



THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ BORDEAUX 1

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET
D'INFORMATIQUE

Par **Mohamed KOUBAA**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

**Tatouage robuste de vidéo basé sur la notion de
régions d'intérêt**

Soutenu le : 23 novembre 2010

Après avis des rapporteurs :

M. Jean-Luc DUGELAY
Mme Dorra SELLAMI

Professeur, Eurecom Sophia Antipolis France
Mdc HDR, Université de Sfax Tunisie

Devant la commission d'examen composée de :

M. Mohamed ABID
M. Jean-Luc DUGELAY
Mme. Dorra SELLAMI
M. Chokri BEN AMAR
M. Henri N ICOLAS

Professeur, Université de Sfax
Professeur, Eurecom Sophia Antipolis
Mdc HDR, Université de Sfax
Mdc HDR, Université de Sfax
Professeur, Université Bordeaux 1

Président
Rapporteur
Rapporteur
co-Directeur de Thèse
Directeur de Thèse

Remerciements

Je tiens tout particulièrement à remercier Mr. Chokri BEN AMAR, Maître de conférences et Directeur Adjoint à l'Ecole Nationale d'Ingénieurs de Sfax (ENIS), qui a co-dirigé cette thèse, pour sa disponibilité, son soutien et ses conseils. Mr. BEN AMAR a beaucoup œuvré pour la mise en valeur de cette thèse. Sa disponibilité illimitée et son suivi minutieux de toutes les particularités de mon travail méritaient d'être ardemment remerciés. Il n'a pas épargné d'effort à m'assurer un environnement de travail exceptionnel. J'ai beaucoup appris à ses côtés et je ne parviens jamais à lui adresser la reconnaissance qu'il méritait.

Mes vives remerciements s'adressent aussi à Mr. Henri NICOLAS, Professeur à l'Université de Bordeaux 1 en France, responsable de l'équipe Image et Son au Laboratoire Bordelais de Recherche en Informatique (LaBRI) et directeur de la MiaGE, pour son suivi minutieux, ses conseils constructifs et ses relectures de ce manuscrit qui m'ont permis de renforcer la rigueur de ce travail. Les discussions fructueuses et les réunions profitables qui nous ont unis m'ont fort servi. Bien plus qu'un directeur de thèse, Mr. NICOLAS restera un grand frère dont la culture et la profondeur d'esprit m'intriguent perpétuellement. Je lui suis considérablement reconnaissant et je manque d'expressions de remerciements digne de tout ce qu'il m'a donné durant ma thèse.

Je souhaite exprimer toute ma gratitude envers les membres du jury qui ont bien voulu consacrer à cette thèse une partie de leur temps extrêmement précieux.

Je commence par remercier Mr. Mohamed ABID, Professeur et Directeur du laboratoire "Computer & Embedded System" à l'ENIS, pour avoir accepté de présider ce jury.

Je remercie également Mr. Liming CHEN, Professeur à l'Ecole Centrale de Lyon, qui a bien accepté de m'honorer par sa présence en qualité d'examineur de ma thèse, pour

l'intérêt qu'il a porté à mon travail et pour les efforts fournis pour pouvoir assister à ma soutenance.

Je remercie en particulier ceux qui ont accepté la lourde charge d'être les rapporteurs de ce travail. J'ai apprécié leur lecture attentive de mon rapport de thèse et leurs remarques pertinentes. Mes remerciements s'adressent donc à Mr. Jean-Luc DUGELAY, Professeur à EURECOM Sophia Antipolis, pour ses remarques constructives, son œil critique et ses encouragements.

Je suis très content par la présence dans ce jury de Mme. Dorra SELLAMI, Maître de Conférences à l'ENIS, d'avoir été mon rapporteur de thèse. Sa présence dans mon jury est pour moi un honneur...

Je tiens aussi à exprimer toute ma reconnaissance au Professeur Adel Mohamed ALIMI, directeur de l'ENIS et directeur de l'unité de recherche "Groupe de Recherche sur les Machines Intelligentes (ReGIM)", qui m'a accueilli dans son unité de recherche et qui m'a fourni l'environnement nécessaire et les meilleures conditions pour mener à bien ce travail. Je le remercie aussi pour tous ses encouragements et ses conseils judicieux. D'ailleurs, je remercie toute la famille "ReGIM" pour tous les bons moments qu'on avait passé ensemble. Un grand Merci à Sodki, Mouda, Habib, Ali, BH, Zied, Maher, Nizar, Tarek, Walid, Mehdi, Olfa, ...

Je ne dois pas aussi oublier mes amis du LaBRI, Mohamed, Julien et Wily avec qui j'ai passé des agréables moments pendant mes séjours en France.

Il m'est aussi impossible de ne pas citer mes amis et frères Slim, Hatem, Nabil et Akram "switch" pour leur soutien morale et pour toute l'affection qu'ils m'ont donné.

Bien évidemment, je remercie mes parents et tous les membres de ma famille, en particulier ma femme et mon fils ADAM pour leur soutien permanent et surtout de m'avoir supporter pendant toutes ces longues et dures années.

Table des matières

Table des matières	iii
Table des figures	x
Liste des tableaux	xiv
Introduction générale	1
I Etat de l’art sur le tatouage	6
1 Généralités sur le tatouage	7
1.1 Introduction	7
1.2 Problématique générale	8
1.2.1 Piratage de documents	8
1.2.2 Impact économique	8
1.2.2.1 Situation en Europe	9
1.2.2.2 Situation en Tunisie	11
1.3 Aspect général du tatouage	11
1.3.1 Position du tatouage	11
1.3.1.1 Cryptographie	12
1.3.1.2 Stéganographie	12
1.3.2 Applications du tatouage	13

1.3.2.1	Droits d’auteurs	13
1.3.2.2	Authentification des documents	13
1.3.2.3	Traçabilité des documents	14
1.3.2.4	Indexation des documents	14
1.3.2.5	Empreinte digitale	14
1.3.3	Contraintes du tatouage	14
1.3.3.1	Capacité	15
1.3.3.2	Imperceptibilité	15
1.3.3.3	Robustesse	15
1.3.4	Attaques innocentes et malveillantes	16
1.4	Principes des schémas de tatouage	17
1.4.1	Domaines d’insertion	17
1.4.1.1	Domaine spatial	18
1.4.1.2	Domaine fréquentiel	18
1.4.1.3	Domaine multirésolution	19
1.4.2	Phase d’insertion	20
1.4.3	Phase de détection	21
1.5	Mesures de la qualité du document tatoué	22
1.5.1	Le PSNR	23
1.5.2	Le wPSNR	23
1.6	Conclusion	24
2	Tatouage vidéo	25
2.1	Introduction	25
2.2	Notion de vidéo	26
2.2.1	Définition	26
2.2.2	Différents formats vidéo	26
2.2.2.1	Formats non compressés	26

2.2.2.2	Formats compressés	28
2.2.3	Contraintes de la vidéo	29
2.3	Tatouage : de l'image à la vidéo	30
2.3.1	Techniques de tatouage d'image	30
2.3.2	Limites des schémas de tatouage d'images dans le contexte de la vidéo	32
2.4	Classification des schémas de tatouage vidéo	32
2.4.1	Schémas dérivés du tatouage d'images fixes	32
2.4.2	Schémas spatio-temporels	33
2.4.3	Schémas purement temporels	33
2.4.4	Tatouage de flux compressés	34
2.5	Motivations	35
2.5.1	Introduction de la notion de régions d'intérêts	35
2.5.2	Notion de mosaïque	35
2.5.3	Zone d'ombre en mouvement	36
2.6	Conclusion	36
II Exploitation des mosaïques pour le tatouage		37
3 Création d'images mosaïques		38
3.1	Principe et hypothèses de base	38
3.2	Techniques de recalage des images vidéo	40
3.2.1	Les modèles de recalage	40
3.2.1.1	Le modèle Translationnel	40
3.2.1.2	Le modèle Affine	40
3.2.1.3	Le modèle Homographique	41
3.2.1.4	Le modèle Quadratique	42
3.2.2	Choix du modèle de recalage	42
3.2.2.1	Choix par défaut	42

3.2.2.2	Choix adaptatif du modèle de recalage	42
3.3	Méthodes d'estimation des paramètres de recalage	43
3.3.1	Block Matching	43
3.3.1.1	Définition	43
3.3.1.2	Problèmes	44
3.3.2	Flot Optique	44
3.3.2.1	Définition et principe	44
3.3.2.2	Problèmes	45
3.3.3	La factorisation SVD (Singular Value Decomposition)	46
3.3.3.1	Définition	46
3.3.3.2	Propriétés	47
3.3.3.3	Utilités de SVD	47
3.3.4	ICP : Iterative Closest Point	48
3.3.4.1	Principe	48
3.4	Approche retenue	48
3.4.1	Principe général	48
3.4.2	Le modèle de Recalage choisi : le modèle Affine	49
3.4.3	Estimation des paramètres de recalage	49
3.4.3.1	Choix des pixels correspondants	50
3.4.3.2	Alignement grossier	51
3.4.3.3	Alignement fin	52
3.4.4	Réécriture de la mosaïque	53
3.5	Conclusion	54
4	Approche de tatouage vidéo basée sur la notion d'image mosaïque	55
4.1	Introduction	55
4.2	Aperçu général de l'approche	55
4.2.1	Sélection des régions d'intérêt	56

4.2.1.1	Localisation des points physiques similaires	57
4.2.1.2	Estimation de l'occurrence des pixels	57
4.2.1.3	Estimation de la vitesse de la caméra	58
4.2.2	Phase d'insertion	59
4.2.3	Phase de détection	62
4.3	Estimation du seuil de détection	65
5	Expérimentations et résultats	66
5.1	Introduction	66
5.2	Expérimentation sans attaque	66
5.3	Évaluation face à l'attaque de Collusion	67
5.4	Évaluation face à la compression MPEG4	70
5.5	Expérimentation face à la suppression d'images	72
5.6	Expérimentation dans le cas d'une caméra fixe	74
5.7	Positionnement par rapport à d'autres travaux	75
5.8	Conclusion	77
III	Exploitation des zones d'ombres pour le tatouage	79
6	Généralités sur les zones d'ombre	80
6.1	Introduction	80
6.2	Modélisation des ombres	81
6.3	Domaines de recherche	87
6.3.1	Imagerie aérienne	87
6.3.2	Détection des ombres dans le cadre général	87
6.3.3	Détection d'ombre en vidéo surveillance	88
6.4	Détection des zones d'ombre en mouvement	89
6.4.1	Schéma de l'approche adoptée	89

6.4.1.1	Calcul du masque de détection de changement initial (CDM)	89
6.4.1.2	Calcul des contours Canny	92
6.4.1.3	Extraction des contours du CDM	92
6.4.1.4	Intégration de plusieurs frames	93
6.4.1.5	Détection des contours Canny en mouvement	93
6.4.1.6	Formation des points de départ de la croissance des régions	94
6.4.1.7	Dilatation morphologique	95
6.4.1.8	Croissance de régions	95
6.4.1.9	Localisation de la zone d'ombre portée sur l'image originale	96
6.5	Conclusion	96
7	Approche de tatouage basée sur les zones d'ombre	98
7.1	Introduction	98
7.2	Schéma général de l'approche	99
7.2.1	Phase d'insertion	100
7.2.1.1	Décomposition en scènes	100
7.2.1.2	Etude des conditions d'illumination	101
7.2.1.3	Insertion de la marque	109
7.3	Phase de détection	110
7.4	Phase d'attaque	110
7.5	Conclusion	111
8	Résultats expérimentaux	114
8.1	Introduction	114
8.2	Choix de la valeur du PSNR	114
8.3	Résultats Expérimentaux	115
8.3.1	Avant adaptation de la marque	115

8.3.2	Résultats d'insertion à différents PSNRs en tenant compte des conditions d'éclairage	116
8.3.3	Attaques par compression	117
8.3.3.1	Résultats de compression avec débit constant et PSNR variable	119
8.3.3.2	Résultats de compression à différents débits et PSNR constant	120
8.4	Conclusion	121
	Conclusion et Perspectives	121
	Bibliographie	125

Table des figures

1.1	Exemples de tatouage visible	15
1.2	Problématique en tatouage	16
1.3	Les différents types d’attaques s’opposant au tatouage	17
1.4	Exemple d’une transformée de fourier	18
1.5	Exemple d’une transformée d’ondelette à 2 niveaux. (a) Image Lena. (b) Image Lena transformée.	19
1.6	Exemple de scalabilité spatiale avec différents facteurs de résolution.	19
1.7	Phase d’insertion de la marque	21
1.8	Phase de détection de la marque	22
2.1	Exemple de patches	31
3.1	Illustration du Bloc Matching	44
3.2	Exemple d’une détection d’un champ dense de vecteurs du mouvement à partir de trois images	46
3.3	Ambiguïté du flot optique	46
3.4	Les différentes étapes de recalage.	50
3.5	Choix des pixels correspondants entre deux images	50
4.1	Diagramme global de l’approche de tatouage proposée	56
4.2	Image mosaïque issue de la séquence de ”Stefan”	58
4.3	Projection des pixels similaires sur le même point au niveau de la mosaïque	58

4.4 Compensation du mouvement entre deux images successives	59
4.5 Schéma d'insertion de la marque	61
4.6 Fonction linéaire	62
4.7 Création des images marques	63
4.8 Schéma de détection de la marque	64
5.1 Séquence de Stefan. Valeurs de corrélation avec différentes clés. (a) utilisant le filtre de Wiener. (b) sans le filtre de Wiener.	67
5.2 Corrélation globale après l'attaque de collusion utilisant les mêmes para- mètres de mouvement dans les phases d'insertion et de collusion	69
5.3 Test de collusion utilisant la meilleure estimation de mouvement dans la phase d'insertion de la marque	69
5.4 Tests de collusion utilisant la meilleure estimation de mouvement dans la phase d'attaque	70
5.5 Séquence de Thabor : valeurs de corrélation après une compression MPEG4 à deux débits différents	71
5.6 Séquence de Stefan : valeurs de corrélation après une compression MPEG4 à deux débits différents	71
5.7 Exemple d'images de vidéos compressées avec mpeg4 à un débit de 200Kbits/s	72
5.8 Séquence de Stefan : valeurs de corrélation utilisant un psnr constant et un psnr dépendant de la vitesse de la caméra	72
5.9 Valeurs de corrélation après la suppression des frames	74
5.10 Détection des frames supprimés et resynchronisation pour la détection . . .	75
5.11 Séquence de Stefan : corrélation après la suppression d'un frame sur deux	76
5.12 Séquence de "Silent" : valeurs de corrélation après une compression mpeg4	76
5.13 Séquence de "Paris" : valeurs de corrélation après une compression mpeg4 .	76
6.1 Zones d'ombre dans le cas d'une source de lumière ponctuelle	82
6.2 Zones d'ombre dans le cas d'une source de lumière étendue	83

6.3	Zones d'ombre dans le cas d'une scène avec deux sources de lumière	84
6.4	Variation de la luminance dans une ombre portée	86
6.5	Algorithme de détection de l'ombre portée dans une séquence vidéo	90
6.6	Exemple de contour de Canny.	92
6.7	Les contours canny en mouvement.	94
6.8	SRE_{153} du frame 153 de la séquence "Silent"	94
6.9	Résultat de la dilatation morphologique.	95
6.10	Élément structurant B_2	95
6.11	Masque de la région d'ombre détectée	96
6.12	Masque de la région d'ombre détectée.	97
7.1	Influence de la lumière sur un même objet	99
7.2	Etapas de la phase d'insertion de la marque	100
7.3	Effet de la variation de la source de la lumière sur l'intensité de l'ombre engendré	101
7.4	Localisation de la zone d'ombre à deux instants différents. (a) Image de référence prise à l'instant t . (b) Image étudiée prise à l'instant $t + n$	102
7.5	Exemples de scènes fortement éclairées	103
7.6	Exemples de scènes moyennement éclairées	103
7.7	Exemples de scènes faiblement éclairées	103
7.8	Choix de la zone d'ombre à étudier. (a) Zone d'ombre totale. (b) Partie de la zone d'ombre portée.	105
7.9	Classement FCM en trois classes(nombre de pixels=500)	106
7.10	Classement FCM en trois classes (nombre de pixels=4096(bloc de 64×64))	107
7.11	Schéma d'insertion de la marque	109
8.1	Corrélation avant et après marquage	116
8.2	Représentation de la corrélation en fonction de porteuses aléatoires	116
8.3	Compression à différents débits et un PSNR=35db	117

8.4	Compression à différents PSNR et un débit de 500kbps	117
8.5	Résultats de corrélation avec PSNR variable dans le cas de la séquence 3 .	118
8.6	Résultats de corrélation avec PSNR variable dans le cas de la séquence 4 .	118
8.7	Résultats de corrélation avec PSNR variable dans le cas de la séquence 5 .	118
8.8	Résultats de corrélation avec PSNR variable dans le cas de la séquence 6 .	119
8.9	Résultats de corrélation après compression à 1000kbps avec PSNR variable dans le cas de la séquence 3	119
8.10	Résultats de corrélation après compression à 1000kbps avec PSNR variable dans le cas de la séquence 5	120
8.11	Résultats de corrélation après compression à différents débits et PSNR=24db dans le cas de la séquence 3	120
8.12	Résultats de corrélation après compression à Débit variable et PSNR=21db dans le cas de la séquence 5	121

Liste des tableaux

1.1	Pertes dues au piratage dans les industries créatives européennes (2008) . . .	9
1.2	Pertes liées au piratage en Europe (2008-2015) – Scénario basé sur les tendances en matière “ d’échange de fichiers”	10
1.3	Pertes d’emplois en Europe (2008-2015) – Scénario basé sur les tendances du “Trafic IP des particuliers”	11
5.1	Comparaison de la robustesse par rapport à d’autres méthodes de tatouage	77
7.1	Matrice d’appartenance des séquences vidéo aux trois classes engendrées .	112
7.2	Résultats du changement du nombre d’entrées du classifieur FCM avec $c=3$	113
7.3	Résultats du changement du nombre d’entrées du classifieur FCM avec $c=4$	113
8.1	Calcul du psnr en fonction des conditions d’éclairage	115

Introduction générale

Contexte général

On assiste aujourd'hui à une convergence totale vers un monde numérique où toute information est numérisée, véhiculée et traitée au moyen de techniques et d'outils très développés. Le développement des réseaux à haut débit, notamment Internet, a facilité la transmission et le partage de l'information. C'est aussi grâce aux nouvelles normes de compression qui ne cessent d'évoluer ces dernières années et qui permettent de réduire d'une manière remarquable la taille des données. En conséquence, il est devenu très aisé de gérer les données volumineuses en terme de stockage et de traitement. En effet, il est possible maintenant de faire des vidéoconférences en temps réel, de regarder des films depuis un serveur distant, d'échanger par e-mail des fichiers de taille importante, etc. Les terminaux de communication ont beaucoup évolués ces dernières années avec l'apparition de nouvelles générations de téléphones portables 3g et 4G (iPhone, BlackBerry, ...), de nouveaux gadgets tels que l'iPad et les tablettes multimédias qui permettent une exploitation plus facile et rapide de l'information.

Cette évolution ne peut être que bénéfique puisqu'elle offre à l'utilisateur l'accès aux images, clips vidéo et d'autres services multimédias. Cependant, de nouveaux enjeux de sécurité se présentent. En effet, la sécurisation des données s'impose afin d'assurer :

- l'intégrité : s'assurer que les données n'ont pas été modifiées, détruits que ce soit d'une manière intentionnelle ou non.
- la confidentialité : s'assurer que les données ne soient accessibles que par les personnes autorisées.
- la traçabilité : suivre le cours de diffusion des données afin de détecter d'éventuelles

manipulations compromettantes et aussi de remonter à la personne responsable de cet acte.

Afin de garantir ces aspects en matière de sécurité, de nouvelles infrastructures informatiques (VPN, proxy, ...) ont été élaborées et de nouveaux logiciels de sécurité (anti-virus, anti-spyware, ...) ont été développés. Les techniques de cryptage des données ont aussi évoluées afin d'assurer une meilleure protection des données transmises sur un réseau en assurant leurs confidentialités. Certes, tous ces dispositifs apportent beaucoup à la sécurité, cependant ils restent insuffisants et limités devant certains contextes particuliers. En effet, dans certains cas, le document peut être visible par tout le monde de façon non cryptée. Par exemple, si une personne possédant un film sur DVD, il peut facilement le partager sur Internet ou même le commercialiser sans l'accord du propriétaire. Dans ce cas, il faut un moyen pour prouver d'abord la propriété de ce document par l'auteur et ensuite essayer de remonter à la personne qui était à l'origine de la distribution illégale du document. Le tatouage s'avère une technique de sécurisation des données qui permet de remédier à ces problèmes. Le principe de cette technique est d'introduire dans le document à protéger une signature invisible qui peut désigner le nom du propriétaire (protection du copyright) ou bien l'identité de la personne propriétaire du document. Cette identité permet de remonter à la personne qui était à l'origine d'une distribution illégale de plusieurs copies de ce document.

Nous nous intéressons dans cette thèse au tatouage vidéo. En effet, notre objectif est de contribuer à l'amélioration et la définition de nouvelles solutions de tatouage vidéo qui soient efficaces et robustes face aux attaques spécifiques à la vidéo. Pour cela, nous nous basons sur la notion de "régions d'intérêt" pour aboutir aux objectifs souhaités. L'idée est de définir, localiser et exploiter des régions d'intérêt dans la vidéo qui favorisent le tatouage. En effet, il est permis d'augmenter l'énergie du tatouage dans de telles régions dans le but d'augmenter la robustesse et ce, sans nuire à l'aspect visibilité du contenu. Ces régions d'intérêt sont localisées tenant compte de la faiblesse et des limites du Système Visuel Humain.

Pour aboutir à la localisation de telles régions, nous proposons, en premier lieu, la création d'image mosaïque qui représente, en une image, une vue globale et panoramique du fond fixe d'une scène donnée. A part le fait que l'image mosaïque fournit un résumé de

toute la scène, elle apporte, en plus, plusieurs informations et connaissances sur le contenu de celle-ci qui peuvent être exploitées pour le tatouage. En effet, et afin de construire la mosaïque, le mouvement de la caméra est compensé, ce qui permet de déterminer les paramètres de mouvement et par conséquent la vitesse de celle-ci. La compensation de mouvement permet aussi de localiser les différents endroits d'un point physique dans la scène. Exploitant cette information, nous envisageons de tatouer tous les points représentant le même point physique par la même marque afin d'échapper à une attaque spécifique à la vidéo qui est la collusion.

En deuxième lieu, nous nous intéressons aux zones d'ombres en mouvement. Ces régions représentent une source d'informations importante relative au contenu de la scène. En effet, à partir de ces régions, on peut estimer les conditions d'illumination d'une scène. Connaissant cette information, on envisage d'adapter l'énergie du tatouage à ces conditions d'illumination de façon à insérer plus d'énergie dans les scènes de faible éclairage tant que l'œil humain est moins sensible par rapport à ces régions.

Contributions

Le travail présenté dans ce mémoire présente deux contributions majeures dans le domaine du tatouage et qui sont :

- *Conception d'une méthode de tatouage basée sur les images mosaïques.* La création d'images mosaïque permet d'une part la localisation des points physiques présents dans une scène, et d'autre part l'estimation du mouvement de la caméra et en particulier sa vitesse. Le fait de localiser les différentes positions d'un point physique tout au long d'une scène permet de tatouer par la même marque tous les pixels représentant ce même point physique. De cette manière, notre schéma de tatouage présente une robustesse à l'attaque de collusion qui consiste à moyennner deux images successives dans le but de détruire la marque.

L'usage de la mosaïque pour optimiser la sélection des zones où l'oeil est moins susceptible de détecter des changement. Il s'agit des régions représentant une texture ou des contours. De telles régions sont localisées à travers l'application d'une filtre, Sobel par exemple. Or, il est nécessaire d'appliquer ce filtre sur toutes les images

de la scène où le tatouage sera introduit. Ce procédé est très coûteux en terme de temps d'exécution. Comme alternative pour optimiser cette tâche, on a proposé d'appliquer le filtre une seule fois sur l'image mosaïque, ensuite, et par projection inverse, on arrive à détecter ces zones tout au long de la scène.

L'estimation de la vitesse de la caméra nous a permis d'adapter l'énergie de la marque à celle-ci. En effet, l'idée consiste à introduire plus d'énergie dans les images où le mouvement de la caméra est assez rapide. Ce choix se base sur le fait que l'œil humain devient moins sensible au fur et à mesure que la vitesse de la caméra augmente. On se permet en quelque sorte de modifier légèrement le contenu des images là où le mouvement de la caméra est rapide.

– *Conception d'une méthode de tatouage basée sur les zones d'ombres en mouvement.*

L'idée consiste à introduire plus d'énergie dans les scènes contenant une source de lumière faible. Afin d'estimer les conditions d'illumination, nous proposons la détection des zones d'ombre en mouvement. En effet, l'intensité de l'ombre peut nous fournir certaines informations à propos des conditions d'illumination d'une scène et en particulier s'il s'agit d'une scène éclairée ou sombre. Après la détection d'une telle zone, la scène est classée selon trois catégories : scène faiblement éclairée, moyennement éclairée et fortement éclairée. Étant donné que la variation de la luminance entre une ombre faible et une ombre sombre n'est pas constante, nous avons adopté un classifieur flou pour décider de la classe d'une scène donnée.

Organisation du mémoire

Afin de détailler les contributions de cette thèse, nous avons choisi d'articuler notre étude autour de huit chapitres regroupés en trois parties.

La **première partie** présente un état de l'art sur le tatouage. Son premier chapitre précise certaines généralités explicitant le besoin et le recours à cette technique de sécurisation. Le deuxième chapitre présente un état de l'art sur les techniques de tatouage vidéo. Cependant, un bref aperçu sur les techniques de tatouage d'images fixes a été proposé afin de citer, de manière non exhaustive, les techniques adaptées au contexte de la vidéo.

Le premier chapitre de la **deuxième partie** explicite la notion d'image mosaïque et

présente les différentes étapes nécessaires pour aboutir à la création d'une telle image. Pour cela, différentes techniques de recalage et d'estimation de mouvement ont été présentées afin d'expliquer le choix de l'approche de mosaïcing réalisée. Quant au deuxième chapitre, il expose l'approche de tatouage basée sur les images mosaïque. Initialement, nous présentons dans ce chapitre la notion de régions d'intérêt et nous explicitons ceux qui ont été sélectionnées, à travers la création de la mosaïque, et exploitées pour le tatouage. Les différentes positions d'un même point physique tout au long d'une scène représente la première région d'intérêt traitée. L'avantage de celle-ci réside dans le fait que tatouer tous les pixels représentant le même point physique par la même marque augmente la robustesse de l'algorithme face à l'attaque de collusion. La deuxième région d'intérêt est l'ensemble d'images de la scène où le mouvement de la caméra est assez rapide. Exploitant la sensibilité de l'œil humain, nous nous permettons d'introduire plus d'énergie de la marque au niveau de ces images. Le troisième chapitre sera consacré à une étude expérimentale, là où on présente les différents tests effectués et résultats obtenus afin de valider notre approche.

La **troisième partie** sera consacrée à l'étude et l'exploitation des zones en mouvement pour le tatouage. Pour cela, le premier chapitre de cette partie expose d'une part certaines généralités sur les zones d'ombres et d'autre part une méthode pour la détection de telles zones dans une séquence vidéo. Le chapitre suivant propose une méthode complète de tatouage vidéo basée sur les zones d'ombre. Une étude des conditions d'illumination sera faite afin d'adapter la marque à la nature de l'éclairage dans une scène. Finalement, et dans le dernier chapitre, une étude expérimentale sera donnée dans le but de valider l'approche proposée et aussi de voir ses limites.

En conclusion, nous ferons une synthèse sur les apports présentés dans ce mémoire, et nous enchainons par la proposition de certaines perspectives de recherche.

Première partie

Etat de l'art sur le tatouage

Chapitre 1

Généralités sur le tatouage

1.1 Introduction

Le développement rapide des moyens de communication et de transmission, notamment le développement d'Internet et des réseaux haut débits, a facilité l'accès à l'information en général. La diffusion et le partage des données numériques sont devenus très aisés et à la portée de tout le monde. Certes, ceci ne peut être que bénéfique puisque l'accès à l'information est devenu instantané. Par contre, le problème de la sécurité se pose de plus en plus. En effet, les serveurs de partage de données, les réseaux P2P ont ouvert un champs très large aux utilisateurs malhonnêtes pour manipuler l'information d'une manière illégale. En plus, il s'est avéré que les techniques existantes de protection tels que les techniques de cryptage sont insuffisantes pour remédier aux nouveaux besoins à savoir la protection du copyright des propriétaires. Une technique assez récente, qui est le tatouage, a été proposée comme une solution qui peut remédier à ce problème. Il s'agit d'introduire dans un médium une signature invisible et qui résiste aux différentes manipulations possibles. Cette signature permet de prouver la propriété du document.

Dans le contexte de la vidéo, il s'agit d'introduire la signature dans les différentes images qui la composent. L'application des techniques de tatouage d'image dans le contexte de la vidéo reste possible. Sauf que la vidéo n'est pas simplement une succession d'images, il est plus intéressant de considérer la vidéo comme un signal $2D + t$ (tenant compte de la composante temporelle) et de concevoir des techniques de tatouage spécifiques à la vidéo

afin de garantir une meilleure robustesse.

Dans ce chapitre, nous allons d'abord expliciter d'avantage la problématique qui a mené à la proposition de cette technique pour justifier le recours à celle-ci. Ensuite, nous présenterons son contexte et sa position par rapport à la science de la cryptologie et puis nous énumérons ses propriétés, ses contraintes et surtout les domaines d'applications qui sont visés par cette nouvelle approche. Un aperçu sur les techniques de tatouage d'images fixes sera donné suivi d'un état d'une classification des techniques spécifiques à la vidéo qui est notre contexte dans cette thèse.

1.2 Problématique générale

1.2.1 Piratage de documents

Les documents numériques quel qu'ils soient sont soumis au problème de piratage. En effet, avec le développement rapide des moyens de communication, des moyens de sauvegarde, des techniques de partage (réseaux P2P, serveurs de téléchargement, ...) et de copie (graveur, flash disk), la procédure de piratage est devenue très simple et très facile à faire. Maintenant, avec les débits très élevés des connexions internet, n'importe qui peut télécharger de la musique ou des films en quelques minutes seulement et sans acquitter des droits des propriétaires.

1.2.2 Impact économique

Le piratage peut avoir une répercussion économique non négligeable. En effet, ce phénomène massif induit une forte perte des chiffres d'affaires et une destruction nette de milliers d'emplois.

Selon l'ESA (Entertainment Software Association) et l'IIPA (International Intellectual Property Alliance), près de 10 millions de jeux vidéo ont été piratés en décembre 2009. Un score élevé, mais qui ne donne pas une lecture complète de la situation du piratage dans le secteur du jeu vidéo. En effet, l'étude ne s'est intéressée qu'à quelques logiciels de P2P, mettant complètement de côté le téléchargement direct via des plates-formes comme

RapidShare ou MegaUpload.

1.2.2.1 Situation en Europe

Une étude réalisée en mois de Mars 2010 par TERA Consultants, un cabinet de conseil indépendant basé à Paris, a porté essentiellement sur l'effet de la piraterie numérique qui recouvre les différentes formes de piratage en ligne, dont l'échange de fichiers ("peer-to-peer P2P"). Cette étude avait pour objectif d'évaluer les conséquences économiques du piratage numérique sur les industries créatives et a visé en particulier son impact sur la perte de chiffre d'affaires et d'emplois.

Comme le montre le tableau 1.1, le piratage a induit, en 2008, un manque à gagner d'environ 10 milliards d'euros et plus de 185 000 destructions d'emplois en Europe dans les industries créatives étudiées.

TAB. 1.1: Pertes dues au piratage dans les industries créatives européennes (2008)

	Perte de chiffre d'affaires (M€)	Perte d'emplois
27 pays membres de l'UE	9.9	186.400
Royaume-Uni	1.4	39.000
France	1.7	31.400
Allemagne	1.2	34.000
Italie	1.4	22.400
Espagne	1.7	13.200

Dans les années à venir, l'essor du haut débit et la numérisation des produits issus des industries créatives vont s'accroître tandis que le piratage physique représentera une part de plus en plus faible du piratage global. En l'absence de mesures efficaces et inscrites dans la durée, ces tendances ne pourront qu'encourager l'expansion du piratage numérique en Europe. La présente étude propose deux scénarios d'estimations de pertes dues au piratage à l'horizon 2015, tous les deux sont basés sur les prévisions de Cisco System concernant le trafic Internet et sur l'hypothèse de l'absence de mesures visant à remédier au piratage. Le Scénario 1, qui repose sur l'hypothèse selon laquelle le piratage numérique va croître à un rythme similaire au trafic lié au partage de fichiers, aboutit à une estimation conservatrice

des pertes. Entre 2008 et 2015, l'échange de fichiers en Europe devrait progresser à un rythme annuel supérieur à 18%. En appliquant un taux de croissance comparable aux pertes dues au piratage, on aboutit à un manque à gagner d'environ 32 milliards d'euros, à l'horizon 2015, dans les secteurs de la musique enregistrée, du film, des séries TV et du logiciel (voir Tableau 1.2).

A défaut de modification significative de l'action des pouvoirs publics et compte tenu de l'augmentation, d'année en année, des pertes dues au piratage, les emplois détruits sur une année le seront définitivement, entraînant un accroissement des pertes d'emplois dans le secteur. Autrement dit, les destructions d'emplois dans l'Union Européenne seront d'environ 610 000 à l'horizon 2015, contre un peu plus de 185 000 en 2008.

TAB. 1.2: Pertes liées au piratage en Europe (2008-2015) – Scénario basé sur les tendances en matière “ d'échange de fichiers”

	2008	2009	2010	2011	2012
Chiffre d'affaires (M€)	10	12	15	19	22
Emplois cumulés	186 600	227 500	276 900	351 500	422 400

Dans le Scénario 2, l'hypothèse retenue est que la croissance du piratage numérique va suivre les tendances du trafic IP global des particuliers en Europe (c'est-à-dire les communications effectuées via le Protocole Internet). Ce scénario prend ainsi en compte le piratage via partage de fichiers, mais également d'autres formes comme le streaming en ligne, fixant dès lors un plafond à l'impact du piratage numérique. Entre 2008 et 2015, le trafic IP des particuliers devrait croître, d'après les prévisions, à un taux supérieur à 24%. A raison d'un taux de progression similaire pour le piratage numérique en Europe, les pertes dans les secteurs de la musique enregistrée, du cinéma, des séries TV et du logiciel s'élèveront à 56 milliards en 2015, contre 10 milliards d'euros environ en 2008. En l'absence de modification significative de l'action des pouvoirs publics et compte tenu de l'augmentation, d'année en année, des pertes dues au piratage, les emplois détruits sur une année le seront définitivement, entraînant un accroissement des pertes d'emplois dans le secteur. Autrement dit, les destructions d'emplois dans l'Union Européenne s'élèveront à 1,2 million à l'horizon 2015, contre un peu plus de 185 000 en 2008 (voir Tableau 1.3).

TAB. 1.3: Pertes d'emplois en Europe (2008-2015) – Scénario basé sur les tendances du “Trafic IP des particuliers”

	2008	2009	2010	2011	2012
Chiffre d'affaires (M€)	10	13	17	24	32
Emplois cumulés	186 600	253 800	345 000	490 200	655 100

1.2.2.2 Situation en Tunisie

La Tunisie fait partie des pays les plus concernés par le piratage. En effet, elle est classée, en 2003, 1ère en piraterie logicielle dans le monde arabe et 3ème en Afrique. Néanmoins, le taux de piratage en Tunisie ne cesse de baisser et affiche une diminution de 11% entre 2004 et 2008. C'est ce qui ressort de l'édition 2008 du rapport de la Business Software Alliance. Ainsi, le taux de piratage informatique en Tunisie est passé de 76% en 2007 à 73% en 2008. La tendance est donc clairement à la baisse puisque ce même taux était de 84% en 2004.

Le président tunisien Zine El Abidine Ben Ali a appelé le 25/02/2010 les responsables concernés de son gouvernement à "prendre toutes les mesures réglementaires adéquates pour lutter contre le piratage et la reproduction illicite", un phénomène qui prend des proportions croissantes en Tunisie.

Les artistes tunisiens, chanteurs et producteurs de cinéma, se plaignent régulièrement du piratage de leurs œuvres sur le marché parallèle, réduisant leurs droits d'auteurs à leur plus simple expression. Les boutiques qui vendent à bas prix les CD gravés de films et d'albums de chanteurs tunisiens et étrangers pullulent à Tunis et dans d'autres villes de l'intérieur du pays.

1.3 Aspect général du tatouage

1.3.1 Position du tatouage

Le tatouage fait parti de la science de la cryptologie. Cette science est en fait l'ensemble des moyens permettant de protéger tout document en assurant sa confidentialité,

son intégrité et son authenticité. On distingue deux sous-classes dans la cryptologie : la cryptographie et la stéganographie dont dérive la discipline de tatouage et le camouflage d'informations. Dans la suite, nous décrivons brièvement chacune de ses disciplines en mettant en évidence les différences qui existent entre elles.

1.3.1.1 Cryptographie

La cryptographie consiste à transmettre un message qui ne soit compréhensible que par le destinataire. Pour cela, celui-ci partage un secret avec l'émetteur du message. Le message "en clair" est transformé à l'aide du "secret" en un message codé. Tout l'art de la cryptographie consiste à faire en sorte que, sans la connaissance du secret, il est extrêmement difficile, voir impossible, de retrouver le message en clair.

1.3.1.2 Stéganographie

La stéganographie est l'art de dissimuler un message à travers d'un autre message. L'objectif de cette dissimulation peut varier et de ce fait découle deux disciplines : le camouflage d'information (Data hiding) et le tatouage (Watermarking).

Camouflage d'information : Le message primaire reste lisible pour tout le monde, tandis que le message secondaire n'est lisible que par les personnes propriétaires de l'information. Par comparaison à la cryptographie, celle-ci permet de chiffrer un message tout en étant accessible. Par contre, dans la stéganographie, le message est discret, caché dans un autre document et donc il est imperceptible. Si la notion de « clef secrète » est commune à la cryptographie et à la stéganographie, la différence essentielle réside dans le fait que, pour la cryptographie, la clef secrète empêchera celui qui n'en a pas la connaissance de déchiffrer le message, alors que la stéganographie empêchera de suspecter son existence même.

Tatouage : Cette technique consiste à la transmission d'un message de manière imperceptible à travers d'un ensemble de données support et ceci d'une manière indissociable. La marque insérée est uniquement connue par le propriétaire. Ses caractéristiques sont uniques et dépendent des clés fournies par celui-ci. Contrairement à la cryptographie, le tatouage protège le médium tout en le laissant accessible. Il faut dire que le tatouage est semblable à la stéganographie puisque les deux sont basés

sur le camouflage d'informations dans un médium mais les deux diffèrent au niveau de leurs objectifs. En effet, dans la stéganographie l'objectif est de transmettre un message en le dissimulant au sein d'un médium qui n'a pas beaucoup d'importance par contre dans le tatouage l'objectif est de protéger le document qui contient l'information dissimulée. Dans le tatouage, l'information cachée a une relation avec le contenu. Une autre différence entre les deux réside dans le fait que dans la détection, il est indispensable d'extraire le message dans la stéganographie, par contre, dans le tatouage, soit on extrait le message soit on détecte simplement sa présence.

1.3.2 Applications du tatouage

1.3.2.1 Droits d'auteurs

Le tatouage offre une alternative intéressante à la cryptographie, car il permet de protéger l'image, même lorsque celle-ci est diffusée. La protection des droits d'auteur représente, quant à elle, l'application la plus courante aujourd'hui. L'objectif est d'incruster une information dans la donnée source, typiquement le copyright du propriétaire, afin de prévenir toute revendication frauduleuse de propriété. Cette signature ne peut être connue que par la personne ou l'organisme de tatouage. Elle dépend donc d'une clé secrète, qui permet son insertion et sa détection. Cette application nécessite la mise en place d'un algorithme de tatouage d'un niveau de robustesse très élevé. En effet, celui-ci ne doit pas être ambigu et doit toujours déterminer l'appartenance du médium, même si d'autres parties insèrent également une marque.

1.3.2.2 Authentification des documents

L'information insérée au sein du document peut aussi permettre de certifier que celui-ci n'a pas été modifié. Il s'agit ici d'une problématique de contrôle d'intégrité des documents. Dans ce cas précis, la signature ajoutée doit être fragile, c'est à dire que la moindre attaque doit altérer la marque. Dans la phase de détection, si la marque a été correctement récupérée alors le document en question n'a pas été manipulé. Dans le cas contraire, on peut affirmer que le document a été l'objet d'une manipulation que ce soit honnête ou malhonnête.

1.3.2.3 Traçabilité des documents

La traçabilité des documents numériques fait face à la pratique de la copie illégale qui devient un phénomène massif vu le développement des moyens de communication et de partage. Le principe est d'insérer une signature personnalisée dans la vidéo dans le contexte de la "video on demand" (VoD) par exemple. Ainsi, si une copie est rediffusée d'une manière illégale, il serait possible grâce à la signature de remonter à l'utilisateur malhonnête qui peut être traduit en justice.

1.3.2.4 Indexation des documents

Le domaine de l'indexation des images consiste à classer de manière automatique des images selon leur contenu. Il permet de faciliter une recherche dans une base de données. Les techniques classiques utilisées consistent à effectuer un traitement automatique de l'image, de manière à dégager les composantes essentielles du contenu. Le tatouage d'un document permet ainsi d'insérer une information (contenant peu de bits) décrivant le contenu. Cela permet de qualifier sommairement l'image, ou d'insérer un pointeur vers une description plus complète.

1.3.2.5 Empreinte digitale

Dans le but de dévoiler les sources des copies illégales, le propriétaire peut insérer différentes signatures ou bien la même signature avec différentes clefs selon les utilisateurs. Ainsi, il peut par exemple affecter un numéro de série pour chaque client. Ceci lui permettra de dévoiler celui qui a été à l'origine de la diffusion des copies illégales présentes sur le marché [59].

1.3.3 Contraintes du tatouage

Concevoir un algorithme de tatouage revient à trouver le meilleur compromis entre trois contraintes qui sont la capacité, l'imperceptibilité et la robustesse.

1.3.3.1 Capacité

C'est la quantité d'information qu'on peut cacher au sein du médium. Il paraît évident que plus on augmente la capacité, plus la marque sera perceptible, et plus la robustesse diminuera (dans le cas où on veut retrouver exactement la marque).

1.3.3.2 Imperceptibilité

C'est l'impact que peut avoir la marque sur le médium. Plus le marquage sera fort, plus il sera visible (et inversement). Garantir l'invisibilité de la signature revient à respecter la qualité visuelle du média. Aucune dégradation ne doit être perçue sur l'objet d'origine. C'est pour cette raison que plusieurs algorithmes de marquage existants ont tendance à effectuer l'insertion dans les zones d'intérêts les moins sensibles à l'œil humain (contour, zone texturée,...). La figure 1.1 ci-dessous expose un exemple de tatouage visible.

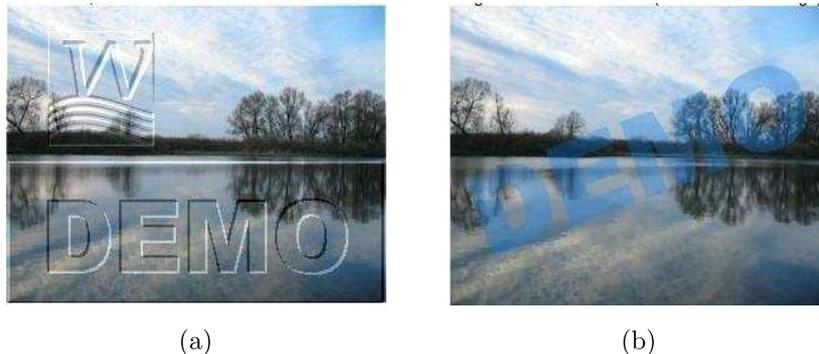


FIG. 1.1: Exemples de tatouage visible

1.3.3.3 Robustesse

C'est la capacité que possède un algorithme de tatouage à résister aux attaques extérieures. Pour la vidéo, il peut s'agir d'attaques simples comme le changement de format de compression, le changement de débit ou tout autre traitement classique. La détection doit être possible quelque soit l'attaque appliquée, cette contrainte définit la robustesse d'un schéma de tatouage en distinguant trois types de tatouage :

- Fragile : ne résistant à aucune attaque
- Semi fragile : ne résistant qu'à quelques types d'attaques

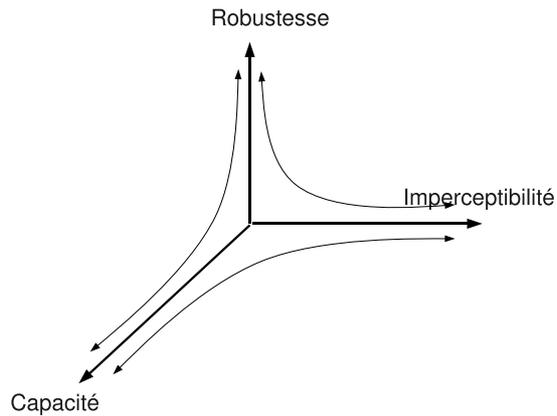


FIG. 1.2: Problématique en tatouage

- Robuste : résistant à tous types d’attaques (théoriquement)

Concevoir un algorithme de tatouage revient à trouver le meilleur compromis entre ces trois principes en fonction de l’application visée. La figure 1.2 présente une bonne illustration de cette problématique. Ainsi, un schéma de tatouage qui est robuste à une très grande variété de traitements ne pourra pas avoir une capacité très importante (les techniques de tatouage pour la protection des droits d’auteurs n’utilisent souvent que 64 bits). À l’inverse, une méthode permettant de cacher plusieurs milliers de bits dans une image de taille standard ne pourra pas être très robuste ([5]).

1.3.4 Attaques innocentes et malveillantes

Comme nous avons vu précédemment, un des points forts d’un tatouage efficace réside dans sa robustesse. Néanmoins, certaines transformations basiques peuvent effacer le marquage, ou du moins potentiellement l’altérer. Toutes ces transformations, volontaires ou involontaires, ayant une influence directe sur le tatouage, sont appelées des attaques. Les attaques peuvent être des traitements visant soit à brouiller soit à enlever la marque de protection dans la vidéo. La figure 1.3 résume les différents types d’attaques. Ceux ci peuvent être aussi classés selon l’intention de leurs utilisations [10]. Dans ce cas, on distingue deux grandes familles :

- Attaques bienveillantes : Il s’agit des traitements qui n’ont pas initialement pour objectif d’empêcher la détection de la marque. Il peut s’agir des dégradations dues à

une compression MPEG2, à des filtrages (réduction de bruit), à un changement de résolution. Un autre traitement couramment utilisé en vidéo est la conversion analogique/numérique, et inversement. Enfin, certaines distorsions géométriques peuvent être utilisées : changement d'échelle, fenêtrage, rotation, etc.

- Attaques malveillantes : De nombreuses attaques initialement développées pour l'image fixe peuvent être facilement adaptées à la vidéo dont le but de retrouver le marquage. Ce qui permet d'altérer le tatouage soit par l'effacement de la signature, soit par la désynchronisation de la signature, soit par l'utilisation des faiblesses de l'algorithme de marquage.

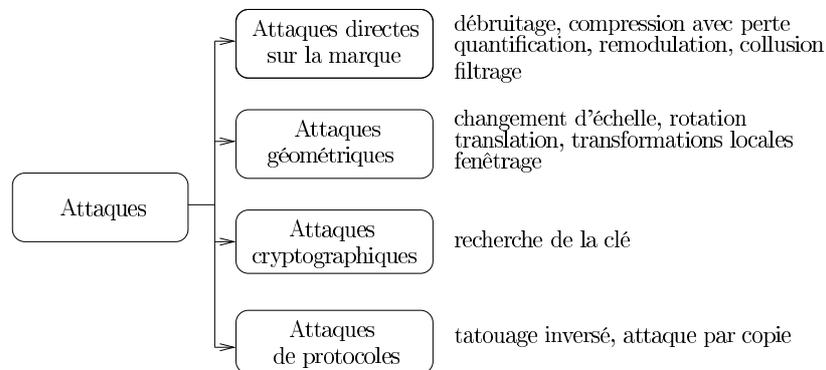


FIG. 1.3: Les différents types d'attaques s'opposant au tatouage

1.4 Principes des schémas de tatouage

Même si les schémas de tatouage sont très variés et semblent être différents les uns des autres, ils s'appuient tous sur un même principe qui se traduit par deux phases importantes qui sont la phase d'insertion de la marque et la phase de détection. Ces deux phases s'appliquent souvent et en général sur un seul espace choisi selon le contexte et l'objectif visé par le schéma en question.

1.4.1 Domaines d'insertion

Dans un schéma de tatouage, le domaine d'insertion est d'une importance majeure. En effet, le choix de celui-ci influe directement sur la fiabilité du schéma en terme de robustesse et de capacité. On peut dire que le choix du domaine dépend de l'application visée puisque

chacun possède ses propres caractéristiques. Dans ce qui suit, nous présentons les différents domaines d'insertion possible ainsi que leurs avantages et inconvénients.

1.4.1.1 Domaine spatial

Il s'agit, dans ce domaine, de transformer directement la luminance des pixels. Les méthodes de tatouage s'appuyant sur ce domaine sont peu coûteuses en terme de temps de calcul puisqu'elles ne nécessitent pas une étape préalable de transformation. De plus, ce domaine est adéquat pour la conception de méthodes de tatouage qui soient robustes aux transformations géométriques. Par contre, ce domaine est peu utilisé puisqu'il n'est pas facile de gérer l'aspect invisibilité de la marque. En plus, ce domaine est sensible aux transformations avec perte telle que la compression.

1.4.1.2 Domaine fréquentiel

C'est l'espace obtenu après une transformée en DCT ¹ ou une transformée FFT ². Le domaine DCT est fréquemment utilisé puisqu'il permet d'avoir des systèmes de tatouage qui résistent à la compression JPEG en image et MPEG en vidéo. Ceci est dû au fait que ces normes de compression se font au niveau de la transformée DCT. Le domaine FFT, par contre, est très peu utilisé même s'il présente l'avantage d'être un espace invariant aux transformations géométriques.



FIG. 1.4: Exemple d'une transformée de fourier

¹Discrete Cosine Transform

²Fast Fourier Transform

1.4.1.3 Domaine multirésolution

C'est un domaine où le signal (image, vidéo...) est représenté sous plusieurs niveaux de résolution. On parle aussi de scalabilité qui peut être fréquentielle (voir Figure 1.5) ou spatiale (voir Figure 1.6). En tatouage, le domaine multirésolution, obtenu après une transformée en ondelettes DWT ³, est fréquemment utilisé pour définir des techniques de tatouage. En fait, cette scalabilité fréquentielle du signal donne une meilleure visibilité du spectre fréquentiel en offrant une classification à plusieurs niveaux des fréquences constituant le signal. En plus, des techniques de compression basées sur la DWT sont apparues ce qui explique entre autre l'émergence des techniques de tatouage vers ce domaine.

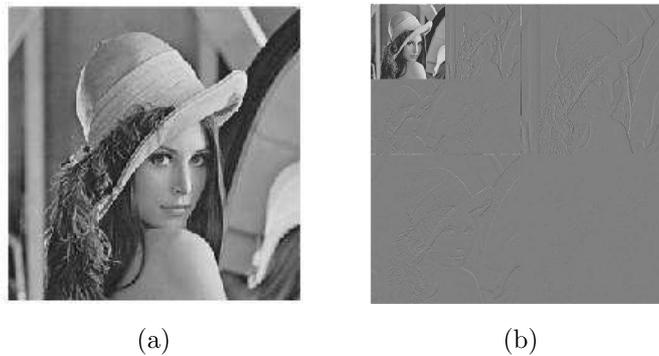


FIG. 1.5: Exemple d'une transformée d'ondelette à 2 niveaux. (a) Image Lena. (b) Image Lena transformée.



FIG. 1.6: Exemple de scalabilité spatiale avec différents facteurs de résolution.

³Discrete Wavelet Transform

1.4.2 Phase d'insertion

Le document original à tatouer passe par une étape de transformation si le domaine d'insertion est différent du domaine spatial. Le signal est noté alors $x = \{x_1, x_2, \dots, x_n\}$ avec les x_i qui représentent les luminances des pixels (domaine spatial) ou les coefficients fréquentiels si par exemple une transformée de DCT est appliquée. Le signal x est combiné avec la marque w qui est obtenue à partir du message original m à cacher et une clé secrète k . Le signal obtenu, noté y passe par une transformée inverse pour aboutir au document original tatoué.

Il existe deux manières pour combiner le signal x avec la marque w :

1. *manière substitutive* : il s'agit de modifier les composantes du signal x afin de représenter le signal w . Pour cela, une relation ou une contrainte doit être définie et imposée sur les composantes x_i . Plusieurs techniques de tatouage substitutives ont été proposées. On peut citer essentiellement la technique de LSB (Last Significant Bit) [83] qui consiste à modifier les bits de poids faible d'une composante du signal x afin d'insérer le message m . L'utilisation d'une clé secrète permet de sélectionner les composantes qui vont être modifiées. Une technique dérivée du LSB est celle proposée dans [81] et qui améliore la robustesse de la technique de base LSB jugée très peu robuste. La quantification ([15]) des composantes de l'image est une technique utilisée pour substituer le message m à cacher. Il s'agit de définir un dictionnaire de symboles pour chaque composante du message m . Toujours dans le domaine spatial, [60, 4] ont proposé des techniques de tatouage substitutives qui consistent à substituer le message m par des composantes géométriques de l'image. Un symbole du message peut être représenté en remplaçant un bloc de l'image par un autre ayant une similarité très élevée. Dans le domaine fréquentiel, on peut citer la méthode proposée par [97]. Elle consiste à imposer une relation entre des coefficients sélectionnés parmi les moyennes fréquences DCT afin de représenter un symbole du message m . Par exemple, pour insérer un bit à 1, les coefficients C_0 et C_1 sont modifiés de tel sorte que l'équation $C_0 < C_1$ soit vérifiée. Le même principe a été appliqué par [49] utilisant le domaine multirésolution, obtenu après une transformée par ondelettes, comme domaine d'insertion.

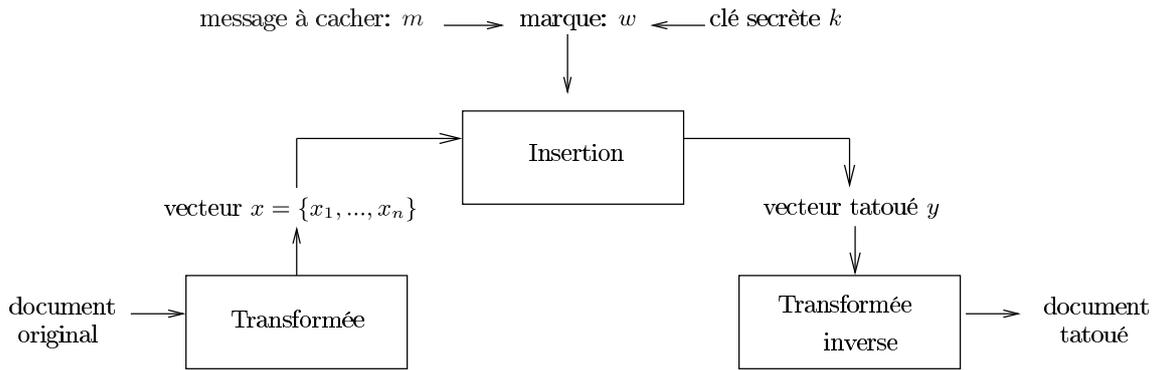


FIG. 1.7: Phase d'insertion de la marque

2. *manière additive* : la manière additive consiste à additionner les deux signaux w et x selon l'équation 1.1.

$$y = x + \alpha w \quad (1.1)$$

y représente le signal tatoué. α est un facteur de pondération qui représente la force de tatouage. Il désigne le maximum d'énergie qui peut être introduit dans le signal hôte sans une altération visible. De nombreuses techniques de tatouage additives ont été proposées. Dans [35], une séquence aléatoire modulée par un message est ajoutée au signal hôte. Chaque bit du message à insérer, de valeur ± 1 , est associé à une région de l'image, et le message complet est alors réparti sur tout le support de l'image. Ce signal est ensuite modulé par une séquence aléatoire, puis pondéré par un masque (pour assurer l'imperceptibilité). Ce signal est alors ajouté à l'image. La technique de Patchwork est une technique additive dans le domaine spatial mais qui présente l'inconvénient d'être non robuste face aux compressions avec perte. Les détails de cette technique ainsi que d'autres qui opèrent dans les domaines transformés seront donnés dans le chapitre suivant à la section 2.3.1.

1.4.3 Phase de détection

Le document tatoué reçu peut être modifié lors de sa transmission ou même par un utilisateur malhonnête qui a éventuellement tenté de détruire la marque. Ce document est transformé dans le domaine où l'insertion a été effectuée. On aboutit au signal \hat{y} . Le message est détecté ou peut être bien extrait du signal \hat{y} utilisant la clé secrète qui a

servi dans la phase d'insertion. Le recours au document original est possible et ce, selon la nature de la détection :

1. **détection aveugle** : le document original n'est pas nécessaire pour extraire la marque. Seulement la clé secrète utilisée lors de l'insertion est nécessaire.
2. **détection semi-aveugle** : la marque originale w est nécessaire.
3. **détection non aveugle** : le signal hôte original x est utilisé pour la détection.

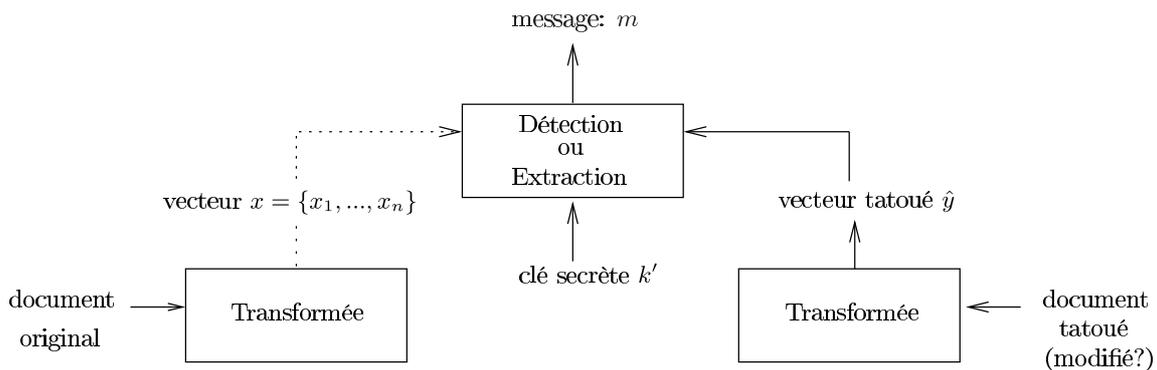


FIG. 1.8: Phase de détection de la marque

En général, le calcul de la corrélation permet de détecter la marque. On distingue deux types :

1. la corrélation linéaire

$$c = \frac{1}{N} \sum_{i=0}^{N-1} w_i \cdot \hat{y}_i \quad (1.2)$$

2. la corrélation normalisée

$$c = \frac{1}{N} \sum_{i=0}^{N-1} \frac{w_i \cdot \hat{y}_i}{\|w\| \cdot \|\hat{y}\|} \quad (1.3)$$

w_i (respectivement \hat{y}_i) désigne la $i^{\text{ème}}$ composante du vecteur w (respectivement \hat{y}).

1.5 Mesures de la qualité du document tatoué

La mesure de la qualité d'une image ou d'une vidéo est nécessaire pour valider un quelconque traitement. Le problème qui se pose est l'objectivité de ces mesures. En effet, une mesure de qualité doit considérer l'aspect visuel qui reste un moyen subjectif pour

décider de la qualité d'un signal manipulé. Plusieurs travaux ([72, 45]) se sont concentrés sur la définition de nouvelles mesures qui soient les plus objectives possible et qui tiennent compte du Système Visuel Humain. Toutes ces métriques proposées sont dérivées d'une métrique très souvent utilisée : le PSNR. Dans le contexte du tatouage, une dérivée de cette métrique qui considère la signature insérée dans l'image ou la vidéo est celle du wPSNR. Nous présentons ci-dessous les définitions de ces deux métriques.

1.5.1 Le PSNR

Ce type de mesure renvoie une valeur numérique indiquant la distorsion causée. La mesure la plus utilisée est l'erreur quadratique moyenne (Mean Square Error, MSE) calculée entre les pixels des deux images à comparer :

$$MSE = \frac{1}{N} \sum_{p \in P} (x_p - \hat{x}_p)^2 \quad (1.4)$$

où P : l'ensemble des N pixels de l'image, x et \hat{x} : les niveaux de gris des images à comparer.

Cependant, on utilise en général le rapport crête signal sur bruit (Peak Signal to Noise Ratio, PSNR). Cette valeur mesure la fidélité puisqu'elle est proportionnelle à la qualité. Elle est donnée par :

$$PSNR = 10 \log_{10} \frac{x_{max}^2}{MSE} \quad (1.5)$$

où x_{max} désigne la luminance maximale possible.

Une valeur de PSNR égale à l'infini (∞) correspond à deux images parfaitement identiques. Elle décroît en fonction de la distorsion et relie donc l'erreur quadratique moyenne à l'énergie maximale de l'image.

1.5.2 Le wPSNR

Le wPSNR (Weighted Peak Signal to Noise Ratio) est une variante du PSNR, qui prend en compte le voisinage de chaque pixel. Cette métrique est basée sur la sensibilité de l'œil humain aux changements des textures et des homogénéités des régions. La valeur

de $wPSNR$ augmente quand la variance est grande et décroît dans le cas contraire. Une nouvelle définition de l'erreur quadratique moyenne (MSE) est alors :

$$wMSE = \frac{1}{N^2} \sum_{i,j=1} \left(\frac{x_p - \hat{x}_p}{1 + var(i, j)} \right)^2 \quad (1.6)$$

et donc :

$$wPSNR = 10 \log_{10} \frac{x_{max}^2}{MSE} \quad (1.7)$$

1.6 Conclusion

Le tatouage est une nouvelle alternative pour sécuriser les documents multimédias tels que l'image, la vidéo et le son. Dans notre cas, nous nous intéressons au cas de la protection des flux vidéo. Dans le chapitre suivant, nous ferons d'abord un tour d'horizon sur les techniques de tatouage d'images fixes puisqu'une vidéo est une suite d'images. Nous montrons la limite de ces techniques dans le contexte de la vidéo, ensuite nous exposons les différentes techniques de tatouage spécifique à la vidéo et qui se basent sur les propriétés d'un flux vidéo considéré comme un signal 3D.

Chapitre 2

Tatouage vidéo

2.1 Introduction

En général, la technique de tatouage permet de protéger un document multimédia des différentes formes de manipulations ou de piraterie. Il s'agit d'introduire de manière robuste et invisible une signature dans le document qui prouve la propriété intellectuelle de l'auteur de ce document. Dans le contexte de la vidéo, le même principe s'applique : il s'agit d'introduire la signature au niveau des images qui composent la vidéo. On peut dire que le tatouage vidéo est une extension du tatouage d'image. Évidemment, c'est le cas et donc l'application des techniques de tatouage d'image sur la vidéo est légitime. Par contre, un signal vidéo, sous sa forme brute, est loin d'être considéré comme simplement une suite d'images. En effet, celui-ci doit être vu comme un signal $2D + t$ si on considère la composante temps et qui fait la caractéristique essentielle d'un signal vidéo. Ainsi, un schéma de tatouage vidéo doit tenir compte de cette nouvelle dimension et de profiter des caractéristiques qu'elle offre. Ceci ne peut qu'augmenter la fiabilité des schémas de tatouage ainsi que leurs robustesses. En plus, et afin d'éviter des problèmes liés à la capacité de stockage et de transmission, un signal vidéo peut se trouver sous une forme compressée. Plusieurs formats de compression sont alors apparus avec une diversité au niveau de leurs structures. Dans ce cas, la vidéo ne doit plus être considérée comme une suite d'images mais plutôt il faut s'adapter à la forme de compression pour concevoir des schémas de tatouage dans le domaine compressé.

Dans ce chapitre, nous commençons par donner un aperçu sur les différents formats, compressés et non compressés de la vidéo numérique. Un bref état de l'art sur les techniques de tatouage d'images suivi d'une classification des techniques existantes de tatouage vidéo. A la fin, nous ferons une synthèse et nous présenterons nos motivations et objectifs dans ce travail de thèse.

2.2 Notion de vidéo

2.2.1 Définition

Une vidéo est un flux composé d'une part d'une suite d'images (25 f/s en Europe, 30 f/s aux USA), donnant l'illusion du mouvement et d'autre part d'une composante sonore. Chaque image est décomposée en lignes horizontales. Chaque ligne est considérée comme une succession de points. La lecture et la restitution d'une image s'effectue donc séquentiellement ligne par ligne comme un texte écrit : de gauche à droite puis de haut en bas.

2.2.2 Différents formats vidéo

Un signal vidéo peut être sous une forme brute (non compressée) ou bien sous une forme compressée.

2.2.2.1 Formats non compressés

1. **Le standard S-Video** : Le standard S-Video (pour « Sony Video »), parfois appelé Y/C, est un mode de transmission vidéo à composantes séparées utilisant des câbles distincts pour faire transiter les informations de luminance (luminosité) et de chrominance (couleur). Une connexion S-Video permet d'offrir une qualité de vidéo optimale en envoyant simultanément les 576 lignes de l'image, sans entrelacement (en une seule fois).

Le signal S-Vidéo est habituellement transporté via un câble comportant un connecteur mini-DIN 4 broches (deux broches distinctes pour chaque composante vidéo).

On trouve ainsi ce type de connecteurs sur les caméscopes ou magnétoscopes haut de gamme S-VHS, les caméras Hi8 ou tout simplement sur la majorité des cartes graphiques possédant une sortie TV. Cependant, le signal S-Vidéo ne permettant de transporter que le signal vidéo, il est nécessaire d'utiliser un câble audio séparé pour le transport des données audio.

Il faut noter toutefois que les prises péritel récentes permettent également de transporter un signal S-Vidéo. Il existe d'ailleurs des adaptateurs permettant de connecter un câble S-Vidéo et des connecteurs RCA audio sur une prise péritel, au prix d'une dégradation supplémentaire du signal. Néanmoins, étant donné que le signal S-Vidéo ne faisait pas partie des spécifications de la prise péritel à l'origine, certains équipements possédant une telle prise ne supportent parfois pas ce type de signal.

2. **Le standard YUV** (appelé aussi CCIR 601), auparavant baptisé YCrCb, est un modèle de représentation de la couleur dédié à la vidéo analogique.

Il se base sur un mode de transmission vidéo à composantes séparées utilisant trois câbles différents pour faire transiter les informations de luminance (luminosité) et les deux composantes de chrominance (couleur). Il s'agit du format utilisé dans les standards PAL (Phase Alternation Line) et SECAM (Séquentiel Couleur avec Mémoire).

Le paramètre Y représente la luminance (c'est-à-dire l'information en noir et blanc), tandis que U et V permettent de représenter la chrominance, c'est-à-dire l'information sur la couleur. Ce modèle a été mis au point afin de permettre la transmission des informations colorées aux téléviseurs couleurs, tout en s'assurant que les téléviseurs noir et blanc existants continuent d'afficher une image en tons de gris.

Voici les relations liant Y à R, G et B, U à R et à la luminance, et enfin V à B et à la luminance :

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ U &= -0.147R - 0.289G + 0.436B = 0.492(B - Y) \\ V &= 0.615R - 0.515G - 0.100B = 0.877(R - Y) \end{aligned} \tag{2.1}$$

Ainsi U est parfois noté Cr et V noté Cb, d'où la notation YCrCb.

3. **Le modèle YIQ** est très proche du modèle YUV. Il est notamment utilisé dans le standard vidéo NTSC (utilisé entre autres aux États-Unis et au Japon). Le para-

mètre Y représente la luminance. I et Q représentent respectivement l'Interpolation et la Quadrature. Les relations entre ces paramètres et le modèle RGB sont les suivantes :

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ I &= 0.596R - 0.275G - 0.321B \\ Q &= 0.212R - 0.523G + 0.311B \end{aligned} \tag{2.2}$$

2.2.2.2 Formats compressés

Le groupe MPEG a été établi en 1988 dans le but de développer des standards internationaux de compression, décompression, traitement et codage d'images animées et de données audio. Il existe plusieurs standards MPEG :

1. **MPEG-1**, développé en 1988, est un standard pour la compression des données vidéos et des canaux audio associés (jusqu'à 2 canaux pour une écoute stéréo). Il permet le stockage de vidéo à un débit de 1.5Mbps dans une qualité proche des cassettes VHS sur un support CD appelé VCD (Vidéo CD).
2. **MPEG-2**, un standard qui est dédié originalement à la télévision numérique (HDTV) offrant une qualité élevée à un débit pouvant aller jusqu'à 40 Mbps, et 5 canaux audio surround. Le MPEG-2 permet de plus une identification et une protection contre le piratage. Il s'agit du format utilisé par les DVD vidéos.
3. **MPEG-4**, un standard qui est destiné à permettre le codage de données multimédia sous formes d'objets numériques, afin d'obtenir une plus grande interactivité, ce qui rend son usage particulièrement adapté au Web et aux périphériques mobiles.
4. **MPEG-7**, un standard visant à fournir une représentation standard des données audio et visuelles afin de rendre possible la recherche d'information dans de tels flux de données. Ce standard est ainsi également intitulé Multimedia Content Description Interface.
5. **MPEG-21**, en cours d'élaboration, dont le but est de fournir un cadre de travail (en anglais framework) pour l'ensemble des acteurs du numériques (producteurs, consommateurs, ...) afin de standardiser la gestion de ces contenus, les droits d'accès, les droits d'auteurs, ...

2.2.3 Contraintes de la vidéo

Dans un flux vidéo, il est possible de ne tatouer que quelques images. De nombreux schémas développés pour les images fixes peuvent être alors appliqués aux séquences vidéo. Cependant, un flux vidéo présente d'autres propriétés, qui peuvent être exploitées pour l'insertion de la signature [11] :

- la taille brute d'une séquence vidéo est beaucoup plus importante que la taille d'une image fixe. L'espace d'insertion de la signature est considérablement augmenté.
- la dimension temporelle du signal traité peut être utilisée pour l'insertion de la signature. Celle-ci peut, par exemple, être insérée dans le mouvement des différents objets de la séquence.

Les séquences vidéo présentent également des contraintes différentes de celles des images fixes :

- la complexité du schéma de tatouage doit être faible. L'insertion et la détection de la signature doivent pouvoir s'effectuer à la volée, dans la plupart des applications. La contrainte de temps réel s'applique essentiellement à la phase de détection.
- le mouvement des objets rend souvent plus intense la visibilité de la marque : ainsi, une marque "fixe" ajoutée sur un objet en mouvement, sera d'avantage perceptible que si l'objet est statique.
- le flux vidéo est souvent compressé de manière à réduire la taille brute des séquences. Rappelons qu'un des standards de compression le plus couramment utilisé dans le monde industriel est MPEG2. L'insertion de la signature peut alors s'effectuer directement lors de la compression. L'insertion sur le format décompressé ne doit pas entraîner, après la compression, une augmentation significative de la taille des données.
- la présence de la signature dans la séquence vidéo peut permettre d'autres attaques que celles liées aux images fixes. Si la signature est redondante dans la séquence, elle peut être estimée en calculant la moyenne des différentes images de la séquence. La signature doit être détectée après une désynchronisation, conséquence d'une perte de quelques images de la séquence par exemple.

2.3 Tatouage : de l'image à la vidéo

Une séquence vidéo est une suite d'images. Pour cela, les premiers schémas de tatouage vidéo étaient inspirés des schémas de tatouages d'images fixes. Nous commençons alors par faire un bref état de l'art sur les schémas de tatouage d'image avant de passer à voir les schémas dédiés à la vidéo.

2.3.1 Techniques de tatouage d'image

Les techniques de tatouage d'images fixes sont très nombreuses et très diversifiées. Elles peuvent être classifiées selon plusieurs critères différents. On peut les classer selon le domaines d'insertion, selon la manière d'insertion (additives ou substitutives) ou encore selon les attaques auxquelles elles résistent. Les premières techniques de tatouage d'images sont des techniques additives dans le domaine spatial. Parmi ces techniques, on distingue surtout celles qui reposent sur l'insertion d'une séquence aléatoire $2D$ dans l'image [85, 88, 35]. Il s'agit d'introduire une séquence aléatoire modulé par le message qui représente l'information à cacher (copyright par exemple). Le message est donc étalé sur toute l'image sous forme d'un signal formé de $+1, -1$ et ce, en tenant compte de l'aspect imperceptibilité de la marque.

La technique de "Patchwork" a elle aussi été traitée dans [77, 7]. La technique consiste à définir, aléatoirement à travers une clé secrète, deux zones (patches) de l'image qui soient de même taille (Voir figure 2.1). L'insertion d'un bit de la marque consiste à augmenter la luminance d'une des zones afin que la différence des moyennes des luminances des deux zones soit supérieure à un seuil préfixé. La détection se fait en calculant cette différence pour décider si l'image est tatouée ou non.

L'avantage de ces techniques est qu'elles s'effectuent dans le domaine image et qu'elles sont faciles à implémenter. Par contre, elles ont l'inconvénient d'être non robustes face à l'attaque de compression avec perte. D'autres techniques sont alors apparues pour faire face à ce type d'attaque. Ces techniques se sont basées sur les propriétés des normes de compression. L'insertion de la marque dans ces techniques se font au niveau de l'espace concerné par la compression : domaine DCT pour la compression JPEG, domaine multirésolution dans le cas de la compression JPEG200. Par exemple, le schéma présenté



FIG. 2.1: Exemple de patches

dans [18], la marque est insérée dans les moyennes fréquences obtenues après une transformée en DCT. La détection se fait en calculant la différence entre l'image originale et l'image tatouée (Schéma non aveugle). Une version améliorée, qui ne nécessite pas l'image originale au moment de la détection, a été présentée par Piva et al. dans [78]. C'est un schéma semi-aveugle qui utilise la corrélation comme mesure de similarité entre la marque originale et la marque récupérée.

On peut également citer [40] qui propose une solution reposant sur la transformée de Fourier. De nombreux schémas de tatouage opérant dans le domaine multirésolution ont été proposés [48, 92, 102], mais [58] est l'un des rares à fonctionner en mode aveugle : le filigrane est inséré dans les trois sous-bandes de détails de la décomposition en ondelettes ; les coefficients des trois sous-bandes sont marqués par l'addition pondérée d'une séquence pseudo aléatoire de même taille que l'ensemble des trois sous-bandes ; la détection s'effectue par une mesure de corrélation.

Récemment, la contrainte de sécurité a été abordée par la communauté de tatouage. En effet, il s'est avéré que la robustesse et l'invisibilité de la signature ne suffisent pas pour garantir la pérennité d'un système de tatouage. Le tatouage asymétrique, qui utilise des paramètres différents pour l'insertion et la détection, est proposé comme une alternative qui offre un niveau de sécurité plus élevé. Les publications suivantes ([27, 28, 89, 87]) présentent un bon état de l'art sur ce type de tatouage. Par ailleurs, l'apparition des schémas basés sur l'information adjacente offre une autre alternative de sécurité tout à fait pertinente. Les travaux publiés dans ce sens sont [61, 63, 62, 54, 17, 64, 23, 73, 75, 26, 30, 16, 57, 2, 1].

2.3.2 Limites des schémas de tatouage d'images dans le contexte de la vidéo

Théoriquement, tout schéma de tatouage d'image peut s'appliquer dans le contexte de la vidéo. Il suffit d'appliquer ce schéma sur toutes les images de la vidéo de manière indépendante. Or, la vidéo est une suite d'images qui sont fortement corrélées et il n'est pas évident de traiter chaque image à part sans considérer ses voisins. La vidéo représente un champ très vaste d'informations. Il offre, par rapport à une image, beaucoup plus de possibilités d'attaques ayant pour objectif de détruire ou même de récupérer la marque. Exploitant la redondance d'informations qui existe entre les images, un utilisateur malhonnête peut remplacer une image par la moyenne de ses voisins, supprimer certaines images en faisant un échantillonnage régulier afin de garantir la cohérence du signal et sans nuire à l'aspect visibilité. Une compression avec perte peut être aussi envisageable pour détruire la marque. Toutes ces attaques qui considèrent l'aspect temporel dans la vidéo, représentent un handicap aux schémas de tatouage d'images fixes. Toutefois, certains de ces schémas peuvent être adaptés au contexte vidéo et ce, en tenant compte de la composante temps dans les phases d'insertion et de détection. Nous donnons un aperçu, dans la section 2.4.1 des schémas de tatouage vidéo dérivés des schémas de tatouage d'image.

2.4 Classification des schémas de tatouage vidéo

2.4.1 Schémas dérivés du tatouage d'images fixes

Le principe est d'appliquer un algorithme de tatouage d'images fixes à chacune des images composant la vidéo, et ce de manière indépendante. Parmi les techniques citées plus haut pour les images fixes, et qui peuvent s'adapter à la vidéo de cette manière, on retrouve [35, 102, 38, 46, 3, 74, 39], dont les plus récentes sont aussi les plus robustes, notamment face aux distorsions asynchrones de l'image. On peut noter que la méthode présentée dans [35] présente l'avantage de pouvoir opérer sur flux compressé (MPEG-2) ou non. Ceci peut éviter des manipulations de décompression-tatouage-recompression si le flux est déjà compressé, ce qui permet non seulement de gagner du temps, mais aussi d'éviter des dégradations de la séquence vidéo. Cependant l'application de cette méthode

telle qu'elle est proposée provoque l'apparition d'artefacts gênants ; mais cet inconvénient peut être contré par l'utilisation d'un procédé de compensation de dérive, c'est-à-dire en modifiant la marque sur une image prédite afin qu'elle annule la contribution de la marque de l'image de référence précédente. Citons aussi des techniques dédiées au tatouage de flux compressé, [68, 6], qui consistent en l'adaptation de l'algorithme de Zhao & Koch [97]. Récemment, [86] ont adapté l'algorithme de [16] pour concevoir une approche de tatouage hiérarchique et informé pour la vidéo et ce en transposant les approches QIM (Quantization Index Modulation).

2.4.2 Schémas spatio-temporels

Si la vidéo est considérée comme un signal 3D, diverses transformées peuvent être appliquées prenant en compte la troisième dimension : le temps. On peut par exemple utiliser une décomposition en ondelettes temporelle, comme proposé dans [65]. Ce schéma ne permet de cacher qu'un seul bit, mais offre l'avantage de générer un filigrane qui dépend à la fois de l'ayant droit de la vidéo, et du contenu même de celle-ci. Les avantages de cette technique sont la grande robustesse aux modifications temporelles (suppression ou insertion d'images, moyenne d'images, interpolation temporelle, ...). Ses inconvénients sont tout d'abord sa faible capacité (on ne peut extraire qu'un bit d'information), mais aussi la nécessité de disposer de la séquence originale lors de l'opération de détection.

Une autre solution proposée repose sur l'utilisation de la transformée de Fourier 3D [24]. Elle permet de sélectionner les fréquences spatio-temporelles adéquates afin d'assurer l'invisibilité du filigrane, et assure également une robustesse face aux décalages (cycliques) spatiaux et temporels. Un motif de synchronisation est également ajouté, afin d'assurer une robustesse face aux changements d'échelle spatiale ou au re-échantillonnage temporel.

2.4.3 Schémas purement temporels

Ces méthodes cachent le filigrane dans le domaine temporel, en modifiant uniquement les très basses fréquences spatiales (moyenne d'image). En conséquence, le filigrane sera robuste aux déformations géométriques de l'image, mais sera sensible aux modifications temporelles (interpolation par exemple). La valeur de la marque sera constante

sur une image donnée, mais variera d'une image à une autre. Ceci peut évidemment avoir pour conséquence l'apparition d'artefacts, de clignotements. Pour éviter ces désagréments, d'une part la marque est pondérée par un masque visuel spatial et/ou temporel, et d'autre part on n'utilise que les basses fréquences spatiales afin de ralentir la variation de la luminance moyenne. La détection consiste en une mesure de corrélation entre la suite des moyennes d'images et le filigrane que l'on s'attend à trouver. On peut citer les techniques de [32, 99] qui reprennent ce principe général en l'améliorant, et [71] qui propose une méthode alternative ne reposant pas sur la luminance moyenne, mais sur des pixels isolés.

2.4.4 Tatouage de flux compressés

Nous avons déjà vu plus haut des techniques permettant de travailler au choix sur les flux compressés ou non. Mais il existe également des méthodes ne s'appliquant qu'aux flux compressés. [90, 91] proposent de s'appuyer sur les vecteurs de mouvement des images prédites dans le format compressé MPEG. Mais le filigrane est sensible aux estimateurs de mouvement utilisés, et risque de disparaître lors d'un transcodage. Dans le contexte de la protection de copie des DVDs, [55] s'appuie sur la structure de GOP utilisée, en attribuant par exemple à une image P la valeur 0 et à une image B la valeur 1. Il faut cependant tenir compte des contraintes imposées par la norme MPEG : la taille du message caché est limitée par la taille du GOP (entre 10 et 20 images par GOP), et si on code un message sur plusieurs GOP il faut faire attention à assurer la synchronisation temporelle lors de la détection ; afin d'assurer l'indétectabilité statistique du filigrane, les statistiques usuelles du GOP doivent être respectées ; il se peut qu'un codeur, adaptant naturellement la structure du GOP à la nature de la séquence vidéo, donne lieu à la détection d'un faux positif. Ces contraintes sont prises en compte dans la technique proposée par [55]. L'avantage de cette technique est que si l'encodage (le tatouage) est coûteux (une attaque, étant équivalente à un ré-encodage complet, serait trop lourde à mener, et risquerait de provoquer la création d'artefacts), la détection ne l'est pas (une réalisation matérielle grand public est possible). Comme pour les autres médias, les publications sont de plus en plus nombreuses, et la notion de tatouage informé a pris beaucoup d'importance ces derniers temps. Parmi les plus récentes, on peut citer [53, 84, 43, 56, 47, 13, 22, 53, 36] ainsi que l'étude sécuritaire de [19].

2.5 Motivations

Le tatouage vidéo est une extension du tatouage d'image. La plupart des techniques s'appliquant sur la vidéo exploitent les caractéristiques propres d'une séquence vidéo et se basent surtout sur l'aspect temporel qui existe dans de tels signaux. Comme on l'a remarqué dans la section 2.4, beaucoup de travaux, présentés dans la littérature, se basent sur la manipulation des vecteurs de mouvement pour insérer la marque. D'autres, se basent sur la structure du flux compressé en s'appuyant sur les images de types P et B pour substituer le message à insérer. Dans ce contexte, notre travail consiste à définir et exploiter d'autres régions de la vidéo qui favorisent le tatouage. Notre idée s'oriente non pas vers l'utilisation de la structure ou de la forme d'un signal vidéo mais plutôt vers le contenu. Il s'agit d'extraire de la vidéo des régions d'intérêt où l'insertion de la marque n'induit aucune dégradation visible même avec une force de tatouage assez élevée.

2.5.1 Introduction de la notion de régions d'intérêts

Une région d'intérêt est une zone de la vidéo qui favorise le tatouage. Dans de telles régions, il est permis d'augmenter l'énergie de la marque sans qu'aucune dégradation ne soit perceptible par l'œil humain. Le choix de telles régions revient donc d'une part à la sensibilité du système visuel humain par rapport à certains critères tels que la vitesse de défilement des images, l'amplitude du mouvement dans une séquence vidéo, etc... et d'autre part aux propriétés du contenu de la vidéo. Afin d'aboutir à ces régions d'intérêt, nous proposons d'abord la création d'image mosaïque et ensuite la détection et l'exploitation des zones d'ombre en mouvement.

2.5.2 Notion de mosaïque

L'idée dans ce travail est de sélectionner les zones des images les plus appropriées à l'introduction du tatouage. Pour cela, plusieurs critères devront être pris en compte. Tout d'abord, lorsqu'une marque est introduite dans une zone de l'image, il est nécessaire d'introduire la même marque au même endroit sur toutes les images de la séquence. En effet, dans le cas contraire, un simple filtrage temporel dans le sens du mouvement

(moyennage d'images successives) permet de détruire la marque. Pour parvenir à une détection efficace de telles régions, une phase d'analyse permettant notamment la création d'images mosaïques (ou sprite dans le contexte MPEG-4) devra être réalisée. Le chapitre 3 sera consacré à la présentation du principe de création d'images mosaïques.

2.5.3 Zone d'ombre en mouvement

Il sera également important de détecter les zones moins importantes visuellement dans lesquelles des modifications pourraient être introduites plus facilement sans qu'elles soient visibles. Il s'agit notamment des zones visibles dans peu d'images différentes, ou des zones où l'œil humain est peu susceptible de détecter des différences. Les zones d'ombres en mouvement font partie de ces zones puisqu'il s'agit de régions plus ou moins sombres.

2.6 Conclusion

Le contexte du tatouage vidéo est très différent de celui de l'image. En effet, un flux vidéo possède ses propres caractéristiques qui diffèrent de ceux de l'image. Ces propriétés sont exploitées pour le tatouage afin d'avoir des schémas assez robustes puisque les techniques de tatouage d'images présentent des limites et sont fragiles par rapport aux attaques spécifiques à la vidéo. Dans la partie suivante, nous proposons une nouvelle méthode de tatouage vidéo basée sur la notion de régions d'intérêt. En effet, il s'agit de définir et de sélectionner les régions de la vidéo qui favorisent le tatouage et où l'insertion de la marque se fait d'une manière invisible. Pour cela, nous proposons la création d'images mosaïques. Dans la dernière partie, nous proposons l'exploitation des conditions d'illumination dans une scène pour concevoir un schéma de tatouage adaptatif. L'estimation des conditions d'illumination nécessite la détection des zones d'ombres en mouvement.

Deuxième partie

Exploitation des mosaïques pour le tatouage

Chapitre 3

Création d'images mosaïques

La création d'une mosaïque d'images à partir d'une séquence vidéo est d'une importance capitale tant pour la production cinématographique et la vidéo-surveillance, que pour les techniques de compression et d'indexation de vidéos. Elle consiste à assembler plusieurs images d'une même scène en un panorama.

L'utilisation de la notion d'images mosaïques a pour but de construire la vue la plus complète possible à partir des informations disponibles dans chacune des images utilisées constituant la séquence. Cela suppose que chaque image de la séquence soit recalée, ou repositionnée, dans un modèle de représentation (ou référentiel) commun.

Dans ce chapitre, nous allons présenter le principe général et les hypothèses de base de création d'images mosaïques et les différentes techniques utilisées pour le recalage (la transformation) des images dans le référentiel de la mosaïque en citant les modèles mathématiques utilisés dans la littérature. Nous présenterons par la suite les différentes approches d'estimation des paramètres de recalage.

3.1 Principe et hypothèses de base

Le principe de construction de l'image mosaïque consiste à recaler les images traitées dans le référentiel de l'image mosaïque. Ces images peuvent être utilisées pour représenter sur une seule image la totalité d'une scène. Cette mosaïque est généralement plus grande qu'une image de la séquence.

Pour obtenir une image mosaïque de bonne qualité, toutes les images doivent être recalées le plus précisément possible dans le référentiel de l'image mosaïque, ce qui suppose la prise en compte de leurs conditions respectives d'acquisition. Dans ce cadre, deux problèmes essentiels peuvent survenir.

Tout d'abord, les images originales ne sont pas acquises au même instant (à moins de disposer de plusieurs caméras ou appareils photo parfaitement synchronisés). Il est alors possible que l'image se soit modifiée dans l'intervalle de temps correspondant, ou que les conditions d'illumination soient différentes. De plus, une modification de l'angle de vue ou de la position de la caméra modifie non seulement la partie visible de l'image, mais aussi sa couleur, et cela, même en l'absence de changement d'illumination dans la scène (par exemple à cause de problèmes de prise de vues à contre-jour). [70].

Pour éviter la prise en compte explicite de ces difficultés, la plupart des méthodes existantes de création d'images mosaïques supposent que toutes les vues disponibles représentent la même réalité physique, ce qui revient implicitement à introduire les deux hypothèses suivantes :

- Objet physique considéré comme rigide : Dans le cas des objets non rigides (typiquement un personnage), il n'est pas possible en théorie de créer des images mosaïques. En pratique et selon les types d'objets et l'application considérée, il est néanmoins possible de construire, dans certains cas, des mosaïques dans le cas d'objets non rigides tout en gardant un bon niveau de qualité et de réalisme (voir par exemple l'image mosaïque représentant le fond de la séquence "Stefan" sur la Figure 3.5 dans laquelle le public n'est pas un objet parfaitement rigide).
- Aucune variation d'éclairage dans les images originales : En pratique, cette hypothèse est le plus souvent remplie dans le cas d'images successives d'une vidéo originale, et cela, à moins qu'il se produise une variation brutale des conditions d'illumination dans la scène, ce qui se produit relativement rarement, que la séquence soit longue, que les positions relatives objets/caméra se modifient très rapidement, ou que la caméra effectue des mouvements d'aller-retour. [69].

3.2 Techniques de recalage des images vidéo

L'étape de recalage consiste à définir, tout d'abord, l'image de référence sur laquelle nous alignerons l'image mobile en lui faisant subir des transformations géométriques. Cette étape est le problème central à résoudre pour construire une image mosaïque qui se base sur le recalage des différentes images de la séquence vidéo dans le référentiel de la mosaïque. La qualité de l'image mosaïque obtenue est fortement liée à la qualité de la projection réalisée qui dépend non seulement des limitations théoriques du modèle de recalage utilisé mais aussi de l'algorithme d'estimation des paramètres de ce modèle.

3.2.1 Les modèles de recalage

Le choix d'un modèle de recalage est un élément important de tout algorithme de création d'images mosaïques dans la mesure où le résultat obtenu en dépend directement. Il est alors essentiel de déterminer les conditions d'utilisation optimales des principaux modèles disponibles. Pour cela, il est nécessaire d'évaluer leurs avantages et inconvénients respectifs.

3.2.1.1 Le modèle Translationnel

Ce modèle suppose un mouvement relatif objet/caméra uniquement translationnel et parallèle au plan image. Cette hypothèse rend ce modèle beaucoup trop restrictif dans la plupart des cas.

$$\begin{aligned}v_x(x, y) &= a_1 \\v_y(x, y) &= a_2\end{aligned}\tag{3.1}$$

$v_x(x, y)$ et $v_y(x, y)$ représentent les deux composantes du vecteur vitesse du pixel de coordonnées (x, y) , les coefficients a_1 et a_2 prenant des valeurs réelles. [96]

3.2.1.2 Le modèle Affine

Il est utilisé dans le cas d'images successives avec un mouvement d'amplitude relativement faible. Ce modèle introduit un paramètre de rotation $2D$. En pratique, le modèle

affine est probablement le modèle le plus couramment utilisé [70]. Le modèle Affine est régi par l'équation suivante :

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} u \\ v \end{bmatrix} \quad (3.2)$$

Avec $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ est la matrice rotation et $\begin{bmatrix} u \\ v \end{bmatrix}$ est le vecteur translation, le vecteur $\begin{bmatrix} x \\ y \end{bmatrix}$ représente les coordonnées du pixel considéré dans l'image n , et $\begin{bmatrix} X \\ Y \end{bmatrix}$ les coordonnées du pixel correspondant à $\begin{bmatrix} x \\ y \end{bmatrix}$ dans l'image $n+1$. Les modèles Affines permettent de caractériser une large classe de mouvements $2D$, comme des translations, rotations, homothéties et leurs combinaisons.

3.2.1.3 Le modèle Homographique

Il est théoriquement correct lorsque la position du centre optique de la caméra ne change pas. Ce modèle est utilisé pour estimer le mouvement entre des images plus éloignées temporellement (pour lesquelles le mouvement de la caméra est de plus forte amplitude).

En pratique, il est difficile d'estimer les paramètres de ce modèle si les points de vue des différentes images sont trop proches, comme c'est le cas pour les images successives d'une vidéo. Ce modèle est donc plutôt adapté aux cas où l'on dispose d'un ensemble de photographies [69].

Les équations du modèle homographique sont :

$$X = \frac{a_1x+a_2y+a_3}{b_1x+b_2y+b_3} \quad Y = \frac{a_4x+a_5y+a_6}{b_1x+b_2y+b_3} \quad (3.3)$$

Le vecteur $\begin{bmatrix} x \\ y \end{bmatrix}$ représente les coordonnées du pixel considéré dans l'image n , et $\begin{bmatrix} X \\ Y \end{bmatrix}$ les coordonnées du pixel correspondant à $\begin{bmatrix} x \\ y \end{bmatrix}$ dans l'image $n + 1$ et $a1, a2, a3, a4, a5, a6, b1, b2$ et $b3$ sont les paramètres du modèles homographique, prenant des valeurs réelles.

3.2.1.4 Le modèle Quadratique

Ce modèle est peu utilisé en pratique car il n'apporte pas une amélioration significative par rapport aux modèles précédents. En plus, l'estimation des paramètres quadratiques est souvent difficile et complexe d'un point de vue opératoire d'après [69].

Les équations du modèle quadratique sont :

$$\begin{aligned} X &= a_1 + a_3x + a_4y + a_7xy + a_9x^2 + a_{10}y^2 \\ Y &= a_2 + a_5x + a_6y + a_8xy + a_{11}x^2 + a_{12}y^2 \end{aligned} \quad (3.4)$$

3.2.2 Choix du modèle de recalage

Le choix du modèle de recalage détermine largement la qualité du mouvement apparent estimé. En théorie, plus le modèle est complexe plus il représente correctement le mouvement apparent. Deux approches principales (pouvant être utilisées conjointement) sont possibles pour choisir le modèle de recalage :

3.2.2.1 Choix par défaut

De nombreuses méthodes utilisent systématiquement le même modèle de recalage. Par exemple le modèle Affine pour les séquences, et le modèle homographique si on utilise des images fixes. A l'inverse, l'utilisation d'un modèle plus complexe permet de prendre en compte des mouvements plus complexes, mais les paramètres correspondants sont plus difficiles à estimer.

3.2.2.2 Choix adaptatif du modèle de recalage

Les paramètres de recalage estimés au préalable pour les images fournissent des informations permettant de choisir efficacement le modèle de recalage pour l'image courante. L'idée consiste principalement à conserver, à complexifier ou à simplifier le modèle de recalage en fonction des paramètres et de la qualité de recalage obtenue aux instants précédents.

3.3 Méthodes d'estimation des paramètres de recalage

L'estimation des paramètres de recalage est un problème incontournable dans le domaine du traitement des séquences d'images. Une méthode d'estimation des paramètres de recalage a pour but d'estimer le plus précisément possible les paramètres de recalage de manière à ce que toutes les images soient correctement positionnées les unes par rapport aux autres dans le référentiel de la mosaïque.

Il existe plusieurs méthodes d'estimation des paramètres de recalage, parmi lesquelles, nous citons :

- Block Matching (Mise en correspondance de bloc).
- Flot optique.
- SVD (Singular Value Decomposition).
- ICP (Iterative Closest Point).

3.3.1 Block Matching

3.3.1.1 Définition

Le principe général du *Bloc Matching* (Mise en correspondance de Blocs) est d'exploiter les redondances temporelles existantes entre des images successives. Pour cela, considérons une séquence vidéo dans laquelle nous voulons estimer le mouvement des différentes images qui y sont contenues. Pour simplifier cette estimation, nous ne considérons que le mouvement présent entre deux images successives : l'image courante et l'image de référence. Chaque image est subdivisée en blocs de taille égale (généralement de 8×8 ou 16×16 pixels) et chaque bloc est considéré comme étant un objet indépendant. Nous faisons l'hypothèse que le mouvement des pixels est uniforme à l'intérieur d'un bloc. De manière informelle, l'algorithme consiste à choisir, pour chaque bloc de l'image courante (image n dans la Figure 3.1), un bloc dans l'image de référence (image $n + 1$ dans la Figure 3.1) et à calculer un critère de comparaison entre ces deux blocs. L'opération est répétée en choisissant un autre bloc jusqu'à ce que tous les blocs d'une zone déterminée de l'image de référence (appelée fenêtre de recherche) soient testés ou jusqu'à arriver à

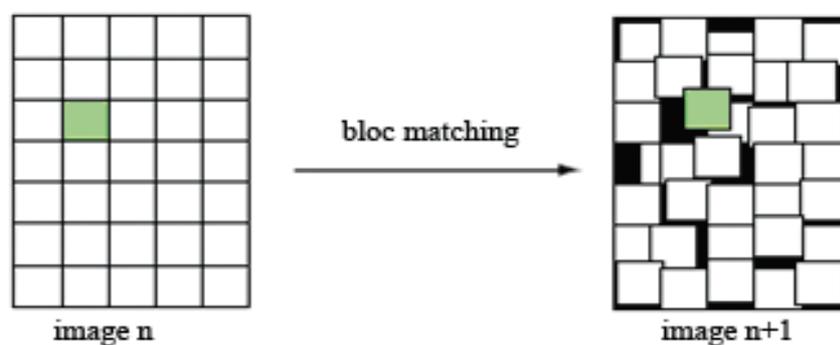


FIG. 3.1: Illustration du Bloc Matching

un critère d'arrêt arbitraire. Le bloc le plus semblable est ainsi identifié dans l'image de référence pour chaque bloc de l'image courante. Nous obtenons de cette manière, pour chaque bloc, un vecteur de déplacement caractéristique du mouvement de ce dernier (voir [29]).

3.3.1.2 Problèmes

La méthode de *Block Matching* facilite l'estimation du mouvement entre deux blocs d'une image à une autre, mais elle utilise malheureusement des vecteurs à valeurs entières, ce qui entre en contradiction directe avec notre conception d'un déplacement continu. En effet, dans le monde réel, il est difficile de concevoir une animation qui évolue uniquement suivant des vecteurs entiers. En plus, la méthode de *Block Matching* paraît très coûteuse en terme de temps de calcul et de ressources utilisées.

3.3.2 Flot Optique

3.3.2.1 Définition et principe

Le *Flot Optique* consiste à estimer le mouvement de chaque pixel à partir des contraintes visuelles. Il se calcule entre deux images. C'est le champ de vecteurs de mouvement, rapportés aux pixels, pour passer d'une image à une autre.

La détection du *Flot Optique* consiste à extraire l'information de la vitesse à partir d'une séquence d'images en faisant l'hypothèse que l'intensité (ou la couleur) des pixels des objets est conservée au cours du déplacement de l'objet.

Soit l'intensité du pixel à la position (x, y) et au temps t . Si $u(x, y)$ et $v(x, y)$ sont les composantes en x et y du vecteur du *Flot Optique* en ce point, nous devons chercher un point dans l'image suivante où l'intensité sera la même au temps $t + dt$ et à la position $(x + dx, y + dy)$:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (3.5)$$

Comme cette équation n'a pas une solution unique, nous devons faire l'hypothèse suivante :
" L'intensité varie lentement avec x , y et t ".

Si cette hypothèse est respectée, nous pouvons développer la partie droite de l'équation 3.5 en série de Taylor :

$$I(x, y, t) = I(x, y, t) + \frac{\partial I}{\partial x} \cdot dx + \frac{\partial I}{\partial y} \cdot dy + \frac{\partial I}{\partial t} \cdot dt + e \quad (3.6)$$

e contenant les termes d'ordres plus élevés.

Si dt tend vers 0 et e est négligeable nous obtenons :

$$\frac{\partial I}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial I}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \quad (3.7)$$

Et puisque,

$$\frac{dx}{dt} = u \quad \text{et} \quad \frac{dy}{dt} = v \quad (3.8)$$

Nous obtenons finalement :

$$\frac{\partial I}{\partial x} \cdot u + \frac{\partial I}{\partial y} \cdot v + \frac{\partial I}{\partial t} = 0 \quad (3.9)$$

L'équation 3.9 est connue sous le nom d'équation de contrainte du flot optique. La Figure 3.2 montre un champ dense de vecteurs (a) extrait à partir de la séquence d'images (b), (c) et (d).

3.3.2.2 Problèmes

Le *Flot Optique* est un problème encore ouvert aujourd'hui puisqu'il est mal posé et le nombre d'inconnues excédant celui des contraintes visuelles. La Figure 3.3 illustre l'ambiguïté qui subsiste en utilisant l'équation de contrainte du *Flot Optique*. Chaque courbe représente un contour d'intensité égale sur deux images consécutives. Ainsi, avec l'équation

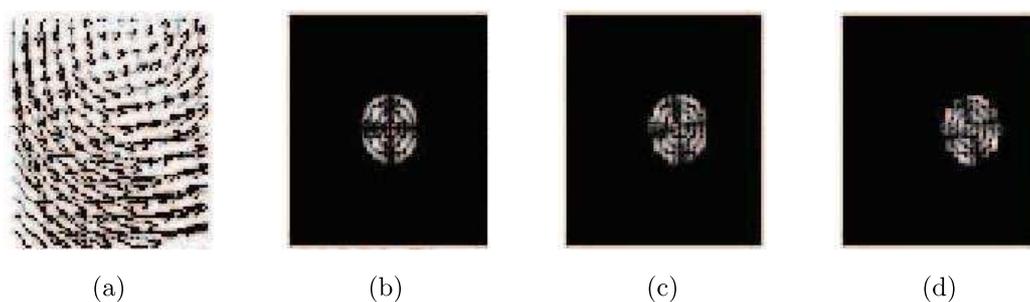


FIG. 3.2: Exemple d'une détection d'un champ dense de vecteurs du mouvement à partir de trois images

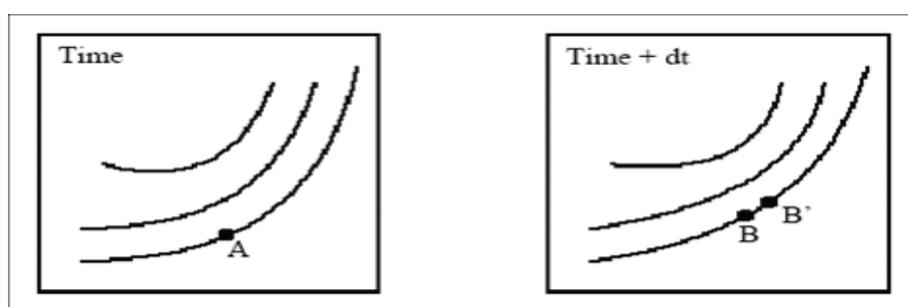


FIG. 3.3: Ambiguïté du flot optique

3.9 dont nous disposons, nous savons uniquement que le point A dans la première image qui s'est déplacé au point B , B' ou même un autre point sur cette même courbe dans la deuxième image. Afin de trouver une solution unique nous avons besoin de contraintes supplémentaires [50].

3.3.3 La factorisation SVD (Singular Value Decomposition)

Elle permet la résolution d'une gamme de problèmes comme par exemple les problèmes des moindres carrés ou des problèmes sous contraintes et bien d'autres, donc SVD sera un outil indispensable.

3.3.3.1 Définition

Soit A , une matrice $m * n$, la décomposition à Valeur Singulière **SVD** (Singular Value Decomposition) est une méthode de décomposition numérique qui permet d'exprimer la

matrice A comme le produit de trois matrices particulières, U , V , et W telles que :

$$A = UWV^T \quad (3.10)$$

U est une matrice $m * n$, orthonormale

W est une matrice $n * n$, diagonale

V est une matrice $n * n$, orthonormale

$$\begin{pmatrix} A \end{pmatrix} = \begin{pmatrix} U \end{pmatrix} \begin{pmatrix} w_1 & 0 & ..0 \\ & w_2 & \\ 0 & .. & w_3 \end{pmatrix} \begin{pmatrix} V \end{pmatrix} \quad (3.11)$$

3.3.3.2 Propriétés

- Les w_i sont les valeurs singulières de A .
- Si la matrice A est singulière, il y a des w_i nuls.
- En général $\text{rang}(A) =$ au nombre de w_i non nuls.
- La décomposition **SVD** est unique.

3.3.3.3 Utilités de SVD

- Première utilité : le traitement des inverses

SVD est un outil indispensable pour le traitement des inverses.

En effet,

$$A^{-1} = (V^T)^{-1}W^{-1}U^{-1} = VW^{-1}U^T \quad (3.12)$$

Car l'inverse d'une matrice orthonormale est sa transposée. W est diagonale, donc W^{-1} est aussi diagonale avec les inverses des w_i .

Mais la matrice A est singulière, donc il existe des valeurs de w_i qui sont nulles, donc l'inverse de W n'existe pas. C'est pourquoi, nous parlerons de pseudo inverse.

Le pseudo inverse de W est la matrice la plus proche de l'inverse de W .

Si $w_i = 0$, nous travaillons avec l'hypothèse : $\frac{1}{w_i} = 0$

- Deuxième utilité : les moindres carrés

Soit le système (S) : $Ax = b$ à résoudre par la méthode des moindres carrés.

$x = \text{pseudo inverse}(A)b$ avec le pseudo inverse de la matrice A est calculé par **SVD**.

3.3.4 ICP : Iterative Closest Point

3.3.4.1 Principe

L'algorithme de plus proche voisin itéré, ou ICP, a été introduit par [8], et utilisé intensivement en imagerie médicale par [25]. Il consiste en une optimisation alternée de correspondances et de transformations.

ICP est utilisé pour raffiner la transformation avant que l'étape finale de vérification confirme ou non la correspondance entre l'image de référence et l'image à transformer. Ainsi, à partir d'une transformation initiale, on réalise les deux étapes suivantes :

1. Mise en correspondance (plus proche voisin) : Cette étape consiste à trouver les points les plus proches entre eux (les points correspondants) entre les deux ensembles de points à recaler.
2. Recalage : Appliquer les transformations géométriques trouvées à l'ensemble des données à recaler en calculant un critère de comparaison (par exemple l'erreur quadratique moyenne MSE, la valeur absolue AV [29] entre les données transformées et les données de référence. Puis, itérer le processus jusqu'à avoir une bonne correspondance entre les points.

3.4 Approche retenue

3.4.1 Principe général

Dans notre travail, nous ne traitons que des scènes statiques c'est à dire nous nous intéressons seulement au fond statique de la scène en faisant abstraction des objets en mouvements. Les hypothèses de base se résument comme suit :

- Pas d'objets physiques en mouvement : nous négligeons le personnage dans notre cas.
- Aucune variation d'éclairage dans les images originales.

3.4.2 Le modèle de Recalage choisi : le modèle Affine

Rappelons que l'étape de recalage consiste à définir, tout d'abord, l'image de référence sur laquelle nous alignerons l'image mobile en lui faisant subir des transformations géométriques. Cette étape est le problème central à résoudre pour construire une image mosaïque qui se base sur le recalage des différentes images de la séquence vidéo dans le référentiel de la mosaïque.

La création d'images mosaïques dans notre cas, s'effectue à partir d'un flux vidéo ou plus précisément, à partir d'une scène vidéo. Une scène est une vidéo ou un extrait d'une vidéo qui représente un événement se déroulant dans un même arrière plan. Nous ne nous intéressons que de l'arrière plan constituant la scène pour la construction de l'image mosaïque. Donc les images extraites de la scène sont très proches. Ce qui explique bien le choix du modèle Affine pour le recalage des images dans le référentiel de la mosaïque puisqu'il est utilisé dans le cas des images successives avec des mouvements faibles entre elles. En plus, le modèle Affine est le modèle le plus couramment utilisé pour les séquences vidéo. Le modèle Affine donne un bon compromis entre une faible complexité et une approximation satisfaisante des mouvements des images de même profondeur.

Nous estimons que chaque pixel voit son mouvement d'une image à l'autre régi par une simple transformation Affine dont l'équation admet six paramètres :

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} u \\ v \end{bmatrix} \quad (3.13)$$

dans laquelle le vecteur $\begin{bmatrix} x \\ y \end{bmatrix}$ représente les coordonnées du point considéré dans l'image n , et $\begin{bmatrix} X \\ Y \end{bmatrix}$ les coordonnées du point correspondant à $\begin{bmatrix} x \\ y \end{bmatrix}$ dans l'image $n + 1$ et a, b, c, d, u et v sont les paramètres du modèle Affine à déterminer.

3.4.3 Estimation des paramètres de recalage

Le problème central dans l'estimation de mouvement est la résolution des paramètres de recalage, ce qui revient à résoudre les paramètres du modèle choisi, qui est dans notre cas, le modèle Affine. Le processus de recalage des images peut ensuite se diviser en

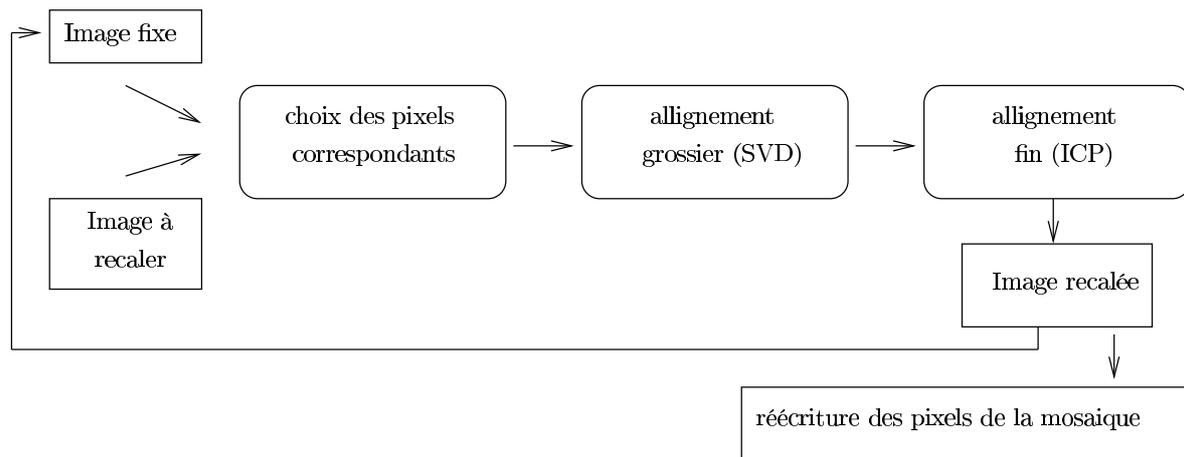


FIG. 3.4: Les différentes étapes de recalage.

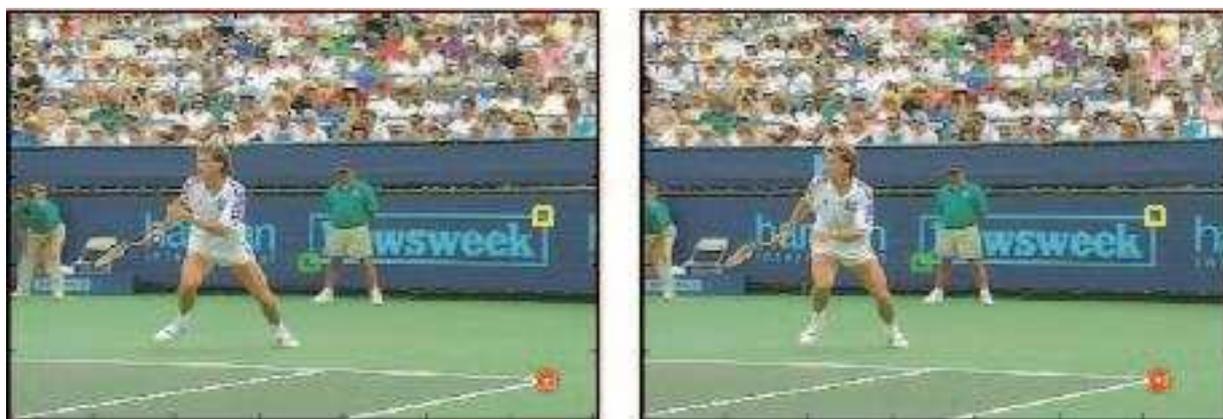


FIG. 3.5: Choix des pixels correspondants entre deux images

trois étapes : choix des pixels correspondants, alignement grossier avec **SVD** et ensuite alignement fin avec **ICP**.

3.4.3.1 Choix des pixels correspondants

Après avoir défini l'image fixe qui va servir comme une image de référence, et l'image à recaler, nous choisissons manuellement les pixels correspondants qui sont au nombre de trois couples correspondants. La Figure 3.5 montre bien comment choisir les points correspondants entre deux images. Le fait de fixer le nombre de couples correspondants à trois est inspiré de notre système qui comporte six inconnus à résoudre. Donc, nous avons besoin de construire un système de six équations par les six points sélectionnés afin de trouver les six paramètres du modèle Affine choisi. Il est à noter que chaque couple forme

deux équations comme suit :

Soient P_1 et P_2 deux pixels de coordonnées respectives (x_1, y_1) et (X_1, Y_1) .

$$\begin{aligned} X_1 &= ax_1 + by_1 + u \\ Y_1 &= cx_1 + dy_1 + v \end{aligned} \quad (3.14)$$

Avec (P_1, P'_1) est un couple de points correspondants.

Une fois les pixels correspondants sont choisis, nous faisons recours à l'alignement grossier qui consiste à déterminer une première approximation des paramètres de recalage du modèle Affine. Ensuite, nous raffinons les résultats obtenus en itérant l'algorithme de l'alignement grossier plusieurs fois jusqu'à atteindre une condition d'arrêt. C'est l'étape de l'alignement fin. En fin, nous appliquons les transformations géométriques obtenues à l'image initiale (Image à Recaler), afin de trouver une image recalée (transformée géométriquement par rapport à l'image de référence).

Revenons au modèle Affine qui peut être défini de la façon suivante :

$$P' = R \times P + t \quad (3.15)$$

avec

$$R = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad (3.16)$$

θ est l'angle de rotation.

R représente la matrice rotation et t le vecteur translation qui font transformer le pixel P vers le pixel P' de coordonnées respectifs (x, y) et (X, Y) .

Dans ce qui suit, nous allons définir les différentes étapes abordées et les différents algorithmes retenus pour estimer les paramètres du modèle Affine proposé.

3.4.3.2 Alignement grossier

L'alignement grossier s'effectue en appliquant la factorisation à valeur singulière SVD pour obtenir une première résolution de la matrice rotation et de vecteur translation du modèle Affine choisi. Soit *imgRef* l'image de référence fixe sur laquelle nous recalons une image appelée *imgRecal*. Soient P_1 , P_2 et P_3 trois pixels de *imgRecal* et P'_1 , P'_2 et P'_3 leurs pixels correspondants dans l'image *imgRef*, avec (x_1, y_1) , (x_2, y_2) , (x_3, y_3) ,

$(X1, Y1), (X2, Y2), (X3, Y3)$ sont les coordonnées respectifs de $P_1, P_2, P_3, P'_1, P'_2$ et P'_3 . Soit P l'ensemble qui regroupe les pixels P_1, P_2 et P_3 et soit P' l'ensemble qui regroupe les pixels P'_1, P'_2 et P'_3 .

Nous définissons la matrice A de la façon suivante :

$$A = (P')^T(Q) \quad (3.17)$$

Avec $Q = P - \text{mean}(P)$ et $P' = P' - \text{mean}(P')$.

"mean" est la fonction moyenne.

D'après la Décomposition à Valeur Singulière, la matrice A peut s'écrire sous la forme :

$$A = UWV^T \quad (3.18)$$

Avec U et V^T sont deux matrices orthonormales et W est une matrice diagonale.

La méthode SVD permet d'obtenir la matrice rotation et le vecteur translation par les formules suivantes :

$$R = V * U^T \quad (3.19)$$

$$t = \text{mean}(P) - R * (\text{mean}(P')) \quad (3.20)$$

C'est une première approximation de la matrice de rotation R et du vecteur de translation t .

3.4.3.3 Alignement fin

L'alignement fin s'effectue en appliquant l'algorithme ICP pour raffiner la transformation avant que l'étape finale de vérification confirme ou rejette la correspondance entre les deux images choisies [79]. Dans ce qui suit, nous énumérons les étapes de l'algorithme de l'alignement fin :

1. Pour chaque point dans l'image à recaler (*imgRecal*), trouver le point le plus proche dans l'image de référence (*imgRef*). Ces deux points forment une correspondance. La figure 13 montre le principe de choix des points correspondants entre eux.
2. Trouver la rotation/translation relative qui minimise la distance entre les deux points de chaque correspondance. C'est le premier recalage qui est initialisé par les valeurs obtenues par SVD.

3. Appliquer la transformation géométrique trouvée à *imgRecal*.
4. Itérer jusqu'à avoir une bonne correspondance entre l'image de référence et l'image à recalcr. Une bonne correspondance peut se traduire par une minimisation de la différence entre l'image de référence et l'image recalée.

Le pseudo algorithme de ICP est le suivant :

```

function ICP(Scene, Model)
begin
  E' ← +∞ (Rot, Trans) ← Initialize - Alignment(Scene, Model); //par SVD
repeat
  E ← E';
  Aligned - Scene ← Apply - Alignment(Scene, Rot, Trans);
  Pairs ← Return - Closest - Pairs(Aligned - Scene, Model);
  (Rot, Trans, E') ← Update - Alignment(Scene, Model, Pairs, Rot, Trans);
Until |E' - E| < Threshold
return (Rot, Trans);
end

```

3.4.4 Réécriture de la mosaïque

Après l'étape de recalage, la construction de l'image mosaïque peut alors débuter, il suffit d'écrire les pixels de l'image de référence et de l'image recalée dans une seule image pour représenter la mosaïque. Le résultat est une mosaïque formée par deux images. Pour itérer le processus de "Mosaïcing", l'image recalée devient la nouvelle image de référence et nous choisissons une nouvelle image à recalcr par rapport à la dernière image déjà recalée. Normalement, le processus de recalage s'effectue en se référant à l'image mosaïque nouvellement construite, c'est à dire c'est l'image mosaïque qui devient l'image de référence. Et puisque la dernière image insérée dans la mosaïque est l'image déjà recalée, il suffit donc d'appliquer le recalage de la nouvelle image à recalcr par rapport à cette dernière image recalée et insérée dans la mosaïque.

3.5 Conclusion

Dans ce chapitre, nous avons appliqué le modèle Affine comme modèle de recalage et nous avons choisi les algorithmes **SVD** et **ICP** comme des algorithmes d'estimation des paramètres de recalage. Nous exploitons les caractéristiques des images mosaïques pour définir une nouvelle approche de tatouage vidéo robuste, tel sera l'objet du chapitre suivant.

Chapitre 4

Approche de tatouage vidéo basée sur la notion d'image mosaïque

4.1 Introduction

L'objectif de cette approche n'est pas de fournir un algorithme "universel", robuste à toutes les attaques mais plutôt d'apporter de nouvelles solutions afin d'améliorer la robustesse des techniques existantes. Ceci peut être fait à travers la proposition de nouvelles régions intéressantes pour le tatouage vidéo.

Dans ce chapitre, nous présentons les différentes régions d'intérêt adoptées. Ensuite, nous mettons en œuvre celles-ci pour l'élaboration d'une approche de tatouage vidéo. Une phase de tests et d'expérimentations sera effectuée dans le but de montrer l'efficacité de ces régions.

4.2 Aperçu général de l'approche

L'approche proposée est une approche additive dans le domaine spatial. Elle est basée sur la technique d'étalement de spectre. Aussi, c'est une approche semi-aveugle puisque la marque originale est nécessaire dans la phase de détection. En fait, et comme le montre la figure 4.1 notre approche est composée de trois modules principaux : un module pour l'insertion de la marque, un module pour la détection et un autre pour la sélection des

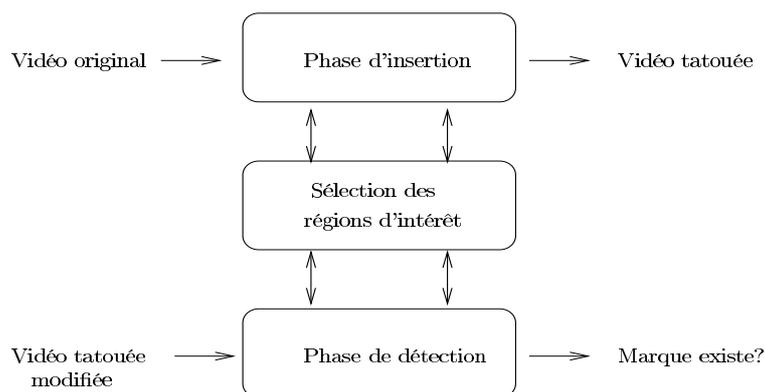


FIG. 4.1: Diagramme global de l'approche de tatouage proposée

régions d'intérêt. En effet, ce dernier module consiste à définir et sélectionner les zones appropriées dans la vidéo où la marque doit être insérée et ensuite détectée. C'est pour cela que ce module est partagé entre les modules d'insertion et de détection. Dans ce qui suit, nous détaillons chacun de ces modules afin d'expliciter les étapes de l'approche proposée.

4.2.1 Sélection des régions d'intérêt

On appelle "région d'intérêt" une zone dans la vidéo qui favorise l'insertion de la marque. En effet, l'insertion de la marque dans ces zones améliore nettement la robustesse de l'algorithme de tatouage en permettant d'augmenter l'énergie de la marque sans altération visible du signal traité. Une séquence vidéo est une suite d'images (25 images/s) qui sont très corrélées entre elles. On peut trouver du son, du mouvement des objets, du mouvement de la caméra, des changements de scènes, différentes conditions d'illumination, etc. Cette diversité d'informations dans une séquence vidéo peut fournir plusieurs régions d'intérêt pour le tatouage. Dans notre cas, on s'intéresse essentiellement aux régions d'intérêt suivantes : les pixels qui représentent le même point physique (point réel), on s'intéresse ici à localiser ces pixels et à estimer leurs occurrences tout au long de la séquence ; les images où le mouvement de la caméra est rapide ainsi que les zones d'ombres en mouvement. Dans la suite nous nous intéressons aux deux premiers critères. La partie suivante sera consacrée à l'étude et l'exploitation des conditions d'illumination, notamment les zones d'ombres en mouvement.

Notre approche a pour objectif d’exploiter la redondance de l’information tout au long de la composante temps dans une séquence vidéo. Cette redondance se manifeste, par exemple, par la présence d’un même point physique à plusieurs images et à plusieurs endroits. Ceci est dû au mouvement de la caméra. Notre idée, par rapport à cela, est de tatouer tous les points représentant le même point physique par la même marque. Avec ce procédé, on peut garantir une robustesse face à une attaque typiquement vidéo qui est la ”Collusion”, largement explicitée dans ([20]). Son principe est de remplacer une image de la séquence vidéo par la moyenne (interpolation) de ses voisins après un recalage dans le sens du mouvement. Nous verrons, dans le chapitre ”Expérimentations et résultats” de plus amples détails sur cette attaque. Pour aboutir à une bonne localisation de ces points, la création d’image mosaïque semble la solution la plus adéquate. En effet, une image mosaïque représente, en une seule image (voir Figure 4.2), le fond fixe de la scène tout au long de la séquence faisant abstraction des objets en mouvement.

4.2.1.1 Localisation des points physiques similaires

Comme expliqué dans le chapitre 3, l’image mosaïque est créée suite à la projection de chaque image F_i de la scène dans un nouveau référentiel suivant un vecteur θ_i associé à chacune de ces images. Ce vecteur n’est autre qu’une estimation de mouvement moyennant un modèle Affine. Comme illustré dans la figure 4.3, les points P et $P + d\theta$ respectivement situés dans les images F_i et F_{i+1} représentent un même point physique. $d\theta$ représente le déplacement de la caméra. Puisqu’ils représentent un même point physique, ils seront donc projetés sur un même point P' au niveau du référentiel de la mosaïque. En résumé, tous les points qui représentent le même point physique seront tous projetés sur un seul point dans la mosaïque. En d’autre terme, et dans un processus inverse, à partir d’un point de la mosaïque on pourra localiser, par projection inverse et moyennant les vecteurs de recalage θ_i , tous les points similaires tout au long de la scène.

4.2.1.2 Estimation de l’occurrence des pixels

Une fois toutes les occurrences d’un même point physique sont localisées, on pourra bien évidemment estimer l’occurrence de chacun tout au long de la scène. L’occurrence d’un point désigne en fait le nombre d’images où apparait ce point. Alors, si un point

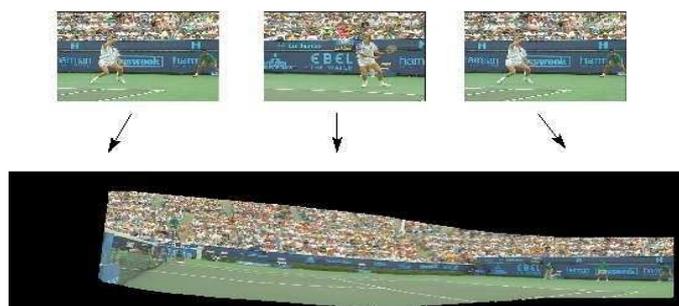


FIG. 4.2: Image mosaïque issue de la séquence de "Stefan"

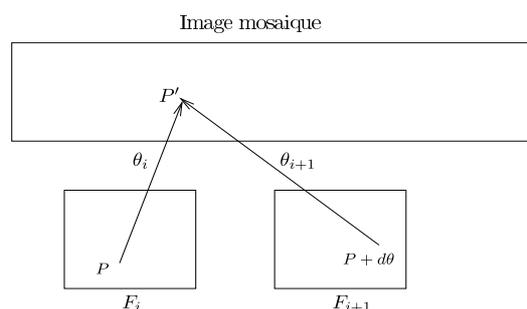


FIG. 4.3: Projection des pixels similaires sur le même point au niveau de la mosaïque

apparaît dans peu d'images, on peut dire que ce point est invisible à l'œil nu puisqu'on dispose de 25 images/s. Par rapport à notre contexte, de tels points sont considérés comme points intéressants pour le tatouage. En effet, une modification ou même une altération au niveau de ces points n'influe pas sur l'aspect visibilité de la séquence. Pour ce fait, nous envisageons d'augmenter la force de tatouage sur ces zones afin d'améliorer la robustesse de l'algorithme. On verra dans la partie d'expérimentation que ce procédé permet de faire face à l'attaque de compression qui est une attaque générant une perte de données. Après de telle attaque, la marque peut toujours être récupérée.

4.2.1.3 Estimation de la vitesse de la caméra

Le mouvement de la caméra peut être considéré dans la sélection de certaines régions d'intérêt pour le tatouage. En général, la caméra, dans une séquence vidéo, peut être fixe ou mobile. Dans ce dernier cas, la caméra peut faire des translations, des rotations ou bien même des effets de zoom pour approcher la cible. Quand le mouvement de la caméra est rapide, il devient difficile au spectateur, en particulier à l'œil humain, de se concentrer sur

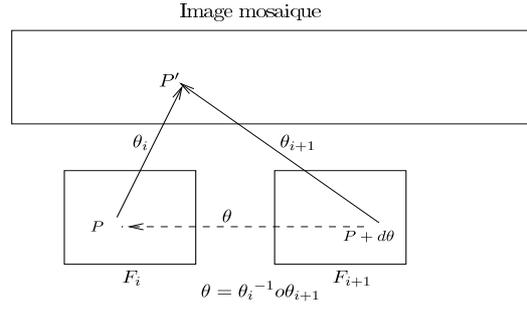


FIG. 4.4: Compensation du mouvement entre deux images successives

le contenu de la scène puisque la capacité d'assimilation de la rétine est limitée. Du coup, celle-ci devient moins sensible à toute modification au niveau du contenu. Nous profitons de ce fait pour augmenter l'énergie de la marque dans les images où le mouvement de la caméra est rapide pour gagner en terme de robustesse. Pour estimer la vitesse de la caméra, le mouvement entre deux images successives doit être compensé. Ceci peut être effectué au moyen des paramètres de mouvement associés à chaque image de la séquence. La figure 4.4 illustre comment l'estimation du mouvement entre deux images est réalisée. Ayant θ_i et θ_{i+1} les paramètres de mouvement associés respectivement aux images F_i et F_{i+1} alors le vecteur estimant le mouvement entre ces deux images est donné par l'équation 4.1 suivante :

$$\theta = \theta_i^{-1} + \theta_{i+1} \quad (4.1)$$

L'inverse du paramètre θ_i est obtenu selon l'équation suivante :

$$\theta_i^{-1} = \begin{pmatrix} t_x^{-1} \\ t_y^{-1} \\ k^{-1} \\ \beta^{-1} \end{pmatrix} = \begin{pmatrix} \frac{-(\beta t_y + t_x(1+k))}{A} \\ \frac{\beta t_x - t_y(1+k)}{A} \\ \frac{-\beta^2 - k(1+k)}{A} \\ \frac{-\beta}{A} \end{pmatrix} \quad (4.2)$$

avec $A = (1 + k)^2 + \beta^2$. t_x et t_y sont les coefficients de translation. k et β représentent respectivement les facteurs de zoom et de rotation.

4.2.2 Phase d'insertion

La phase d'insertion (Figure 4.5) peut être décrite par les trois étapes suivantes : création de la mosaïque, conception de la marque et l'insertion de la marque

1. *Conception de la marque.* La marque est créée selon les étapes suivantes :
 - **Création de la Mosaïque Marque.** La mosaïque marque est une suite binaire générée aléatoirement au moyen de la technique d'étalement de spectre [18] et utilisant comme graine un code secret. Celui-ci garantit l'unicité du signal généré. La taille du signal est égal à la taille de la mosaïque.
 - **Création du masque.** L'objectif de créer un masque est d'avoir une marque adaptative au contenu, particulièrement aux caractéristiques de chaque pixel. En effet, dans le but d'augmenter la robustesse de l'algorithme, il est possible d'augmenter l'énergie du tatouage dans les zones appropriées où l'œil humain est moins sensible aux modifications éventuelles. Pour cela, certains critères ont été adoptés :
 - (a) Le premier critère consiste à sélectionner les pixels appartenant aux régions texturées et aux contours. Ces régions sont obtenues en appliquant un filtre de Sobel qui estime l'activité local autour de chaque pixel tenant en compte la sensibilité du Système Visuel Humain. Pour minimiser la complexité des calculs, le filtre de Sobel est appliqué une seule fois sur l'image mosaïque et non pas sur toutes les images de la scène.
 - (b) Le deuxième critère considéré est la fréquence d'apparition de chaque pixel au sein de la séquence vidéo. En effet, l'idée consiste à introduire plus d'énergie dans les pixels qui n'apparaissent que rarement dans la séquence, c'est à dire dans peu d'images. Toujours en exploitant la faiblesse du Système Visuel Humain, il est permis de modifier ou même dégrader ces zones qui sont imperceptibles à l'œil nu.
 - **Création des images marques.** La dernière étape consiste à générer la marque correspondante à chaque image de la séquence. Tout d'abord, la mosaïque marque et le masque défini précédemment sont mixés pour aboutir à ce qu'on appelle "Mosaïque Marque Adaptative". Ensuite, ayant les vecteurs de recalage des images de la séquence, les images marques sont obtenues en faisant une projection inverse partant du référentiel de la mosaïque vers le niveau image (voir Figure Figure 4.7). De cette manière, tous les points représentant le même point physique auront la même marque.

$$IM_i(P) = IM_{i+1}(P + d\theta) \quad (4.3)$$

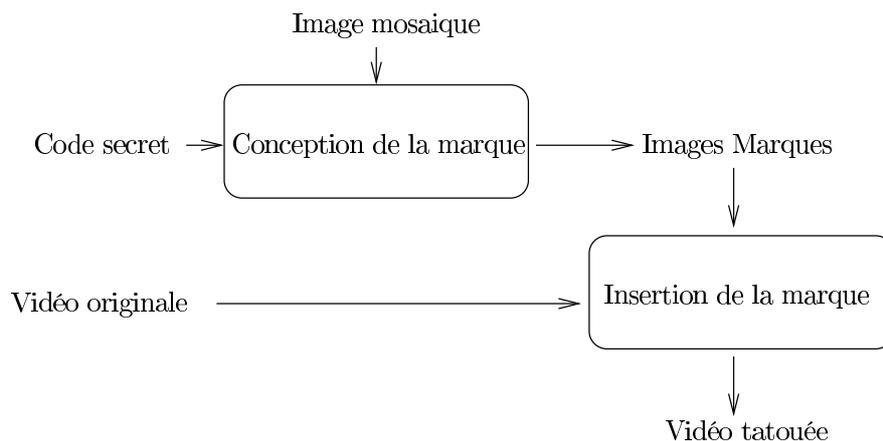


FIG. 4.5: Schéma d'insertion de la marque

Aussi et à ce niveau, un autre critère peut être considéré et pris en compte ; il s'agit de la vitesse de la caméra. Comme on l'a expliqué dans la section 4.2.1.3, il est possible de modifier d'avantage les images où le mouvement de la caméra est rapide. En effet, plus le mouvement de la caméra est rapide, moins l'œil est capable de détecter les éventuelles dégradations dans le contenu.

Pour chaque image de la séquence, nous associons un coefficient γ_i qui représente le taux de dégradation acceptable dans l'image d'indice i et il est calculé selon les équations 4.4 et 4.5.

$$eqm = \frac{1}{M \times N} \times \sum \sum IM_i^2 \quad (4.4)$$

$$\gamma_i = \sqrt{\frac{255^2}{eqm \times 10 \frac{PSNR}{10}}} \quad (4.5)$$

$M \times N$ représente la taille de chaque image. eqm désigne l'Erreur Quadratique Moyenne.

L'idée, en fait, est de spécifier une valeur de $PSNR$ à chaque image qui reflète la qualité qu'on souhaite obtenir après la phase de tatouage. La valeur de $PSNR$ dépendra de la vitesse de la caméra de manière inversement proportionnelle. Maintenant reste à voir comment classer les différentes valeurs de vitesse et dire si une vitesse est faible, moyenne ou rapide. Dans ce cas, on est devant un problème de

logique flou. Pour donner un sens à la vitesse de la caméra, une analyse statistique a été présentée dans [52]. Cette analyse consistait à donner une vingtaine de séquences ayant chacune un niveau de vitesse bien particulier et c'est à eux de les classer. Le résultat de cette analyse est qu'une vitesse est classée moyenne si elle est comprise entre 12 et 52 (nombre de pixels déplacés entre deux images successives). Tenant compte de cette étude, on peut définir les valeurs de $PSNR$ selon la fonction linéaire définie par la Figure 4.6.

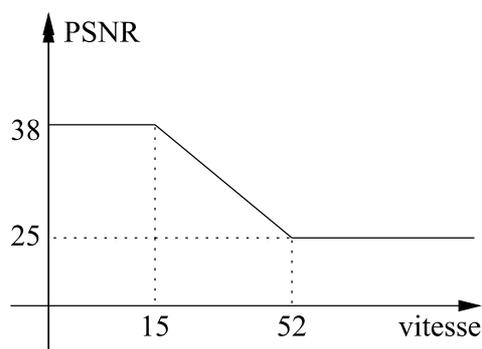


FIG. 4.6: Fonction linéaire

2. *Insertion de la marque.* Une fois les images marques sont créées, l'insertion se fait d'une manière additive selon l'équation suivante :

$$\hat{F}_i = F_i + IM_i \quad (4.6)$$

4.2.3 Phase de détection

La détection de la marque nécessite la connaissance de la clé secrète associée à la séquence vidéo. Cette clé contient deux données essentielles :

1. Le code permettant de générer la marque originale.
2. Les paramètres liés au processus de compensation de mouvement. Selon l'application visée, deux cas sont alors possibles :
 - Seulement les paramètres de l'algorithme utilisés pour la création de la mosaïque sont fournis. Dans ce cas, cet algorithme est nécessaire pour le calcul des paramètres de mouvement. Un utilisateur malhonnête peut bien évidemment générer

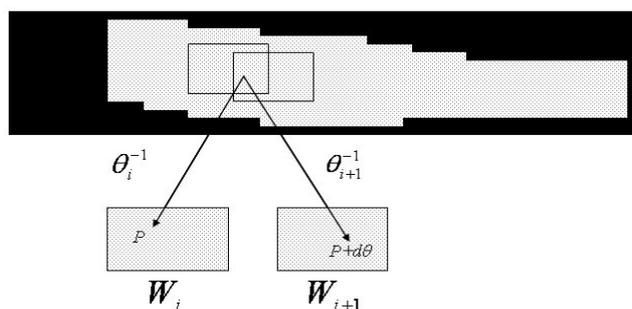


FIG. 4.7: Création des images marques

une image mosaïque utilisant n'importe quel algorithme de mosaïcing, mais il n'aura jamais exactement les mêmes paramètres de mouvement utilisés pour l'insertion de la marque. Par conséquent, la marque ne peut pas être localisée tout au long de la séquence et elle ne peut donc pas être récupérée. Cette propriété offre une meilleure protection à la séquence vidéo tatouée. Il faut noter que ceci est possible seulement si la séquence vidéo a été manipulée. Effectivement, si par exemple, la séquence subit une compression alors les paramètres de mouvement ne peuvent plus être récupérés. Le cas suivant se présente alors.

- Les paramètres de mouvement (un vecteur par image) sont inclus dans la clé secrète. Dans ce cas, la connaissance de la clé décryptée est suffisante pour récupérer la marque, même si la vidéo originale a été modifiée par un processus de compression par exemple.

Une fois la clé secrète est décryptée, la phase de détection de notre approche est similaire à celle de la phase d'insertion. D'abord, les paramètres de mouvement sont extraits de la clé secrète, ou recalculés en utilisant les informations contenues dans la clé. Un vecteur de mouvement θ_i est alors associé à chaque image F_i de la séquence vidéo.

En utilisant le code secret, la mosaïque marque W est générée. Ayant les vecteurs de mouvement θ_i et la mosaïque marque W , nous pouvons déterminer les images marques originales W_i insérées dans chaque image F_i .

La détection est basée sur l'hypothèse suivante :

$$H_0 : \widehat{F}_i - F_i = N_i \text{ (image non tatouée)}$$

$$H_1 : \widehat{F}_i - F_i = N_i + \widehat{W}_i \text{ (image tatouée)}$$

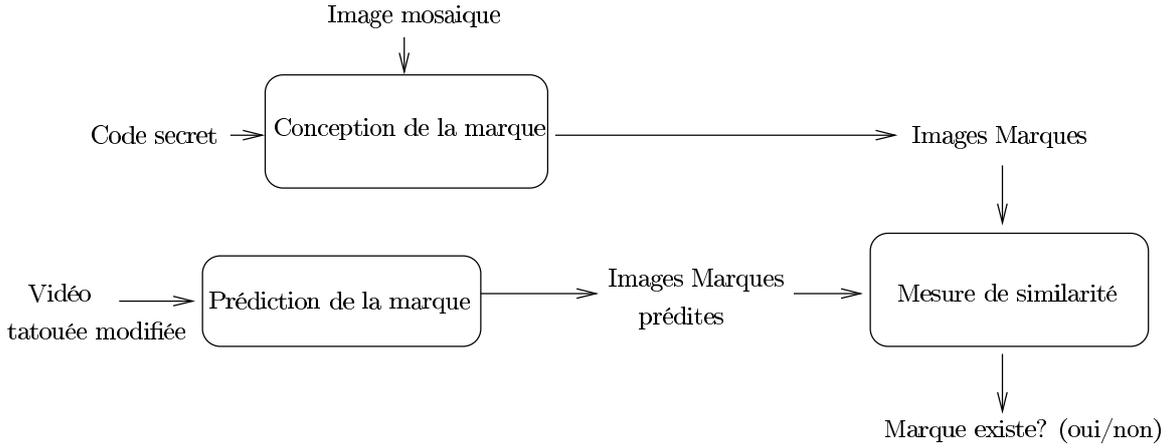


FIG. 4.8: Schéma de détection de la marque

où \hat{F}_i est l'image test d'indice i , \hat{W}_i est la marque récupérée à partir de l'image et N_i un éventuel bruit dû à une transmission ou une quelconque manipulation de la séquence vidéo.

L'hypothèse de décision est obtenue par le calcul de la corrélation (similarité) qui existe entre chaque image tatouée et sa correspondante image marque originale W_i .

$$C = \frac{1}{N} \sum_i \text{Corr}(\hat{F}_i - F_i, W_i) = \frac{1}{N} \sum_i (\hat{F}_i - F_i) \cdot W_i \quad (4.7)$$

C est la corrélation totale entre la scène tatouée et la marque originale. Elle est comparée à un seuil λ pour déterminer si la séquence de test est tatouée ou non. Pour améliorer la détection de la marque, nous avons utilisé un filtre de Wiener[44] pour prédire l'image marque \hat{W}_i de l'image de test. L'image marque est définie comme suit :

$$\hat{W}_i = \frac{V_l[W_i]}{V_l[W_i] + V_l[\hat{I}_i]} (\hat{I}_i - M_l[\hat{I}_i]) \quad (4.8)$$

où V_l et M_l représentent respectivement la variance locale et la moyenne locale du signal en question. Elle sont calculées sur des blocs de taille 3*3 autour de chaque pixel. Alors l'équation 4.7 devient :

$$C = \frac{1}{N} \sum_i \text{Corr}(\hat{W}_i, W_i) = \frac{1}{N} \sum_i (\hat{W}_i) \cdot W_i \quad (4.9)$$

4.3 Estimation du seuil de détection

La décision finale est obtenue à partir de la somme des valeurs de corrélation de toutes les images de la séquence vidéo. La valeur globale est comparée à un seuil qui dépend de la probabilité de fausse alarme (dire qu'un contenu réellement non tatoué est tatoué) qui est désigné par $\lambda(p_{fa})$.

$$corr = \frac{1}{N} \sum_1^N corr(frame_i) > \lambda(p_{fa}) \quad (4.10)$$

N désigne le nombre d'images dans la séquence.

Selon les travaux cités dans [82], la valeur du seuil peut être estimée après une analyse statistique. en effet, la loi de probabilité de la variable aléatoire $X = corr(k1, k2)$ est estimée. $k1$ et $k2$ représentent deux codes distincts qui sont utilisés respectivement dans la phase d'insertion et dans la phase de détection de la marque.

Par une théorie statistique, la loi de probabilité de X est :

$$P_{fa} = \frac{1}{2} erfc\left(\frac{\lambda(P_{fa})}{\sqrt{2 \cdot corr^2}}\right) \quad (4.11)$$

Quand P_{fa} égale à 10^{-8} , l'équation (4.11) devient :

$$\lambda(P_{fa}) = 3.97\sqrt{2 \cdot corr^2} \quad (4.12)$$

La valeur de $corr$ est égal à la moyenne des valeurs de corrélation obtenues avec 200 codes différents.

Chapitre 5

Expérimentations et résultats

5.1 Introduction

La robustesse de l'approche de tatouage proposée est testée par rapport à certaines attaques spécifiques à la vidéo à savoir la collusion, la suppression d'images et la compression utilisant le codec MPEG4. Nous avons considéré deux séquences vidéo qui sont "Stefan" et "Thabor". Ce sont des séquences au format CIF¹.

5.2 Expérimentation sans attaque

Avant de passer à l'évaluation de la robustesse de l'approche face aux différentes attaques, nous allons d'abord appliquer la phase de détection sur des séquences vidéo tatouées mais non attaquées. Ceci permet d'une part de valider l'approche, c'est à dire que la marque peut être récupérée sans problème si aucune perturbation n'est parvenue au niveau de la séquence tatouée. D'autre part, les résultats obtenus serviront comme références pour évaluer et comparer ceux obtenus après une phase d'attaque. Mise à part la validation de l'approche, nous mettons en évidence dans cette partie l'effet du filtre de Wiener sur la détection. En effet, ce type de filtre améliore nettement les résultats de détection en faisant une prédiction de la marque insérée dans la séquence vidéo de test. La Figure 5.1 montre la performance de ce filtre. On voit bien qu'avec l'utilisation de ce

¹Commun Interface Format 288x352

filtre, le pic de corrélation au niveau de la clé qui a servi à l'insertion est nettement mis en évidence. Par contre, dans la deuxième courbe où le filtre n'est pas appliqué, même si le pic de corrélation existe, on a du mal à le distinguer parmi d'autres pics.

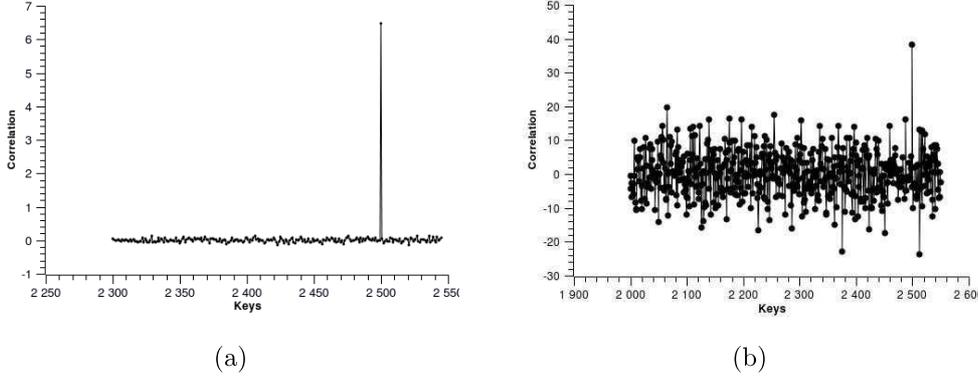


FIG. 5.1: Séquence de Stefan. Valeurs de corrélation avec différentes clés. (a) utilisant le filtre de Wiener. (b) sans le filtre de Wiener.

5.3 Évaluation face à l'attaque de Collusion

L'attaque de collusion consiste essentiellement à moyenner les images successives dans le but d'éliminer la marque sans nuire à la qualité de la séquence. Avant de faire le moyennage, il est nécessaire d'estimer et de compenser le mouvement relatif de la caméra par rapport à la scène entre deux frames successives. Ceci est fait dans le but de déterminer la région de couverture, c'est-à-dire la zone similaire entre ces deux frames où le moyennage est appliqué. Si la marque n'est pas appliquée de manière identique sur chaque pixel représentant le même point physique alors la marque peut facilement être détruite.

Trois étapes sont nécessaires pour la collusion :

1. Compensation de mouvement. Dans cette étape, le mouvement est compensé entre F_{i+1} et F_i en utilisant $\theta = \theta_i \circ \theta_{i+1}^{-1}$. Ceci permet d'avoir une image compensée I .
2. moyennage des images. Cette étape consiste à moyenner les deux images I et F_i en utilisant la fonction m comme suit :

$$m : \begin{cases} \tilde{I}(k, l) = \frac{1}{2}(\hat{F}_i(k, l) + I(k, l)) \\ 0 \leq k \leq M, 0 \leq l \leq N \end{cases} \quad (5.1)$$

\tilde{I} est le résultat. $M \times N$ désigne la taille des images de la séquence.

3. Compensation inverse de mouvement. L'image tatouée et attaquée est obtenue par une compensation inverse du mouvement.

En pratique, deux cas sont possibles :

1. Si les paramètres de mouvement utilisés durant la phase d'insertion correspondent exactement au mouvement de la caméra, alors l'attaque de collusion ne peut pas détruire la marque sans dégradation visible des images de la séquence.
2. En pratique, ce cas ne peut jamais exister du fait que le modèle de mouvement ne peut pas représenter parfaitement le mouvement 3D de la caméra, et la détection dépend dans ce cas des paramètres de mouvement utilisés dans la phase de collusion.

Trois cas sont possibles :

- Les paramètres de mouvement utilisés dans la phase d'insertion et la phase de collusion sont identiques : la marque ne peut pas être détruite.
- Les paramètres de mouvement utilisés dans la phase d'insertion sont plus précis que ceux utilisés dans la phase de collusion : la marque peut être récupérée. Autrement, la vidéo attaquée doit être trop endommagée.
- Les paramètres de mouvement utilisés dans la phase de collusion sont mieux estimés que ceux utilisés dans la phase d'insertion : dans ce cas, une partie de la marque peut être détruite mais pas toute la marque. En effet, du fait que les paramètres de mouvement utilisés dans la phase d'insertion ont servi à créer une image mosaïque de qualité très acceptable, ceci veut dire qu'une grande partie de pixels ont été bien estimés et donc la marque sur ces pixels ne sera pas détruite. En plus, il est important de mentionner que même si la marque a été détruite sur certains images de la séquence, elle peut être sauvegardée dans d'autres. La totalité de la vidéo reste donc tatouée.

Pour évaluer l'approche de tatouage proposée, nous avons appliqué l'attaque de collusion sur les vidéos tatouées de "Stefan" et "Thabor" en suivant les cas discutés ci-dessus. Évidemment, le premier cas n'est pas testé puisque nous ne pouvons jamais faire une estimation idéale du mouvement et avoir les vecteurs de mouvements réels. Dans le premier cas, où les paramètres de mouvement utilisés dans la phase d'insertion et dans la phase d'attaque de collusion sont identiques, nous obtenons le résultat illustré dans la figure

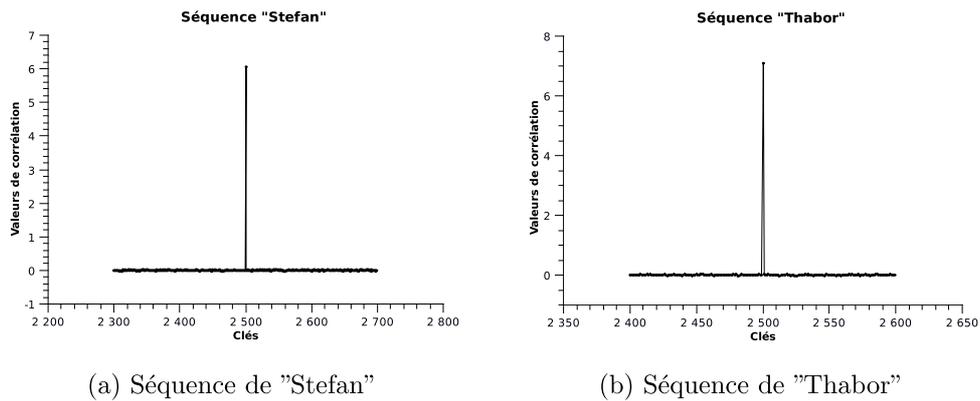


FIG. 5.2: Corrélation globale après l'attaque de collusion utilisant les mêmes paramètres de mouvement dans les phases d'insertion et de collusion

5.2 qui montre les valeurs de corrélation après la phase de détection en utilisant 200 clés différentes. Pour les deux vidéos de test, une valeur pic est obtenue au niveau de la clé utilisée pendant la phase d'insertion. Nous concluons que la marque peut toujours être facilement récupérée après l'attaque.

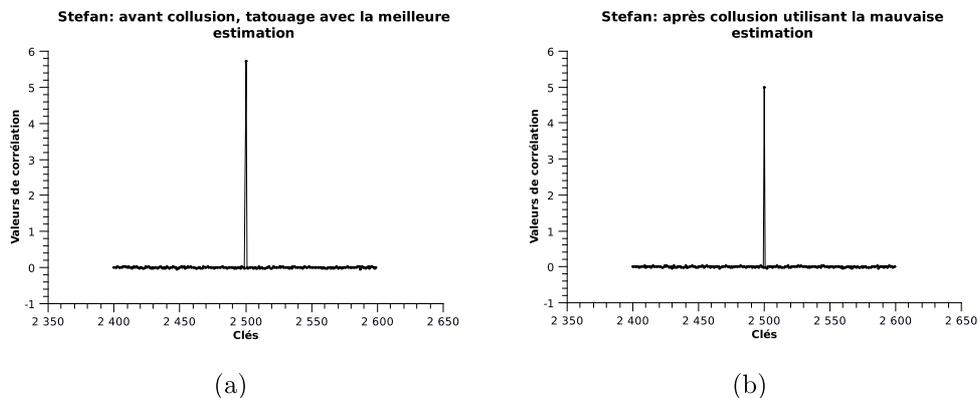


FIG. 5.3: Test de collusion utilisant la meilleure estimation de mouvement dans la phase d'insertion de la marque

Dans le deuxième cas, où les paramètres de mouvement utilisés dans la phase d'insertion sont mieux estimés que ceux utilisés dans la phase d'attaque, on peut dire aussi que la marque existe encore et qu'elle peut être détectée. C'est ce qu'on peut remarquer sur la figure 5.3 où le pic de corrélation est très clair. De même, et d'après la figure 5.4, le pic de corrélation est clair quand les paramètres de mouvement utilisés pour la phase d'attaque sont mieux estimés que ceux utilisés pendant la phase d'insertion de la marque.

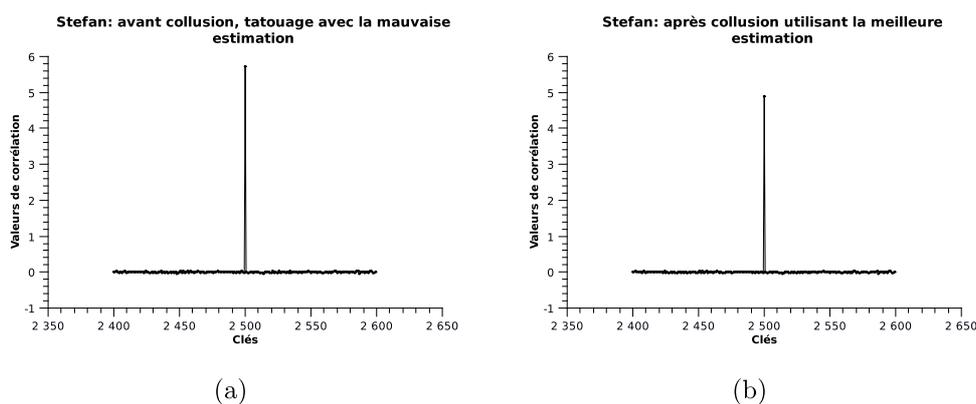


FIG. 5.4: Tests de collusion utilisant la meilleure estimation de mouvement dans la phase d'attaque

En résumé, même si les vecteurs de mouvements utilisés dans les phases d'insertion et d'attaque sont différents, la marque peut toujours être récupérée.

5.4 Évaluation face à la compression MPEG4

Par nature, la compression vidéo tend à éliminer les informations non visibles dans une vidéo dans le but de la compresser. Même si ce n'est pas une attaque intentionnelle, la marque invisible peut potentiellement être éliminée par le processus de compression. Puisque la plupart des séquences vidéos sont compressées, un système de tatouage fiable doit résister à une telle attaque. Évaluation, plus le taux de compression augmente, plus la détection de la marque devient plus difficile, mais au même temps, on tend à avoir une séquence de très mauvaise qualité ou même inexploitable.

Dans le but de tester la fiabilité de notre schéma de tatouage, nous avons compressé les séquences vidéos de test avec un codeur H264 avec différents débits : 1 Mbits/sec (bonne qualité), 500 kbits/sec (qualité moyenne) et 200 kbits/sec (qualité basse). Les Figures 5.5 et 5.6 montrent qu'avec tous les différents débits, un pic de corrélation existe dans les valeurs de corrélation qui correspondent à la clé utilisée pendant la phase d'insertion de la marque. Ces pics de corrélation sont nettement supérieurs au seuil de décision pour les débits de 1 Mbits/sec et 500 kbits/sec. Par conséquent, la marque peut être facilement récupérée après de tels débits de compression. Au niveau des vidéos compressées à 200

kbits/sec, le pic de corrélation est inférieur au seuil de décision et ne permet pas de détecter clairement la marque. Par contre, à ce débit de compression, la qualité vidéo obtenue est très médiocre comme on peut le constater sur les images (voir Figure 5.7) de vidéos compressées à 200 kbits/sec. Il est clair qu'à tel taux de compression, la vidéo ne devient plus exploitable. Alors, même si la marque ne peut être récupérée, la totalité de la vidéo ne peut être exploitée.

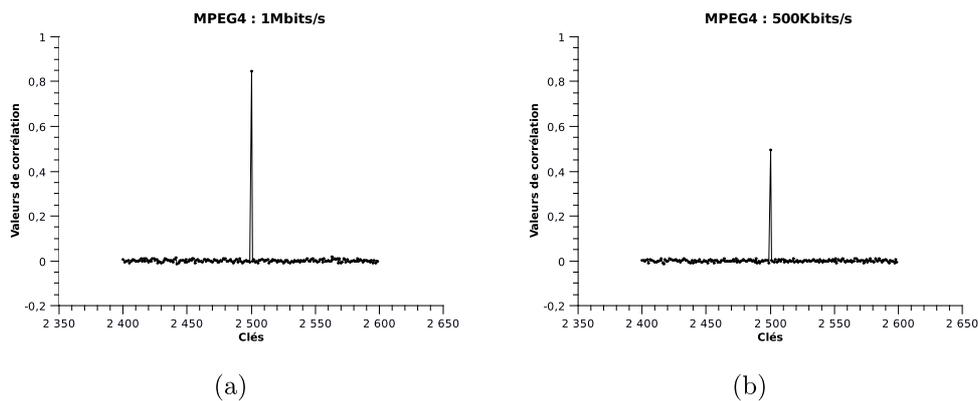


FIG. 5.5: Séquence de Thabor : valeurs de corrélation après une compression MPEG4 à deux débits différents

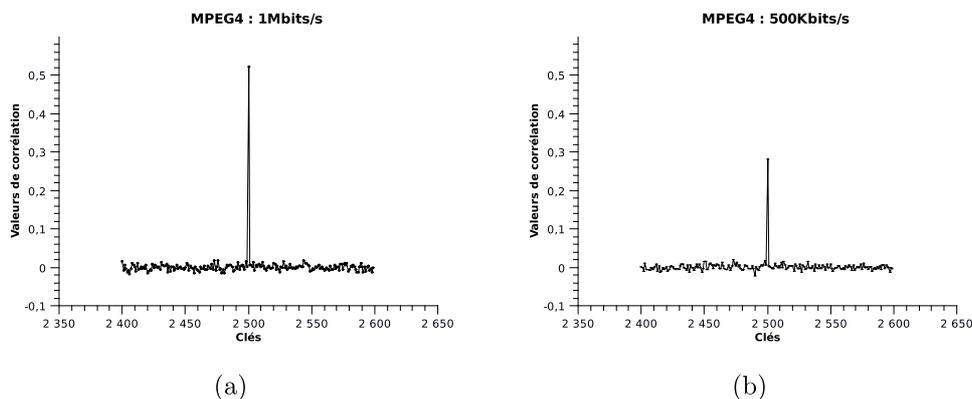


FIG. 5.6: Séquence de Stefan : valeurs de corrélation après une compression MPEG4 à deux débits différents

Pour améliorer la robustesse de l'approche face à la compression MPEG4, nous avons exploité la faiblesse du Système Visuel Humain qui devient beaucoup moins sensible aux modifications et distorsions éventuelles sur les images où le mouvement de la caméra est rapide. Au niveau de telles images, nous optons pour une insertion de la marque avec plus

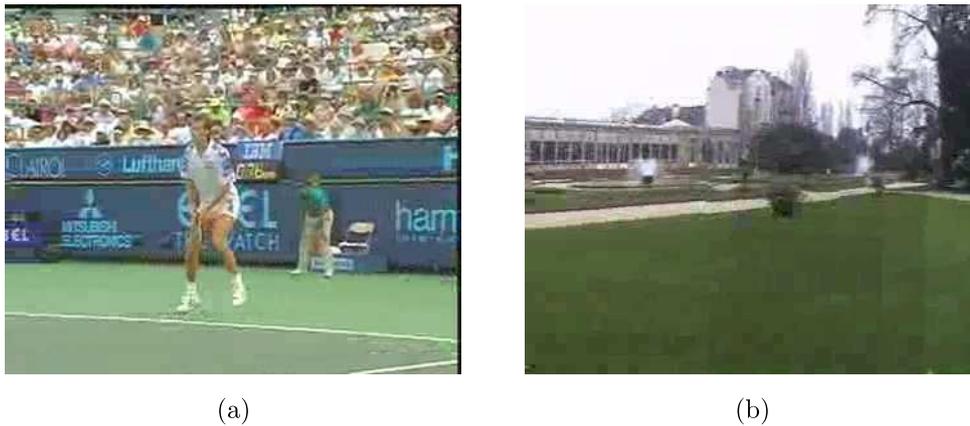


FIG. 5.7: Exemple d'images de vidéos compressées avec mpeg4 à un débit de 200Kbits/s

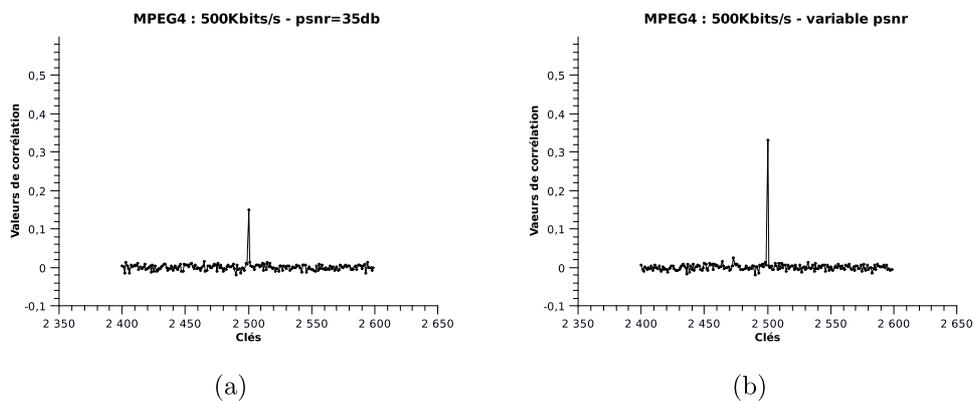


FIG. 5.8: Séquence de Stefan : valeurs de corrélation utilisant un psnr constant et un psnr dépendant de la vitesse de la caméra

d'énergie pour renforcer le tatouage et assurer ainsi une meilleure détection. Pour cela, nous fixons la valeur du psnr, qui reflète la qualité souhaitée de l'image après tatouage, selon la vitesse de la caméra. Dans la Figure 5.8, nous remarquons que la corrélation est beaucoup plus évidente si la vitesse de la caméra est prise en considération au moment de l'insertion de la marque.

5.5 Expérimentation face à la suppression d'images

Les tailles temporelles des séquences de tests peuvent être différentes de celles des séquences originales. Ceci peut être dû aux différents cas suivants :

- La frame de départ de la séquence est différente.
- Certaines frames ont été perdues au cours d'un processus de compression (scalabilité temporelle)
- Certaines frames ont été aléatoirement supprimées (ou perdues au cours d'une transmission) comme une tentative de supprimer ou de désynchroniser la marque.

Dans ces cas, un processus doit être défini dans le but de récupérer quand même la marque.

En pratique, deux cas sont possibles :

1. Pas de connaissance a priori à propos des frames supprimées. Deux cas se présentent :
 - la valeur de corrélation est supérieure au seuil de décision : nous pouvons conclure que la marque existe encore et peut être récupérée.
 - la valeur de corrélation est inférieure au seuil de décision. Dans ce cas, le premier frame supprimé peut être détecté par la sélection du numéro du frame où la valeur de corrélation chute d'une manière brusque. Nous pouvons voir, dans la Figure 5.9, que les frames numéro 260 et numéro 30 correspondent aux premiers frames supprimés respectivement dans les séquences de Stefan et Thabor. Une fois les frames supprimés sont détectés, nous pouvons resynchroniser (faire correspondre) les frames avec leurs vecteurs de mouvement correspondants. Après la resynchronisation, nous appliquons de nouveau le processus de détection. Cette étape est répétée jusqu'à ce qu'on obtienne une valeur de corrélation acceptable, c'est à dire supérieure au seuil de décision. La Figure 5.10 montre un exemple du processus de resynchronisation en cas de suppression de frames. Nous pouvons remarquer, sur cette figure, que le pic de corrélation est progressivement récupéré par le processus de synchronisation, ce qui permet la détection de la marque tout au long de la séquence. Nous devons mentionner aussi que s'il y a beaucoup de frames qui ont été supprimés alors cette démarche n'aboutira pas à une détection fiable, par contre la vidéo devient inexploitable du fait qu'une grande partie a été supprimée. Dans le cas particulier où le frame de départ est différent, les premiers frames de la vidéo originale seront successivement non considérés jusqu'à obtenir une valeur de corrélation signifiante.
2. Connaissance a priori sur les frames supprimés. Pour avoir une séquence vidéo exploitable, la suppression de frames doit être régulière à raison d'un frame sur deux

ou sur quatre qui soit supprimé (scalabilité temporelle). Dans ce cas, si les étapes de synchronisation citées ci-dessus échouent, le processus de détection peut être lancé tenant compte qu'un frame sur deux ou sur quatre a été supprimé. La figure 5.11 montre les valeurs de corrélation après la suppression de quelques frames de la séquence de Stefan à raison d'un frame sur deux. Le pic de corrélation obtenu montre que la marque peut toujours être détectée.

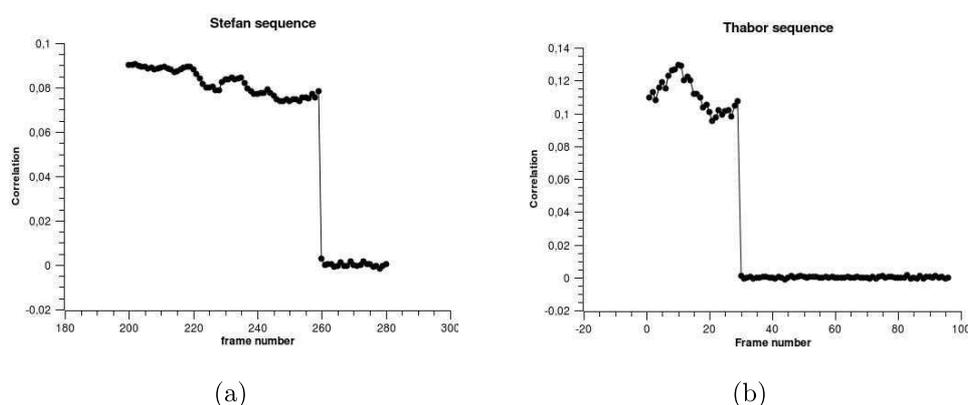


FIG. 5.9: Valeurs de corrélation après la suppression des frames

5.6 Expérimentation dans le cas d'une caméra fixe

Quand il s'agit d'une séquence vidéo où la caméra est fixe, le fond fixe de la scène abstraction faite sur les objets en mouvement, est le même sur chaque image. Dans ce cas, nous n'avons pas besoin de créer la mosaïque puisque tous les points physiques sont fixes et peuvent être bien évidemment localisés simplement. L'approche de tatouage devient simple et consiste à insérer la même marque dans toutes les images de la scène. En conséquence, le problème de la collusion ne se pose plus puisque tous les pixels représentant le même point physique seront tatoués par la même marque. Par contre, d'autres régions d'intérêt ne seront plus considérées dans la phase d'insertion de la marque à savoir la vitesse de la caméra, qui est nulle dans ce cas, et la fréquence d'apparition des pixels qui est ici constante. La non considération de ces régions d'intérêt affecte considérablement la robustesse de l'approche face à la compression MPEG4. C'est ce qu'on peut remarquer dans les Figures 5.12 et 5.13 où le pic de corrélation décroît d'une manière significative

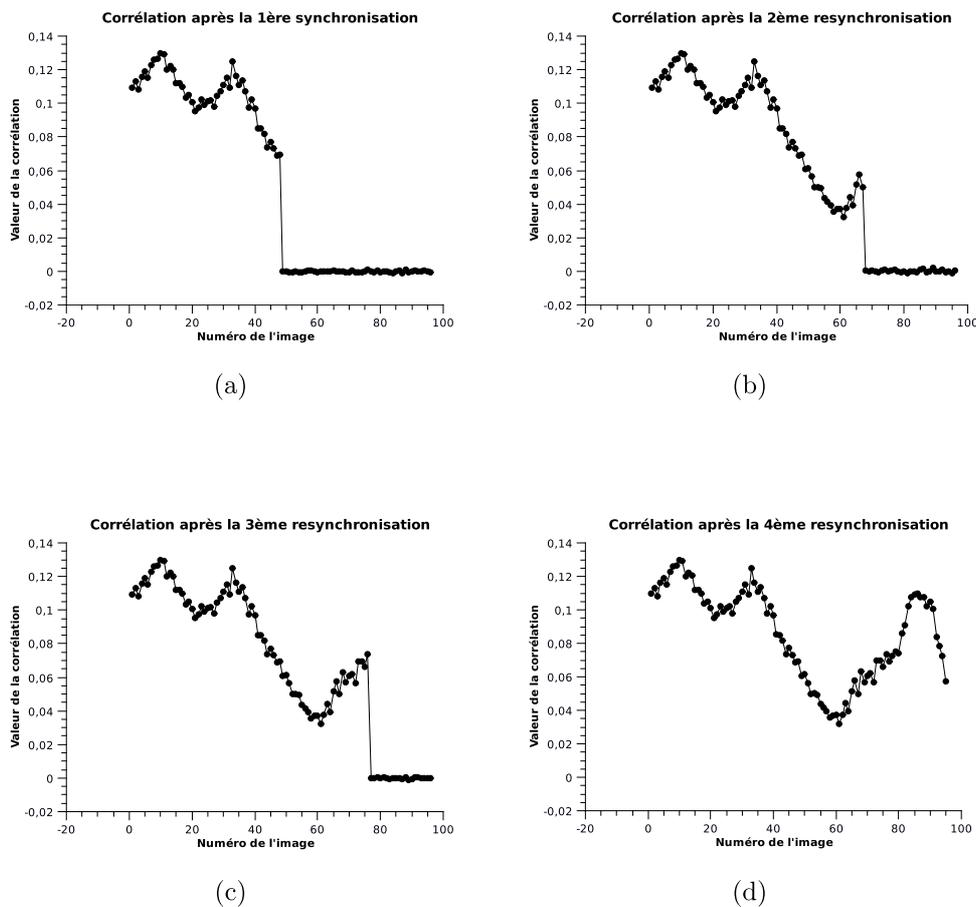


FIG. 5.10: Détection des frames supprimés et resynchronisation pour la détection

après une étape de compression à 200Kbits/s.

5.7 Positionnement par rapport à d'autres travaux

L'approche proposée est comparée à des travaux précédents afin de se situer par rapport à eux. Ces travaux sont présentés par Zhaowan et al.[100], Hao-Xian et al.[33] et Rajab et al.[51].

Les premiers, par Zhaowan et al.[100], proposent une approche de tatouage vidéo basée sur la détection de mouvement au sein de la séquence. En effet, pour garantir une invisibilité de la marque, celle ci est insérée dans les images contenant beaucoup de mouvement et où l'œil humain est beaucoup moins sensible aux modifications. Cette approche est robuste face à la compression MPEG2, la suppression d'image et le cadrage

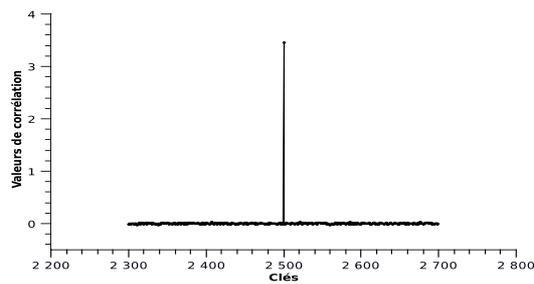


FIG. 5.11: Séquence de Stefan : corrélation après la suppression d'un frame sur deux

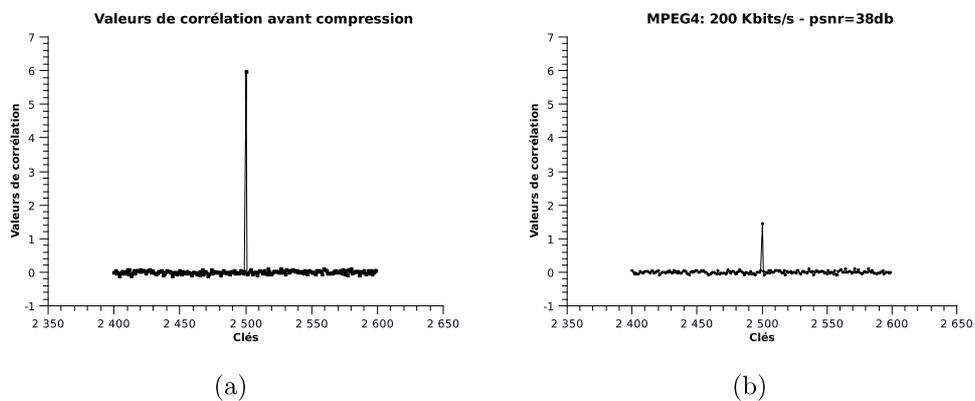


FIG. 5.12: Séquence de "Silent" : valeurs de corrélation après une compression mpeg4

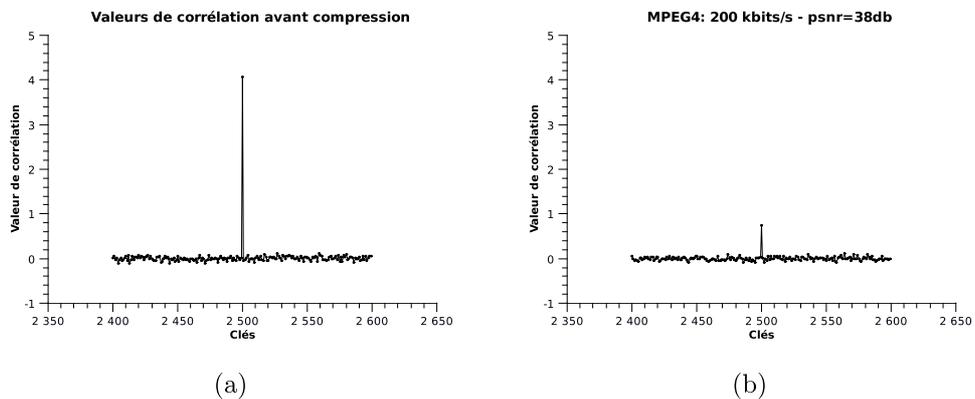


FIG. 5.13: Séquence de "Paris" : valeurs de corrélation après une compression mpeg4

d'image. La détection se base sur le calcul de la corrélation normalisée entre la marque originale et la marque prédite.

La deuxième approche, par Hao-Xian et al.[33], est une approche de tatouage dans le domaine spatial. Son principe est d'exploiter les avantages issus des images intra et

inter pour garantir une invisibilité et en plus une robustesse de la marque. Le taux des bits erronés (BER) est utilisé comme mesure de détection de la marque. Elle présente une robustesse face à certaines attaques à savoir la compression MPEG2, la suppression d'images, la collusion et aussi face à l'ajout de bruit.

La dernière approche, présentée par [51], consiste à insérer la marque dans la diagonale des matrices S, U et V obtenues après une décomposition en valeurs singulières (SVD). La corrélation normalisée est utilisée dans la phase de détection.

En se comparant par rapport à ces travaux (voir Tableau 5.1), notre approche possède des résultats similaires. Par contre, nous considérons la compression MPEG4 au lieu de MPEG2. Réellement, il est plus intéressant d'avoir une approche robuste face à la compression MPEG4 puisque cette norme est actuellement très utilisée mais qui n'est pas abordée dans les travaux cités auparavant. Aussi, l'attaque de collusion est peu discutée dans les travaux de Hao-Xian et Rajab et nullement traitée dans l'approche de Zhaowan.

TAB. 5.1: Comparaison de la robustesse par rapport à d'autres méthodes de tatouage

Attaques	MPEG2	MPEG4	collusion	Supp. d'images	Bruitage	Shuffling
méth. de Zhaowan	✓	-	-	✓	-	✓
méth. de Hao-Xian	✓	-	✓	✓	✓	-
méth. de Rajab	✓	-	✓	✓	-	-
méth. proposée	-	✓	✓	✓	-	-

5.8 Conclusion

Au cours de ce chapitre, une étude expérimentale visant à prouver la robustesse de notre approche a été présentée. Les différents tests effectués ont montré l'efficacité de certaines régions d'intérêt pour le tatouage vidéo. En effet, appliquer la même marque sur tous les points représentant un même point physique permet d'éviter les conséquences d'une attaque de collusion. En plus, augmenter l'énergie de la marque sur certaines zones sans nuire à la visibilité de la séquence permet aussi d'augmenter la robustesse de l'approche

face à la compression MPEG4. Ces zones sont intéressantes pour le tatouage puisque le système visuel humain y est beaucoup moins sensible aux modifications.

Troisième partie

Exploitation des zones d'ombres pour le tatouage

Chapitre 6

Généralités sur les zones d'ombre

6.1 Introduction

On s'intéresse dans cette partie à une autre région d'intérêt qui favorise le tatouage : il s'agit des zones d'ombres en mouvement. Une ombre se définit comme l'obscurité que cause un corps opaque en interceptant la lumière. Elle est une conséquence particulière de l'illumination dans une scène. Les ombres sont le résultat d'une interaction entre une composante géométrique et une composante photométrique.

Les ombres sont présentes dans les séquences vidéo contenant une source de lumière ponctuelle ou étendue. Elles sont souvent considérées comme des éléments perturbateurs lors des prises de vues. Elles possèdent en effet un caractère imprévisible pouvant altérer, en cinéma par exemple, l'esthétique d'une scène. Les différents éclairages sont alors souvent organisés de manière à minimiser le plus possible les ombres dans les images. Les ombres sont aussi problématiques pour plusieurs applications de traitement et d'analyse automatiques d'images. Par contre, elles contiennent aussi en réalité un grand nombre d'informations sur la scène. Elles permettent notamment à l'œil une meilleure compréhension des scènes par l'évaluation des profondeurs relatives. En peinture par exemple, les ombres sont assimilées à des couleurs obscures employées dans un tableau pour représenter les parties des objets les moins éclairées et qui servent à donner du relief aux autres.

Dans le cadre des méthodes automatiques de traitement ou d'analyse d'images, la

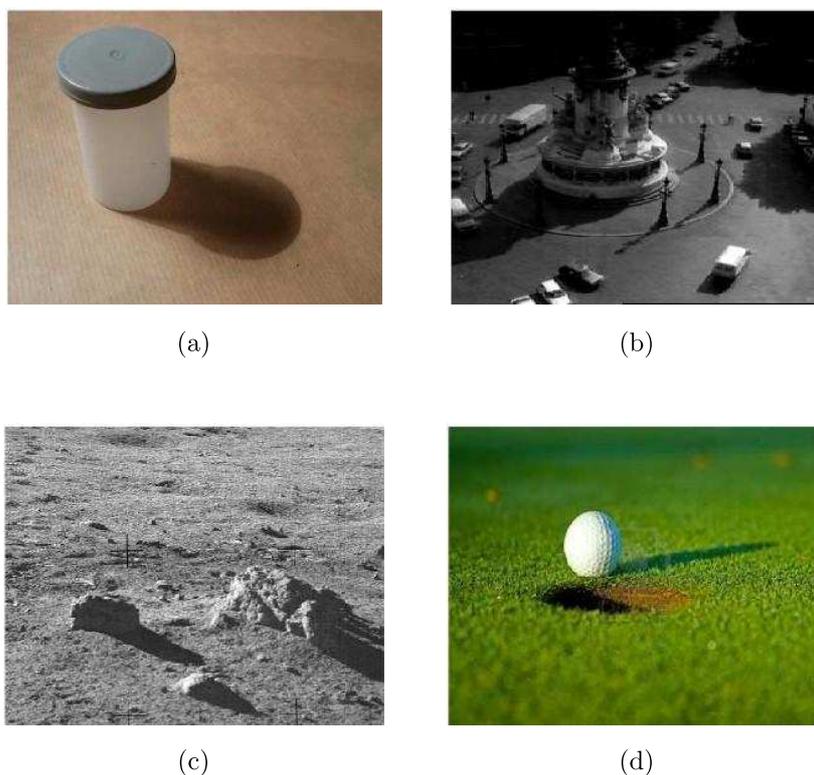


FIG. 6.0: Exemple de scènes présentant des zones d'ombres.

difficulté majeure consiste à différencier les objets de leurs ombres [76]. Beaucoup de méthodes de détection ou d'analyse des ombres existent aujourd'hui et leurs résultats sont initialement liés à la nature des images traitées et aux hypothèses formulées. Pour résoudre ce type d'erreurs, il est tout d'abord nécessaire de définir et de modéliser les ombres de manière à mieux appréhender le problème posé. Il s'agit dans ce chapitre de présenter une étude générale sur les zones d'ombres, de citer les domaines de recherche et de présenter l'approche adoptée pour la détection des ombres portées dans une séquence vidéo et qui va être utilisée dans le schéma général de tatouage.

6.2 Modélisation des ombres

Les ombres sont une conséquence de l'obstruction d'une source lumineuse quelconque par un obstacle (un objet). Une multitude d'ombres différentes peuvent cependant être créées. Cette multitude est dû aux différentes formes que peut avoir l'objet, à la nature de la source lumineuse, sa forme (ponctuelle ou étendue), sa distance par rapport à l'objet, la

couleur de la lumière émise et par la forme de la surface sur laquelle est projetée l'ombre. Toutefois, les ombres peuvent être classées en trois catégories :

- **L'ombre propre** partie de l'objet obstruteur qui n'est pas éclairée par la source lumineuse. Cette zone d'ombre prend la forme de l'objet dans sa partie non éclairée. (voir figure 6.1)

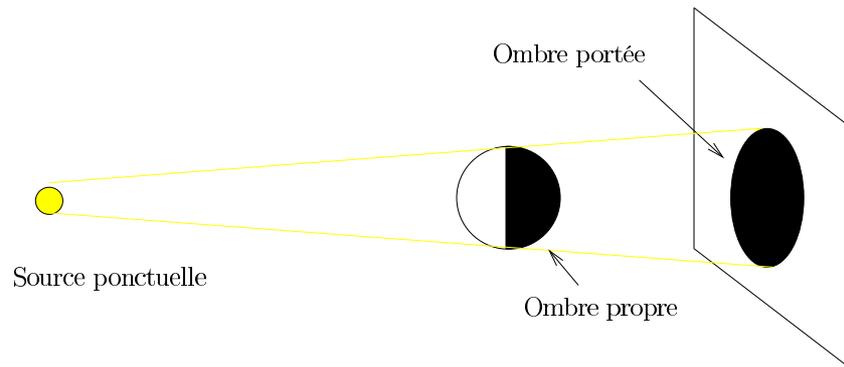


FIG. 6.1: Zones d'ombre dans le cas d'une source de lumière ponctuelle

- **L'ombres portée** région de la surface sur laquelle est projetée l'ombre. Cette surface est un écran qui, placé derrière un objet par rapport à la source de lumière, ne reçoit pas de rayon. La taille et la forme de l'ombre portée dépend de la forme, la taille et la position de l'objet par rapport à la source, mais aussi de l'emplacement et de l'inclinaison de l'écran.
- **La pénombre** région de la surface placé derrière l'objet obstruteur et qui est éclairée partiellement par une source lumineuse étendue (voir figure 6.2). Elle se situe aux à l'entour des ombres portées, càd qu'elle constitue la transition entre l'ombre et la zone qui est totalement éclairée dans la surface de projection.

Dans notre contexte, nous allons nous intéresser seulement aux ombres portées. Les pénombres sont des zones très limitées en surface et donc elles ne favorisent pas trop le tatouage. En plus, il n'est pas évident de trouver ces zones là puisqu'elles nécessitent la présence d'au moins une source de lumière étendue et assez forte par rapport à la lumière ambiante. Les ombres propres, quand à elles, sont difficiles à modéliser (du point de vue géométrique et photométrique) et à localiser dans une séquence puisqu'elles sont très dépendantes de l'objet sur lequel elles se trouvent.

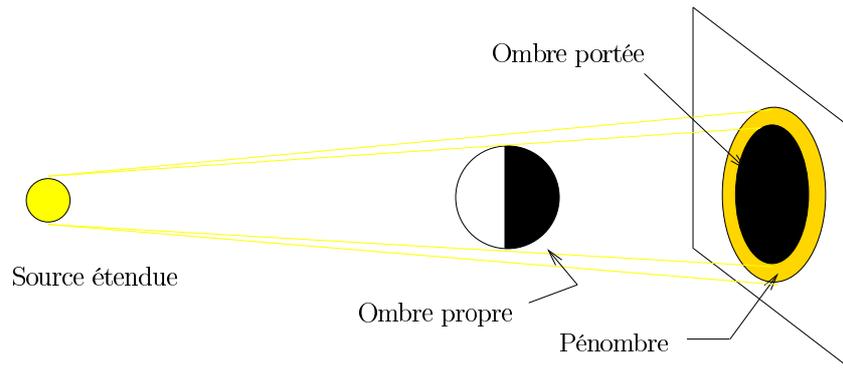


FIG. 6.2: Zones d'ombre dans le cas d'une source de lumière étendue

La modélisation des ombres portées nécessite la définition de deux composantes principales :

- **Une composante géométrique** : il s'agit de déterminer la forme géométrique éventuelle de l'ombre portée. Cette forme dépend en fait de plusieurs paramètres à savoir la forme géométrique de l'objet obstructeur de la lumière, la forme géométrique de la surface sur laquelle se projette l'ombre et par la localisation des différentes sources de lumières existantes dans la scène. La connaissance a priori de telles informations n'est pas toujours disponible et leur estimation n'est pas facile à réaliser. Pour cette raison, la modélisation géométrique des ombres est très peu abordée. Néanmoins, certains cas, où plusieurs hypothèses a priori sont fixées, la modélisation géométrique est exploitée comme par exemple dans le cas de traitement d'images aériennes où le sol est plat et la source de lumière est le soleil.
- **Une composante photométrique** : il s'agit de déterminer l'intensité de la lumière tenant compte des sources de lumières existantes dans la scène, de la lumière ambiante et des propriétés photométriques de l'environnement à savoir la surface sur laquelle se trouve l'ombre portée. Même si la zone d'ombre n'est pas éclairée par la source de lumière principale, elle n'est pas supposée être toujours une région totalement sombre puisqu'elle est éclairée par la lumière ambiante. Plus la lumière ambiante est faible, plus l'ombre est sombre. En se basant sur le modèle d'illumination de *Phong* et le modèle *ambient*, Pinel [76] a proposé une modélisation complète qui modélise les reflets à la surface des objets.

Dans le cas d'une scène éclairée par une seule source de lumière et si $I(p)$ désigne

la luminance d'un pixel p dans une image alors :

$$I(p) = \begin{cases} k_a I_a + k_d I_s < \vec{N}(p), \vec{L}(p) > + k_s I_s < \vec{V}(p), \vec{L}'(p) >^n & \text{si illuminé} \\ k_a I_a + \alpha(p) [k_d I_s < \vec{N}(p), \vec{L}(p) > + k_s I_s < \vec{V}(p), \vec{L}'(p) >^n] & \text{si pénombre (6.1)} \\ k_a I_a & \text{si ombre} \end{cases} \quad (6.2)$$

avec I_a et I_s sont respectivement les intensités de la lumière ambiante et de la source de lumière. k_a , k_d et k_s sont respectivement le coefficient de réémission de la lumière ambiante, le coefficient de diffusion du matériau et la spécularité intrinsèque du matériau. $\vec{L}(p)$ est la direction de la source, $\vec{L}'(p)$ la direction symétrique de $\vec{L}(p)$ par rapport à la normale à la surface $\vec{N}(p)$ et $\vec{V}(p)$ la direction d'observation. n est un facteur servant à caractériser la spécularité de la surface. $\alpha(p) \in [0, 1]$ décrit ici la transition au sein de la pénombre. La taille de la pénombre et l'illumination dans la pénombre dépendent de la lumière qui n'est généralement pas ponctuelle et de la géométrie de la scène.

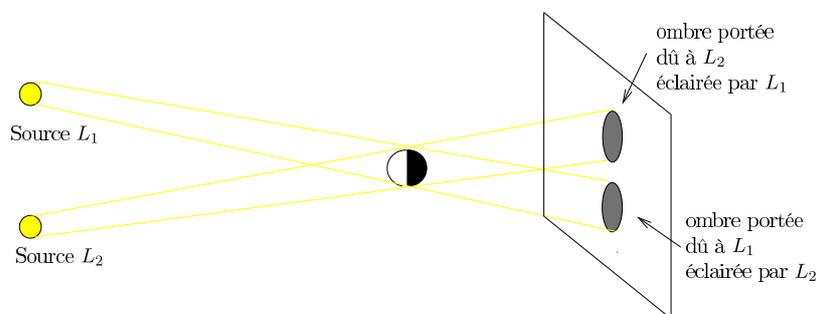


FIG. 6.3: Zones d'ombre dans le cas d'une scène avec deux sources de lumière

Dans le cas d'une scène avec deux sources de lumière distinctes (voir figure 6.3), plusieurs combinaisons de régions d'ombre peuvent survenir. On peut avoir un chevauchement de deux ombres, d'une ombre et d'une pénombre et bien même de deux pénombres. Partant du modèle d'illumination avec une seule source de lumière, le modèle relatif à deux sources est le suivant :

$$I(p) = \left\{ \begin{array}{ll}
 k_a I_a + k_{1_d} I_{1_s} < \vec{N}(p), \vec{L1}(p) > + k_{1_s} I_{1_s} < \vec{V}(p), \vec{L1}'(p) >^n \\
 + k_{2_d} I_{2_s} < \vec{N}(p), \vec{L2}(p) > + k_{2_s} I_{2_s} < \vec{V}(p), \vec{L2}'(p) >^n & \text{si illuminé} \\
 \\
 k_a I_a + k_{1_d} I_{1_s} < \vec{N}(p), \vec{L1}(p) > + k_{1_s} I_{1_s} < \vec{V}(p), \vec{L1}'(p) >^n & \text{si ombre2} \\
 \\
 k_a I_a + k_{2_d} I_{2_s} < \vec{N}(p), \vec{L2}(p) > + k_{2_s} I_{2_s} < \vec{V}(p), \vec{L2}'(p) >^n & \text{si ombre1} \\
 \\
 k_a I_a + \alpha_1(p) [k_{1_d} I_{1_s} < \vec{N}(p), \vec{L1}(p) > \\
 + k_{1_s} I_{1_s} < \vec{V}(p), \vec{L1}'(p) >^n] & \text{si pénombre2, ombre1} \\
 \\
 k_a I_a + \alpha_2(p) [k_{2_d} I_{2_s} < \vec{N}(p), \vec{L2}(p) > \\
 + k_{2_s} I_{2_s} < \vec{V}(p), \vec{L2}'(p) >^n] & \text{si pénombre1, ombre2} \\
 \\
 k_a I_a + \alpha_1(p) [k_{1_d} I_{1_s} < \vec{N}(p), \vec{L1}(p) > \\
 + k_{1_s} I_{1_s} < \vec{V}(p), \vec{L1}'(p) >^n] + k_{2_d} I_{2_s} < \vec{N}(p), \vec{L2}(p) > \\
 + k_{2_s} I_{2_s} < \vec{V}(p), \vec{L2}'(p) >^n & \text{si pénombre1} \\
 \\
 k_a I_a + k_{1_d} I_{1_s} < \vec{N}(p), \vec{L1}(p) > + k_{1_s} I_{1_s} < \vec{V}(p), \vec{L1}'(p) >^n \\
 + \alpha_2(p) [k_{2_d} I_{2_s} < \vec{N}(p), \vec{L2}(p) > + k_{2_s} I_{2_s} < \vec{V}(p), \vec{L2}'(p) >^n] & \text{si pénombre2} \\
 \\
 k_a I_a + \alpha_1(p) [k_{1_d} I_{1_s} < \vec{N}(p), \vec{L1}(p) > \\
 + k_{1_s} I_{1_s} < \vec{V}(p), \vec{L1}'(p) >^n] + \alpha_2(p) [k_{2_d} I_{2_s} < \vec{N}(p), \vec{L2}(p) > \\
 + k_{2_s} I_{2_s} < \vec{V}(p), \vec{L2}'(p) >^n] & \text{si 2 pénombres} \\
 \\
 k_a I_a & \text{si 2 ombres}
 \end{array} \right. \quad (6.3)$$

En général, une scène peut comporter plusieurs sources de lumière. On peut citer l'exemple d'un terrain de football qui est éclairé en général par quatre sources de lumière placées dans les quatre coins du terrain. En conséquence, un joueur présente quatre ombres différentes. Si o_1 désigne l'ombre générée suite à l'éclairage de la source s_1 , alors o_1 est éclairée par les trois autres sources s_1 , s_2 et s_3 . Dans ce

cas, pour déterminer l'intensité de la lumière au niveau de l'ombre o_1 , les intensités des sources s_1 , s_2 et s_3 doivent être prises en considération. Dans ce cas, la modélisation photométrique des ombres est une généralisation du cas à deux sources cités précédemment. Toutefois, ces modèles restent insuffisants et limités en terme de fidélité par rapport à la réalité. En effet, la lumière ambiante n'est pas constante pour tous les points d'une scène et donc il n'est pas évident de la modéliser par une seule valeur. La figure 6.4 montre une zone d'ombre, sans pénombre, agrandie pour remarquer le fait que les valeurs de luminance ne sont pas constantes.



FIG. 6.4: Variation de la luminance dans une ombre portée

D'après cette étude des modélisations géométriques et photométriques des ombres portées, la détection et l'analyse des ombres dans les images peuvent être simplifiées. Les ombres font en effet partie des zones les plus sombres d'une image, mais la difficulté majeure consiste à savoir si cette zone sombre est le résultat d'une obstruction de la lumière ou simplement d'un matériau foncé. C'est pourquoi afin d'obtenir de bons résultats, la prise en compte de plusieurs de leurs caractéristiques devront être utilisées. Parmi celles-ci, on trouve :

- les ombres sont parmi les zones les plus sombres de l'image ;
- les ombres possèdent une pénombre, donc des contours flous ;
- les ombres ne modifient pas la texture de la surface sur laquelle elles se projettent ;
- les ombres diminuent la luminance de la surface sur laquelle elles se projettent ;
- les ombres ne modifient généralement pas la couleur de la surface sur laquelle elles se projettent.

6.3 Domaines de recherche

Les ombres sont une manifestation particulière de l'illumination dans une scène. Plusieurs axes de recherche se sont intéressés à la détection, au traitement ou à l'analyse des ombres dans les images. Il peut s'agir de supprimer les ombres pour faciliter une phase ultérieure d'analyse. A l'inverse, la détection des ombres peut faciliter l'obtention d'informations sur la scène. Il peut être également nécessaire de recréer ou synthétiser les ombres pour des applications de type réalité augmentée.

6.3.1 Imagerie aérienne

Des méthodes de détection des ombres portées ont été développées en imagerie aérienne. Historiquement, elles sont parmi les premières méthodes qui ont été développées pour détecter et utiliser explicitement les ombres. Elles permettent principalement de détecter dans ces images les objets possédant une hauteur. Dans ce cadre, Nagao, Matsuyama et Ikeda ([66]) étaient les premiers qui ont proposé en 1979 une méthode de détection des immeubles dans les zones urbaines en utilisant les ombres. Elles ont été utilisées pour distinguer les zones possédant une hauteur, comme les maisons ou les arbres, des objets plats comme les routes ou la mer. Les informations issues de prises de vues infrarouges sont mélangées aux informations couleur pour calculer un histogramme sur lequel un seuil est calculé afin de détecter les ombres. La direction du soleil, est supposée connue, et elle permet d'extraire les régions adjacentes produisant ces ombres.

6.3.2 Détection des ombres dans le cadre général

Le domaine de la segmentation d'images reste particulièrement sensible aux ombres. En effet, une segmentation classique considère le plus souvent une ombre comme un objet à part de l'image. Selon les méthodes et les buts visés, cela peut s'avérer assez problématique. En effet, les ombres peuvent fausser les résultats de la détection des objets mobiles dans une séquence MPEG-4 par exemple, puisque les ombres sont segmentés et assimilés aux objets. Plusieurs méthodes de détection des ombres dans les images ont alors été développées. Elles utilisent certaines caractéristiques connues des ombres et posent

des hypothèses en relation avec la nature des images. Jiang et Ward ([41, 42]) proposent une méthode de détection et de classification des ombres dans une image. Les ombres portées se projettent sur un fond plan non texturé, les bords de l'image appartiennent au fond, les occlusions entre les objets et les ombres sont minimales et enfin une seule source de lumière est présente. Trois processus s'enchaînent alors. Dans un premier temps, le processus de bas niveau extrait les régions qui sont plus sombres que leur voisinage. Une deuxième étape consiste à détecter dans les zones sombres certaines caractéristiques, comme les contours des régions, la pénombre, les ombres propres et les ombres portées. Enfin le processus de haut niveau intègre les hypothèses et confirme la consistance des zones identifiées.

6.3.3 Détection d'ombre en vidéo surveillance

La segmentation des objets mobiles s'avère être très efficace en vidéo surveillance, notamment grâce à une caméra fixe et à la connaissance d'une image de référence ne possédant aucun objet mobile. Cette image pourra être mise à jour pour prendre en compte les variations lentes d'illumination, notamment dans les scènes d'extérieur. Plusieurs méthodes de détection des ombres ont alors été proposées dans ce cadre de la vidéo surveillance. Les hypothèses plus fortes présentes dans ces cas spécifiques permettant de meilleurs résultats que dans un cadre général. La caméra est en effet supposée fixe et le fond statique. Le principe de base est identique pour toutes ces méthodes : une image de référence ne contenant aucun objet mobile est calculée au préalable. Elle permet une détection beaucoup plus facile des objets en mouvement ainsi que de leurs ombres dans l'image. La difficulté consiste à séparer les modifications dues à la présence d'un objet à celles résultantes d'un changement photométrique local comme une ombre portée par exemple. En résumé, toutes ces méthodes présentent trois étapes importantes :

- Création de l'image de référence ;
- Extraction des modifications par rapport à l'image de référence ;
- Identification des ombres.

6.4 Détection des zones d'ombre en mouvement

Il s'agit dans cette partie de détecter l'ombre portée d'un objet en mouvement dans une séquence vidéo. La méthode de [93] a été adoptée. Cette méthode permet, tout d'abord de générer un masque de détection de changement initial contenant l'objet en mouvement et son ombre. Puis une carte de contours Canny [14] est calculée. Et par la suite, la région de l'ombre est détectée en appliquant "l'intégration de multi-frames", et la croissance de régions.

Donc en résumé, il s'agit premièrement de détecter l'objet en mouvement dans la scène avec son ombre portée. Ensuite, de séparer l'objet de son ombre. Les hypothèses prises en considération dans ce travail sont :

- Une caméra fixe,
- une source de lumière forte causant une ombre portée,
- Un fond statique et un seul objet en mouvement sur la scène.

La méthode est appliquée sur des séquences vidéo qcif (176x144) de 300 frames et de fréquence de 30 images par secondes.

6.4.1 Schéma de l'approche adoptée

Les différentes étapes de l'algorithme (Figure 6.5) adopté pour la détection d'ombre portée dans une séquence vidéo sont illustrées par la figure suivante. On passera par la suite à décrire les différentes parties, étape par étape.

6.4.1.1 Calcul du masque de détection de changement initial (CDM)

La détection des changements dans une séquence d'images constitue une source d'information très importante pour la détection de l'objet en mouvement. En effet, le calcul du masque de détection de changements contient souvent l'objet en mouvement avec son ombre portée. Les approches de détection de changement dans les séquences d'images peuvent être classées en deux catégories :

- approches basées sur la détection et l'analyse du fonds de scène
- approches basées sur l'estimation du mouvement

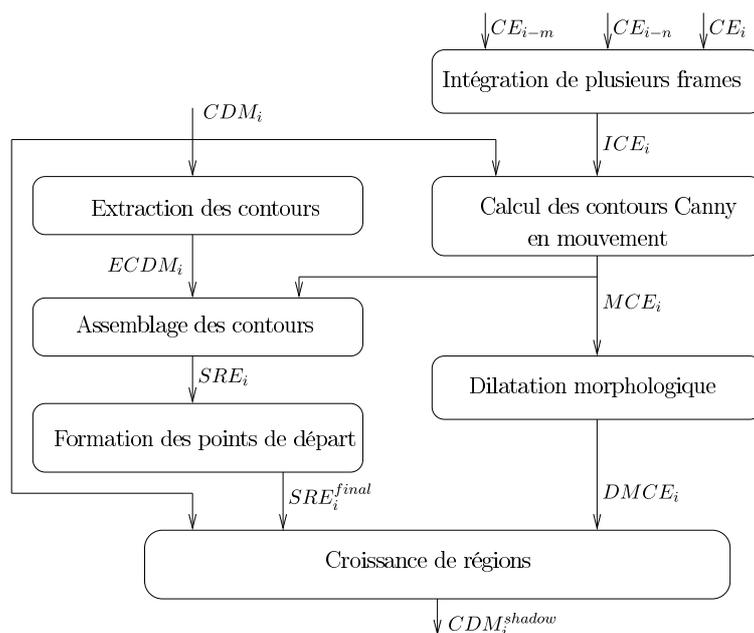


FIG. 6.5: Algorithme de détection de l'ombre portée dans une séquence vidéo

L'approche adoptée dans ce travail, pour la détection des objets en mouvement dans une scène, consiste dans une première étape, à identifier le fond de la scène, puis, dans une seconde étape, de le soustraire de la scène entière afin d'isoler l'avant de la scène. Cette approche, couramment appelée "background subtraction", est très efficace lorsqu'on est en présence d'objets en mouvement sur un fond statique.

Pour cela, la méthode d'enregistrement de fond proposée par [95] est appliquée ici pour calculer le masque de détection de changements (CDM). Le CDM est calculé en utilisant la différence d'intensités entre les frames consécutives d'une même séquence vidéo. Premièrement, on détecte les variations d'intensité des pixels en utilisant la différence entre deux images consécutives. Si on considère l'image courante I_n et l'image précédente I_{n-1} , la différence entre les images FD est donnée par :

$$FD(p) = |I_n(p) - I_{n-1}(P)| \quad (6.4)$$

Cette différence est classiquement seuillée pour distinguer les variations d'intensité des pixels dues aux objets en mouvement de celles dues au bruit. Soit BW le résultat de seuillage de l'image FD . Un pixel p , avec $BW(p) = 1$ est un pixel classifié comme changeant, c'est à dire que pour ce pixel, la variation d'intensité entre deux images est supérieure à un seuil. Ensuite, on passe à une étape de calcul du nombre d'occurrence de

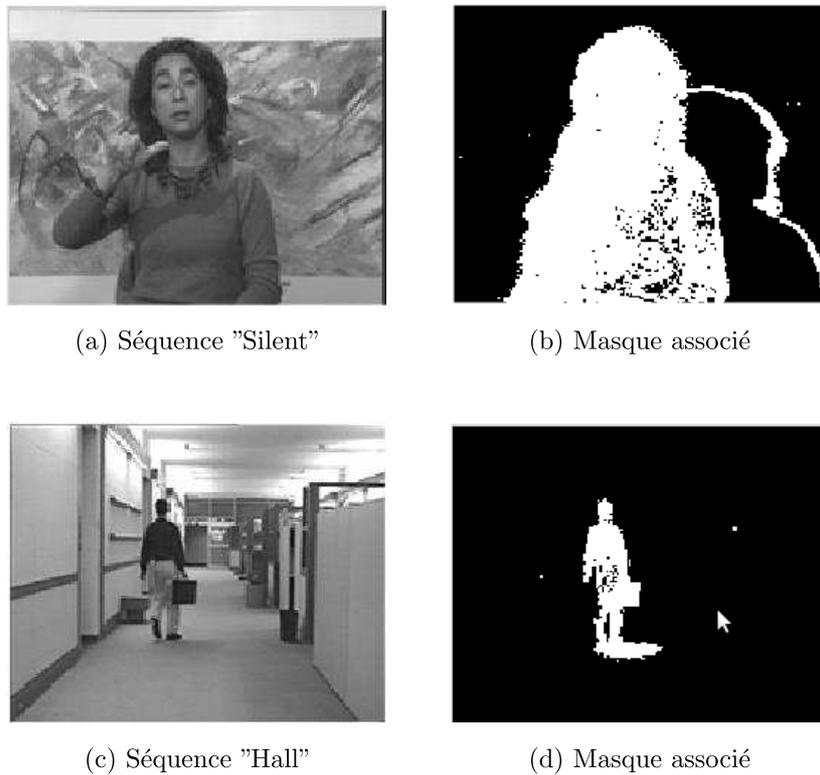


FIG. 6.5: Exemple de masque de détection de changement.

chaque pixel du frame dans une séquence de frames. Si le nombre d'occurrences du pixel est important, donc ce pixel appartiendra automatiquement au fond de la scène. C'est-à-dire il s'agit d'un point fixe, non changeant. Cette matrice d'occurrence sera par la suite parcourue, en comparant l'occurrence de chaque pixel par rapport à une valeur choisie comme un seuil maximal au dessous duquel le pixel est considéré comme un pixel en mouvement, à ce moment là ce pixel prendra la valeur 0 (point noir). Sinon le pixel va être considéré comme un point du fond de la scène. Une fois le fond de la scène est récupéré, une soustraction sera faite entre le frame initial et le résultat de la détection du fond afin d'obtenir l'objet en mouvement avec son ombre portée.

Une étape de seuillage de ce résultat est effectuée afin d'obtenir le masque de détection de changements (*CDM*) de l'objet en mouvement et son ombre portée. Le résultat obtenu est un masque binaire où les pixels blancs (le niveau gris est 255), indiquent l'avant de la scène, tandis que les Pixel noirs, où le niveau gris est 0, indiquent l'arrière plan de la scène. Dans cette étape l'ombre fait partie de l'avant de la scène.

Le résultat de cette étape est présenté dans la Figure 6.5.

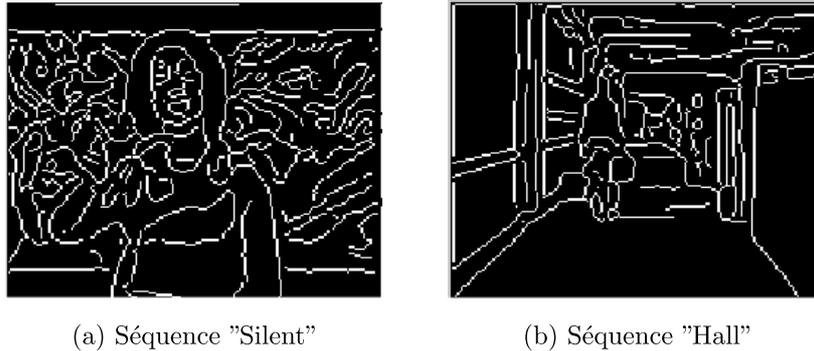


FIG. 6.6: Exemple de contour de Canny.

6.4.1.2 Calcul des contours Canny

Définition Pour le calcul du filtre de Canny [14], une approche analytique plus élaborée est employée. Il s'agit d'une technique de filtrage optimal. Canny a en effet cherché à déterminer de façon analytique en 1986 un filtre à partir de trois critères :

- un critère de bonne détection garantissant une réponse forte en sortie du filtre même en présence de faibles contours sur l'image d'entrée,
- un critère de bonne localisation du contour,
- un critère d'unicité de la réponse permettant d'assurer une seule détection pour un contour et ainsi d'éviter les effets de rebond.

Application du Canny L'algorithme Canny est utilisé pour la détection des contours car il garantit une bonne détection, c'est-à-dire une réponse forte même à des faibles contours. Le résultat est sous forme d'une carte binaire, où les pixels blancs, dont la valeur en niveaux de gris est de 255 représentent les contours Canny. Les résultats de l'application de cet algorithme sur le frame 153 de l'exemple Silent et le frame 53 de l'exemple du Hall sont illustrés dans la Figure 6.6.

6.4.1.3 Extraction des contours du CDM

Dans cette étape il s'agit d'extraire les contours du CDM initial déjà calculé par application de la méthode de Canny. Le résultat de cette étape, noté ECDMi, est un masque binaire où les pixels blancs représentent les contours de l'avant de la scène tout

en considérant qu'au moins un de ses voisins est dans le fond de la scène.

$$CE_i = \{p/ECDM_i(p) = 255, \exists q = NG(P), ECDM_i(q) \neq 255\} \quad (6.5)$$

où p et q sont des pixels de l'image, et la fonction NG permet de calculer les voisins du pixel p (q est l'un des 4 pixels voisins de p).

6.4.1.4 Intégration de plusieurs frames

Les contours Canny dans une scène peuvent être classés en deux catégories : les contours statiques et les contours en mouvement. Les contours en mouvement sont les plus intéressants car ils vont être la base de l'étape de l'algorithme de croissance de régions discuté plus tard. Pour éliminer les contours qui font partie du fond de la scène, et ne préserver que les contours de l'objet en mouvement, une méthode basée sur trois frames est utilisée, d'où le nom "Intégration de plusieurs frames". La méthode adoptée est donnée par l'équation suivante :

$$ICE_i(p) \in \begin{cases} \text{Pas de contour si } CE_i(p) = 0 \\ \text{Contour statique si } CE_i(p) = CE_{i-n}(p) = CE_{i-m}(p) = 255 \\ \text{Contour en mouvement sinon} \end{cases} \quad (6.6)$$

Où p désigne un pixel.

Les points du contour dans la carte des contours Canny du frame courant sont considérés des contours statiques s'ils ont la même valeur dans les frames précédents ($i - n$) et ($i - m$). Les valeurs de n et m ont pour valeurs respectivement 3 et 5 et non pas 1 et 2, et ceci pour mieux différencier les deux catégories de contours surtout dans le cas où le mouvement de l'objet est lent.

6.4.1.5 Détection des contours Canny en mouvement

En utilisant comme entrées le CDM_i et le ICE_i , on peut facilement détecter les contours en mouvement par intersection des deux résultats. La sortie de cette étape est notée : MCE_i . Autrement :

$$MCE_i = \{p/ICE_i = \text{contourcannyenmouvement}, p \in CDM_i\} \quad (6.7)$$

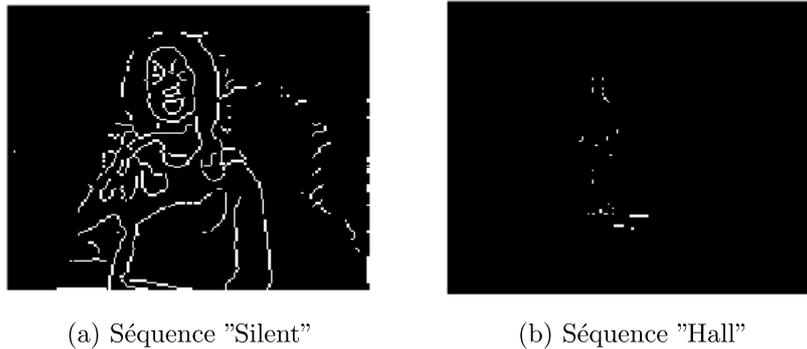


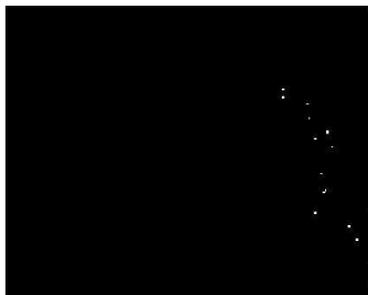
FIG. 6.7: Les contours canny en mouvement.

La Figure 6.7 illustre des exemples de contours Canny en mouvement détectés dans deux séquences vidéo différentes.

6.4.1.6 Formation des points de départ de la croissance des régions

Puisque la méthode de croissance de régions sera adoptée pour détecter l'ombre portée, alors un ensemble de points de départ doit être généré. Les points de contours de la région d'ombre dans le CDM_i peuvent résoudre le problème. Tout d'abord on doit séparer les points de contours de la région d'ombre dans $ECDM_i$ des contours de l'objet en mouvement.

Dans cette partie on a commencé par trouver les points des contours communs dans $ECDM_i$ et MCE_i , en appliquant l'intersection entre ces deux résultats. Ensuite l'algorithme classique de connexion des composants de [34] a été utilisé afin de connecter les points de départ trouvés. Le résultat de cette étape est noté SRE_i qui est présenté dans la figure 6.8.

FIG. 6.8: SRE_{153} du frame 153 de la séquence "Silent"

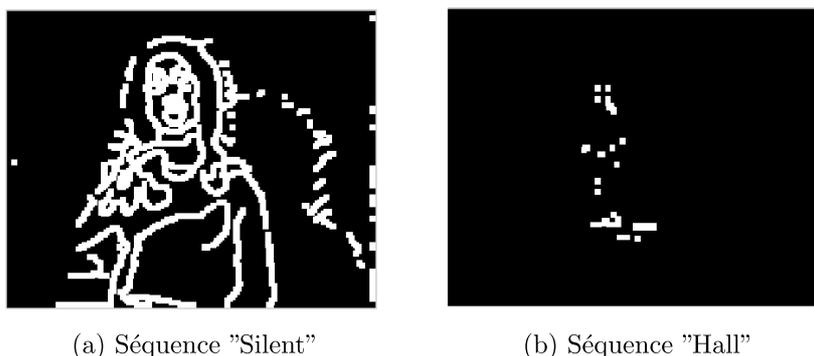


FIG. 6.9: Résultat de la dilatation morphologique.

6.4.1.7 Dilatation morphologique

Une dilatation morphologique [67] est appliquée au résultat MCE_i contenant les contours canny en mouvement, en utilisant un élément structurant carré 3x3. Et ceci dans le but d'accentuer ces contours.

$$DMCE_i = MCE_i \oplus B_1 \quad (6.8)$$

Où B_1 est l'élément structurant (carré 3x3) et $DMCE_i$ est le MCE_i dilaté.

Le résultat obtenu en appliquant la dilatation morphologique est donné par la figure 6.9.

6.4.1.8 Croissance de régions

En utilisant les résultats $DMCE_i$, SRE_i et le CDM_i , une dilatation conditionnelle peut être appliquée pour extraire la région d'ombre. L'élément structurant B_2 utilisé pour cette dilatation est donné par la figure 6.10 où seulement les quatre voisins directes sont considérés.

0	1	0
1	1	1
0	1	0

FIG. 6.10: Élément structurant B_2

Cette étape est dite étape d'étiquetage des objets qui permet de référencer chaque objet de l'image par une valeur entière appelée étiquette ce qui facilite leur distinction.

L'étiquetage est appliqué avec un voisinage d'ordre 4, et la condition S est définie par :

$$S = \{p/p \in ICDM_i, p \notin DMCE_i\} \quad (6.9)$$

L'opération de la dilatation conditionnelle est décrite comme suit :

Mettre $C_0 = SRE_i, m = 0$

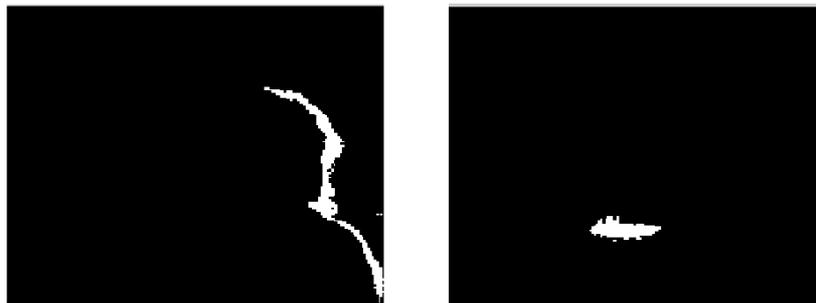
Répéter

$$m = m + 1$$

$$C_m = (C_{m-1} \oplus B_1) \cap S$$

Jusqu'à $(C_m = C_{m-1})$

Le résultat final de la croissance de région est donné par la figure 7.3.



(a) Séquence "Silent"

(b) Séquence "Hall"

FIG. 6.11: Masque de la région d'ombre détectée

6.4.1.9 Localisation de la zone d'ombre portée sur l'image originale

Finalement et après avoir localisé la région d'ombre, on est capable maintenant de délimiter cette zone sur l'image originale, comme l'illustre la figure 6.12. L'ombre est présentée par un contour blanc.

6.5 Conclusion

L'existence des ombres dans les images d'une manière générale est très importante. Ces manifestations photométriques à la surface des objets d'une scène sont considérées dans de



(a) Séquence "Silent"



(b) Séquence "Hall"

FIG. 6.12: Masque de la région d'ombre détectée.

nombreux domaines. Dans ce chapitre, des modélisations géométriques et photométriques des ombres portées ont été étudiées. Plusieurs méthodes s'intéressant aux ombres dans la littérature ont alors été présentées. La détection des ombres a commencé dans le cadre contraint de l'imagerie aérienne, puis s'est étendue dans des cas plus généraux, comprenant notamment la vidéo surveillance. Une approche de détection d'ombre en mouvement dans une séquence vidéo a été présentée. Cette région d'ombre sera étudiée dans le chapitre suivant afin de l'exploiter par la suite dans la phase de tatouage vidéo.

Chapitre 7

Approche de tatouage basée sur les zones d'ombre

7.1 Introduction

Les conditions d'illumination d'une scène prennent une importance assez considérable dans la plupart des méthodes d'analyse d'images. La lumière est une composante qui influence beaucoup sur la façon de voir les choses. En effet, elle peut rendre très différents deux objets identiques. La figure 7.1 montre un exemple illustratif de l'effet de la lumière sur les objets d'une scène. Un même objet a été éclairé par une source lumineuse, posée de différentes manières à différents angles, mais en gardant à chaque fois la même distance. L'objet semble à chaque fois prendre une allure différente. Dans certains cas, il est sombre et une partie de sa texture et/ou sa couleur devient invisible. Face à ces effets, les méthodes d'analyse d'images se fixent en général des connaissances a priori sur le contenu à analyser et se basent sur des hypothèses afin d'améliorer leurs performances. Par exemple, [37] se base sur l'estimation de la couleur d'une surface indépendamment des conditions d'illumination. [94] se base sur la teinte "chair" pour la détection de visages sauf que les résultats dépendent fortement de la variation de la lumière dans la scène. En résumé, les conditions d'illumination influent directement sur la perception de l'environnement en général. Nous exploitons cet effet dans le contexte du tatouage vidéo. En effet, notre approche tend à estimer les conditions d'illumination d'une scène donnée et ensuite adapter

la marque à celles-ci. Autrement dit, si la scène présente un éclairage faible, alors l'oeil humain devient moins sensible aux modifications. Dans ce cas, l'augmentation de l'énergie de la marque dans les images de la scène serait possible. En conséquence, la robustesse sera assurée surtout face aux attaques avec perte telle que la compression.

Pour parvenir à estimer les conditions d'illumination d'une scène, nous proposons la détection des zones d'ombre en mouvement. En effet, une zone d'ombre est une conséquence directe de la lumière existante dans la scène. En effet, l'intensité de l'ombre dépend considérablement des conditions d'illumination : si la scène est éclairée par une lumière ambiante, aucune ombre ne sera créée. Par contre, si la scène est éclairée par une source de lumière, ponctuelle ou étendue, d'intensité forte, on remarquera des ombres assez sombres, sinon ce sont des ombres de faibles intensités qui seront créées.



FIG. 7.1: Influence de la lumière sur un même objet

7.2 Schéma général de l'approche

Comme toute approche de tatouage, il faut spécifier les deux phases d'insertion et de détection de la marque. Nous décrivons successivement ces deux phases en explicitant leurs différentes étapes.

7.2.1 Phase d'insertion

La figure 7.2 présente les différentes étapes de la phase d'insertion. Ayant la vidéo originale à tatouer, celle-ci est décomposée en scènes. Pour chaque scène, on détecte, dans la limite du possible, les zones d'ombres en mouvement. Ces zones serviront ensuite à classifier la scène selon son intensité d'éclairage. Connaissant les conditions d'illumination de la scène, la marque est formatée et adaptée à ces conditions puis elle sera insérée dans la vidéo.

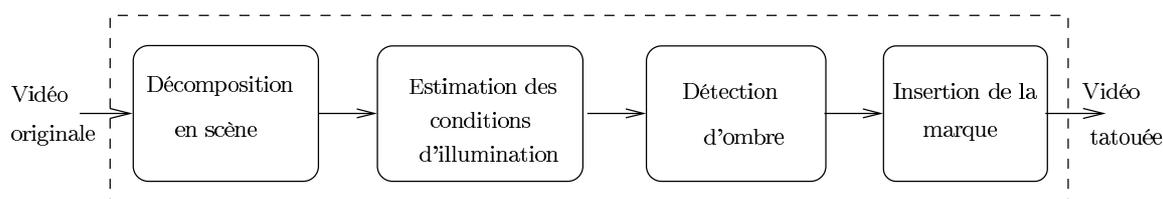


FIG. 7.2: Etapes de la phase d'insertion de la marque

7.2.1.1 Décomposition en scènes

Une vidéo est composée de plusieurs scènes. Une scène représente en fait l'ensemble d'images successives qui représentent un même contexte (même fond, mêmes objets, même intensité de mouvement ...). Dans notre approche, il est impératif de passer par une étape de décomposition de la vidéo en différentes scènes vu que les conditions d'éclairage, en général, changent d'une scène à une autre. La décomposition d'une séquence vidéo en scènes est une tâche complexe. En effet, les approches existantes ([101, 12, 9, 12, 98]) sont objectives et peuvent donner des résultats différents sur une même séquence tant qu'ils se basent sur des critères et des hypothèses différents. Par exemple, une scène définie par le même fond (background) peut être divisée en deux scènes différentes si l'intensité du mouvement des objets de celle-ci varie d'une manière importante. D'une manière subjective, une séquence vidéo peut être décomposée en scènes en se basant sur l'oeil humain. Dans ce cas, la décomposition se base essentiellement sur le choix de l'utilisateur. Dans notre cas, on utilise la manière subjective tenant compte des conditions d'éclairage comme critère de décision.

7.2.1.2 Etude des conditions d'illumination

L'objectif de cette étude est de voir les conditions d'illumination d'une scène donnée en exploitant les zones d'ombres détectées a priori dans celle-ci. En d'autres termes, il s'agit de chercher la classe d'appartenance d'une scène donnée et en particulier son degré d'appartenance à cette classe. La notion de classe est définie par la variation d'éclairage de la scène. Pour cela, on distingue 3 classes : classe des scènes fortement éclairées, classe des scènes moyennement éclairées et classe des scènes faiblement éclairées (sombres). Le nombre de classe peut bien évidemment varier selon le besoin. Pour aboutir à une telle classification, il suffit de voir l'intensité des zones d'ombres dans la scène. En effet, l'intensité d'une ombre dépend d'une manière très forte des conditions d'éclairage : plus la source lumineuse est forte, plus l'ombre est sombre et vice versa. Ces deux exemples

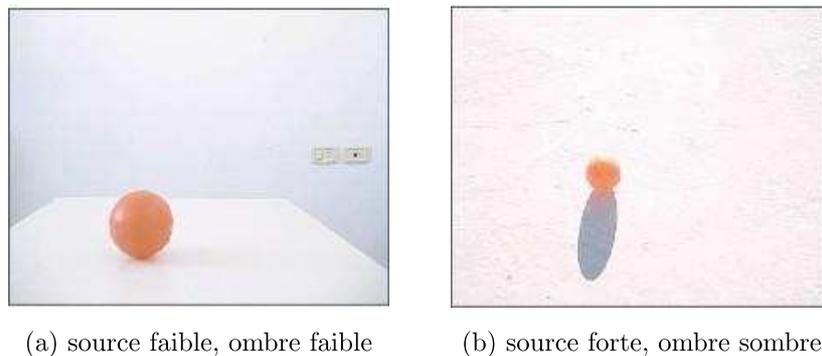


FIG. 7.3: Effet de la variation de la source de la lumière sur l'intensité de l'ombre engendré montrent bien que si les conditions d'éclairage changent, l'intensité de la région d'ombre change aussi. A partir de la zone d'ombre détectée, une scène peut être alors jugée comme fortement, moyennement ou faiblement éclairée.

Etude des variations de la luminance L'idée est d'estimer la variations de la luminance entre une zone dans le cas où elle est ombrée et la même zone dans le cas où elle est non ombrée. Pour cela, il faut choisir deux images à deux instants différents de la même séquence qui ont le même background et où un objet, avec son ombre, qui s'est déplacé. Dans un premier temps, la luminance dans un frame, utilisé comme 'frame de référence', est calculée où la zone d'étude de l'ombre est en dehors de cette image, autrement dit ce frame est considéré comme le background de la scène. Ensuite il faut calculer la lumi-

nance des différents pixels de l'ombre détectée. Les valeurs des différences de luminance sont calculées de la façon suivante :

$$I(p) = I_{ref}(p) - I_{t+n}(p) \quad (7.1)$$

Où $I_{ref}(p)$ représente la luminance du point appartenant à l'image de référence à l'instant t , $I_{t+n}(p)$ est la luminance du pixel correspondant dans l'image présentant la zone d'ombre à étudier.

La figure 7.4 présente un exemple de deux images d'une séquence vidéo, l'une est considérée comme image de référence prise à un instant t , et la deuxième présente l'objet en mouvement avec son ombre extraite à un instant $t + n$.



FIG. 7.4: Localisation de la zone d'ombre à deux instants différents. (a) Image de référence prise à l'instant t . (b) Image étudiée prise à l'instant $t + n$.

A partir du calcul des différences de luminance, on va essayer de tirer les conclusions suivantes : si la variation est importante, alors l'éclairage est intense. Et si la variation est modérée alors l'éclairage est ambiant. Par contre si la variation de luminances est faible, alors il s'agit d'une scène faiblement éclairée. L'intensité de l'ombre varie donc selon l'éclairage de la scène étudiée. Les figures 7.5, 7.6 et 7.7 montrent quelques exemples de scènes présentant des conditions d'illumination différentes. Les valeurs des différences de luminances trouvées, pour chaque scène en appliquant la relation 7.1, sont des valeurs très différentes. Mais le problème qui se pose ici, c'est qu'on n'a pas d'information sûre pour décider qu'une séquence quelconque est fortement, moyennement ou faiblement éclairée. Pour cela, une classification ou un regroupement de ces différentes séquences est nécessaire, dans le but d'avoir un certain seuil à partir duquel, une séquence peut être



FIG. 7.5: Exemples de scènes fortement éclairées



FIG. 7.6: Exemples de scènes moyennement éclairées



FIG. 7.7: Exemples de scènes faiblement éclairées

jugée appartenant à une telle classe. A partir des données qu'on dispose et qui sont sous forme de valeurs de différences de luminosité, on se trouve dans une situation un peu floue puisque certaines valeurs sont très proches par exemple, entre une séquence faiblement éclairée et une séquence moyennement éclairée. On ne peut pas dire ici qu'une séquence appartienne à une classe et une seule, mais par contre, une séquence peut effectivement appartenir simultanément à plusieurs classes à des degrés d'appartenance différents. Pour cela, une classification floue des différentes séquences a été adoptée. Et puisqu'il ne s'agit d'aucune information ou base de référence étiquetée a priori sur les objets à traiter pour effectuer un apprentissage ; c'est-à-dire une nouvelle séquence ne peut pas être jugée à quelle classe elle appartient, et selon quel critère la classer, une classification non supervi-

sée a été choisi. L'algorithme retenu pour la classification est, l'algorithme des c-moyennes floues de Bezdek.

Classification Fuzzy C-Means (FCM) On va commencer par une description brève de quelques propriétés fondamentales des algorithmes de classification floue classiques. On présentera par la suite les résultats de l'application de cette méthode de classification sur les données traitées dans notre contexte.

1. L'algorithme des c-moyennes floues

(a) Définition

L'algorithme des C-moyennes floues (fuzzy c-means : FCM) est une extension directe de l'algorithme classique des c-moyennes (c-means), où l'on a introduit la notion d'ensemble flou dans la définition des classes. Cet algorithme a été développé essentiellement par Bezdek, à partir des idées originales de Ruspini (clustering flou) ([80]) et de Dunn (ISODATA flou) ([21]). Le principe de base est de former à partir des individus non étiquetés c-groupes qui soient les plus homogènes et naturels possible. "Homogène" et "naturel" signifient que les groupes obtenus doivent contenir des individus les plus semblables possible, tandis que des individus de groupes différents doivent être les plus dissemblables possible ([31]).

(b) L'ensemble des données

Le regroupement est une division des données en groupes ou classes d'objets similaires. Chaque groupe est constitué d'objets similaires entre eux et non similaires aux objets d'autres groupes. Les données sont typiquement des observations de certains processus physiques. Chaque observation est constituée de n variables mesurées, groupées en un vecteur colonne de dimension n,

$$z_k = [z_{1k}, \dots, z_{nk}]^T, z_k \in \mathfrak{R} \quad (7.2)$$

Un ensemble de N observations est donné par :

$$Z = \{z_k | k = 1, \dots, N\} \quad (7.3)$$

(c) L'ensemble des sorties

Les sorties de l'algorithme FCM sont sous forme d'un nombre bien défini de

clusters, choisit arbitrairement dès le début, et une matrice d'appartenance contenant les degrés d'appartenance de chaque observation à chaque classe aboutie. Les classes ou clusters sont définies par leurs centres.

2. Application de l'algorithme FCM

La méthode FCM est une méthode qui a l'avantage d'être non supervisée, simple, rapide. Elle nécessite une connaissance sur les données et le regroupement que l'on souhaite obtenir. En fait, certains paramètres doivent être spécifiés en entrée : le nombre de classes, le coefficient de pondération, la tolérance d'arrêt, et la matrice de la norme induite. De plus, la matrice de partition floue, doit être initialisée.

(a) Données en entrées

Les données en entrées sont un ensemble de vecteurs colonnes dont les valeurs sont des différences de luminance des pixels dans la zone d'ombre détectée calculées en utilisant toujours la relation 7.1. Le choix de la région d'ombre est basé sur le principe d'éviter les zones éclairées par une partie de la source lumineuse qui présentent la pénombre, et s'intéresser à la zone de l'ombre portée. La figure 7.8 montre un exemple d'extraction d'une surface de l'ombre portée à partir de la zone d'ombre totale.

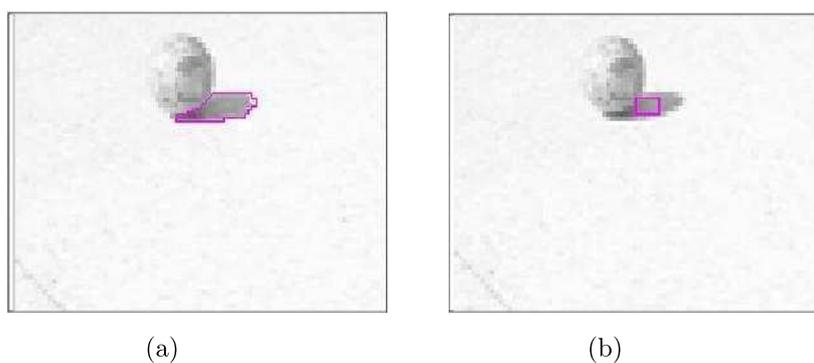


FIG. 7.8: Choix de la zone d'ombre à étudier. (a) Zone d'ombre totale. (b) Partie de la zone d'ombre portée.

Deux vecteurs de tailles différentes sont utilisés comme entrées de l'algorithme FCM. Le premier ensemble est de taille 500, et le deuxième est un bloc de 64×64 , soit 4096 valeurs.

(b) Sorties du classifieur FCM

Les sorties du classifieur FCM sont sous forme de groupes de valeurs (des clusters) qui sont décrits par leurs centres, et une matrice de degrés d'appartenances de chaque entrée à chacune des clusters résultants. Comme il est montré dans la partie qui suit.

– Classement en trois classes

Dans un premier cas, les vecteurs d'entrée sont de dimensions 500. Donc la matrice en entrée est de 500 colonnes et le nombre de lignes correspondent aux nombre de séquences étudiées. En appliquant l'algorithme FCM, les valeurs de luminance des différentes séquences étudiées sont groupées en trois classes. Les centres des trois classes trouvées sont 20.79 ; 23.17 et 48.94. On peut dire qu'une zone d'ombre où les valeurs de luminance sont autour de la valeur la plus faible peut être jugée comme une zone appartenant à une séquence faiblement éclairée. Par contre si les valeurs des différences de luminance sont autour de la valeur 23.17, donc on peut dire que la zone d'ombre appartient à une séquence moyennement éclairée. Finalement si ces valeurs sont plus importantes, et sont autour de la valeur 48.94, donc la séquence à laquelle appartient la zone d'ombre présente un éclairage plus important.

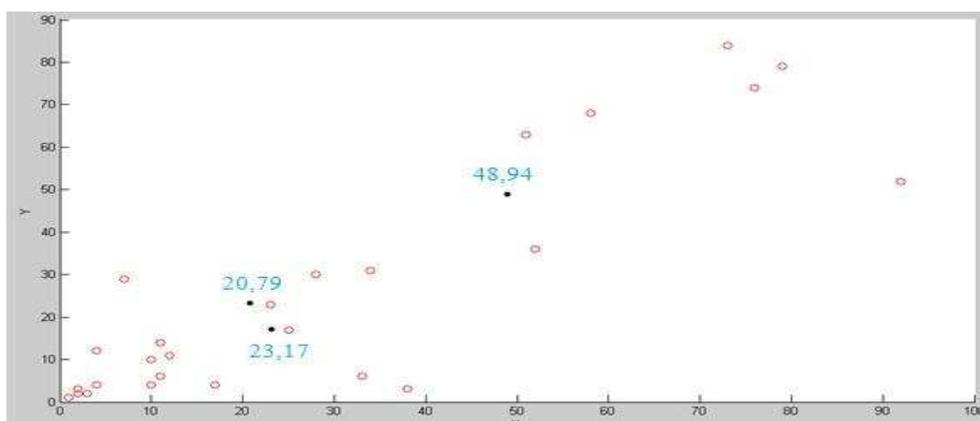


FIG. 7.9: Classement FCM en trois classes(nombre de pixels=500)

En utilisant un vecteur dont la taille est de 4096 valeurs, comme entrée du classifieur FCM, le résultat du classement trouvé est donné par la figure 7.10. Les classes résultantes sont respectivement de centres 23.72 ; 27.37 et 44.06. A partir des séquences étudiées, les classes résultantes correspondent aux

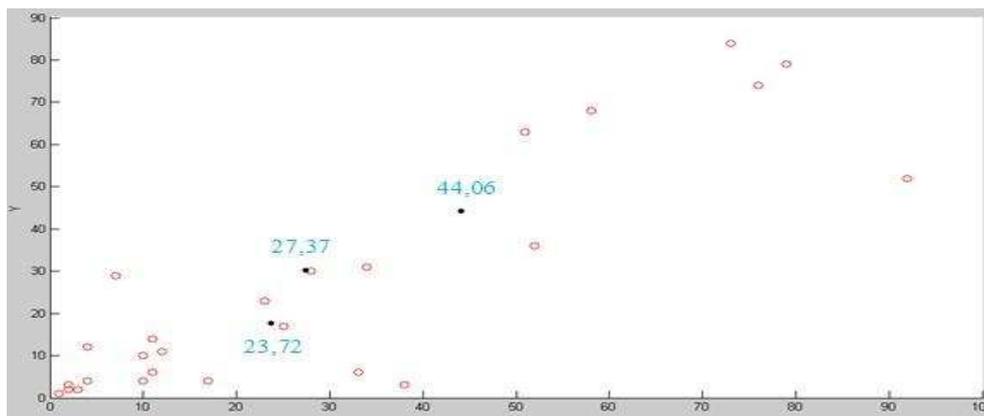


FIG. 7.10: Classement FCM en trois classes (nombre de pixels=4096(bloc de 64×64))

centroïdes trouvés. Ces séquences partagent un degré d'appartenance aux différentes classes avec une somme des degrés égale à 1. Le cluster auquel est associée une séquence est celui dont le degré d'appartenance sera le plus élevé. Ces degrés sont représentés sous forme d'une matrice permettant de définir les appartenances floues de chaque vecteur d'entrée à chacune des trois classes résultantes. Le tableau 7.1 présente l'exemple de matrice d'appartenance résultante de la classification FCM de 31 séquences vidéo dont la zone d'ombre étudiée est de 4096 pixels, en trois classes.

– Classement en quatre classes

Dans ce cas, les valeurs de luminance des différentes séquences étudiées sont groupées en quatre classes. Pour un vecteur de données de 500 pixels, les centres des quatre classes trouvées sont 17.34 ; 25.62 ; 30.61 et 54.31 comme le montre la figure ci-dessous. Pour un vecteur d'entrée de taille 4096 pixels, les quatre classes résultantes sont respectivement de centres : 18,76 ; 27,55 ; 30,55 et 49,68.

D'après ce regroupement, les différentes séquences peuvent être classées en quatre classes : séquences faiblement éclairées, moyennement éclairées, fortement éclairées et très fortement éclairées.

– Prise de décision

En travaillant sur un nombre d'entrées différent, les centroïdes des différentes classes changent. Le tableau 7.2 montre le résultat de la variation du nombre

de vecteurs d'entrée sur les centres des classes en fixant le nombre de classes à 3. D'après ces résultats, on peut dire que le centre de la première classe varie autour de la valeur 20, le centre de la deuxième classe est autour de 30 et pour la troisième classe elle présente un centroïde s'approchant de la valeur 40. L'effet de la modification du nombre d'entrées sur un nombre de clusters de 4 et en jouant sur le même nombre d'entrées est donné par le tableau 7.3.

On peut dire que la première classe est centrée au niveau de la valeur 20, la deuxième admet pour centre une valeur autour de 25. Une troisième classe a pour valeur variant autour d'un centre de 35, et une quatrième classe est de centre variant autour de 50 à peu près. Donc d'après ces résultats, on peut dire que généralement on peut prendre la décision suivante : si les valeurs de luminance de la zone d'ombre sont autour de la valeur 20, alors cette séquence peut être jugée faiblement éclairée. Si ces valeurs sont autour de la valeur 25 à 30 donc on peut dire qu'il s'agit d'une séquence moyennement éclairée. Si les valeurs de luminance sont autour de 35 à 40 donc cette séquence est fortement éclairée. Et si les valeurs de luminance atteignent la valeur 50 alors on peut dire ici qu'il s'agit d'une séquence très fortement éclairée.

(c) Traiter une nouvelle séquence

Pour classifier une nouvelle séquence inconnue, où la zone d'ombre doit être déjà détectée et les valeurs de luminances engendrées par l'ombre sont déjà calculées à l'aide de la relation 7.1, cette séquence doit être affectée au centroïde le plus proche des classes obtenues. Autrement dit si les valeurs de luminance calculées d'une séquence donnée sont autour d'un centre d'une classe des classes résultantes, alors on peut juger que cette séquence appartient automatiquement à cette classe. En utilisant la méthode FCM, on peut ajouter un nouveau vecteur de données contenant bien évidemment les valeurs de différences de luminances calculées dans la zone d'ombre détectée de cette séquence. Et on relance notre programme. En sortie l'algorithme donne comme résultat les degrés d'appartenance de cette séquence à chacune des classes obtenues. Le degré d'appartenance le plus important présentera la classe à laquelle appartient la

séquence désirée. Mais cette séquence peut appartenir aux autres classes à un degré bien défini comme il est déjà montré ci-dessus.

7.2.1.3 Insertion de la marque

Le schéma de tatouage utilisé est un schéma additif dans le domaine spatial. Il comporte 4 étapes :

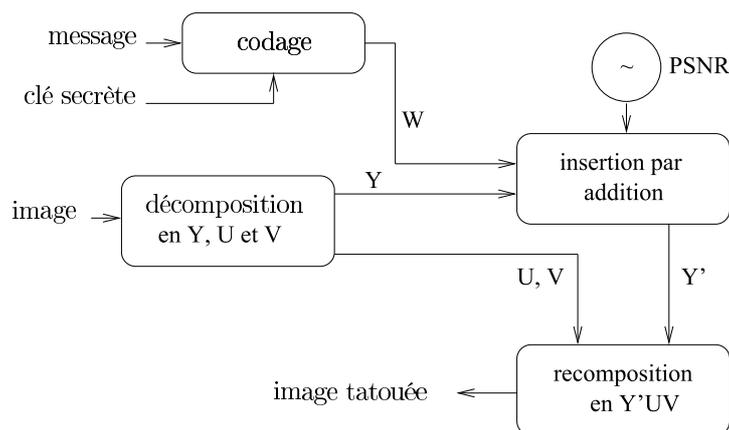


FIG. 7.11: Schémà d'insertion de la marque

1. Décomposition en YUV : chaque image de la scène est décomposée en trois composantes (Y, U et V). L'insertion de la marque se fait au niveau de la composante Y puisqu'elle représente 70% de l'information utile.
2. Codage : cette étape consiste à créer la marque w qui va être insérer dans les images de la vidéo. Pour cela, nous utilisons la technique d'étalement de spectre. En effet, le message, qui représente l'information à cacher, est étalé sur la taille de l'image à travers une séquence aléatoire générée à partir d'un code secret.
3. Insertion additive : la marque w est insérée dans la composante Y de l'image par addition selon la fonction suivante : $Y' = Y + \alpha \cdot m_i \cdot w$ Avec α un facteur de pondération (force du tatouage) qui dépend de la valeur de PSNR fournie par la phase d'étude des conditions d'illumination. m_i est un masque psychovisuel qui met en evidence les zones hétérogènes de l'image à savoir la texture et les contours.
4. Recomposition : une fois les composantes Y sont tatouées, on procède à la reconstruction des images pour avoir la scène tatouée.

7.3 Phase de détection

Le schéma de tatouage adoptée est un schéma semi aveugle, c'est-à-dire qu'il nécessite seulement la connaissance de la marque originale au moment de la détection. La détection consiste à déterminer si la marque originale existe toujours dans la vidéo tatouée. On est donc amené à détecter sa présence ou son absence. Pour cela, nous calculons une valeur de similarité (corrélation) entre la marque originale et les composantes Y de la scène tatouée. Si la similarité dépasse un seuil (dans notre cas le seuil est défini d'une manière expérimentale) alors on peut affirmer la présence de la marque, sinon la marque est probablement détruite suite à une attaque intentionnelle ou malhonnête. Dans notre étude, nous utilisons une corrélation linéaire définie comme suit :

$$sim = \langle w, \hat{w} \rangle \quad (7.4)$$

7.4 Phase d'attaque

Comme nous l'avons vu précédemment, un des points forts d'un tatouage efficace réside dans sa robustesse. Néanmoins, les attaques basiques peuvent effacer le marquage, ou du moins potentiellement l'altérer. La robustesse du tatouage appliquée dans cette approche est évaluée face à des attaques spécifiques à la vidéo visant à détruire la marque entière ou une partie de la marque. Dans notre contexte, notre évaluation s'est limitée sur l'attaque de compression utilisant le codec mpeg4 en variant plusieurs débits allant d'une bonne qualité de compression (2000 Kbits/s) jusqu'à une mauvaise qualité (500 et même 200 kbits/s). Le choix de cette attaque s'explique par le fait que notre approche a pour objectif d'augmenter l'énergie de la marque en exploitant les conditions d'illumination de la scène. Et pour montrer l'efficacité de cette stratégie, nous devons adopter une attaque qui vise à atténuer ou même détruire le signal qui représente la marque. La compression semble donc adéquate pour ce genre d'évaluation.

7.5 Conclusion

Nous avons présenté, dans ce chapitre, une approche de tatouage basée sur la détection des zones d'ombre en mouvement. En effet, l'ombre représente une source d'information sur la scène puisque, à partir de ces zones, nous pouvons estimer les conditions d'illumination. Le tatouage est alors adapté à ces conditions dans le but d'augmenter son énergie au niveau des images de la scène qui représentent une lumière faible. En effet, l'oeil humain est moins sensible aux modifications au niveau des images plus ou moins sombres. Nous montrons dans le chapitre suivant, une étude expérimentale afin de valider cette approche.

TAB. 7.1: Matrice d'appartenance des séquences vidéo aux trois classes engendrées

Numéro séquence	classe 1	classe 2	classe 3
1	0,432237127	0,388512559	0,179250313
2	0,009831492	0,028742223	0,961426284
3	0,026585536	0,092356615	0,881057849
4	0,027789245	0,071101493	0,901109262
5	0,166249972	0,63619161	0,197558419
6	0,740184759	0,191756269	0,068058972
7	0,271176419	0,603588758	0,125234822
8	0,016935558	0,063742303	0,91932214
9	0,011511669	0,033122509	0,955365821
10	0,655692058	0,262174027	0,082133915
11	0,153767425	0,611330283	0,234902292
12	0,10442582	0,34436074	0,55121344
13	0,116712156	0,435703667	0,447584177
14	0,010442204	0,032692364	0,956865431
15	0,039099656	0,145639276	0,815261069
16	0,021856595	0,060907537	0,917235868
17	0,029193175	0,084942503	0,885864321
18	0,058697534	0,198667275	0,742635191
19	0,013590137	0,039216199	0,947193664
20	0,028231229	0,072078684	0,899690087
21	0,058697534	0,198667275	0,742635191
22	0,041917387	0,131506452	0,826576162
23	0,034141935	0,120980499	0,844877565
24	0,13044592	0,391031522	0,478522558
25	0,334336357	0,466634337	0,199029306
26	0,651890795	0,242585055	0,10552415
27	0,278663675	0,508419538	0,212916787
28	0,167794426	0,63744923	0,194756345
29	0,304046846	0,552189907	0,143763247

TAB. 7.2: Résultats du changement du nombre d'entrées du classifieur FCM avec $c=3$

	31 séquences	26 séquences	23 séquences	20 séquences
Centre1	23,72	26,65	24,75	28,9
Centre2	27,37	33,4	28,52	31,1
Centre3	44,06	36 ,12	39,53	41,37

TAB. 7.3: Résultats du changement du nombre d'entrées du classifieur FCM avec $c=4$

	31 séquences	26 séquences	23 séquences	20 séquences
Centre1	18,76	20,33	17,79	22,33
Centre2	27,55	34,41	25,32	25,1
Centre3	30,55	35,69	36,15	37,07
Centre4	49,68	40,19	49,01	54,03

Chapitre 8

Résultats expérimentaux

8.1 Introduction

Dans cette partie, nous allons exploiter les conditions d'illumination, étudiées préalablement, dans la phase de tatouage. En effet, la force de tatouage sera adaptée à l'intensité lumineuse de la scène en question. Nous allons montrer l'effet de cette adaptation sur la robustesse du schéma de tatouage surtout face aux attaques de compression.

8.2 Choix de la valeur du PSNR

La signature à insérer dans une séquence vidéo est considérée comme un bruit additif qui influe sur la qualité du signal original. Le PSNR est une métrique qui donne une idée sur la qualité du signal, ou en d'autre terme le taux de dégradation. En effet, les résultats expérimentaux ont prouvé qu'une image modifiée de PSNR égal à 30db est considérée comme étant une image de bonne qualité. Dans notre contexte, nous nous permettons d'avoir une vidéo tatouée de qualité moyenne (PSNR faible) dans les scènes faiblement éclairées et ce dans le but d'augmenter l'énergie de la marque. Pour cela nous devons définir une équation qui représente la valeur du PSNR en fonction de la luminance de la scène. Sachant que pour une très bonne qualité du signal nous avons supposé une valeur de PSNR de 35db et pour une qualité dégradée nous avons pris comme valeur 20db.

L'équation résultante est alors une fonction linéaire affine :

$$PSNR = 15x + 20 \quad (8.1)$$

Où x : la valeur d'appartenance à la classe des scènes fortement éclairées ; $0 < x < 1$
 Si la valeur de x s'approche de la valeur zéro (scène sombre) alors on se trouve dans le cas d'un PSNR de 20db, et si par contre x s'approche de 1 (scène fortement éclairée) le PSNR s'approche de 35db. Nous avons effectué une série de tests sur un ensemble de séquences vidéo dont les conditions d'éclairage sont différentes. Ces séquences sont composées d'un ensemble de frames de 176x144 pixels. D'après l'étude des conditions d'illumination réalisée dans le chapitre précédent et le résultat de classification *fcm*, nous avons pris deux exemples de séquences pour chaque classe. Les deux premières séquences appartiennent à la classe 1, qui présente la classe des scènes fortement éclairées où la différence de luminosité de la zone d'ombre et le fond fixe de la scène est importante. Les séquences 2 et 3 appartiennent à la classe moyennement éclairée, et les deux dernières séquences appartiennent à la classe 3 ou les scènes sont faiblement éclairées. Le tableau 8.1 présente les résultats numériques de notre méthode de calcul du *PSNR* :

TAB. 8.1: Calcul du psnr en fonction des conditions d'éclairage

	Séq1	Séq2	Séq3	Séq4	Séq5	Séq6
Appartenance à C3	0,9614	0,9569	0,2349	0,1948	0,0821	0,1055
PSNR (db)	34	34	24	23	21	21

8.3 Résultats Expérimentaux

8.3.1 Avant adaptation de la marque

Avant de voir l'effet de l'adaptation de la marque aux conditions d'illumination de la scène, nous montrons dans la figure 8.1 d'abord les valeurs de corrélations avant et après marquage pour voir en fait le seuil de détection. Ces valeurs seront des références pour comparer et évaluer d'abord l'effet l'adaptation de la marque aux conditions d'éclairage et ensuite l'effet de la compression à différents débits. Aussi, nous montrons dans la figure

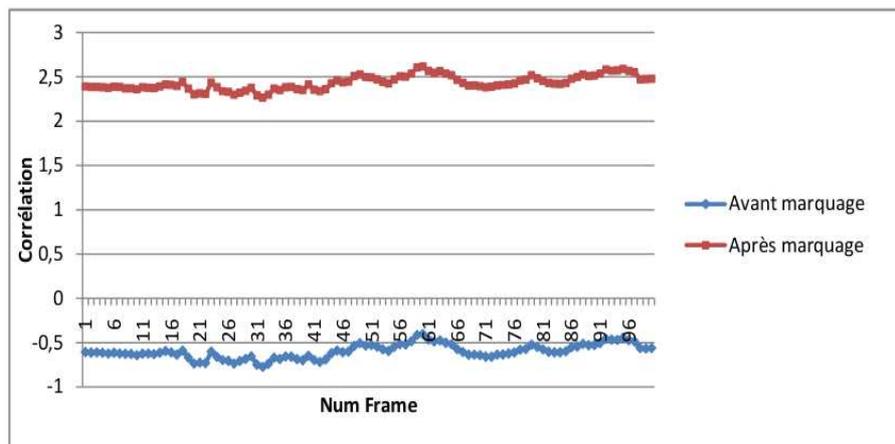


FIG. 8.1: Corrélation avant et après marquage

8.2, le résultat de la phase de tatouage utilisant un jeu de codes secrets. Le but est de montrer que le pic de corrélation est obtenu seulement au niveau du code qui a servi à l'insertion de la marque (ici le code est 2500).

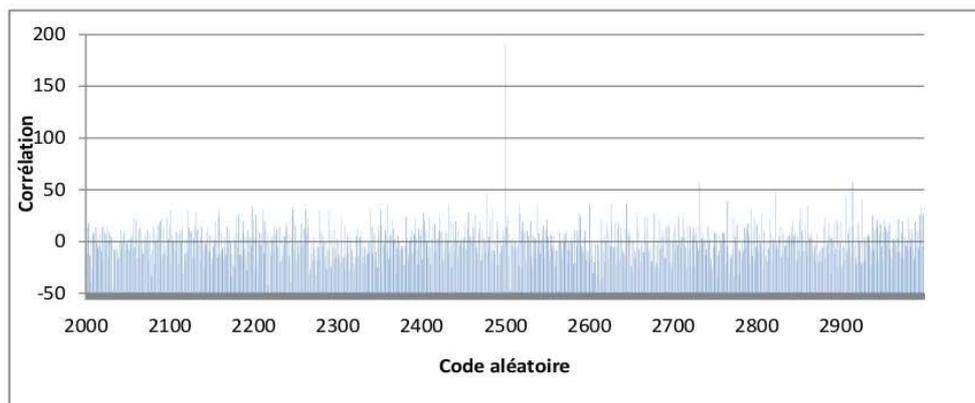


FIG. 8.2: Représentation de la corrélation en fonction de porteuses aléatoires

8.3.2 Résultats d'insertion à différents PSNRs en tenant compte des conditions d'éclairage

Nous montrons dans cette partie l'effet de l'adaptation de la marque aux conditions d'illumination. Dans le cas où la scène est fortement éclairée, on adopte un PSNR égal à 35db, c'est-à-dire qu'on ne se permet pas d'augmenter l'énergie de la marque puisqu'elle sera visible. Par contre, dans les scènes moyennement et faiblement éclairées, on adopte des

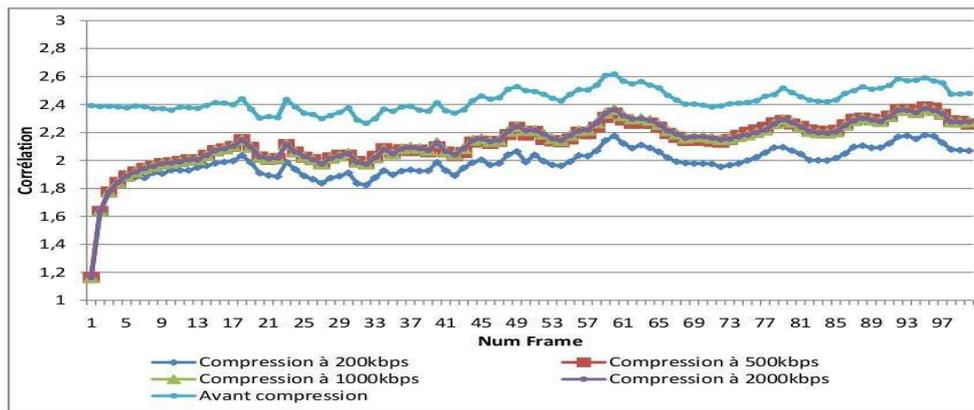


FIG. 8.3: Compression à différents débits et un PSNR=35db

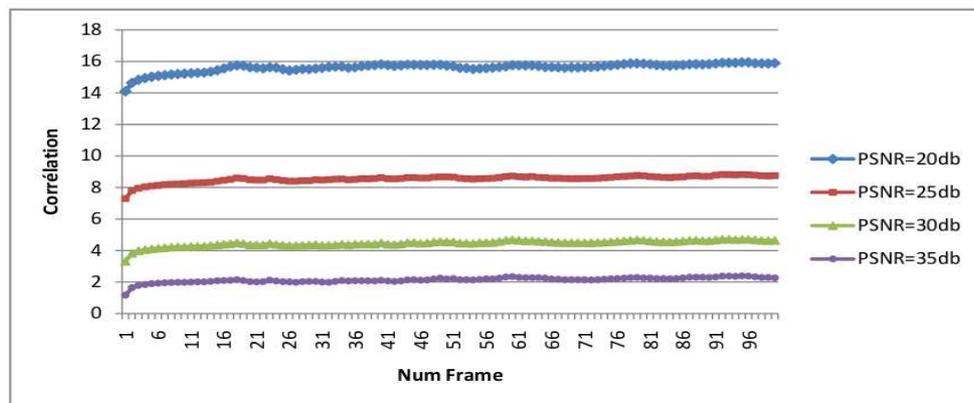


FIG. 8.4: Compression à différents PSNR et un débit de 500kbps

valeurs de PSNR plus faibles proportionnellement à l'intensité lumineuse. Bien évidemment, dès qu'on augmente l'énergie de la marque, la corrélation au niveau de la détection devient plus importante. Ceci peut être remarqué dans les figures 8.5, 8.6, 8.7 et 8.8.

8.3.3 Attaques par compression

La robustesse du tatouage appliqué dans cette approche doit être évaluée face à des attaques spécifiques à la vidéo visant à détruire ou à affaiblir la signature. La compression, utilisant le codec MPEG4, est une attaque très utilisée. Nous testons et évaluons la robustesse de notre schéma de tatouage face à la compression en variant plusieurs débits.

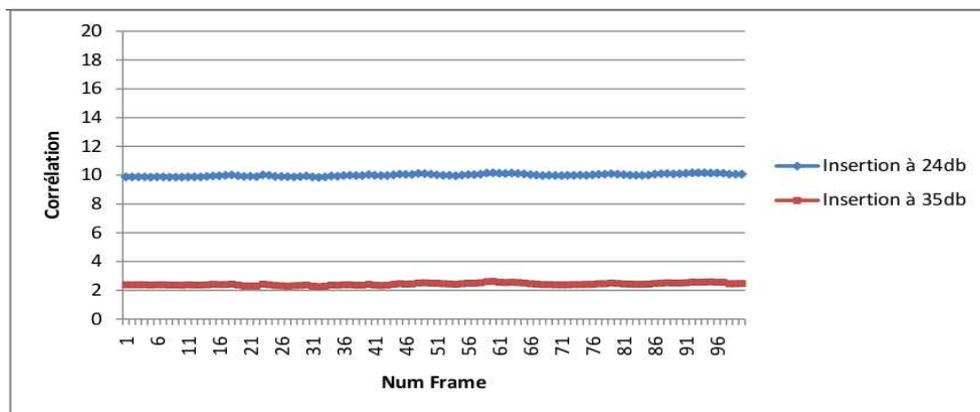


FIG. 8.5: Résultats de corrélation avec PSNR variable dans le cas de la séquence 3

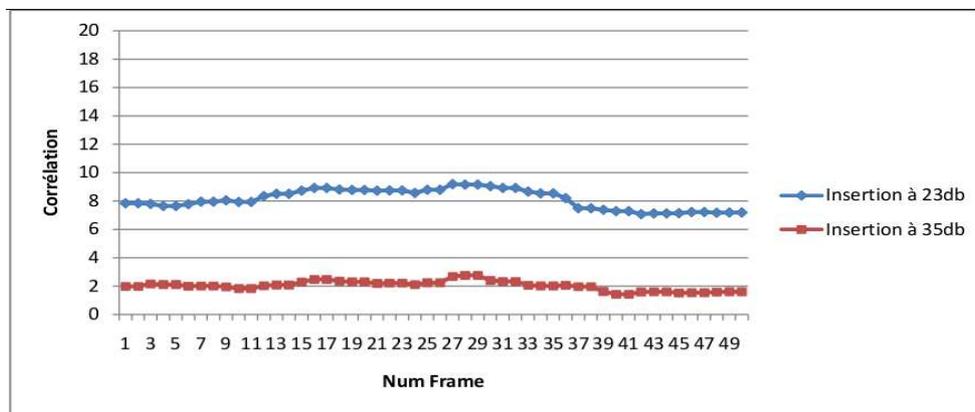


FIG. 8.6: Résultats de corrélation avec PSNR variable dans le cas de la séquence 4

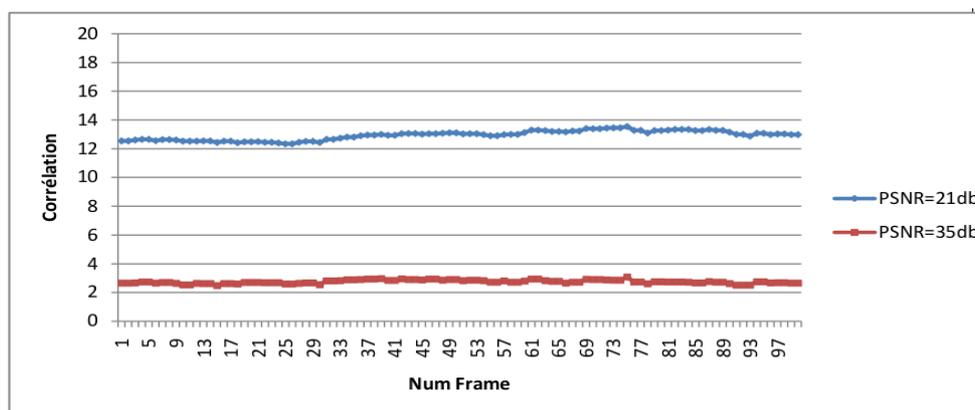


FIG. 8.7: Résultats de corrélation avec PSNR variable dans le cas de la séquence 5

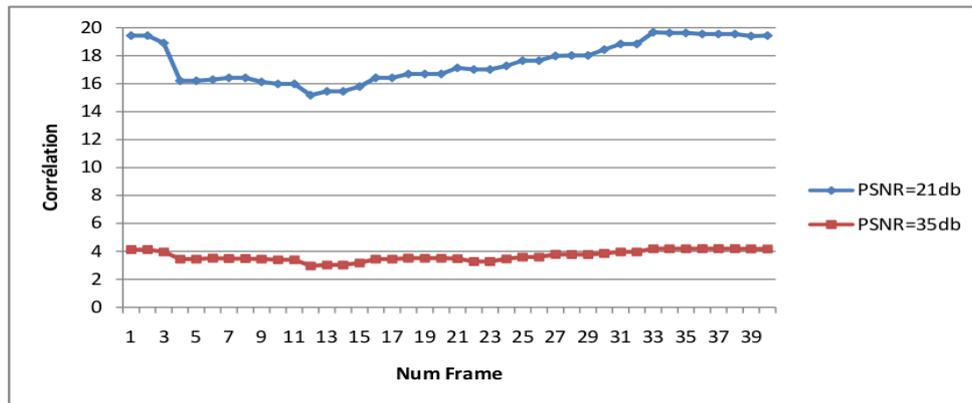


FIG. 8.8: Résultats de corrélation avec PSNR variable dans le cas de la séquence 6

8.3.3.1 Résultats de compression avec débit constant et PSNR variable

Les figures ci-dessous montrent d'une part l'effet de la compression sur les valeurs de corrélations au niveau de la détection et d'autre part l'effet de l'adaptation de la marque aux conditions d'illumination. D'après ces figures, on voit bien qu'après une compression à 1000kbps/s, on perd, même légèrement, au niveau de la corrélation. Par contre, dans le cas où on considère la luminosité de la scène et on insère la marque à 24db dans le cas d'une scène moyennement éclairée et 21db dans le cas d'une scène faiblement éclairée, la corrélation devient plus évidente même après la compression. C'est-à-dire que la marque peut toujours être récupérée même après attaque.

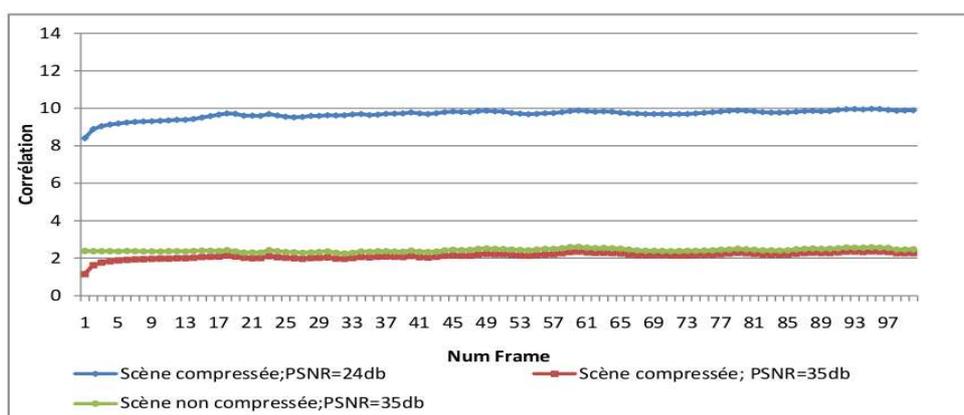


FIG. 8.9: Résultats de corrélation après compression à 1000kbps avec PSNR variable dans le cas de la séquence 3

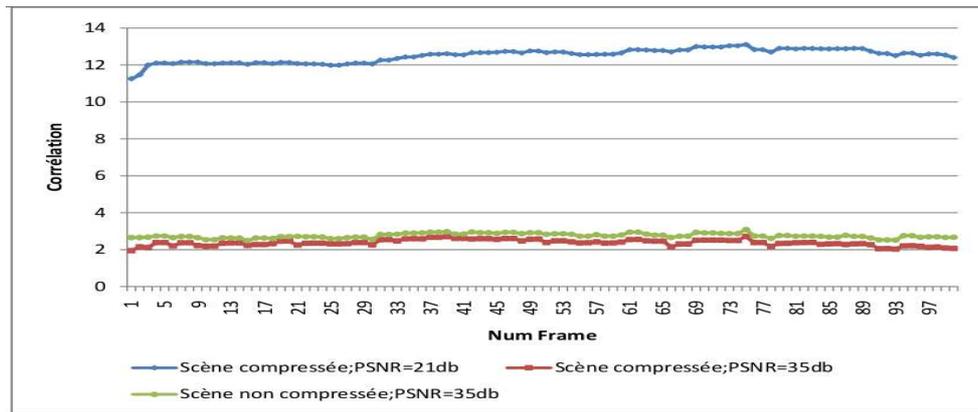


FIG. 8.10: Résultats de corrélation après compression à 1000kbps avec PSNR variable dans le cas de la séquence 5

8.3.3.2 Résultats de compression à différents débits et PSNR constant

Dans cette partie, nous essayons de voir l'effet de la compression à différent débits et non seulement à 1000kbps/s. Pour cela, nous avons appliqué la compression allant d'une bonne qualité (2000kbps/s) à une mauvaise qualité (200kbps/s). Dans le cas où la luminosité de la scène est considérée, on remarque que la marque peut toujours être récupérée puisque la corrélation à 200kbps/s est toujours supérieure à celle avant compression. (voir figure 8.11 et 8.12).

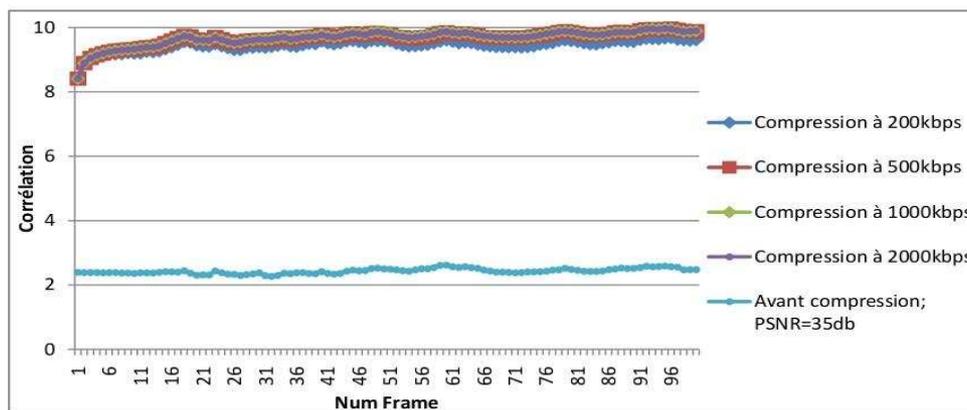


FIG. 8.11: Résultats de corrélation après compression à différents débits et PSNR=24db dans le cas de la séquence 3

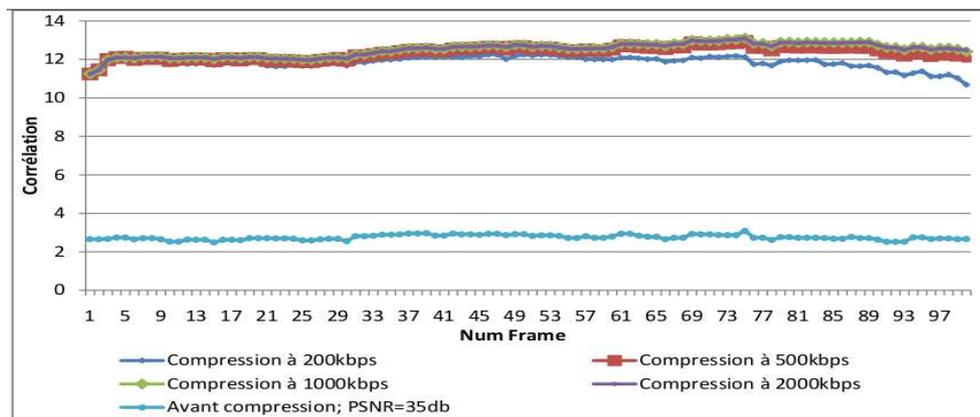


FIG. 8.12: Résultats de corrélation après compression à Débit variable et PSNR=21db dans le cas de la séquence 5

8.4 Conclusion

Ce chapitre a été consacré à la présentation des différents résultats expérimentaux de notre approche de tatouage adoptée. Nous avons réussi à exploiter les conditions d'éclairage d'une scène vidéo quelconque dans la phase de tatouage et nous avons pu étudier l'effet de l'adaptation de la marque insérée sur la robustesse de la méthode de tatouage face aux attaques de compression à différents débits.

Conclusion et Perspectives

L'évolution spectaculaire des moyens de communication a engendré plusieurs failles de sécurité. Le problème de la copie illégale, de la protection du copyright en font partie. Le tatouage s'est avéré une alternative pour la protection des documents multimédia, en particulier l'image, le son et la vidéo. En effet, le tatouage consiste à introduire dans le document à protéger certaines informations utiles permettant de prouver la propriété de l'auteur du document et de détecter l'origine d'une copie illégale. De nouvelles approches de tatouage permettent même de détecter ceux qui ont participé à créer une copie piratée à travers la combinaison de plusieurs copies originales.

Dans ce contexte se présente notre travail. En effet, notre objectif était de contribuer à l'amélioration des techniques de tatouage existantes en proposant de nouvelles méthodes. Améliorer la performance d'une technique de tatouage revient, en partie, à augmenter sa robustesse, c'est-à-dire sa capacité à résister aux différentes attaques que peut subir un document tatoué. Pour atteindre cet objectif, nous nous sommes basés sur la notion de régions d'intérêt dans la vidéo. Une région d'intérêt se définit comme étant une zone ou un espace dans la vidéo qui favorise le tatouage. L'insertion de la marque dans ces zones apporte plus en terme de robustesse. Pour cela, nous avons proposé deux idées :

- concevoir une méthode de tatouage basée sur les images mosaïques. En effet, une image mosaïque permet de fournir certaines informations relatives au contenu de la vidéo et qui sont utiles pour le tatouage. On a vu, qu'à partir de la mosaïque, la localisation des différents points physiques dans la scène peut être faite. Cette localisation a permis de tatouer par la même marque chaque point physique dans ses différentes positions dans la scène. D'après les tests expérimentaux, cette approche a montré une efficacité importante surtout qu'elle apporte une robustesse face à l'attaque de collusion. En plus, tatouer par la même marque tous les pixels

représentant le même point physique permet d'étaler la marque sur toute la scène. Par conséquence, la suppression de certaines images n'influe pas sur le processus de détection. Les tests expérimentaux ont montré que la marque peut toujours être récupérée même après une telle attaque.

D'un autre coté, le tatouage utilisant une énergie plus importante au niveau des images où le mouvement de la caméra est rapide a été un point positif dans le sens où l'insertion dans ces images ne nuit pas à l'aspect visibilité du contenu. En effet, du fait que l'œil humain devient moins sensible quand le mouvement de la caméra est rapide, on a pu introduire plus d'énergie de la marque. Par conséquence, la robustesse est améliorée surtout face à l'attaque de compression avec perte.

- concevoir une méthode de tatouage basée sur les zones d'ombre en mouvement. L'idée consistait à exploiter les conditions d'illumination d'une scène pour former un tatouage plus robuste. Il s'agit de détecter, à travers la détection des zones d'ombre, les scènes faiblement éclairées ou même sombres qui favorisent l'insertion de la marque avec plus d'énergie, se basant sur le fait que l'œil humain est moins sensible aux modifications dans de telles zones. La technique de détection d'ombre utilisée a été couplé avec un classifieur flou pour classer une scène donnée selon son degré de luminosité. Le tatouage est enfin adapté à l'intensité de la source lumineuse de la scène. Les études expérimentales ont montré que le fait d'adapter la marque aux conditions d'illumination augmente considérablement la robustesse de la méthode de tatouage face aux attaques de compression avec perte.

Perspectives

Nos perspectives concernent d'abord l'amélioration de ce qui a été proposé pendant ce travail afin de rendre plus performantes les méthodes proposées. Pour cela, on propose de :

- Exploiter le mouvement de la caméra tels que le mouvement de rotation et les effets de zoom et non seulement la translation.
- Poursuivre l'exploitation des zones d'ombre pour le tatouage. Une alternative serait d'introduire la marque dans les ombres détectées. Une pré-analyse de ces zones serait

nécessaire pour savoir si elles favorisent le tatouage ou non : une ombre de faible intensité projetée sur une région homogène ne sera pas intéressante. Par contre, si l'ombre est sombre et se trouvant sur une zone texturée par exemple, on pourrait éventuellement introduire la marque au niveau de cette zone avec une force de tatouage assez importante.

- Trouver le bon compromis entre les différentes régions d'intérêt pour optimiser l'énergie à insérer. En effet, si on dispose de plusieurs régions d'intérêt, il faudra trouver la meilleure combinaison à faire. On se trouve dans ce cas devant un problème d'optimisation multicritères.

Comme autre perspective, nous visons l'exploitation des résultats obtenus dans ce travail pour concevoir une approche complète de tatouage dans le contexte de la traçabilité des documents multimédias. En effet, dans le contexte de la "Video on Demand" (VoD) par exemple, chaque utilisateur reçoit une version personnalisée du document, contenant un identifiant personnel inséré grâce à une technique de tatouage robuste. Ainsi, si une copie est rediffusée telle quelle de manière illégale, il sera alors possible de remonter à l'utilisateur malhonnête qui peut être traduit en justice. Pour cela, un score est calculé pour chaque utilisateur pour déterminer quelles personnes ont participé dans la génération de la copie illégale. Dans ce cadre, on propose d'optimiser le coût de calcul des scores. Ceci peut être réalisé en optant pour une nouvelle forme pour les mots de code, qui soit hiérarchique selon une structure arborescente. Il s'agit de rassembler les clients en différents groupes selon certains critères. Dans ce cas, le mot de code ne désigne plus seulement l'identité du client mais aussi son appartenance à un groupe. De cette manière, le score ne sera pas calculé pour chaque client mais plutôt pour chaque groupe, ensuite pour chaque sous-groupe du groupe ayant le score le plus élevé, etc.

Bibliographie

- [1] A. Abrardo and M. Barni. Informed watermarking by means of orthogonal and quasi-orthogonal dirty paper coding. *IEEE Transactions on Signal Processing*, 53 :824–833, 2005.
- [2] O. Alattar and A. Alattar. A fast hierarchical watermark detector for real-time software or low-cost hardware implementation. In *IEEE International Conference on Image Processing*, 2005.
- [3] Patrick Bas, N. Boulgouris, F. Koravos, J.-M. Chassery and M. Strintzis, and Benoit Macq. Robust watermarking of video objects for mpeg-4 applications. In *Applications of Digital Image Processing XXIV*. SPIE, 2001.
- [4] Patrick Bas, Jean M. Chassery, , and Franck Davoine. A geometrical and frequential watermarking scheme using similarities. In *Security and Watermarking of Multimedia Contents*, pages 264–272. SPIE, 1999.
- [5] Patrick Patrick Bas and Chassery Jean-Marc. Tatouage couleur adaptatif fondé sur l'utilisation d'espaces perceptifs uniformes. *Traitement du signal*, 21(5) :517–531, 2004.
- [6] S. Baudry. *Stratégies de codage de canal pour le tatouage de vidéo*. PhD thesis, ENST Paris, 2002.
- [7] W. Bender and D. Morimoto. Techniques for data hiding. In *IBM Systems Journal*, volume 2420. SPIE, 1995.
- [8] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2) :239–256, 1992.

-
- [9] G Boccignone, A Chianese, V Moscato, and A Picariello. Foveated shot detection for video segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, (15) :365–377, 2005.
- [10] Yann Bodo. *Elaboration d'une technique d'accès conditionnel par tatouage et embrouillage vidéo basée sur la perturbation des vecteurs de mouvement*. Signal et images, Ecole National Supérieur des Télécommunication, 2003.
- [11] Yann Bodo, Nathalie Laurent, and Jean-Luc Dugelay. Tatouage vidéo par marquage hiérarchique des vecteurs mouvements. In *CORESA 2003, 8èmes journées d'études et d'échanges Compression et Représentation des Signaux Audiovisuel*, Lyon, France, 2003.
- [12] J S Boreczky and L A Rowe. Comparison of video shot boundary detection techniques. In *in Proc. IS&T/SPIE Intern. Symposium Electronic Imaging*, 1996.
- [13] P. Campisi and A. Neri. Video watermarking in the 3d-dwt domain using perceptual masking. In *IEEE International Conference on Image Processing*. IEEE, 2005.
- [14] J Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6) :679–698, 1986.
- [15] B. Chen and G. Wornell. An information-theoretic approach to the design of robust digital watermarking systems. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999.
- [16] Brian Chen and Gregory W. Wornell. Quantization index modulation : A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4) :1423–1443, 2001.
- [17] L. Coria-Mendoza, P. Nasiopoulos, and R. Ward. A robust watermarking scheme based on informed coding and informed embedding. In *IEEE International Conference on Image Processing*. IEEE, 2005.
- [18] Ingemar J. Cox, Joe Kilian, Tom Leighton, and Talal Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12) :1673–1687, 1997.
- [19] Gwenael Doerr. *Security issue and collusion attacks in video watermarking*. PhD thesis, EURECOM, 2004.

-
- [20] Gwenaél Doerr and Jean-Luc Dugelay. Problématique de la collusion en tatouage vidéo. *Traitement du Signal*, 22(6) :563–574, 2005.
- [21] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3 :32–57, 1973.
- [22] E. Sayrol, M. Soriano, M. Fernandez, J. Casanellas, and J. Tomas. Development of a platform offering video copyright protection and security against illegal distribution. In *IS&T/SPIE International Symposium on Electronic Imaging of Proceedings of the SPIE, Security, Steganography, and Watermarking of Multimedia Contents VII.*, volume 5681, pages 76–83. SPIE, 2005.
- [23] J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod. Scalar costa scheme for information embedding. *IEEE Transactions on Signal Processing*, 51(4) :1003–1019, 2003.
- [24] F. Deguillaume, G. Csurka, J. O’Ruanaidh, and T. Pun. Robust 3d dft video watermarking. In *Security and Watermarking of Multimedia Contents I*, volume 3657, pages 113–124. SPIE, 1999.
- [25] Feldmar. *Recalage rigide, non rigide et projectif d’images médicales tridimensionnelles*. PhD thesis, Ecole Polytechnique, Décembre 1995.
- [26] R. Fischer, R. Tzschoppe, and R. Bauml. Lattice costa schemes using subspace projection for digital watermarking. *European Trans. Telecommunications*, 15(4) :351–362, 2004.
- [27] T. Furon. *Application du tatouage numérique à la protection de copie*. PhD thesis, ENST Paris, 2002.
- [28] T. Furon and P. Duhamel. An asymmetric watermarking method. *IEEE Transactions on Signal Processing*, . numéro spécial ”Signal processing for data hiding in digital media and secure content delivery”, 51(4) :981–995, 2003.
- [29] Vincent Garcia. Estimation de mouvement subpixelique par blocs adaptée à la couleur avec modèle de mouvement. Master’s thesis, DEA Image Vision. Laboratoire I3S - Equipe CReATIVE : Université de Nice - Sophia Antipolis, septembre 2004.
- [30] G. Le Guelvouit. *Tatouage robuste par étalement de spectre avec prise en compte de l’information adjacente*. PhD thesis, INSA Rennes, 2003.

-
- [31] Fairouz Hadi and Khier Benmahammed. Etude comparative entre la morphologie mathématique floue et le regroupement flou. In *International Conference : Sciences of Electronic, Technologies of Information and Telecommunications*, pages 27–31, Tunisia, March 2005.
- [32] L. Haitzma and T. Kalker. A watermarking scheme for digital cinema. In *IEEE International Conference on Image Processing*, volume 2, pages 487–489. IEEE, 2001.
- [33] Wang Hao-Xian, Lu Zhe-Ming, Pan Jeng-Shyang, and Sun Sheng-He. Robust blind video watermarking with adaptive embedding mechanism. *International Journal of Innovative Computing*, 1 :247–259, 2005.
- [34] Robert Haralick and Linda Shapiro. Computer and robot vision, 1992.
- [35] F. Hartung and B. Girod. Watermarking of uncompressed and compressed video. *IEEE Transaction on Signal Processing*, 3 :283–333, 1998.
- [36] E. Hauer and M. Steinebach. Robust digital watermark solution for intercoded frames of mpeg video data. In *Security, Steganography, and Watermarking of Multimedia Contents VII*, volume 5681, pages 381–390. SPIE, 2005.
- [37] J. Ho, B.V. Funt, and M.S. Drew. Separating a color signal into illumination and surface reflectance components : theory and applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12 :966 – 977, 1990.
- [38] Setyawan Iwan, G. Kakes, and Lagendijk Reginald L. Synchronization-insensitive video watermarking using structured noise pattern. In E. J. Delp and P. W. Wong, editors, *Security and Watermarking of Multimedia Contents IV*, volume 4675, pages 520–530. SPIE, April 2002.
- [39] J. Dittmann, T. Fiebig, and R. Steinmetz. New approach for transformation-invariant image and video watermarking in the spatial domain : self-spanning patterns (ssp). In *Security and Watermarking of Multimedia Contents II*, volume 3971, pages 176–185. SPIE, 2000.
- [40] J.-F. Delaigle, C. De Vleeschouwer, and B. Macq. A psychovisual approach for digital picture watermarking. *Journal of Electronic Imaging*, 7(3) :628–640, 1998.

-
- [41] C. Jiang and M.O. Ward. Shadow identification. In *Computer Vision and Pattern Recognition*, pages 606 – 612, 1992.
- [42] Caixia Jiang and Matthew O. Ward. Shadow segmentation and classification in a constrained environment. *CVGIP : Image Underst.*, 59(2) :213–225, 1994.
- [43] H. Joumaa and F. Davoine. Performance of an application video watermarking scheme using informed techniques. In *IEEE International Conference on Image Processing*, 2005.
- [44] Hernandez Juan R. and Perez-Gonzalez Fernando. Statistical analysis of watermarking schemes for copyright protection of images. In *Proceedings of the IEEE*, volume 87, pages 1142–1166, 1999.
- [45] Habiba Loukil Hadj Kacem and Mohamed Salim Bouhleb. Mesure de la qualité des images par l’utilisation de la loi de weber. In *International Conference : Sciences of Electronic, Technologies of Information and Telecommunications*, 2005.
- [46] T. Kalker, G. Depovere, J. Haitsma, , and M. Maes. A video watermarking system for broadcast monitoring. In *Security and Watermarking of Multimedia Contents I*, volume 3657, pages 103–112. SPIE, 1999.
- [47] A. Koz and A. Alatan. Oblivious video watermarking using temporal sensitivity of hvs. In *IEEE International Conference on Image Processing*. IEEE, 2005.
- [48] D. Kundur and D. Hatzinakos. A robust digital image watermarking scheme using the wavelet based function. In *IEEE International Conference on Image Processing*, volume 1, pages 544–547. IEEE, 1997.
- [49] D. Kundur and D. Hatzinakos. Digital watermarking using multiresolution wavelet decomposition. In *IEEE ICASSP’98*, volume 1, pages 2659–2662. IEEE, 1998.
- [50] Marc Kunze. Développement d’un module de reconnaissance de mouvement et de controle d’imitation pour le robot humanoïde robota. Technical report, Laboratoire de Système Autonome, Faculté de Science et Technique de l’Ingénieur, 2004.
- [51] Rajab Lama, Al-Khatib Tahani, and Al-Haj Ali. Video watermarking algorithms using the svd transform. In *European Journal of Scientific Research*, volume 30, pages 389–401, 2009.

-
- [52] Thitiporn Lertrusdachakul, Terumasa Aoki, and Hiroshi Yasuda. A sense of speed in camera motion. In *IMSA'06 : Proceedings of the 24th IASTED international conference on Internet and multimedia systems and applications*, pages 86–90, Anaheim, CA, USA, 2006. ACTA Press.
- [53] E. Lin and E. Delp. Temporal synchronization in video watermarking. In *numéro spécial "Supplement on Secure Media"*, volume 52(10), pages 3007–3022, 2004.
- [54] L. Lin, I. Cox, G. Doerr, , and M. Miller. An efficient algorithm for informed embedding of dirty-paper trellis codes for watermarking. In *IEEE International Conference on Image Processing*. IEEE, 2005.
- [55] J.-P. Linnartz and J. Talstra. Mpeg pty-marks : cheap detection of embedded copyright data in dvd-video. In *Computer Security - ESORICS'98*, number 1485 in Lecture Notes in Computer Science, pages 221–240. Springer-Verlag, 1998.
- [56] Lian-Shan Liu, Ren-Hou Li, and Qi Gao. A robust video watermarking scheme based on dct. In *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, volume 8, pages 5176–5180, Guangzhou, China, 2005.
- [57] C.S. Lu, S.W. Sun, and P.C. Chang. Robust hash-based image watermarking with resistance to geometric distortions and watermark-estimation attack. In *Security, Steganography, and Watermarking of Multimedia Contents VII.*, volume 5681, pages 147–163. SPIE, 2005.
- [58] M. Barni, F. Bartolini, V. Cappellini, A. Lippi, and A. Piva. A dwt-based technique for spatio-frequency masking of digital signatures. In *Security and Watermarking of Multimedia Contents I*, volume 3657, pages 31–39. SPIE, 1999.
- [59] M. Kutter, S. Voloshynovskiy, and A. Herrigel. Watermark copy attack. In *Security and Watermarking of Multimedia Contents II*, volume 3971. SPIE, 2000.
- [60] M. Maes and C. van Overveld. Digital watermarking by geometric warping. In *IEEE International Conference on Image Processing*, volume 2, pages 424–429. IEEE, 1998.
- [61] H. Malvar and D.A.F. Florencio. Improved spread spectrum : A new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing*, nu-

-
- méro spécial Signal processing for data hiding in digital media and secure content delivery*, 51(4) :898–905, 2003.
- [62] J. Mayer and J. Bermudez. Multi-bit informed embedding watermarking with constant robustness. In *IEEE International Conference on Image Processing*, 2005.
- [63] M. Miller, I. Cox, and J.A. Bloom. Informed embedding : exploiting image and detector information during watermark insertion. In *IEEE International Conference on Image Processing*, volume 3, pages 1–4. IEEE, 2000.
- [64] M. Miller, G. Doerr, and I. Cox. Applying informed coding and embedding to design a robust, high capacity watermark. *IEEE Transactions on Image Processing*, 13(6) :792–807, 2004.
- [65] Swanson Mitchell D., Zhu Bin, and Tewfik Ahmed H. Multiresolution scene-based video watermarking using perceptual model. *IEEE journal on selected areas in communications*, 16(4) :540–550, 1998.
- [66] Makoto Nagao, Takashi Matsuyama, and Yoshio Ikeda. Region extraction and shape analysis in aerial photographs. *Computer Graphics and Image Processing*, 10(3) :195 – 223, 1979.
- [67] Laurent Najman and Michel Schmitt. Ligne de partage des eaux. In Michel Schmitt and Juliette Mattioli, editors, *Morphologie Mathématique*, pages 121–140. Masson, 1994.
- [68] P. Nguyen and S. Baudry. Tatouage direct de flux mpeg-2. In *CORESA '00*, 2000.
- [69] Henri Nicolas. *Contribution à la création et à la manipulation des objets vidéo*. PhD thesis, Université de Rennes 1, Institut de Formation Supérieure en Informatique et en Communication, 2001.
- [70] Henri Nicolas. new methods for dynamic mosaicing. *IEEE Transaction on Image Processing*, 10(8) :1239–1251, 2001.
- [71] X. Niu, M. Schmucker, , and C. Busch. Video watermarking resisting to rotation, scaling and translation. In *Security and Watermarking of Multimedia Contents IV*, volume 4675, pages 512–519. SPIE, 2002.

-
- [72] Thrasyvoulos N. Pappas and Robert J. Safranek. Perceptual criteria for image quality evaluation. In *in Handbook of Image and Video Processing*, pages 669–684. Academic Press, 2000.
- [73] S. Pateux and G. Le Guelvouit. Practical watermarking scheme based on wide spread spectrum and game theory. *IEEE Transaction on Signal Processing : Image Communication*, 18 :283–296, 2003.
- [74] Bas Patrick. *Méthodes de Tatouage d’Images Fondées sur le Contenu*. PhD thesis, INPG- Grenoble, 2000.
- [75] L. Pérez-Freire and F. Pérez-Gonzalez. Spread-spectrum vs. quantization-based data hiding : misconceptions and implications. In *SPIE International Symposium on Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VII*, volume 5681, pages 341–352. SPIE, 2005.
- [76] Jean-Marie Pinel. *Etude des conditions d’éclairément dans une séquence d’images et application à la composition et au codage vidéo*. PhD thesis, Thèse de l’université de Rennes I, nov 2002.
- [77] I. Pitas. A method for signature casting on digital images. In *IEEE International Conference on Image Processing*, volume 3, pages 215–218. IEEE, 1996.
- [78] A. Piva, M. Barni, F. Bartolini, , and V. Capellini. Dct based watermark recovering without resorting to the uncorrupted original image. In *IEEE International Conference on Image Processing*, pages 520–523. IEEE, 1997.
- [79] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *3DIM*, pages 145–152, 2001.
- [80] Enrique H. Ruspini. A new approach to clustering. *Information and Control*, 15(1) :22–32, 1969.
- [81] Hassen Seddik, Mounir Sayadi, and Farhat Fnaiech. Nouveau schéma de tatouage par substitution s’appliquant aux techniques spatiales robuste aux attaques asynchrones. In *International Conference : Sciences of Electronic, Technologies of Information and Telecommunications*, 2005.
- [82] Tai Sheng-Chuan, Wang Chuen-Ching, and Chen Ying-Ru. Robust watermarking for mpeg-2 videos. In *International Computer Symposium*, page 85, December 2002.

-
- [83] Mitchell D. Swanson, Bin Zhu, and Ahmed H. Tewfik. Transparent robust image watermarking. In *IEEE International Conference on Image Processing*, pages 211–214, 1996.
- [84] M. Taniguchi. Scaling detection method for watermarked mpeg content. In *IEEE International Conference on Image Processing*. IEEE, 2005.
- [85] Andrew Z. Tirkel, G.A. Rankin, Ron G. van Schyndel, W.J. Ho, and C.F Osborne. Electronic watermark. In *DICTA*, pages 666–672, 1993.
- [86] N. Tournier, M. Chaumont, and W. Puech. Tatouage informé hiérarchique d’un message hiérarchique (en vue de la protection vidéo). In *COmpression et REprésentation des Signaux Audiovisuels*, 2009.
- [87] J. Tzeng, W. Hwang, and I.L. Chern. An asymmetric subspace watermarking method for copyright protection. *IEEE Transactions on Signal Processing, numéro spécial "Supplement on Secure Media II"*, 53(2) :784–792, 2005.
- [88] Ron G. van Schyndel, Andrew Z. Tirkel, and C.F Osborne. A digital watermark. In *IEEE International Conference on Image Processing*, volume 2, pages 86–90. IEEE, 1994.
- [89] J.E. Vila-Forcen, O. Koval, S. Voloshynovskiy, and T. Pun. Asymmetric spread spectrum data-hiding for laplacian host data. In *IEEE International Conference on Image Processing*. IEEE, 2005.
- [90] T. Vynne and F. Jordan. Embedding a digital signature in a video sequence using motion vectors. In *IEEE International Conference on Image Processing*. IEEE, 1996.
- [91] T. Vynne and F. Jordan. Embedding a digital signature in a video sequence. In *US patent*, volume 5, pages 960,081, septembre 1999.
- [92] X. Xia, C. Boncelet, and C. Arce. A multiresolution watermark for digital images. In *IEEE International Conference on Image Processing*, volume 1, pages 548–551. IEEE, 1997.
- [93] Dong Xu, Xuelong Li, Zhengkai Liu, and Yuan Yuan. Cast shadow detection in video segmentation. *Pattern Recognition Letters*, 26(1) :91 – 99, 2005.

-
- [94] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja. Detecting faces in images : A survey. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 24(1) :34–58, 2002.
- [95] Shao yi Chien, Shyh yih Ma, and Liang-Gee Chen. Efficient moving object segmentation algorithm using background registration technique. *IEEE Trans. Circuits Syst. Video Technol*, 12 :577–586, 2002.
- [96] Titus Zaharia. *Modèles paramétriques de mouvement pour la description des contenus vidéo dans le cadre du futur standard MPEG-7*. PhD thesis, Unité de Projets ARTEMIS, Institut National des Télécommunications, 2002.
- [97] J. Zhao and E. Koch. Embedding robust labels into images for copyright protection. In *International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies - KnowRight'95*, volume 82, pages 242–251, Oldenbourg, 1995. Schriftenreihe der osterreichischen Computer Gesellschaft.
- [98] W Zhao, J Wang, D Bhat, K Sakiewicz, N Nandhakumar, and W Chang. Improving color based video shot detection. *IEEE Int. Conf. On Multimedia Computing and Systems*, (2) :752–756, 1999.
- [99] Y. Zhao and R. Lagendjik. Video watermarking scheme resistant to geometric attacks. In *IEEE International Conference on Image Processing*, volume 2, pages 145–149. ICIP, IEEE, 2002.
- [100] Sun Zhaowan, Liu Ju, Sun Jiande, Sun Xinghua, and Ling Jie. A motion location based video watermarking scheme using ica to extract dynamic frames. *Neural Computing & Applications*, 18(5) :507–514, Juin 2009.
- [101] Di Zhong, , Di Zhong, and Shih fu Chang. Video shot detection combining multiple visual features.
- [102] W. Zhu, Z. Xiong, , and Y. Zhang. Multiresolution watermarking for images and video : a unified approach. In *IEEE International Conference on Image Processing*, volume 1, pages 465–469. IEEE, 1998.

Tatouage robuste de vidéo basé sur la notion de régions d'intérêt

Mohamed KOUBAA

Résumé

Le travail présenté dans ce mémoire a pour objectif le développement de nouvelles approches permettant d'introduire de manière robuste des marques dans une vidéo. Il s'agira tout d'abord de sélectionner les zones des images les plus appropriées à l'introduction du tatouage. Pour cela, plusieurs critères devront être pris en compte. Tout d'abord, lorsqu'une marque est introduite dans une zone de l'image, il est nécessaire d'introduire la même marque au même endroit sur toutes les images de la séquence. En effet, dans le cas contraire, un simple filtrage temporel dans le sens du mouvement permet de détruire la marque. Il sera également important de détecter les zones moins importantes visuellement dans lesquelles des modifications pourraient être introduites plus facilement sans qu'elles soient visibles. Il s'agit notamment des zones visibles dans peu d'images différentes, ou des zones où l'œil humain est peu susceptible de détecter des différences, telles que les zones d'ombres en mouvement. Pour parvenir à une détection efficace de telles régions, une phase d'analyse permettant notamment la création d'images mosaïques (ou sprite dans le contexte MPEG-4) et la détection des ombres en mouvement devra être réalisée.

Abstract

The work presented in this thesis aims to develop new approaches to introduce, in a robust way, a marks in a video. The areas of the images which are most appropriate to the introduction of the tattoo should be selected. For that, several criteria must be taken into account. Firstly, when a mark is introduced into an area of the image, it is necessary to introduce the same mark at the same location on all images in the sequence. It is also important to detect an invisible areas where changes could be introduced easily without being seen. These are areas where the human eye is less sensitive to changes, such as shadows region of moving objects. To achieve the effective detection of such regions, a phase of analysis including the creation of mosaic images and the detection of shadows of moving objects will be realized.

Mots clés : Tatouage vidéo, Image mosaïque, zones d'ombre.

Key-words : Video watermarking, mosaic image, Shadow region.