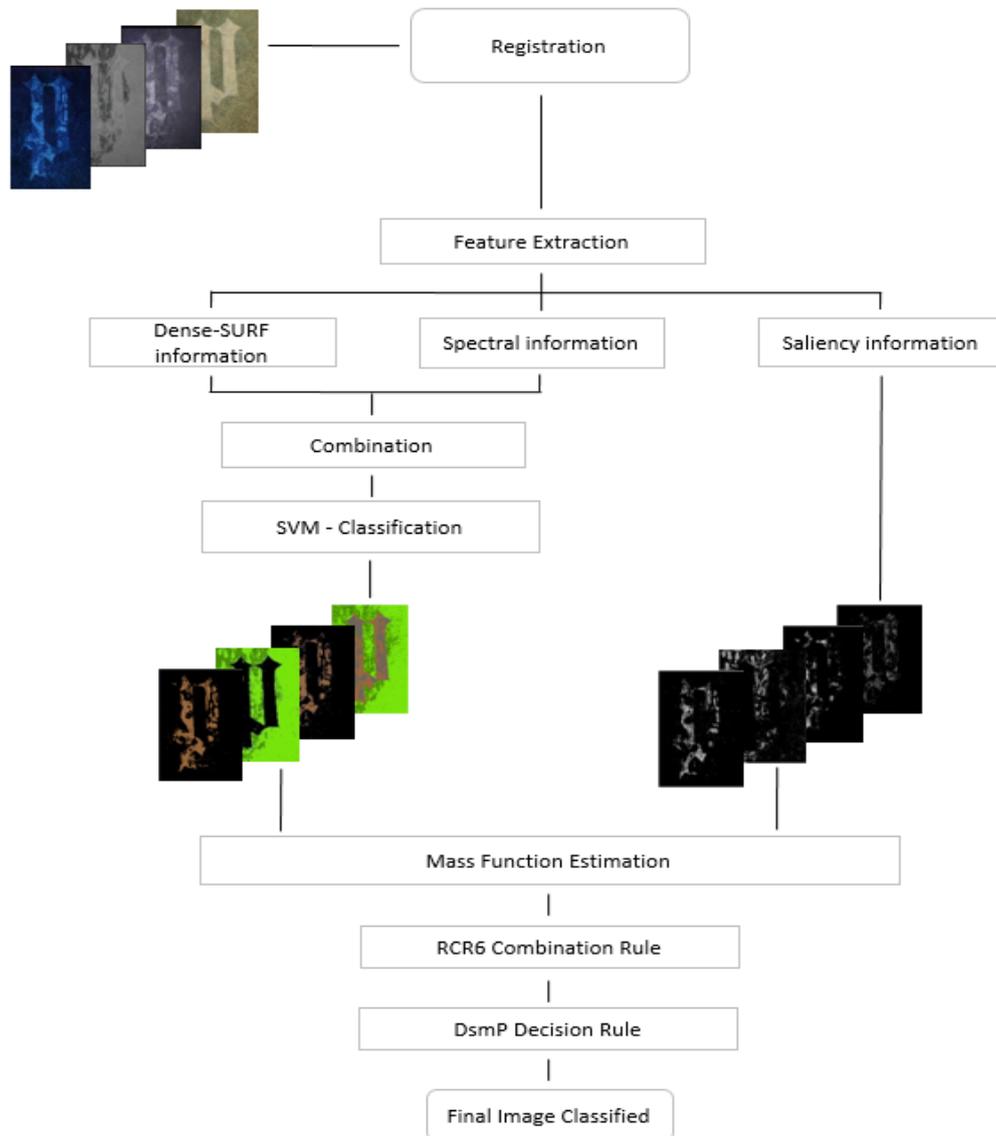


Figure 61 le system global de la fusion des images multimodales basé sur la DSMT intégrant la saillance

La méthodologie adoptée (pour la classification basée sur la fusion des images) peut ainsi être divisée en plusieurs étapes :

- L'étape de prétraitement qui consiste à aligner toutes les images de telle sorte qu'un pixel soit associé à la même position physique sur toutes les modalités.
- L'étape d'extraction des caractéristiques qui concerne l'extraction des primitives spatiales, spectrales et perceptuelles.
- L'étape de la classification qui vise à fournir les matrices de probabilités de chaque modalité en se basant sur l'utilisation conjointe des caractéristiques spatiales et spectrales.
- L'étape de la fusion qui consiste à combiner toutes les matrices de probabilités issues de la pré-classification avec les cartes de la saillance issues de la phase de l'extraction des caractéristiques en utilisant un formalisme adéquat.

Ces étapes sont présentées dans les sections suivantes.

6.2.1. Le prétraitement : le recalage

Généralement, le prétraitement qui précède la classification vise à éliminer les imperfections qui altèrent les informations par un ensemble d'actions telles que le filtrage, les opérations sur les gradients, la segmentation, etc. Cependant, dans la classification basée sur les théories de l'incertain, ces imperfections sont protégées, modélisées et combinées pour aider à prendre la décision.

Le recalage reste cependant un prétraitement généralement nécessaire, afin d'établir les correspondances locales entre les données d'une scène obtenue à partir d'un ou plusieurs capteurs, potentiellement à différentes positions et échelles spatiales, en utilisant des transformations spatiales et radiométriques optimales entre les images. Dans le cas des images multimodales, le recalage présente une problématique à cause de la différence significative parfois importante en terme d'intensités entre les images [53] [54]. Pour effectuer cette étape, nous avons appliqué la deuxième contribution détaillée dans le Chapitre 4, qui se base sur l'exploitation des descripteurs SURF pour l'identification et la description des points d'intérêt associé à un filtrage basé sur la distance entre les points de correspondances identifiés et leurs orientations par rapport aux points de références (voir section 3 Chapitre 4).

6.2.2. L'extraction des caractéristiques

L'extraction de caractéristiques est une étape cruciale du processus de fusion et de classification des images. Il convient de noter qu'un choix pertinent des primitives à extraire peut améliorer significativement les performances de la classification.

Dans ce travail, les caractéristiques extraites sont liées à l'information spectrale, l'information spatiale ainsi que des caractéristiques liées à la saillance visuelle locale :

- L'information spectrale/intensité

L'information spectrale est largement utilisée dans plusieurs applications de classification. Dans ce travail, nous avons extrait les valeurs spectrales de chaque pixel en tant que vecteur d'attributs, puis nous les avons converties dans l'espace de représentation CIELab pour une meilleure corrélation des couleurs à la vision humaine.

➤ L'information spatiale

Pour extraire les informations spatiales, nous avons utilisé la méthode Speeded up robust feature SURF proposé par *Herbert Bay* [175], et qui consiste initialement en deux phases : détection et description des points d'intérêts. Nous avons appliqué la contribution détaillée au chapitre 5 qui consiste à effectuer la description SURF sur chaque pixel de l'image. Ceci est fait, en commençant par attribuer à chaque pixel l'orientation dominante calculée en combinant les résultats d'ondelettes de Haar dans un voisinage circulaire autour de chaque pixel, puis en créant 4×4 sous-régions autour du pixel. Dans chaque sous-région, des réponses en ondelettes de Haar exprimées en pixels sont calculées, qui sont à leur tour additionnées pour former un descripteur à 64 éléments.

➤ La saillance

○ Expérimentations

Dans notre cas d'application, nous avons mis en œuvre quatre modalités d'imagerie (UV, IR, UVF et visible). Les zones ayant fait l'objet de ces acquisitions contiennent des traces de peinture d'origine et de région repeinte, ce qui constitue dans ce cas les deux classes que nous souhaitons différencier. L'idée sous-jacente est que sur ces deux classes, la saillance est potentiellement significativement différente, tout au moins pour certaines des modalités mises en œuvre.

Dans ce chapitre, nous présentons de façon comparative les résultats d'application des différentes méthodes d'estimation de la saillance visuelle sur les images (présentées dans le Chapitre 2), afin de retenir l'approche qui apparaît la plus appropriée dans ce cas d'application. La performance de chaque méthode est évaluée en utilisant la métrique de similarité NSS (*Normalized Scanpath Saliency*) [176]. Cette mesure quantifie les valeurs de la carte de la saillance en les comparant aux cartes de fixation de l'œil. Ces cartes de fixation sont générées en se basant sur la saillance visuelle basée sur le graphe [107], en faisant référence au travail de *Yu* [177] dans lequel l'auteur a choisi GBVS pour obtenir la carte de fixation de leur jeu de donnée. Ce modèle est considéré comme étant plus fiable que des algorithmes plus classiques qui sont devenus des références. La Figure 62 montre une comparaison visuelle des différentes méthodes d'estimation de la saillance visuelle, appliquée sur la scène « P » (Figure 5 – Château de Germolles).

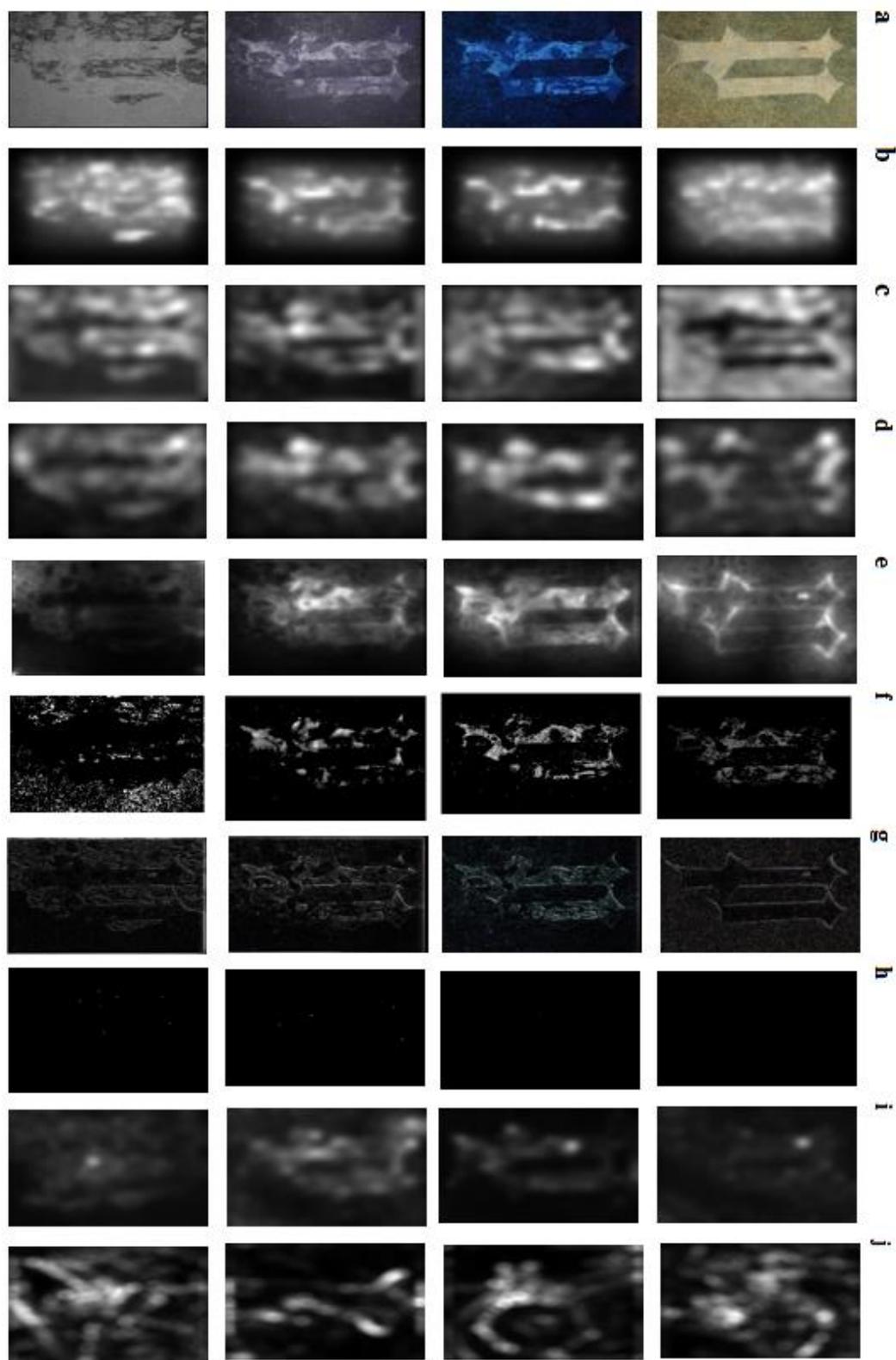


Figure 62 Comparaisons de cartes de saillance dans la base de données d'images multimodales:
 (a) l'image originale, (b) Carte de fixation, (c) [101],(d)[111], (e) [112], (f) [113], (g) [80], (h) [110], (i) [100],
 (j) [105].

Les résultats (Tableau 7) montrent que le modèle *Itti* [74] est performant pour les modalités UVF et IRR. La méthode proposée par *Goferman* dans [112] obtient un meilleur score en modalité UV, tandis que celle proposée par *Rahtu* et al. dans [113] obtient le meilleur score pour les modalités VIS, UVF et UVR alors que la détection de saillance spectrale de [80] est plus performante pour la modalité visible.

Les méthodes	Koch [101]	Bharath [110]	Hou [90]	Zhang [100]	Goferman [112]	Rahtu [113]	Le Moan [80]	Zhan [105]
VIS	0.7200	0.0523	0.6000	-0.0573	0.4200	1.0100	5.1600	0.2800
UVF	1.2000	0.0945	0.9600	0.6900	1.0000	1.1720	17.960	0.6500
UVR	0.5100	0.0617	0.5700	0.0700	0.4700	1.0100	37.923	0.3300
IRR	1.0600	0.0217	0.2400	0.1900	0.4200	0.4400	0.2800	0.0572

Tableau 7 Métrique NSS des méthodes comparées

Chaque modalité contient donc des informations sur les régions d'origine et les régions repeintes de la même zone, et chacune met l'accent sur une région spécifique en termes de couleur, de contraste, d'intensité, de réflectance, de texture ou de luminance, ce qui peut attirer l'attention de l'homme, c.à.d. générer une saillance importante. La métrique NSS est largement utilisée pour la comparaison des cartes de la saillance. Les résultats montrent que certaines méthodes d'estimation de saillance ne sont pas appropriées à notre cas d'application.

Dans le Tableau 7, on peut remarquer que la segmentation des régions saillantes à partir d'images et de vidéos (*Rahtu* et al.) [113] permet de générer une carte de saillance plus fine sur les trois modalités UVR, UVF et VIS, par rapport aux autres méthodes. A notre sens, cette méthode a une meilleure performance du fait de la variété des fonctionnalités utilisées, du lissage des paramètres et des aspects statistiques mis en œuvre. La détection spectrale de saillance proposée dans [80] est plus efficace pour distinguer les régions saillantes de la modalité visible en raison de l'introduction de l'attribut de réflectance qui caractérise cette modalité. Le modèle *Itti* [101] donne un bon score dans les modalités IRR et UVF. La méthode basée sur le contexte [112] fournit un bon score dans la modalité UVF, mais un score faible dans d'autres modalités, car elle a tendance à favoriser les limites plutôt que la région intérieure. Le ROI [110] et X. a *Hou* [178] donnent un score faible pour le problème traditionnel de la séparation de l'arrière-plan. Alors que BMS [100] donne un score faible, cette méthode visant plutôt à extraire l'objet saillant plutôt que le pixel saillant. La méthode basée sur le résidu spectral a également un score faible, car elle n'interprète pas la saillance locale. En se basant sur cette étude comparative, nous avons choisi d'extraire ces caractéristiques en utilisant la méthode proposée par *Rahtu* et al. [113], détaillée ci-après.

- la méthode de *Rahtu* et al. [113]

Cette méthode utilise les contrastes de luminances locales et la notion de la couleur mappée à un espace des caractéristiques $F(x)$ divisé en bacs disjoints $Q_{F(x)}$.

Le calcul de saillance commence par l'application d'une fenêtre coulissante w divisée en fenêtres intérieures noté K , considéré comme étant le noyau, et en frontière nommé B , dans laquelle les points dans K sont considérés saillants et les points de B font partie de l'arrière-plan. Une illustration est présentée dans la Figure 63. La mesure de saillance d'un point x est présentée sous forme d'une probabilité conditionnelle (nombre entre 0 et 1).

Considérant Z une variable aléatoire dans la fenêtre W qui décrit la distribution des pixels dans W . La mesure de la saillance peut être définie comme suit :

$$S_0(x) = P(Z \in K | F(Z) \in Q_{F(x)}) \quad 6.1$$

Avec $Q_{F(x)}$ présente le bac qui contient $F(Z)$.

Un point x est qualifié saillant :

- Si la mesure calculée est proche de 1.
- Si les caractéristiques de x sont similaires aux caractéristiques des points de la zone B (fenêtre intérieure) et différentes de celles de la zone K .

La mesure peut être définie comme une probabilité conditionnelle et calculée à l'aide de la formule de Bayes suivante (équation 1.7) :

$$S_0(x) = \frac{h_K(x)p_0}{h_K(x)p_0 + h_B(x)(1-p_0)} \quad 6.2$$

Avec

- $0 < p_0 < 1$
- $p(x) = p_0(x)/|K|$ pour $x \in K$.
- $p(x) = (1 - p_0)/|B|$ pour $x \in B$.

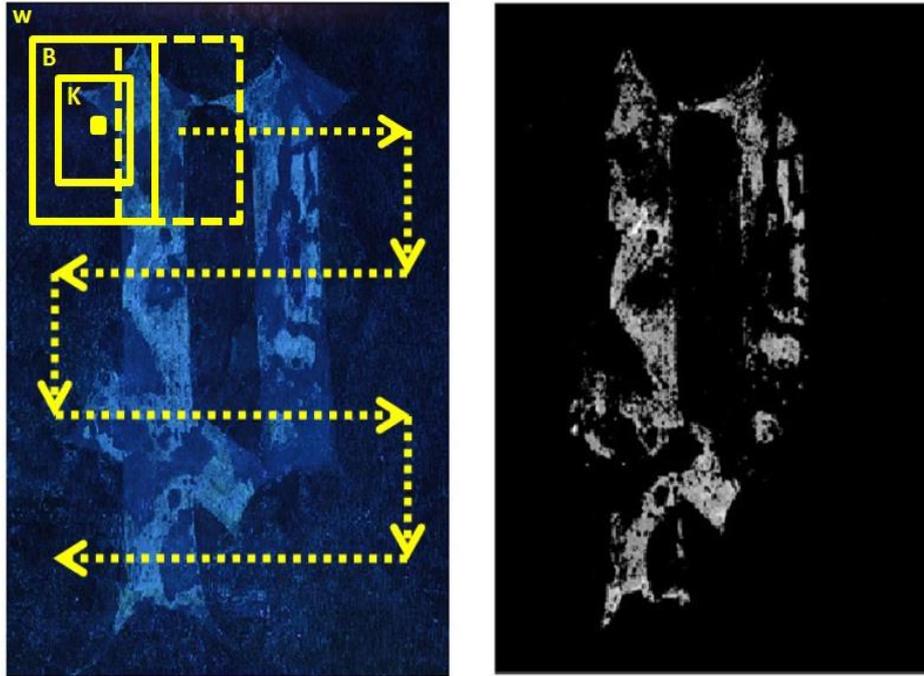


Figure 63 le calcul de la carte de la saillance selon la méthode de Rahtu *et al.*

Une mesure de saillance régularisée est ensuite introduite pour la rendre plus robuste au bruit.

6.2.3. La pré-classification SVM

Après l'étape de prétraitement (recalage), nous établissons des échantillons d'apprentissage, puis nous construisons deux fichiers de descripteurs en extrayant des caractéristiques spatiales Dense SURF et les primitives spectrales. Une combinaison entre ces deux informations est ensuite effectuée, le premier fichier est pour les échantillons d'apprentissage et le deuxième pour l'image de test. Un modèle SVM est créé en se basant sur le premier fichier.

Afin de générer les matrices de probabilités qui vont être utilisées par la théorie de Dezert-Smarandache, nous avons effectué une pré-classification en utilisant le deuxième fichier construit. Les données multimodales traitées sont non linéaires et de très grande dimension. Pour cette raison, nous avons choisi le classifieur SVM avec le noyau RBF qui a montré son efficacité en traitement de données de telle dimension.

Ce problème est considéré comme étant multi classe puisque nous disposons de quatre classes. Nous avons adopté l'approche d'un contre le reste (One-Against-All OAA) pour traiter les informations incomplètes fournies par les différentes modalités. Les quatre classes sont alors traitées avec quatre classifieurs SVM binaires : chaque classe est présentée à 4 classifieurs, et chaque classifieur renvoie sa décision.

Le label ou la décision retenue est celle associée au classifieur qui répond le mieux.

Ensuite nous récupérons les probabilités d'appartenance des pixels aux différentes classes, qui seront conservées dans des matrices comme suit :

$$\text{VIS} = \begin{pmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1m} \\ \cdot & \cdot & & & \cdot \\ \cdot & & x_{ij} & & \cdot \\ \cdot & & & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{nm} \end{pmatrix} \quad \text{IRR} = \begin{pmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1m} \\ \cdot & \cdot & & & \cdot \\ \cdot & & x_{ij} & & \cdot \\ \cdot & & & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{nm} \end{pmatrix}$$

$$\text{UVR} = \begin{pmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1m} \\ \cdot & \cdot & & & \cdot \\ \cdot & & x_{ij} & & \cdot \\ \cdot & & & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{nm} \end{pmatrix} \quad \text{UVF} = \begin{pmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1m} \\ \cdot & \cdot & & & \cdot \\ \cdot & & x_{ij} & & \cdot \\ \cdot & & & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{nm} \end{pmatrix}$$

Pour un ensemble de classe $C = \{\theta_1 \dots \theta_n\}$, chaque pixel $I(i, j)$ de l'image est associé à un vecteur de probabilité $x_{ij} = [P(I(i, j)/\theta_1), \dots, P(I(i, j)/\theta_n)]$.

6.2.4. La classification/Fusion basé sur DSMT

➤ Le cadre de discernement et l'estimation des fonctions de masse

L'étape d'estimation des fonctions de masse est très importante dans le processus de fusion, car les imperfections telles que l'incertitude, l'imprécision, le paradoxe seront introduites à cette phase. La génération la plus utilisée pour ces masses est constituée des probabilités issues de la pré-classification SVM.

Chaque source (modalité / carte de saillance) notée S_i^b ($i = 1, \dots, K$) donne la probabilité d'appartenance à une ou deux classes, ainsi que leurs classes complémentaires qui présentent la masse de l'ignorance partielle, sauf la modalité VIS qui fournit les probabilités d'appartenance à chacune des classes. En se basant sur [19], nous notons $\Theta = \{\theta_1 \dots \theta_n\}$ et la masse GBBA de chaque source est donnée par :

$$\begin{cases} m_S(\theta_i) = \frac{P(x|\theta_i)}{z} & \forall \theta_i \in \Theta, \\ m_S(\bar{\theta}_i) = \frac{P(x | \bigcup_{0 < j < n, j \neq i} \theta_j)}{z} & \forall \theta_j \in \Theta, \\ m_S(\emptyset) = 0 \end{cases} \quad 6.3$$

Où $z = \sum_{j=0}^n p(x|\theta_j)$ est un terme de normalisation que nous avons utilisé afin de nous assurer que $\sum m = 1$.

➤ Règle de combinaison

Les masses estimées doivent être combinées avec la règle appropriée qui traite le conflit généré par différentes sources S_i^b . Dans ce travail, nous avons utilisé la règle PCR6 [37] dans la phase de la combinaison, car elle donne la meilleure performance par rapport à toutes les règles de combinaison citées dans la section précédente et testées avec nos jeux de données.

Le PCR6 est calculé comme suit :

En considérant N sources indépendantes, les masses $m_{PCR_6}(\cdot)$ Combinées acquises à partir de $N > 2$ sources sont calculées comme suit :

$$m_{PCR_6}(X) = m_{12\dots S}(X) + \sum_{\substack{X_1, X_2, \dots, X_S \in D^\Theta \setminus \{\emptyset\} \\ X_1 \cap X_2 \dots X_S = \emptyset}} [\sum_{r=1}^S \delta_{X_r}^X \cdot m_r(X_r)] \cdot \frac{m_1(X_1)m_2(X_2)\dots m_S(X_S)}{m_1(X_1)+m_2(X_2)+\dots+m_S(X_S)} \quad 6.4$$

$$m_{PCR_6}(\emptyset) = 0, \forall X \in D^\Theta \setminus \{\emptyset\}, \quad 6.5$$

Où

$$\delta_{X_r}^X \triangleq \begin{cases} 1, & \text{if } X = X_r \\ 0, & \text{if } X \neq X_r \end{cases} \quad 6.6$$

Avec les masses $m_{12\dots S}(X) \equiv \bigcap m(X)$ correspondant au consensus conjonctif sur X entre $N > 2$ sources.

➤ Règle de décision

Une fois l'étape de combinaison réalisée, nous calculons les fonctions de croyance généralisée, telles que la fonction de crédibilité, la fonction de plausibilité et celle de la transformation pignistique qui sont estimées à partir de la fonction de masse généralisée m

Dans ce travail, nous avons utilisé une transformation probabiliste DSmp. Cette fonction consiste à convertir la mesure de masse combinée en une mesure de probabilité en utilisant Eq (6.8) pour prendre la décision finale.

$$\left\{ \begin{array}{l} DSmP_{\varepsilon}(\emptyset) = 0 \\ DSmP_{\varepsilon}(X) = \sum_{Y \in G^{\theta}} \frac{\sum_{\substack{Z \subseteq X \cap Y \\ C(Z)=1}} m(Z) + \varepsilon \cdot C(X \cap Y)}{\sum_{\substack{Z \subseteq Y \\ C(Z)=1}} m(Z) + \varepsilon \cdot C(Y)} \times m(Y) \quad \forall X \in G^{\theta} \setminus \{\emptyset\} \end{array} \right.$$

G^{θ} peut représenter l'ensemble D^{θ} au complet ou un ensemble réduit avec contrainte, selon le modèle utilisé (Shafer, libre ou Hybride). ε étant un paramètre de réglage et $C(X \cap Y)$ et $C(Y)$ sont respectivement les cardinalités de $X \cap Y$ et Y .

6.3. Résultat et discussion

La méthodologie proposée est appliquée sur la base de données décrite dans le premier chapitre (château de Germolles). Dans cette section, nous comparons les performances de différentes méthodes de classification.

Pour l'illustration, nous avons choisi deux ensembles d'images multimodales de différentes chambres du château de Germolles à savoir.

3. Le premier ensemble utilisé présente une scène sur le mur sud du dressing de la comtesse.

La zone illustrée sur la Figure 64 est acquise selon quatre modalités. VIS, UVF, UVR et IRR. Chaque modalité a une résolution de 3744×5616 pixels.

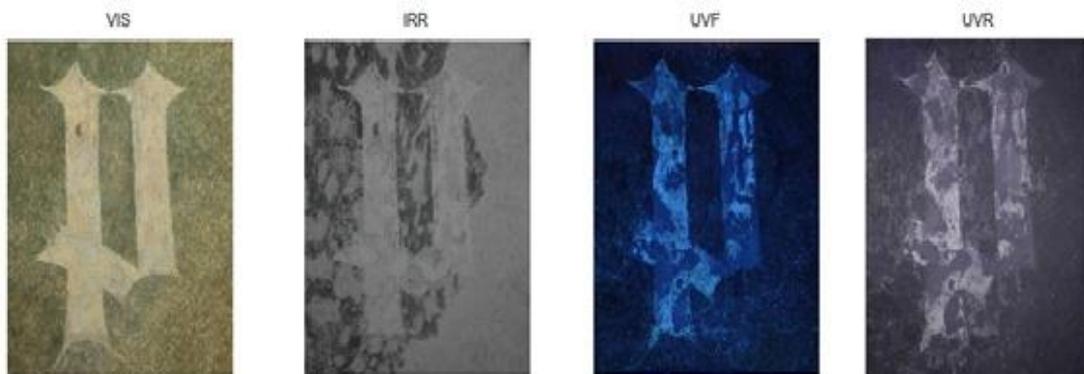


Figure 64 Cas d'application 1 - scène sur le mur sud du dressing de la comtesse du Château de Germolles

4. Le deuxième ensemble illustré sur la Figure 65 est associé à quatre modalités d'acquisition : VIS, UVF, UVR et IRR, avec une résolution de 4744×3616 pixels.

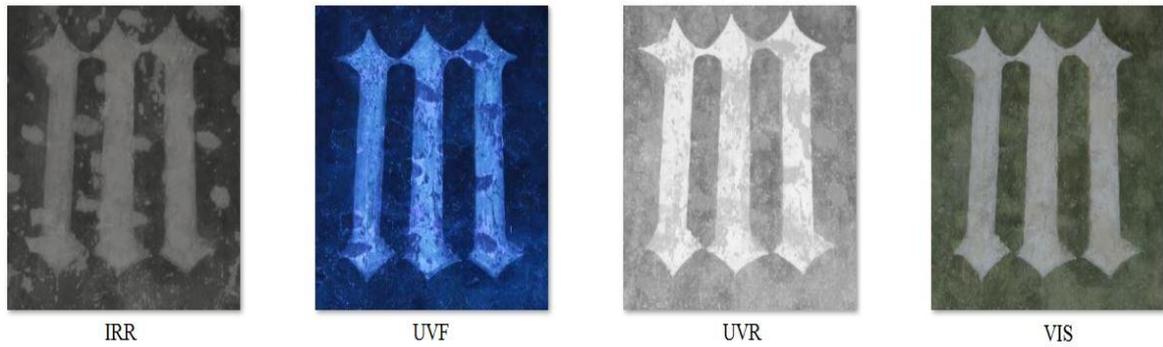


Figure 65 Cas d'application 2 – scène sur le mur sud du dressing de la comtesse du Château de Germolles

Nous définissons tout d'abord le cadre de décernement qui prend en compte la connaissance préalable de la scène étudiée. Nous avons identifié 4 classes, qui constituent le cadre de décernement Θ . Celles-ci sont le blanc original (*Original White* -WO), le vert original (*Original green* -GO), le blanc repeint (*Repainted White* - WR) et le vert repeint (*Repainted Green* -GR).

Θ est ainsi défini par :

$$\Theta = \{WO, WR, GO, GR\}$$

Les informations spectrale et Dense-SURF sont ensuite extraites et utilisées conjointement comme entrée du classifieur SVM à l'aide du noyau RBF. Les cartes de la classification des modalités IRR, UVR et UVF des deux scènes illustrées sont présentées dans les Figures 66 et 69. Pour la modalité VIS, les cartes de classification sont présentées dans les Figures 68a et 71a. En parallèle, les informations sur la saillance sont extraites à l'aide de la méthode proposée dans [179], les cartes obtenues sont présentées en Figures 67 et 70.

En exploitant les informations de la zone d'étude ainsi que celles obtenues par classification SVM, certains éléments de l'ensemble hyper-power-set D^Θ sont exclusifs. Certaines contraintes peuvent être prises en compte pour mieux s'adapter à la situation réelle et réduire le D^Θ , par exemple $WO \cap GR = \emptyset$ (modèle hybride de DSMT).

La fonction de masse associée à la classe soulignée et sa complémentarité dans chaque modalité est alors calculée (Equation 6.3), puis la règle PCR6 est utilisée pour combiner les masses calculées sur la base de l'équation (6.4). Enfin, la décision est prise en utilisant le critère DsmP maximum. Les cartes finales de classification pour les deux cas d'applications, fournies par DSMT (sans saillance), sont présentées dans la Figure 68b et 71b. Les cartes de classification obtenue à l'aide de DSMT-Saillance sont illustrées en Figure 68c et 71c.



(a)

(b)

Legend    
WO WR GO GR



(c)

Figure 66 : les cartes de classification SVM de la première scène, (a) Classification SVM de la modalité UVF, (b) Classification SVM de la modalité UVR, (c) Classification SVM de la modalité IRR



(a)



(b)



(c)



(d)

Figure 67 : Les cartes de la saillance correspondant à la première scène, (a) la carte de la saillance de la modalité VIS, (b) la carte de la saillance de la modalité UVF, (c) la carte de la saillance de la modalité UVR, (d) la carte de la saillance de la modalité IRR



Legend



(a)



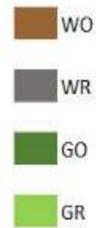
Legend



(b)

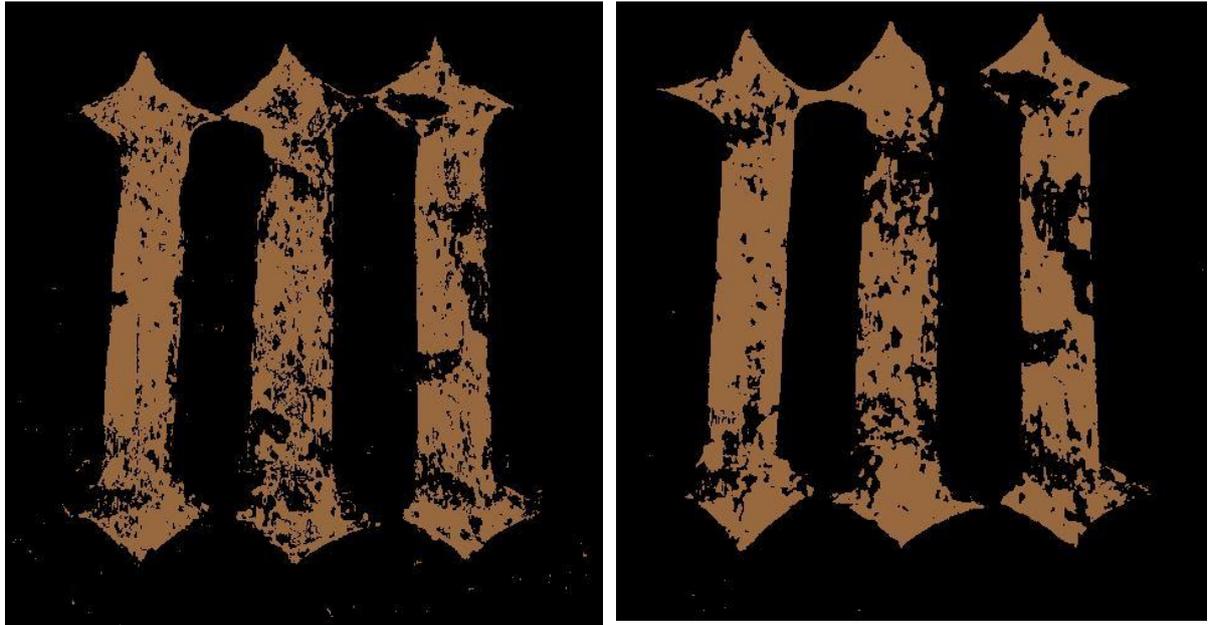


Legend



(c)

Figure 68 : la classification SVM de la modalité VIS (a), Classification DSMT des images multimodales (b), Classification en se basant sur DSMT et la saillance des images multimodales (c)



(a)

(b)

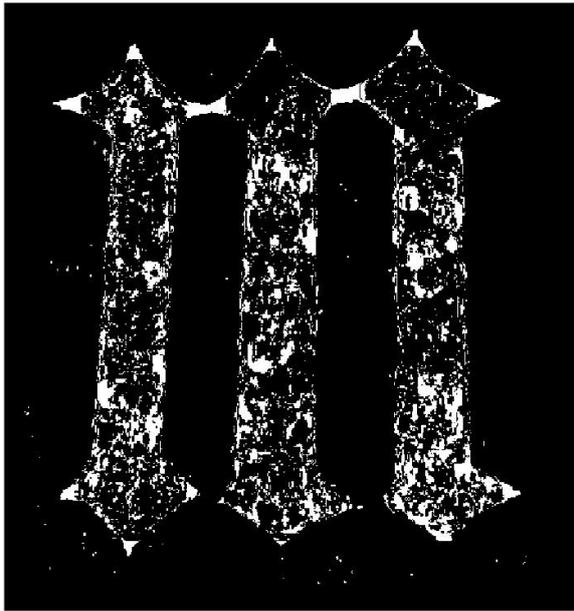
Legend

			
WO	WR	GO	GR



(c)

Figure 69 : les cartes de classification SVM de la deuxième scène, la carte de la classification SVM de la modalité UVF (a), la carte de la classification SVM de la modalité UVR (b), la carte de la classification SVM de la modalité IRR (c).



(a)



(b)

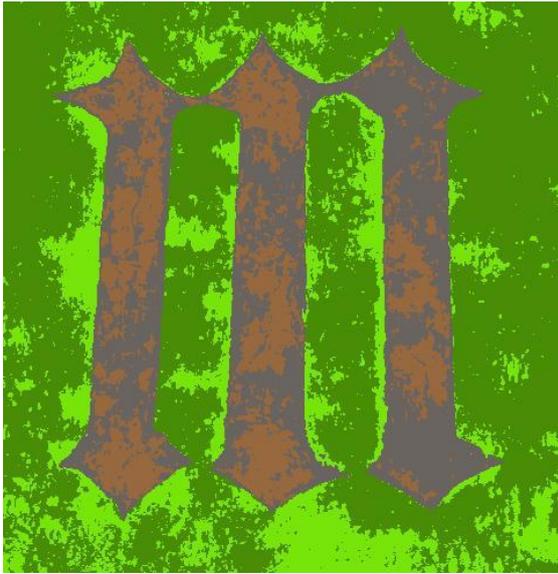


(c)



(d)

Figure 70 : Les cartes de la saillance correspondant à la première scène, (a) la carte de la saillance de la modalité VIS, (b) la carte de la saillance de la modalité UVEF, (c) la carte de la saillance de la modalité UVR, (d) la carte de la saillance de la modalité IRR



(a)



(b)

Legend

			
WO	WR	GO	GR



(c)

Figure 71 : la carte classification SVM de la modalité VIS (a), la carte de la classification DSmT des images multimodales (b), la carte de classification en se basant sur DSmT et la saillance des images multimodales (c).

Les résultats montrent une amélioration des performances avec l'intégration d'aspects perceptuels (saillance) dans le processus DSMT. L'analyse visuelle des cartes de classification montre que le résultat de la méthode proposée est meilleur (au regard de la vérité terrain dont nous disposons) sur les classes WR et WO que le résultat obtenu en utilisant DSMT seul (sans saillance) pour ces mêmes classes. Alors que la carte obtenue en utilisant une image unimodale présente un résultat dégradé en termes de lissage et de connectivité entre les classes.

Afin d'évaluer les performances des méthodes utilisées et de comparer les résultats, nous avons utilisé également la précision globale (OA) qui présente le pourcentage de pixels correctement classés et le taux d'erreur moyen (MER) qui présente le pourcentage d'erreurs pour les pixels mal classés. Le Tableau 8 présente la moyenne obtenue en utilisant les différentes méthodes sur la totalité des tests. La méthode proposée produit une meilleure précision globale de 95,39% par rapport à la classification DSMT qui fournit une précision globale de 91,46% et la SVM classification qui donne une précision globale de 86,43%. En termes de taux d'erreur, la méthode proposée donne un faible score MER de 4,225% par rapport à DSMT-Classification et SVM-Classification qui fournit un MER de 8,53% et 14,73% respectivement.

En conclusion, l'utilisation de la théorie DSMT avec la règle de combinaison PCR6 fournit des meilleures performances grâce à son efficacité à gérer correctement les informations de conflit fournies par les différentes sources et montre une amélioration significative de la classification par rapport à la classification unimodale SVM. Ainsi, l'intégration des informations de saillance dans le processus de fusion présente un avantage réel.

METHODS	OA	MER
SVM-Classification	85,506%	15,74%
DSMT- Classification	92.886%	8,53%
DSMT-Saliencie Classification	95,50%	4,225%

Tableau 8 Précision et erreurs de classification résultant de différentes méthodes

6.4. Conclusion

Dans ce chapitre, nous avons proposé une nouvelle méthode de classification des images multimodales. Dans un premier temps, nous avons extrait des informations spatiales (Dense-SURF), spectrales et saillantes. Les informations spatiales et spectrales extraites sont combinées et transmises au classifieur SVM pour l'étape de pré-classification.

Ensuite, les résultats de la classification SVM obtenus à partir de chaque modalité sont fusionnés en utilisant la théorie DSMT, l'utilisation conjointe de DSMT et de SVM offre de meilleures performances par rapport à la classification unimodale SVM.

Dans la deuxième étape, les informations de la saillance extraites sont modélisées et combinées avec les résultats issus de la classification SVM à l'aide du processus DSMT basé sur la règle de combinaison PCR6 et la règle de décision DsmP.

La méthode proposée permet d'obtenir les meilleures performances en termes de précision et de taux d'erreur par rapport à la classification basé sur DSMT-SVM et la classification SVM fondé sur une seule image unimodale.

7. Conclusion et perspectives

Dans cette thèse, nous voulions évaluer la contribution potentielle de la fusion des images multimodales dans le domaine du patrimoine. Partant du principe que la classification fondée sur une seule image/modalité ne donne qu'un aperçu partiel et/ou déformé de la réalité, l'utilisation de plusieurs sources peut permettre de fournir une classification plus fiable et de bonne qualité. L'objectif de ce travail était donc de modéliser les différentes imperfections qui entachent les informations issues de différentes sources/modalités de haute résolution spatiale, puis ensuite combiner ces informations en se basant sur les méthodes de l'incertain et plus précisément sur la théorie de Dezert-Smarandache afin d'obtenir une classification commune.

D'un point de vue méthodologique, nous avons étudié la possibilité de mettre en place de nouvelles techniques intervenant dans les différentes phases de la réalisation de ce processus de fusion, telles que la phase du prétraitement, la phase de la pré-classification et la phase de la fusion elle-même. Plusieurs questions ont été posées au début de ce manuscrit et ont constitué le fil conducteur de ce travail de recherche. Pour y répondre, nous avons réalisé une étude de la littérature, ce qui nous a amenés à proposer quatre contributions originales, qui seront résumées dans ce qui suit.

Les deux premières méthodes proposées, notamment au chapitre 5, sont consacrées à l'alignement/recalage des images que nous cherchons à fusionner en se basant sur la méthode géométrique SURF. Le recalage de ces différentes images soulève le problème de la présence des points aberrants, vu la nature et les conditions d'acquisition différentes entre les modalités, ce qui nous a poussés à réfléchir à une méthode permettant de filtrer les points d'appariement détectés.

La première contribution vise la suppression des points aberrants en se basant sur la distribution gaussienne des données. Cette méthode intègre deux filtres qui exploitent la distribution gaussienne des distances et des orientations entre les points putatifs sur les deux images. La méthode a montré sa performance face aux déformations liées à la translation et au changement d'échelle. Néanmoins, cette efficacité a des limites dans le cas de l'existence des déformations de rotation importantes.

Pour surmonter ces limitations, nous avons proposé une deuxième stratégie basée dans un premier lieu sur un filtrage automatique utilisant la distribution gaussienne des distances calculées entre les points putatifs détectés sur les deux images (modèle et requête). Ensuite, et en se basant sur le résultat de ce premier filtrage, nous avons fixé deux points comme référence sur chacune des images (modèle et requête) et puis nous avons calculé les distances entre ces deux points et le point en question ainsi que leurs orientations. Par la suite, un critère strict a été mis en place en utilisant les distances calculées et les orientations afin de supprimer tout point aberrant.

Après l'étape du recalage, nous avons effectué l'étape de pré-classification dans le processus de la fusion adopté. Dans cette étape, nous avons soulevé deux points importants sur

lesquels se base notre troisième contribution. Le premier point est le choix des différentes caractéristiques qui vont être utilisées comme descripteurs spatiaux ainsi que le processus d'extraction d'information spatiale. Le deuxième point, est sur la façon selon laquelle nous allons combiner l'information spectrale et l'information spatiale dans la classification.

Pour ce faire, nous nous sommes intéressés en premier lieu à représenter les différentes modalités dans un espace de représentation commun en utilisant le descripteur spatial SURF, en raison de son efficacité et sa capacité en termes d'invariance au changement d'échelle et à la rotation et aussi pour sa rapidité de calcul.

La méthode proposée extrait les caractéristiques SURF pour chaque pixel dans chacune des images appelé Dense SURF, en exploitant uniquement la partie de description du descripteur SURF. Puis, nous nous sommes concentrés sur l'utilisation conjointe de l'information Dense SURF avec l'information spectrale et les incorporer dans le processus de la classification.

Nous avons utilisé la méthode SVM pour effectuer cette classification en raison de sa performance prouvée en se basant sur une étude comparative entre plusieurs classificateurs où elle s'est avéré la mieux adaptée pour classifier des données importantes, même lorsque le nombre d'échantillons d'apprentissage est relativement faible, ainsi que pour sa facilité d'utilisation et le coût réduit de la mise en place de ses algorithmes. Ceci malgré la difficulté rencontrée lors du choix du noyau approprié et la détermination empirique de ses paramètres en fonction de la distribution des données d'apprentissage.

Dans la quatrième contribution, présentée dans le chapitre 6, nous avons utilisé conjointement les résultats issus du classifieur SVM en se basant sur la méthode DSMT, pour améliorer la classification. Ceci s'est avéré un outil pratique pour inclure l'information perceptuelle issue des cartes de la saillance générées dans le processus de la fusion via la modélisation des jeux de masses.

Les perspectives de ces travaux sont nombreuses. Elles concernent notamment l'amélioration des méthodes proposées. Par exemple, au niveau de la pré-classification, la réduction de la dimensionnalité des vecteurs fournis par la combinaison de l'information spectrale et spatiale va permettre une réduction du temps de calcul. Nous proposons aussi l'utilisation des noyaux composites où chaque type de caractéristiques sera associé à un noyau différent.

Au niveau de la fusion, nous proposons l'utilisation du modèle d'Appriou pour l'estimation des fonctions des masses au lieu des probabilités conditionnelles utilisée dans ce travail. À moyen terme, il serait intéressant d'introduire la conjonction entre les classes pour mieux modéliser les imperfections et le paradoxe des données issues de différentes sources, pour mieux cerner le conflit important entre les données.

ANNEXE 1

La classification SVM

Introduction

La représentation de différentes caractéristiques extraites de chaque modalité dans un espace commun va permettre une régularisation de classification de ces images, et qui vise généralement à séparer l'image en plusieurs classes en attribuant à chaque pixel une étiquette représentant un thème dans la zone d'étude réelle[180]. Les données extraites de ces images sont d'un côté considérées comme des problèmes non linéaires en ce qui concerne la relation entre elles, d'autres part, elles sont de très grandes dimensions et nécessitent des outils mathématiques, qui garantissent la rapidité du calcul. C'est pour cette raison que nous avons choisi la méthode SVM (Support Vector Machine).

Ce chapitre est organisé comme suit : nous commençons par donner le principe de la méthode de SVM, ensuite nous traitons les différents cas des SVMs à savoir le cas linéaire, le cas non linéaire et le cas de multiclasse, puis nous citons quelques applications des SVMs dans la classification des images et nous terminons par une conclusion.

Généralité sur la classification SVM

Les méthodes de la classification dans le domaine d'imagerie, d'une manière générale, permettent de donner une vision globale[180] sur la composition de l'image, en regroupant les données qui les composent en sous-ensembles homogènes avec des caractéristiques similaires. Ces méthodes peuvent être divisées, selon le contexte, en deux principales catégories :

- Les méthodes de classification supervisées : une phase d'apprentissage est tout d'abord effectuée dans cette catégorie, en disposant d'une information à priori sur le nombre de classes possibles et en construisant des échantillons pour chaque classe. La classification est effectuée en se servant de la base d'apprentissage et en attribuant à chaque objet/pixel sa classe la plus adaptée. Pour ce faire, plusieurs algorithmes ont été

développés, tels que les méthodes du maximum de vraisemblance[181], les réseaux de neurones, les approches probabilistes, les machines à Support Vecteurs (SVM) ... etc.

- Les méthodes de classification non supervisées : dite aussi les méthodes de classification automatiques. Cette catégorie ne nécessite pas une intervention de l'utilisateur qu'à la phase d'interprétation des résultats. Ces méthodes sont itératives et cherchent à regrouper les données en optimisant des distances ou des pseudo-distances et en se basant sur les observations. La méthode K-means (K moyennes), la méthode de la logique floue (FCM) et bien d'autres sont les plus connues dans cette catégorie.

Un autre type de classement divise ces méthodes en deux groupes [182]:

- Les méthodes paramétriques : ces méthodes modélisent les données par des lois de probabilité (normale ou bien multi normale) et estiment leurs paramètres. La classification s'effectue en affectant à chaque pixel/objet une probabilité d'appartenance à une classe donnée.
- Les méthodes non paramétriques : ces méthodes ont été mises au point pour résoudre le problème des paramètres statistiques lorsqu'on essaie de calculer la séparation des classes et ne demandent aucune utilisation de l'hypothèse concernant la distribution statistique des données. Les réseaux de neurones artificiels (CNN), les arbres de décision et les machines à Support Vecteurs (SVM) sont les techniques les plus utilisées de cette catégorie.

Dans ce travail, nous utilisons le SVM qui est largement utilisée dans les applications de classification grâce à ses performances permettant, son efficacité et son bon fonctionnement. Introduit par Vladimir Vapnik[169] [170][171], son principe consiste à trouver une séparation des classes en maximisant la marge séparatrice entre elles. Il comporte deux modèles : le modèle linéaire et le modèle non linéaire que nous détaillons dans cette section. SVM a été utilisé au tout début pour les classifications binaires, mais une extension aux cas multi-classe a été mise en place [183][169][184] très rapidement.

SVM linéaire

La classification SVM dans un problème simple/linéaire est réalisée en construisant un hyperplan ou un ensemble d'hyperplans dans un espace de dimension élevée ou infinie séparant les différentes étiquettes de classe. La bonne séparation est assurée avec l'hyperplan qui

maximise la marge entre le point de données d'entraînement le plus proche d'une classe. Pour définir cet hyperplan, il existe deux cas principaux : le cas où les données sont linéairement séparables et le cas où les données sont non linéairement séparables.

- **Données linéairement séparables**

A l'origine, SVM a été utilisé pour la séparation/classification de deux classes[185][186] dans un espace approprié. Son principe consiste à prédire le label d'éléments de R^n en se basant sur une base d'apprentissage $S = \{(x_i, y_i)/i \in [1, l]\}$ qui contient les objets $x_i \in R^n$ et les labels ($y_i = \pm 1$) correspondant à la classe d'appartenance des objets.

Cette base d'apprentissage S est dite linéairement séparable s'il existe une surface de séparation H appelé l'hyperplan d'équation $f(x) = 0$, avec $f(x) = w'x + b$ (voir Figure 72), qui sépare l'espace en deux parties (positive et négative) en cherchant une fonction f qui attribue le label $y_i = +1$ aux éléments vérifiant $f(x_i) > 0$ ou bien le label $y_i = -1$ aux éléments vérifiant

$$f(x_i)y_i > 0 \quad \forall i \in [1, l] \tag{A1.1}$$

La distance entre un vecteur d'apprentissage x_i et l'hyperplan H est définie comme suit :

$$d(x_i, H) = \frac{f(x_i)y_i}{\|w\|} \tag{A1.2}$$

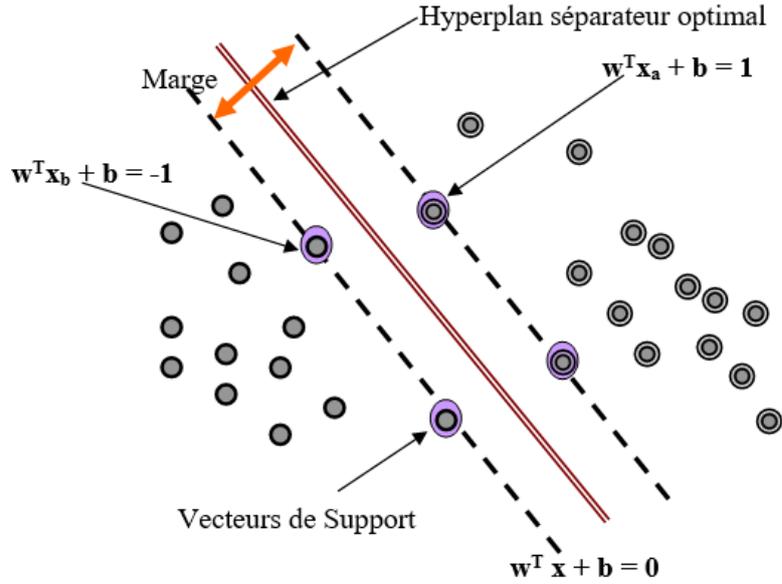


Figure 72 Classification linéaire : Données linéairement séparables [187]

Dans ce cas, nous cherchons un hyperplan séparateur qui maximise la marge de H . la marge est définie par la plus petite distance entre un vecteur d'apprentissage et l'hyperplan H qui est notée par :

$$\rho = \min_i d(x_i, H) \tag{A1.3}$$

Généralement, une règle de décision ayant une probabilité d'erreur faible, par rapport aux autres règles linéaires, correspond à un tel hyperplan, et ainsi considéré comme étant un problème de maximisation sous contraintes :

$$\begin{cases} \max \rho \\ \frac{y_i(w'x_i + b)}{\|w\|} \geq \rho, \forall i \in [1, l] \end{cases} \tag{A1.4}$$

Ce problème peut être simplifié en fixant la norme w à $\|w\| = \frac{1}{\rho}$ et le problème d'optimisation devient :

$$\begin{cases} \min_{w,b} \frac{1}{2} \|w\|^2 \\ y_i(w'x_i + b) \geq 1, \forall i \in 1 \dots l \end{cases} \quad \text{A1.5}$$

Le problème d'optimisation est présenté sous une forme duale qui est résolu en utilisant les multiplicateurs Lagrangiens, qui seront ensuite attribués aux contraintes de manière suivante :

$$L(w,b,\alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i \{(w'x_i + b)y_i - 1\} \quad \text{A1.6}$$

Avec $\alpha_i \geq 0 \forall i$,

Qui peut être développé afin d'être indépendant de w et de b , et en utilisant les contraintes $\frac{\partial L}{\partial w} = 0$, on obtient l'expression suivant :

$$L(w,b,\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i' x_j), \quad \text{A1.7}$$

Après, une maximisation de ce problème par rapport à α , sous les contraintes (2.10), est effectuée en utilisant n'importe quel algorithme de programmation quadratique.

$$\begin{aligned} \alpha_i &\geq 0 \quad i = 1, \dots, l \\ \sum_{i=1}^l \alpha_i y_i &= 0 \end{aligned} \quad \text{A1.8}$$

Pour résoudre ce problème, les conditions de Kuhn et Tucker (KT) (2.12, 2.13, 2.14, 2.15) doivent être vérifiées pour des nombres α_i^* , $\forall i \in [1, l]$:

$$\frac{\partial L}{\partial w}(w^*, \alpha^*) = 0 \quad \text{A1.9}$$

$$c_i(w^*) \geq 0, \quad \forall i \in [1, l] \quad \text{A1.10}$$

$$\alpha_i^* [y_i(w^{*'}x_i + b^*) - 1] = 0, \quad \forall i \in [1, l] \quad \text{A1.11}$$

$$\alpha_i^* \geq 0, \quad \forall i \in [1, l] \quad \text{A1.12}$$

- x_i est considéré comme étant un vecteur de support, si $\alpha_i^* \geq 0$, et la contrainte $y_i(w^*x_i + b^*) = 1$ est active,
- x_i n'est pas un vecteur de support, si $\alpha_i^* = 0$ et $y_i(w^*x_i + b^*) > 1$.

Avec $D = \{i \in [1, l] / \alpha_i^* > 0\}$ est l'ensemble des indices des vecteurs de support.

La solution (w^*, b^*) du problème initial est comme suit :

$$w^* = \sum_{i=1}^l \alpha_i^* y_i x_i = \sum_{i \in D} \alpha_i^* y_i x_i, \quad \text{A1.13}$$

Pour tout $i \in D$, la solution du problème initial en fonction des vecteurs de support uniquement est :

$$y_i(w^*x_i + b^*) = 1 \Rightarrow b^* = \frac{1}{y_i} - w^*x_i = y_i - w^*x_i \quad \text{A1.14}$$

L'hyperplan séparateur optimal a donc pour équation $f^*(x) = 0$ avec

$$f^*(x) = \sum_{i \in D} \alpha_i^* y_i x_i' x + b^* \quad \text{A1.15}$$

Pour conclure, le SVM est une règle de décision correspondante $g^*(x) = \text{sgn}(f^*(x))$ et qui s'exprime uniquement à l'aide des produits scalaires entre x et les vecteurs de support.

• Données non-linéairement séparables

Il se peut que les données que nous souhaitons classifier soient entachées de bruit et d'erreurs, ce qui rend la recherche d'un hyperplan optimale qui sépare ces données, qui sont non linéairement séparables, en deux classes. Une tâche difficile et voir même impossible.

Pour ce faire, une idée de pénalisation des séparatrices admettant des données mal classées est adapté dans ce cas, en utilisant une fonction de coût qui affecte une pénalité pour toutes les données qui ne sont pas du bon côté des marges et qui sont à une distance ξ de la marge qu'ils devraient respecter, ce cas est illustré sur la Figure 73.

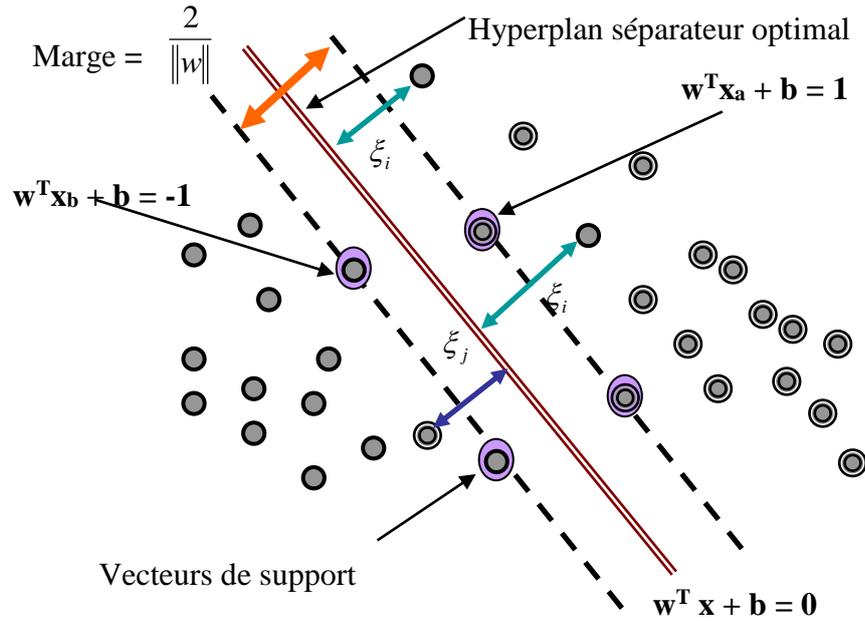


Figure 73 Cas des données non-linéairement séparables [187]

Pour ce faire, des variables d'écart $\xi_i, i = 1, \dots, l$, appelées aussi des variables molles, sont introduites de la manière suivante :

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i, \quad \text{A1.16}$$

Sous les contraintes
$$\begin{cases} y_i(w'x_i + b) \geq 1 - \xi_i, & \forall i: 1, \dots, l \\ \xi_i \geq 0, & \forall i: 1, \dots, l \end{cases}$$

Ensuite, une pondération des erreurs de classification est effectuée en utilisant le terme de régularisation de marge afin d'optimiser la fonction séparatrice finale.

Ce problème d'optimisation est résolu en utilisant les multiplicateurs de Lagrangiens comme suit :

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i \{ (w'x_i + b) y_i - (1 - \xi_i) \} - \sum_{i=1}^l \beta_i \xi_i \quad \text{A1.17}$$

Et en vérifiant les conditions KT suivantes :

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0 \quad \text{A1.18}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^l \alpha_i y_i = 0 \quad \text{A1.19}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad \forall i:1,\dots,l \quad \text{A1.20}$$

Avec

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad \text{A1.21}$$

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad \text{A1.22}$$

$$\beta_i = C - \alpha_i \quad \forall i:1,\dots,l \quad \text{A1.23}$$

En se basant sur ces contraintes, le problème d'optimisation peut être indépendant de w et β devient :

$$L(w, b, \xi, \alpha, \beta) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (\mathbf{x}_i' \mathbf{x}_j) \quad \text{A1.24}$$

Pour le maximiser par rapport à α , le problème dual est utilisé sous les contraintes suivantes :

$$\alpha_i \geq 0 \quad \forall i:1,\dots,l$$

$$\beta_i \geq 0 \quad \forall i:1,\dots,l$$

$$\beta_i = C - \alpha_i \quad \forall i:1,\dots,l \quad \text{A1.24}$$

$$\sum_{i=1}^l \alpha_i y_i = 0$$

- Si α^* est la solution de ce problème, les conditions de *KT* deviennent :

$$\alpha_i^* [y_i (w^{*'} x_i + b^*) - 1 + \xi_i^*] = 0, \quad \forall i \in [1, l] \quad \text{A1.25}$$

$$\beta_i^* \xi_i^* = 0, \quad \forall i \in [1, l] \quad \text{A1.26}$$

Dont $\beta_i^* = C - \alpha_i^*$ les vecteurs de supports x_i correspondent aux α_i^* strictement positifs et satisfaisant :

$$y_i (w^{*'} x_i + b^*) = 1 - \xi_i^* \quad \text{A1.27}$$

Si $1 - \xi_i^* < 0$, deux cas sont envisagés, soit ces données sont mal classées, soit les vecteurs sont à l'intérieur de la marge avec une distance de l'hyperplan inférieure ou égale à $\rho = 1/\|w\|$. Les vecteurs qui sont sur le bord de la marge ont une distance égale à ρ avec de plus $\beta_i^* > 0$, ce qui signifie que $0 \leq \alpha_i \leq C$, alors $\xi_i^* = 0$.

Alors

$$w^* = \sum_{i=1}^l \alpha_i^* y_i x_i = \sum_{i \in D} \alpha_i^* y_i x_i \quad \text{A1.28}$$

Avec $D = \{i \in [1, l] / \alpha_i^* > 0\}$, l'ensemble des indices des vecteurs de support

b^* est alors obtenu en considérant un vecteur de support tel que $0 \leq \alpha_i^* \leq C$ de la manière suivante :

$$b^* = \frac{1}{y_i} - w^{*'} x_i = y_i - w^{*'} x_i \quad \text{A1.29}$$

Tout comme dans le cas linéairement séparable, la fonction de décision est exprimée en fonction des produits scalaires x avec les vecteurs de support et s'écrit comme suit :

$$g^*(x) = \text{sgn}\left(\sum_{i \in D} \alpha_i^* y_i (x_i' x) + b^*\right) \quad \text{A1.30}$$

SVM non-linéaire

• Formalisation de problème

Dans la plupart des applications, et surtout dans les applications des images multimodales, les données traitées ne sont pas toujours linéairement séparables (voir Figure 74), et du coup la base d'apprentissage demande une séparatrice de décision plus complexe qu'un simple hyperplan linéaire. Pour cette raison, une généralisation des SVMs linéaires était introduite en utilisant la notion de noyaux.

L'idée de SVM non linéaire est de mapper les données à un nouvel espace de représentation intermédiaire H de grande dimension et pourra être aussi de dimension infinie :

$$\Phi: \mathbb{R}^n \rightarrow H \quad \text{A1.31}$$

Et puis, rechercher l'hyperplan séparateur optimal dans l'espace H .

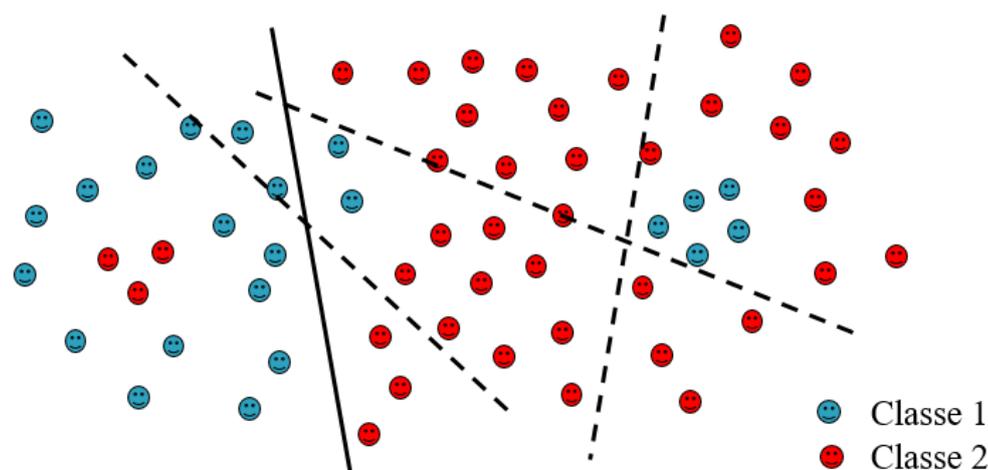


Figure 74 Cas de SVM non linéaire

Pour ce faire, une application Φ de \mathbb{R}^n est utilisée muni d'un produit scalaire dans H , dont une *fonction noyau* $K: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ est définie en vérifiant :

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle, \quad \text{A1.32}$$

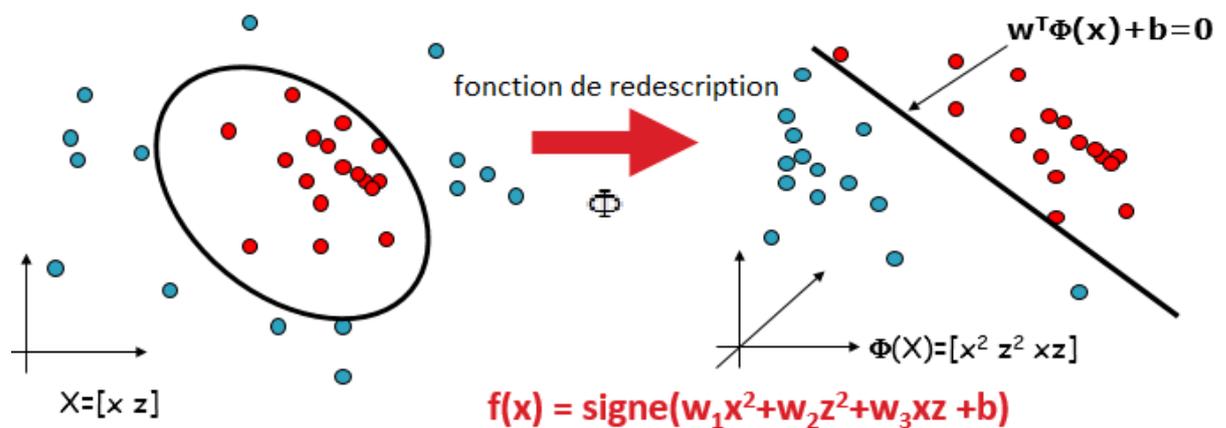


Figure 75 : Exemple de séparation linéaire par hyperplan optimale dans un nouvel espace de description

La fonction noyau peut être considérée comme étant une mesure de similarité entre entrées x_i et x_j , en supposant que deux entrées similaires doivent être associées à des sorties similaires.

Après la définition de la fonction noyau utilisée, le problème d'optimisation s'écrit comme suit :

$$\max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad \text{A1.33}$$

$$0 \leq \alpha_i \leq C \quad \forall i: 1, \dots, l \quad \text{A1.34}$$

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad \text{A1.35}$$

- Si α^* est la solution de ce problème. La fonction discriminante $f^*(x)$ est :

$$f^*(x) = \sum_{i \in D} \alpha_i^* y_i K(\mathbf{x}, \mathbf{x}_i) + b^* \quad \text{A1.36}$$

Et la fonction de décision correspond à :

$$g^*(x) = \text{sgn}(f^*(x)) \quad \text{A1.37}$$

- **La notion des noyaux**

Une fonction noyau K symétrique correspond à un produit scalaire dans un espace H , si elle vérifié la condition *de Mercer* suivante:

$$\forall f : R^n \rightarrow R \text{ t.q. } \int f(x)^2 dx < \infty, \int K(x, y) f(x) f(y) dx dy \geq 0 \quad \text{A1.38}$$

Il existe plusieurs types de fonctions noyaux, dont nous citons celles d'usage courant :

- *Noyau polynomial*

$$K(x, y) = (x' y + c)^r, \text{ avec } c \in R_+ \text{ et } r \in N \quad \text{A1.39}$$

Ce noyau est une méthode populaire pour la modélisation non linéaire dans laquelle une séparation par polynomiale de degré au plus de r est recherchée.

Avec l'ordre de $\frac{(r+n)!}{r!n!}$ qui représente la dimension de l'espace transformé induit par un noyau polynomial, où n est la dimension de l'espace de départ.

- *Noyau gaussien*

Les fonctions de base radiales (RBF) vérifient la *condition de Mercer*, et utilisent les frontières radiales pour chercher une séparatrice optimales, en se basant sur la Gaussienne suivante :

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad \text{A1.40}$$

La similarité est le point clé de mappage en utilisant la fonction $\Phi(x)$ induite par ce noyau et dont un exemple va être mappé sur une fonction Gaussienne représentant la similarité de l'exemple avec tous les autres.

Le paramètre σ permet de contrôler la capacité de la classification tout en réglant la largeur de la Gaussienne de sorte que :

- Si σ grand, la similarité d'un exemple par rapport à ceux qui l'entourent sera assez élevée,

- Si σ tendant vers 0, l'exemple ne sera similaire à aucun autre.

Du coup, le fait de resserrer la Gaussienne, permet au classifieur d'apprendre n'importe quel ensemble d'apprentissage sans commettre d'erreur.

SVM multiclasse

À l'origine, SVM était une classification binaire, mais de nombreuses applications nécessitaient une classification multiclasse (voir Figure 76). La classification multiclasse peut être considérée comme étant une généralisation de la classification binaire en examinant presque tout le temps une combinaison de plusieurs classifications binaires dans lesquelles chaque paire de classes est réalisée séparément. Plusieurs approches ont été proposées pour résoudre ce problème [188]. Ces approches sont regroupées en deux catégories principales [189] :

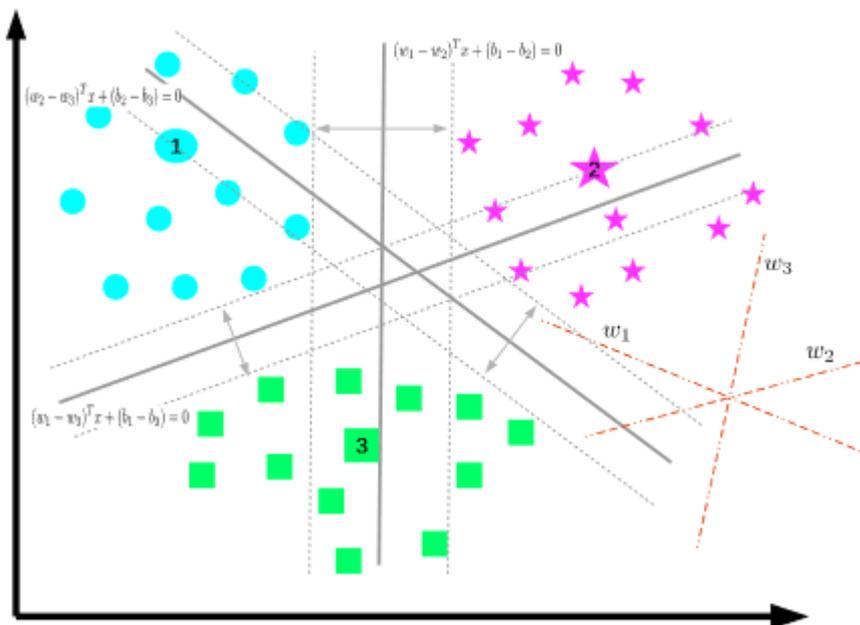


Figure 76 Cas de SVM multiclass

- **Un-contre-tous**

La stratégie "un contre tous" est la plus simple et la plus ancienne des méthodes de décomposition. Son idée consiste à transformer le problème de multi classification donc à k classes en k classifieurs binaires. Le k -ième classifieur distingue la catégorie d'indice k de tous

les autres. Chaque exemple est présenté donc à k classifieurs, et chaque classifieur renvoie sa décision. Le label ou la décision retenue est celle associée au classifieur qui répond le mieux.

La fonction discriminante h_k^* obtenue en opposant la classe w_k aux autres :

$$g(x) = \arg \max_k (h_k^*(x)). \quad A1.41$$

• Un-contre-un

La stratégie "un contre un" est introduite par *Knerr* et al. [190] et consiste en une adaptation de SVM pour combler les problèmes multiclassés, en utilisant la classification par couple de catégories donc à C_Q^2 classifieurs. Le problème de classification multiclassé est transformé en $\frac{k.(k-1)}{2}$ dont le classifieur est indicé par le couple (k, l) (avec $1 \leq k < l \leq Q$), et chaque classe k est comparée avec celle d'indice l . La décision retenue est basée soit sur le vote majoritaire, soit sur le graphe acyclique.

L'intérêt principal de la méthode SVM est, sa facilité de mise en œuvre de ses algorithmes qui est peu coûteuse[191]. Cette méthode est largement utilisée dans divers domaines, surtout dans le domaine des images satellitaires et de télédétection dans la dernière décennie [192], dont l'utilisation de SVM présente de nombreux avantages, tels que son efficacité de traitement des données de grande dimension, ce qui simplifie le schéma traditionnel de la classification proposé par Landgrebe[193] tout en évitant la procédure de l'extraction/sélection de caractéristiques complexes.

Plusieurs recherches ont été basées sur l'utilisation de SVM pour la tâche de classification. Dans [194], les auteurs ont combinés le K-means et le SVM pour la classification de l'utilisation des sols et de la couverture végétale. *Yu* et al. [195] ont utilisé SVM à la classification lithologique automatisée dans une zone d'étude dans le nord-ouest de l'Inde en utilisant des données de télédétection. L'utilisation de SVM donne de meilleures performances par rapport à la MLC.

Wu et al. [196] ont proposé une nouvelle approche intégrant les points d'intérêts SURF avec le SVM. Pour segmenter l'image en classe modifiée et non modifiée, un détecteur et un descripteur SURF sont utilisés pour extraire deux images de type point d'intérêt. Les points

extraits sont ensuite comparés et filtrés à l'aide de Ransac. Les points d'intérêts conservés sont considérés comme des échantillons d'apprentissage pour une classe inchangée et le reste comme une classe changée. *Bekkari et al.* [191] combinent des caractéristiques de Haralick avec des informations spectrales pour la classification des images satellites haute résolution à l'aide du SVM.

Auparavant, de nombreuses recherches sur la classification d'images multimodales dans plusieurs domaines ont été abordées. Dans le domaine des images médicales, *Juergen Dukart et al.* [197] ont proposé une méthode pour améliorer la détection et la différenciation de la démence de la maladie d'Alzheimer et du lobar fronto-temporel. En effectuant une classification basée sur le SVM aux informations combinées issues de la tomographie par émission de positrons d'imagerie par résonance magnétique (IRM) et de [F18] fluorodéoxyglucose (TEP-FDG). Les auteurs ont été motivés par l'amélioration de la classification dans [198][199], et qui combine les informations provenant de la TEP-FDG et de l'IRM en utilisant les méthodes statistiques telles que les analyses multivariées de modèles en utilisant SVM, et qui permet une considérable amélioration de la détection ainsi que la différenciation de la démence de la maladie d'Alzheimer et de la dégénérescence lobaire fronto-temporale [200].

Toujours dans le même domaine, *Molina et al.* [201] ont utilisé le SVM avec la texture et des caractéristiques fonctionnelles sur des images RM multimodales, afin d'estimer les structures du cancer après une application d'un ensemble d'opérations de prétraitement en tant qu'interpolation avec B-Spline et normalisation de l'intensité entre les images multimodales utilisées.

Dans la reconnaissance biométrique, *Zhou et al.* [202] présentent un nouveau système de reconnaissance oculaire multimodal permettant d'obtenir une précision de reconnaissance élevée en appliquant des algorithmes de filtre de Gabor et de filtre log-Gabor 1-D comme reconnaissance de l'iris et une segmentation automatique de la sclère, génération de modèle de motif comme reconnaissance de sclérotique.

Geethu et al. [203] ont développé une nouvelle approche biométrique en utilisant SVM afin de permettre un système efficace de reconnaissance des empreintes digitales et de l'iris. Cette approche repose sur la combinaison de caractéristiques provenant de données unimodales, iris et empreintes digitales, au niveau de score correspondant pour une authentification biométrique multimodale.

Dans le domaine de la stéréovision, *Apatean* et al. [204] ont proposé une méthode de reconnaissance d'obstacles basée sur SVM utilisant un vecteur fondu de texture globale extrait à la fois d'images infrarouges (IR) et visibles.

Brehar et al. [205] ont développé un alignement de données spatio-temporelles pour un mécanisme de fusion de classification afin de combiner des détections infrarouges et en niveaux de gris sur un flux piétonnier unifié.

Dans le domaine hyperspectral, *Pomrehn* et al. [206], ont montré l'avantage de l'utilisation des données multimodales dérivées microspectroscopie Raman et IR qui fournissent des informations complémentaires dans l'ordre d'améliorer le taux de classification des signatures de matériaux hyperspectraux très complexes en se basant sur le SVM. Les caractéristiques extraites par des techniques de DR basées sur des statistiques sont fusionnées par deux approches spectroscopiques et montrent une significative amélioration de la tâche de classification.

Dans le domaine du patrimoine culturel, *Yasser* et al. [207] ont présenté un nouveau cadre de classification permettant de sauvegarder le patrimoine culturel grâce à un faux-semblant numérique. Le cadre incorporait une analyse de données multimodale et une extraction de données augmentée basée sur le contenu en utilisant un paquet de fonctionnalités (BOF) et un réseau de neurones à convolution CNN.

Conclusion

Dans ce chapitre, nous avons présenté le principe de fonctionnement du classifieur SVM, que nous avons retenu dans nos travaux en raison de sa performance montrée sur les données telles que celles dont nous disposons. Ensuite nous avons décrit son extension pour le cas multiclassés basées sur les méthodes de noyau, en expliquant le fonctionnement des deux stratégies : la stratégie "un contre tous" et la stratégie "un contre un". Nous avons conclu par les principaux travaux qui traitent le problème de classification en utilisant le classifieur SVM dans différents domaines.

Bibliographie

- [1] M. J. P. R. Weissleder, "Imaging in the era of molecular oncology," *Nature*, no. 452, pp. 580–589, 2008.
- [2] S. A. J. Culver, W. Akers, "Multimodality molecular imaging with combined optical and SPECT/PET modalities"," *J. Nucl. Med.*, no. 49, pp. 169–172, 2008.
- [3] M.Z Abdullah, "Image Acquisition Systems," *Comput. Vis. Technol. Food Qual. Eval.*, pp. 3–35, 2008.
- [4] V. C. É. Belin, D. Rousseau, T. Boureau, "Thermography versus chlorophyll fluorescence imaging for detection and quantification of apple scab," *Comput. Electron. Agric.*, vol. 90, pp. 159–163, 2013.
- [5] B. M. X. Wang, W. Yang, A. Wheaton, N. Cooley, "Efficient registration of optical and IR images for automatic plant water stress assessment," *Comput. Electron. Agric.*, vol. 74, no. 2, pp. 230–237, 2010.
- [6] F. B. C. Simon Chane, A. Mansouri, F.S. Marzani, "Integration of 3D and multispectral data for cultural heritage applications: Survey and perspectives," *Image Vis. Comput.*, vol. 31, no. 1, pp. 91–102, 2013.
- [7] G. Luis, S. Member, D. Tuia, and S. Member, "Multimodal Classification of Remote Sensing Images : A Review and Future Directions," pp. 1–52, 2015.
- [8] A. B. A. M. Hanan Anzid Gaetan Le Goic and D. Mammass, "Improving point matching on multimodal images using distance and orientation automatic filtering," *Comput. Syst. Appl. (AICCSA), 2016 IEEE/ACS 13th Int. Conf.*, pp. 1–8, 2016.
- [9] H. Anzid, G. Le Goic, A. Bekkari, A. Mansouri, and D. Mammass, "Multimodal Images Classification using Dense SURF, Spectral Information and Support Vector Machine," *Procedia Comput. Sci.*, vol. 148, pp. 107–115, 2019.
- [10] H. Anzid, G. le Goic, A. Mansouri, and D. Mammass, "Improvement of Multimodal Images Classification Based on DSMT Using Visual Saliency Model Fusion With SVM," *Int. J. Comput. Technol.*, vol. 18, pp. 7418–7430, 2018.
- [11] S. C. A. Thomopoulos, R. Viswanathan, and D. C. Bougoulas, "Optimal decision fusion in multiple sensor systems," *IEEE Trans. Aerosp. Electron. Syst.*, no. 5, pp. 644–653, 1987.
- [12] A. White, "Data Fusion Lexicon, Joint Directors of Laboratories, Technical Panel for C3," *Nav. Ocean Syst. Center, San Diego, Tech. Rep*, 1987.
- [13] A. Dromigny-Badin, "Fusion d'images par la théorie de l'évidence en vue d'applications médicales et industrielles." Lyon, INSA, 1998.
- [14] P. L. Rothman and R. V Denton, "Fusion or confusion: knowledge or nonsense?," in

Data Structures and Target Classification, vol. 1470, pp. 2–13, 1991.

- [15] I. Bloch, “Information combination operators for data fusion: a comparative review with classification,” in *Image and signal processing for Remote Sensing*, vol. 2315, pp. 148–160, 1994.
- [16] I. Bloch, “Fusion d’informations en traitement du signal et des images,” *Hermes Sci. Publ.*, vol. 2, 2003.
- [17] I. Bloch and H. Maître, “Fusion de données en traitement d’images: modèles d’information et décisions,” *TS. Trait. du signal*, vol. 11, no. 6, pp. 435–446, 1994.
- [18] A. Martin, “La fusion d’informations,” *Polycopié de cours ENSIETA-Réf*, vol. 1484, p. 117, 2005.
- [19] A. ELHASSOUNY, “Fusion d’images par la théorie de Dezert-Smarandache (DSmT) en vue d’applications en télédétection.” Université Ibn Zohr, 2013.
- [20] D. Dubois, “Théorie des possibilités; applications a la représentation des connaissances en informatique,” 1988.
- [21] P. Smets, “Imperfect information: Imprecision and uncertainty,” in *Uncertainty management in information systems*, Springer, pp. 225–254, 1997.
- [22] Y. Yan, “Fusion de mesures de déplacement issues d’imagerie SAR: Application aux modélisations séismo-volcaniques.” Université de Grenoble, 2011.
- [23] E. Lefevre, O. Colot, and P. Vannoorenberghe, “Belief function combination and conflict management,” *Inf. fusion*, vol. 3, no. 2, pp. 149–162, 2002.
- [24] F. Smarandache and J. Dezert, “Advances and applications of DSmT for information fusion-Collected works-Volume 3.” American Research Press, 2009.
- [25] P. Smets, “Varieties of ignorance and the need for well-founded theories.,” *Inf. Sci. (Ny)*, vol. 57, no. 58, pp. 135–144, 1991.
- [26] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
- [27] B. V Dasarathy, “Sensor fusion potential exploitation-innovative architectures and illustrative applications,” *Proc. IEEE*, vol. 85, no. 1, pp. 24–38, 1997.
- [28] J. Zhang, “Multi-source remote sensing data fusion: status and trends,” *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 5–24, 2010.
- [29] Y. Lemeret, E. Lefevre, and D. Jolly, “Fusion de données provenant d’un laser et d’un radar en utilisant la théorie de Dempster-Shafer,” *MAJECSTIC’04, Fr. 2004*, 2004.
- [30] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, “Multimodal classification of remote sensing images: A review and future directions,” *Proc. IEEE*, vol. 103, no. 9,

pp. 1560–1584, 2015.

- [31] M. Mangolini, T. Ranchin, and L. Wald, “Procédé et dispositif pour augmenter la résolution spatiale d’images à partir d’autres images de meilleure résolution spatiale.” 1992.
- [32] W. Dou, Y. Chen, X. Li, and D. Z. Sui, “A general framework for component substitution image fusion: An implementation using the fast image fusion method,” *Comput. Geosci.*, vol. 33, no. 2, pp. 219–228, 2007.
- [33] A. Dromigny-Badin, “Image fusion using evidence theory: applications to medical and industrial images.” Thèse EEA. Lyon: INSA de Lyon, 1998, 158p, 1998.
- [34] I. Hammami, “Fusion of heterogeneous remote sensing images by credibilist methods.” Ecole nationale supérieure Mines-Télécom Atlantique, 2017.
- [35] J. Dong, Z. Dafang, H. Yaohuan, and F. Jinying, “Survey of Multispectral Image Fusion Techniques in Remote Sensing Applications, Image Fusion and Its Applications, Yufeng Zheng, Alcorn State University, USA (Ed.), ISBN: 978-953-307-182-4, InTech.” 2011.
- [36] C.-C. Chu and J. K. Aggarwal, “Image interpretation using multiple sensing modalities,” *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 8, pp. 840–847, 1992.
- [37] C. Degrigny *et al.*, “Technical Study of Germolles’ wall paintings : the input of imaging techniques,” *Virtual Archaeol. Rev.*, vol. 7, pp. 1–8, Jul. 2016.
- [38] J.-M. M. N. Sabater and A. Almansa, “How accurate can block matches be in stereovision,” *SIAM J. Imaging Sci.*, vol. 4, pp. 472–500, 2011.
- [39] H. Sawhney and R. Kumar, “True multi-image alignment and its application to mosaicing and lens distortion correction,” *PAMI*, vol. 21, pp. 235–243, 1999.
- [40] F. A. S. Leprince S. Barbot and J.-P. Avouac, “Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements,” *Commun. Comput. Inf. Sci.*, vol. 127, pp. 15–28, 2007.
- [41] A. T. J. T. Evanthia Faliagka George Matsopoulos and G. Tzimas, “Registration and Fusion Techniques for Medical Images: Demonstration and Evaluation,” *GRSS*, vol. 45, pp. 1529–1558, 2011.
- [42] L. G. Brown, “A survey of image registration techniques,” *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.
- [43] J. C. Gee, C. Barillot, L. Le Briquer, D. R. Haynor, and R. K. Bajcsy, “Matching structural images of the human brain using statistical and geometrical image features,” in *Visualization in Biomedical Computing 1994*, 1994, vol. 2359, pp. 191–205.
- [44] J. Z. Renbo Xia and Y. Liu, “A robust feature-based registration method of multimodal image using phase congruency and coherent point drift,” *SPIE*, vol. 8919,89190, pp. 1–

8, 2013.

- [45] C. Rominger, A. Martin, C. Rominger, and A. Martin, “Recalage et fusion d’images sonar multivues : utilisation du conflit,” 2012.
- [46] J. W. X. W. Jingga Zhang and D. Feng, “Multi-modal image registration with joint structure tensor and local entropy,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, pp. 1765–1775, 2015.
- [47] A. Boucher and L. U. P. Descartes, “Recalage et analyse d’un couple d’images : application aux mammographies,” 2013.
- [48] J. Chen and J. Tian, “Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor,” *Prog. Nat. Sci.*, vol. 19, pp. 643–651, 2008.
- [49] A. Sotiras, C. Davatzikos, and N. Paragios, “Deformable medical image registration: A survey,” *IEEE Trans. Med. Imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.
- [50] J. N. Ulysses and A. Conci, “Measuring Similarity in Medical Registration,” 2010.
- [51] G. M. Alexis Roche and N. Ayache, “Unifying Maximum Likelihood Approaches in Medical Image Registration,” *INRIA*, vol. 11, pp. 71–80, 2000.
- [52] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain,” *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.
- [53] Q. Z. Xiaoyong Shen LiXu and JiayaJia, “Multi-modal and multi-spectral registration for natural images,” *ECCV 2014*, pp. 309–324, 2014.
- [54] A. S. Konstantinos Karantzas and N. Paragios, “Efficient and automated multimodal satellite data registration through MRFs and linear programming,” *CVPR*, pp. 329–336, 2014.
- [55] H. Xin, “Modélisation et recalage d’images protéomiques.” Institut National des Sciences Appliquées de Lyon, 2008.
- [56] D. V. G. M. F. Maes A. Collignon and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Trans. Med. Imaging*, vol. 16, pp. 187–198, 1997.
- [57] G. M. A. Roche and X. Pennec, “The correlation ratio as a new similarity measure for multimodal image registration,” *Proc. Med. Image Comput. Comput. Interv.*, pp. 1115–1124, 1998.
- [58] F. Barrera F. Lumbreras and A. D. Sappa, “Multispectral piecewise planar stereo using Manhattan-world assumption,” *Pattern Recognit. Lett.*, vol. 34, pp. 52–61, 2013.
- [59] J. Zhang, J. Wang, X. Wang, and D. Feng, “Multimodal image registration with joint structure tensor and local entropy,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 11,

pp. 1765–1775, 2015.

- [60] C. H. Chen, *Signal and image processing for remote sensing*. Boca Raton: CRC Press, 2007.
- [61] R. Bouchiha and K. Besbes, “Automatic remote-sensing image registration using SURF,” *Int. J. Comput. Theory Eng.*, vol. 5, pp. 88–92, 2013.
- [62] A. Sotiras, “Discrete image registration: a hybrid paradigm.” Ecole Centrale Paris, 2011.
- [63] P. J. Besl and N. D. McKay, “Method for registration of 3-D shapes,” in *Sensor Fusion IV: Control Paradigms and Data Structures*, 1992, vol. 1611, pp. 586–607.
- [64] D. G. Lowe, “Object recognition from local scale-invariant features,” *Proc Int. Conf. Comput. Vision, Corfu*, vol. 5, pp. 1150–1157, 1999.
- [65] M. B. D. Firmenichy and S. Susstrunk, “Multispectral interest points for RGB-NIR image registration,” *Proc. IEEE Int. Conf. Image Process.*, vol. 110(3), pp. 181–1849, 2011.
- [66] T. T. H. Bay and L. V Gool, “SURF: speeded up robust features,” *Proc. Eur. Conf. Comput. Vis.*, pp. 404–417, 2006.
- [67] J. Chen and J. Tian, “Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor,” *Prog. Nat. Sci.*, vol. 19, no. 5, pp. 643–651, 2009.
- [68] Z. J. Dong Zhao Yan Yang and X. Hu, “Rapid multimodality registration based on MM-SURF,” *Neurocomputing*, vol. 131, pp. 87–97, 2014.
- [69] G. Wang, Z. Wang, Y. Chen, and W. Zhao, “Biomedical Signal Processing and Control Robust point matching method for multimodal retinal image registration,” *Biomed. Signal Process. Control*, vol. 19, pp. 68–76, 2015.
- [70] R. Roopalakshmi and G. R. M. Reddy, “A novel spatio-temporal registration framework for video copy localization based on multimodal features,” *Signal Processing*, vol. 93, no. 8, pp. 2339–2351, 2013.
- [71] M. I. Patel, V. K. Thakar, and S. K. Shah, “Image Registration of Satellite Images with Varying Illumination Level Using HOG Descriptor Based SURF,” *Procedia Comput. Sci.*, vol. 93, pp. 382–388, 2016.
- [72] P. V. Lukashevich, B. A. Zalesky, and S. V. Ablameyko, “Medical image registration based on SURF detector,” *Pattern Recognit. Image Anal.*, vol. 21, no. 3, p. 519, 2011.
- [73] M. Su, C. Zhang, Z. Chen, and S. Jiang, “Registration of multimodal brain images based on optical flow,” in *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017 10th International Congress on*, pp. 1–5, 2017.
- [74] S. Saleem and R. Sablatnig, “A robust SIFT descriptor for multi-spectral images,” *IEEE Signal Process. Lett.*, vol. 21, pp. 400–403, 2014.

- [75] R. Bajcsy and C. Broit, "Matching of deformed images," in *Sixth International Conference on Pattern Recognition (ICPR '82)*, pp. 351–353, 1982.
- [76] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *IEEE Trans. image Process.*, vol. 5, no. 10, pp. 1435–1447, 1996.
- [77] G. E. Christensen, "Consistent linear-elastic transformations for image matching," in *Biennial International Conference on Information Processing in Medical Imaging*, pp. 224–237, 1999.
- [78] M. Semchedine, "Système coopératif hybride de classification dans un SMA: application à la segmentation d'images IRM." 2018.
- [79] S. Susstrunk, R. Buckley, and S. Swen., "Standard RGB Color Spaces," *IS T/SID 7th Color Imaging Conf.*, vol. 7, 1999.
- [80] S. a Le Moan, "Saliency for spectral image analysis," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, pp. 2472–2479, 2012.
- [81] M. J. J. E. Cubillasa, "THE APPLICATION OF SUPPORT VECTOR MACHINE (SVM) USING CIELAB,COLOR MODEL, COLOR INTENSITY AND COLOR CONSTANCY AS FEATURES FOR ORTHO IMAGE CLASSIFICATION OF BENTHIC HABITATS IN HINATUAN,SURIGAO DEL SUR, PHILIPPINES," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, pp. 189–194, 2016.
- [82] M. Singha and K. Hemachandran, "Content based image retrieval using color and texture," *Signal Image Process.*, vol. 3, no. 1, p. 39, 2012.
- [83] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," *Int. J. Sci. Res. Publ.*, vol. 3, no. 5, p. 1, 2013.
- [84] E. Miyamoto and T. Merryman, "Fast calculation of Haralick texture features," *Hum. Comput. Interact. institute, Carnegie Mellon Univ. Pittsburgh, USA. Japanese Restaur. Off.*, 2005.
- [85] S. Zhang, J. Huang, Y. Huang, Y. Yu, H. Li, and D. N. Metaxas, "Automatic image annotation using group sparsity," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 3312–3319, 2010.
- [86] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 2548–2555, 2011.
- [87] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 510–517, 2012.
- [88] D. Kim, "Face Components Detection using SURF Descriptors and SVMs," pp. 51–56, 2008.

- [89] D. Schmitt and N. Mccooy, "Object Classification and Localization Using SURF Descriptors," pp. 1–5, 2011.
- [90] A. Alfanindya, N. Hashim, and C. Eswaran, "Content Based Image Retrieval And Classification Using Speeded-Up Robust Features (SURF) and Grouped Bag-of-Visual-Words (GBoVW)," pp. 77–82, 2013.
- [91] S. Matuska, R. Hudec, P. Kamencay, M. Benco, and M. Zachariasova, "Classification of wild animals based on SVM and local descriptors," *AASRI Procedia*, vol. 9, pp. 25–30, 2014.
- [92] J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. Van Gool, "Hough transform and 3D SURF for robust three dimensional classification," in *European Conference on Computer Vision*, pp. 589–602, 2010.
- [93] F. A. P. Sharma, "Saliency map for human gaze prediction in images," *Sixt. Color Imaging Conf. Portland, Oregon, USA*, 2008.
- [94] C. K. L. Itti, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *PAMI*, pp. 1254–1259, 1988.
- [95] Y. T. Z Chen, "An improved saliency detection algorithm based on Itti's model," *Teh. Vjesn.*, pp. 1337–1344, 2014.
- [96] K. N. J. Han, "Unsupervised extraction of visual attention objects in color images," *Trans Circuits Syst. Video Technol.*, pp. 141–145, 2006.
- [97] D. W. U. Rutishauser, "Is bottom-up attention useful for object recognition?," *CVPR*, pp. 37–44, 2004.
- [98] F. W. Y. Wei, "Geodesic saliency using background priors," *ECCV*, 2012.
- [99] V. V M, "Saliency-based discriminant tracking," *CVPR*, 2009.
- [100] C. G. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process*, pp. 185–198, 2010.
- [101] I. a Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Res.*, 2000.
- [102] A. O. A. Torralba, "Contextual guidance of attention and eye movements in real-world scenes: the role of global features in object search," *Psychol. Rev.*, 2006.
- [103] K. E. T. Judd, "Learning to predict where humans look," *IEEE 12th Int. Conf. Comput. Vis.*, pp. 2106–2113, 2009.
- [104] A. Torralba, "Modeling global scene factors in attention," *JOSA*, 2003.
- [105] X. H. Zhan, "Saliency Detection: A Spectral approach," *ieee Conf. Comput. Vis. pattern*

- Recognit.*, 2007.
- [106] N. Bruce and J. Tsotsos, “Saliency based on information maximization,” in *Advances in neural information processing systems*, pp. 155–162, 2006.
- [107] J. a Harel, “Graph-based visual saliency,” *Adv. Neural Inf. Process. Syst.*, pp. 545–552, 2007.
- [108] S. a He, “Exemplar-driven top-down saliency detection via deep association,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 5723–5732, 2016.
- [109] C. S. Francesca Murabito, “Top-Down Saliency Detection Driven by Visual Classification,” *arXiv Prepr. arXiv*, pp. 1709–5307, 2017.
- [110] R. a Bharath, “Scalable scene understanding using saliency-guided object localization,” *Control Autom. (ICCA), 2013 10th IEEE Int. Conf.*, pp. 1503–1508, 2013.
- [111] X. a Hou, “Image signature: Highlighting sparse salient regions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 194–201, 2012.
- [112] S. a.-M. Goferman, “Context-aware saliency detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1915–1926, 2012.
- [113] R. E. a, “Segmenting salient objects from images and videos,” *Eur. Conf. Comput. Vis.*, pp. 366–379, 2010.
- [114] L. Wald, “Some terms of reference in data fusion,” *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1190–1193, 1999.
- [115] J. J. Clarke and A. L. Yuille, “Data fusion for sensory information processing,” *Boston, MA Kluwer Acad. doi*, vol. 10, pp. 971–978, 1990.
- [116] L. A. Zadeh, “Probability measures of fuzzy events,” *J. Math. Anal. Appl.*, vol. 23, no. 2, pp. 421–427, 1968.
- [117] G. Shafer, *A mathematical theory of evidence*, vol. 42. Princeton university press, 1976.
- [118] F. Smarandache and J. Dezert, “Applications and Advances of DS_mT for Information Fusion, Collected Works.” American Research Press, Rehoboth, 2004.
- [119] A. P. Dempster, “Upper and lower probabilities induced by a multivalued mapping,” in *Classic Works of the Dempster-Shafer Theory of Belief Functions*, Springer, pp. 57–72, 2008.
- [120] A. Samet, E. Lefevre, and S. Ben Yahia, “Belief function classification with conflict management: application on forest image,” in *Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on*, pp. 14–20, 2014.
- [121] C. Lian, “Information fusion and decision-making using belief functions: application to therapeutic monitoring of cancer.” Université de Technologie de Compiègne, 2017.

- [122] I. Hammami, “Fusion of heterogeneous remote sensing images by credibilist methods Imen Hammami,” 2018.
- [123] L. Comtet, “Advanced Combinatorics: The Art of Finite and Infinite Expansions, enlarged ed. Reidel, Dordrecht,” *Math. Rev. MR460128 Zentralblatt MATH*, vol. 283, 1974.
- [124] R. Dedekind, “Über Zerlegungen von Zahlen durch ihre grössten gemeinsamen Theiler,” in *Fest-Schrift der Herzoglichen Technischen Hochschule Carolo-Wilhelmina*, Springer, pp. 1–40, 1897.
- [125] Quellec, Gwénolé, et al. "Multimodal medical case retrieval using the dezert-smarandache theory." *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008.
- [126] P. Smets and R. Kennes, “The transferable belief model,” *Artif. Intell.*, vol. 66, no. 2, pp. 191–234, 1994.
- [127] P. Smets, “Data fusion in the transferable belief model,” in *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, vol. 1, pp. PS21-PS33, 2000.
- [128] D. Dubois and H. Prade, “Fuzzy sets and systems: Theory and applications,” *Am. Math. Soc.*, vol. 7, no. 3, pp. 603–612, 1982.
- [129] R. R. Yager, “On the Dempster-Shafer framework and new combination rules,” *Inf. Sci. (Ny)*, vol. 41, no. 2, pp. 93–137, 1987.
- [130] R. R. Yager, “Hedging in the combination of evidence,” *J. Inf. Optim. Sci.*, vol. 4, no. 1, pp. 73–81, 1983.
- [131] J. Dezert and F. Smarandache, “Proportional conflict redistribution rules for information fusion,” *Adv. Appl. DSMT Inf. Fusion-Collected Work.*, vol. 2, pp. 3–68, 2006.
- [132] P. Djiknavorian and D. Grenier, “Reducing DSMT hybrid rule complexity through optimization of the calculation algorithm,” *Adv. Appl. DSMT Inf. Fusion*, p. 365, 2006.
- [133] A. Nassim, “Développement de modèles de fusion et de classification contextuelle d’images satellitaires par la théorie de l’évidence et la théorie du raisonnement plausible et paradoxal.”, 2009.
- [134] A. Elhassouny, S. Idbraim, A. Bekkarri, D. Mammass, and D. Ducrot, “Multisource Fusion / Classification Using ICM and DSMT with New Decision Multisource Fusion / Classification Using ICM and DSMT with New Decision Rule,” no. September, 2012.
- [135] F. Smarandache and J. Dezert, “*Advances and Applications of DSMT for Information Fusion, Vol. IV: Collected Works*”. Infinite Study, 2015.
- [136] T. Lee, J. A. Richards, and P. H. Swain, “Probabilistic and evidential approaches for

- multisource data analysis,” *IEEE Trans. Geosci. Remote Sens.*, no. 3, pp. 283–293, 1987.
- [137] E.-H. Zahzah, “Contribution à la représentation des connaissances et à leur utilisation pour l’interprétation automatique des images satellite.” Toulouse 3, 1992.
- [138] H. Kim and P. H. Swain, “Evidential reasoning approach to multisource-data classification in remote sensing,” *IEEE Trans. Syst. Man. Cybern.*, vol. 25, no. 8, pp. 1257–1265, 1995.
- [139] N. Milisavljević and I. Bloch, “Improving mine recognition through processing and Dempster-Shafer fusion of multisensor data,” *Comput. Intell. Recognition, Tech. Appl.*, vol. 17, pp. 319–343, 2005.
- [140] A.-L. Joussetme, É. Bossé, and A. Jouan, “Analysing an identity information fusion algorithm based on evidence theory,” DEFENCE RESEARCH AND DEVELOPMENT CANADA VALCARTIER (QUEBEC), 2004.
- [141] H. Chu, X. Guisong, and S. Hong, “SAR images classification method based on Dempster-Shafer theory and kernel estimate,” *J. Syst. Eng. Electron.*, vol. 18, no. 2, pp. 210–216, 2007.
- [142] I. Hammami, J. Dezert, and G. Mercier, “Kohonen-Based Credal Fusion of Optical and Radar Images for Land Cover Classification,” in *2018 21st International Conference on Information Fusion (FUSION)*, pp. 1623–1630, 2018.
- [143] M. Liu and S. Chen, “SAR target configuration recognition based on the Dempster–Shafer theory and sparse representation using a new classification criterion,” *Int. J. Remote Sens.*, pp. 1–19, 2019.
- [144] F. Yang, H. Wei, and P. Feng, “A hierarchical Dempster-Shafer evidence combination framework for urban area land cover classification,” *Measurement*, 2018.
- [145] W. Ouerghemmi, A. Le Bris, N. Chehata, and C. Mallet, “A TWO-STEP DECISION FUSION STRATEGY: APPLICATION TO HYPERSPECTRAL AND MULTISPECTRAL IMAGES FOR URBAN CLASSIFICATION,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 42, 2017.
- [146] H. A. Moghaddam and S. Ghodrathnama, “Toward semantic content-based image retrieval using Dempster–Shafer theory in multi-label classification framework,” *Int. J. Multimed. Inf. Retr.*, vol. 6, no. 4, pp. 317–326, 2017.
- [147] R. Khedam *et al.*, “Improvement of Land Cover Map from Satellite Imagery using DST and DSMT Improvement of Land Cover Map from Satellite Imagery Using DST and DSMT,” no. September, 2006.
- [148] C. Lian, S. Ruan, T. Dencœux, and P. Vera, “Outcome prediction in tumour therapy based on Dempster-Shafer theory,” in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pp. 63–66, 2015.

- [149] C. Lian, S. Ruan, T. Dencœux, F. Jardin, and P. Vera, “Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction,” *Med. Image Anal.*, vol. 32, pp. 257–268, 2016.
- [150] S. Razi, M. R. Karami, and J. Ghasemi, “A novel method for classification of BCI multi-class motor imagery task based on Dempster-Shafer theory,” *Inf. Sci. (Ny)*, 2019.
- [151] I. Bloch, “Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account,” *Pattern Recognit. Lett.*, vol. 17, no. 8, pp. 905–919, 1996.
- [152] Y. M. Zhu, L. Bentabet, O. Dupuis, V. Kaftandjian, D. Babot, and M. Rombaut, “Automatic determination of mass functions in Dempster-Shafer theory using fuzzy c-means and spatial neighborhood information for image segmentation,” *Opt. Eng.*, vol. 41, no. 4, pp. 760–771, 2002.
- [153] M. Wafa and E. Zagrouba, “Tumor extraction from multimodal MRI,” in *Computer Recognition Systems 3*, Springer, pp. 415–422, 2009.
- [154] S. Corgne, L. Hubert-Moy, J. Dezert, and G. Mercier, “Land cover change prediction with a new theory of plausible and paradoxical reasoning,” in *Proc. of Fusion*, pp. 8–11, 2003.
- [155] A. Bouakache, A. Belhadj-Aissa, and G. Mercier, “*Satellite image fusion using Dezert-Smarandache theory*”. Infinite Study, 2009.
- [156] A. Bouakache, “Fusion des images satellitaires par la théorie d’évidence et la théorie du raisonnement plausible at paradoxal.” 2005.
- [157] A. Elhassouny, S. Idbraim, A. Bekkari, D. Mammass, and D. Ducrot, “*Application of DS_mT-ICM with Adaptive decision rule to supervised classification in multisource remote sensing*”. JOURNAL OF COMPUTING, VOLUME 5, ISSUE 1, ISSN 2151-96, JANUARY 2013.
- [158] A. Elhassouny, S. Idbraim, and D. Ducrot, “Change Detection by Fusion / Contextual Classification based on a Hybrid DS_mT Model and ICM with Constraints Change Detection by Fusion / Contextual Classification based on a Hybrid DS_mT Model and ICM with Constraints,” no. November, 2011.
- [159] J.-C. Okaingni, S. Ouattara, A. Kouassi, W. J. Vangah, A. K. Koffi, and A. Clement, “*Modeling and Characterization of Vegetation, Aquatic and Mineral Surfaces Using the Theory of Plausible and Paradoxical Reasoning from Satellite Images: Case of the Toumodi-Yamousoukro-Tiébissou Zone in V Baoulé (Côte d’Ivoire)*”. Infinite Study, 2017.
- [160] B. Pannetier and J. Dezert, “Multiple ground target tracking and classification with DS_mT,” *Adv. Appl. DS_mT Inf. Fusion*, p. 337, 2015.

- [161] G. Quellec, M. Lamard, G. Cazuguel, C. Roux, and B. Cochener, "Multimodal medical case retrieval using the deZert-smarandache theory," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 394–397, 2008.
- [162] A.-L. Joussetme, A. Martin, and P. Maupin, "Gestion de l'information paradoxale contrainte par des requêtes pour la classification de cibles dans un réseau de capteurs multi-modalités," in *Colloque Systèmes Complexes d'Information et de Gestion des Risques pour l'Aide à la Décision, Brest, France*, pp. 24–25, 2008.
- [163] N. Abbas *et al.*, "The effective use of the DSMT for multi-class classification," 2017.
- [164] N. Abbas, Y. Chibani, B. Hadjadji, and Z. A. Omar, "A DSMT Based System for Writer-Independent Handwritten Signature Verification." *2016 19th International Conference on Information Fusion (FUSION)*. IEEE, 2016.
- [165] J. M. J. Pluim and M. Viergever, "Image registration by maximization of combined mutual information and gradient information," *Proc. Med. Image Comput. Comput. Interv.*, pp. 103–129, 2000.
- [166] & B. Fischler M. A. and R. C., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, pp. 381–395, 1981.
- [167] L. S. S. N. & A. FBeckouche S., "Robust outliers detection in image point matching," *Comput. Vis. Work. (ICCV Work. IEEE Int. Conf.)*, pp. 180–187, 2011.
- [168] H. Anzid, G. Le Goic, A. Bekkari, A. Mansouri, and D. Mammass, "An automatic filtering algorithm for SURF-based registration of remote sensing images," in *Advanced Technologies for Signal and Image Processing (ATSIP), 2017 International Conference on*, pp. 1–7, 2017.
- [169] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. neural networks*, vol. 10, no. 5, pp. 988–999, 1999.
- [170] L. Chapel, "Maintenir la viabilité ou la résilience d'un système : les machines à vecteurs de support pour rompre la malédiction de la di-mensionnalité," Thesis of doctorate, university of Blaise Pascal - Clermont II, 2007.
- [171] S. Aseervatham, "Apprentissage à base de Noyaux Sémantiques pour le traitement de données textuelles," Université Paris-Nord-Paris XIII, 2007.
- [172] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, vol. 37, no. 1, pp. 35–46, 1991.
- [173] A. Bekkari, S. Idbraim, A. Elhassouny, D. Mammass, M. El Yassa, and D. Ducrot, "Spectral and Spatial Classification of High Resolution Urban Satellites Images using Haralick features and SVM with SAM and EMD distance Metrics," vol. 46, no. 11, pp.

28–37, 2012.

- [174] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Advances in spectral-spatial classification of hyperspectral images,” *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
- [175] T. T. Herbert Bay Andreas Ess and L. Van Gool, “Speeded-up robust features (SURF),” *Comput. Vis. image Underst.*, vol. 110(3), pp. 346–359, 2008.
- [176] A. I. R. J. Peters, “Components of bottom-up gaze allocation in natural images,” *Vision Res.*, pp. 2397–2416, 2005.
- [177] J.-G. a.-S. Yu, “A computational model for object-based visual saliency: Spreading attention along gestalt cues,” *IEEE Trans. Multimed.*, pp. 273–286, 2016.
- [178] X. a Hou, “Saliency detection: A spectral residual approach,” *Comput. Vis. Pattern Recognition, CVPR’07*, pp. 1–8, 2007.
- [179] E. Rahtu, J. Kannala, M. Salo, and J. Heikkil, “Segmenting Salient Objects from Images and Videos,” *Eur. Conf. Comput. Vis.*, pp. 366–379, 2010.
- [180] C. Samson, “Contribution à la classification des images satellitaires par approche variationnelle et équation aux dérivées partielles,” university of Nice-sophia Antipolis, 2000.
- [181] S. F. Gull, “Developments in maximum-entropy data analysis,” in *Maximum Entropy and Bayesian Methods*, J. Skilling, Ed. Kluwer Academic, Dordrecht, pp. 53–71, 1989.
- [182] P. Kamavisdar, S. Saluja, and S. Agrawal, “A survey on image classification approaches and techniques,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 1, pp. 1005–1009, 2013.
- [183] J. Weston and C. Watkins, “Multi-class support vector machines,” Citeseer, 1998.
- [184] J. H. Friedman, “Another approach to polychotomous classification,” *Tech. Report, Stat. Dep. Stanford Univ.*, 1996.
- [185] V. N. Vapnik, “Estimation of dependencies based on empirical data Springer,” *New York*, 1982.
- [186] O. Bousquet, “Introduction au Support Vector Machines (SVM),” *Cent. Math. applied, Polytech. Sch. Palaiseau*, 2001.
- [187] A. Bekkari, M. El, and S. Idbraim, “SVM Classification of Urban High-Resolution Imagery Using Composite Kernels and Contour Information SVM Classification of Urban High-Resolution Imagery Using Composite Kernels and Contour Information,” no. April, 2013.
- [188] S. Majumdar and J. Bharadwaj, “Feature Level Fusion of Multimodal Images Using Haar Lifting Wavelet Transform,” *World Acad. Sci. Eng. Technol. Int. J. Comput. Electr.*

Autom. Control Inf. Eng., pp. 1023–1027, 2014.

- [189] A. E. D. M. M. E. Y. D. D. Aissam Bekkari Soufian Idbraim, “spectral and spatial information classification of high resolution urban satellite Images using haralick features and svm with sam and EMD distance metrics,” *Int. J. Comput. Appl.*, pp. 975–8887, 2012.
- [190] S. Knerr, L. Personnaz, and G. Dreyfus, “Single-layer learning revisited: a stepwise procedure for building and training a neural network,” in *Neurocomputing*, Springer, pp. 41–50, 1990.
- [191] A. Bekkari, S. Idbraim, A. Elhassouny, D. Mammass, D. Ducrot, and others, “Svm and Haralick features for classification of high resolution satellite images from urban areas,” in *International Conference on Image and Signal Processing*, pp. 17–26, 2012.
- [192] J. G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco and and P. G. Inglada, E. Christophe, J. Chanussot, “Decision fusion for the classification of hyperspectral data,” *IEEE Trans. Geos. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, 2009.
- [193] D. A. Landgrebe., “Hyperspectral image data analysis,” *IEEE Signal processing magazine*, vol. 19, no 1, p. 17-28, 2002.
- [194] J. Al-doski, S. B. Mansorl, and H. Z. M. Shafri, “Image classification in remote sensing,” *Dep. Civ. Eng. Fac. Eng. Univ. Putra, Malaysia*, 2013.
- [195] L. Yu, A. Porwal, E.-J. Holden, and M. C. Dentith, “Towards automatic lithological classification from remote sensing data using support vector machines,” *Comput. Geosci.*, vol. 45, pp. 229–239, 2012.
- [196] L. Wu, Y. Wang, and J. Long, “An Unsupervised Change Detection Approach for Remote Sensing Image Using SURF and SVM,” in *Chinese Conference on Pattern Recognition*, pp. 537–551, 2016.
- [197] J. Dukart *et al.*, “Meta-analysis based SVM classification enables accurate detection of Alzheimer’s disease across different clinical centers using FDG-PET and MRI,” *Psychiatry Res. Neuroimaging*, vol. 212, no. 3, pp. 230–236, 2013.
- [198] C. Hinrichs *et al.*, “Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset,” *Neuroimage*, vol. 48, no. 1, pp. 138–149, 2009.
- [199] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, and A. D. N. Initiative, “Multimodal classification of Alzheimer’s disease and mild cognitive impairment,” *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.
- [200] J. Dukart *et al.*, “Combined evaluation of FDG-PET and MRI improves detection and differentiation of dementia,” *PLoS One*, vol. 6, no. 3, p. e18111, 2011.
- [201] J. F. G. Molina, L. Zheng, M. Sertdemir, D. J. Dinter, S. Schönberg, and M. Rädle, “Incremental learning with SVM for multimodal classification of prostatic

- adenocarcinoma,” *PLoS One*, 2014.
- [202] Z. Zhou, Y. Du, N. L. Thomas, and E. J. Delp, “Multimodal eye recognition,” *Mob. Multimedia/Image Process. Secur. Appl.*, pp. 770–806, 2010.
- [203] G. S. Kumar and C. J. Devi, “A Multimodal SVM Approach for Fused Biometric Recognition,” *Int. J. Comput. Sci. Inf. Technol.*, pp. 3327–3330, 2014.
- [204] A. Apatean, A. Rogozan, and A. Benschair, “Svm-based obstacle classification in visible and infrared images,” no. Eusipco, pp. 293–297, 2009.
- [205] R. Brehar, C. Vancea, and T. Marit, “Pedestrian Detection in the Context of Multiple-Sensor Data Alignment for Far-Infrared and Stereo Vision Sensors,” *2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*. IEEE, pp. 385–392, 2015.
- [206] C. Pomrehn, D. Klein, A. Kolb, P. Kaul, and R. Herpers, “Supervised classification of monomodal and multimodal hyperspectral data in vibrational microspectroscopy: A comprehensive comparison,” *Chemom. Intell. Lab. Syst.*, 2018.
- [207] Y. Mustafa, K. M. Clawson, and C. Bowerman, “Saving Cultural Heritage with Digital Make- Believe : Machine Learning and Digital Techniques to the Rescue,” no. July, 2017.