UNIVERSITE BLIDA 1

Faculté de Technologie

Département d'Electronique

THESE DE DOCTORAT

Spécialité : Electronique

PROTECTION DU CONTENU DANS LE STANDARD DE COMPRESSION VIDEO H.264/AVC

Par

Karima AIT SADI

Devant le jury composé de :

N. BENBLIDIA	Professeur, U. de Blida	Présidente
Y. CHERFA	Professeur, U. de Blida	Examinateur
L. MITICHE-HAMAMI	Professeur, E.N.P. Alger	Examinatrice
Y. SMARA	Professeur, U.S.T.H.B., Alger	Examinateur
F. OULEBSIR-BOUMGHAR	Professeur, U.S.T.H.B., Alger	Examinatrice
A. GUESSOUM	Professeur, U. de Blida	Directeur de thèse

Blida 10 Juillet 2017

RESUME

Le tatouage numérique, processus consistant à insérer une empreinte invisible dans le contenu multimédia, a été utilisé pour protéger le contenu vidéo compressé par la norme H264/AVC. Deux systèmes de tatouage numérique sont proposés dans cette thèse. Le premier système est basé sur le tatouage robuste et vise la protection des droits d'auteur du contenu H.264/AVC. Il opère dans le domaine fréquentiel avec une prise en charge minutieuse des contraintes sécurité, robustesse et capacité. Le deuxième système porte sur l'authentification stricte du contenu H.264/AVC, où deux versions sont présentées, chacune étant basée sur une fonction de Hachage différente. Les approches de tatouage réalisées sont basées sur l'insertion d'une marque fragile dans le domaine temporel et sont sensibles à toute altération. Tandis que l'authentification dans la première version est menée sans localisation des régions manipulées, la seconde est plus élaborée puisqu'elle est pourvue de cette fonctionnalité de localisation des régions manipulées et vérifie un maximum de critères d'efficacité d'authentification tout en conservant la qualité visuelle et le flux binaire après insertion.

ABSTRACT

Digital watermarking, a process involving the embedding of an invisible fingerprint into the media content, has been used to protect the content of the H264/AVC compressed video. Two digital watermarking systems are proposed in this thesis, the first system or copyrights' protection involves robust watermarking and shows a viable performance. The second one addresses the hard authentication of H.264/AVC content, where two fragile watermarking-based versions are proposed, each incorporating a different Hashing function. Sensitivity to any content modification was a major objective to enhance in these approaches, and the second version is more elaborated than the first one, since it offers the additional localization functionality of the manipulated regions. Both are seen to provide quite high authentication efficiency while maintaining the visual quality and the bitstream's rate of the original video.

ملخص

أصبح الوشم الرقمي، الذي هو تقنية تكمن في إدخال بصمة غير مرئية داخل المنظومة المتعددة الوسائط، يستعمل لحماية مضمون الفيديو المضغوط وفق معيار H.264/AVC. و يتم اقتراح نظامين للوشم الرقمي ضمن هذه الأطروحة، حيث تعتمد التقنية الأولى المقترحة على الوشم المتين و ترمي إلى حماية حقوق المؤلف لمحتوى H.264/AVC. و يعمل ضمن المجال الترددي مع تكفل دقيق بالعوائق الأمنية ، المتانة و التمكن .

و أما النظام الثاني H.264/AVC ، حيث يتم عرض صيغتين، و تعتمد كل واحدة منهما على بعثرة فهو يتعلق بالمصادقة الصارمة المضمون و لكن بطريقة مختلفة. هذا و يعتمد النهج المتبع للوشم على إدماج علامة هشة في المجال الزمني ، التي هي حساسة لأي اضطراب. إن المصادقة في الصيغة الأولى تتم بدون تحديد مواقع المناطق المراد معالجتها ، و أما الصيغة الثانية فهي أكثر فعالية بما أنها تتوفر على وظيفة تحديد مواقع المناطق هذه، و تحقق أكبر عدد من معايير الفعالية و المصادقة مع الحفاظ على النوعية البصرية و كذا التدفق الثنائي بعد الإدماج.

DEDICACES

A la mémoire de mon Père Arab,

Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour lui. Il a su m'inculquer le sens de la responsabilité, de l'optimisme et de la confiance en soi face aux difficultés de la vie. Ses conseils ont toujours guidé mes pas vers la réussite. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit des sacrifices qu'il a consentis pour mon éducation et ma formation. Que Dieu, le miséricordieux, l'accueille dans son éternel paradis.

A ma très chère mère Nadjia,

Abordable, honorable, aimable : Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi. Tes prières et la bénédiction m'ont été d'un grand secours pour mener à bien mes études. Aucune dédicace ne saurait être assez probante pour exprimer ce que tu mérites pour tous les sacrifices que tu n'as cessé de me faire depuis ma naissance, durant mon enfance et même étant adulte. Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études. Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur.

A mon très cher mari Redha,

Ton encouragement était la bouffée d'oxygène qui me ressourçait dans les moments pénibles, de solitude et de souffrance. Tes sacrifices, ton soutien moral et matériel, ta gentillesse sans égal, ton profond attachement m'ont permis de réussir mes études. Sans ton aide, tes conseils et tes encouragements, ce travail n'aurait vu

le jour. Que ce travail est le témoignage de ma reconnaissance et de mon amour sincère et fidèle.

A mon très cher frère Mahmoud et mes chères sœurs. En souvenir d'une enfance dont nous avons partagé les meilleurs et les plus agréables moments. Pour toute la complicité et l'entente qui nous unissent, ce travail est un témoignage de mon attachement et de mon amour.

A mes enfants Ramy et Mehdi, à mes neveux et nièces, à mes belles-sœurs et mes beaux-frères, les mots ne suffisent guère pour exprimer l'attachement, l'amour et l'affection que je porte pour vous. Je vous dédie ce travail avec tous mes vœux de bonheur, de santé et de réussite.

REMERCIEMENTS

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 Mme Nadjia BENBLIDIA, Professeur au département d'Electronique de l'Université Saad Dahlab de Blida, pour avoir accepté de présider ce jury.

Je remercie également les Professeurs Fatima OULBSIR, Youcef SMARA de l'U.S.T.H.B., le Professeur Yazid CHERFA de l'Université Saad Dahlab de Blida et le Professeur Latifa MITICHE de l'E.N.P. d'Alger qui ont bien accepté de m'honorer par leur présence en qualité d'examinateurs de ma thèse, pour l'intérêt qu'il ont porté à mon travail et pour les efforts fournis pour examiner ma thèse.

Je tiens tout particulièrement à remercier Mr. Abdelrazek Guessoum, professeur au département d'Electronique de l'Université Saad Dahlab de Blida, qui a dirigé cette thèse, pour sa disponibilité, son soutien et ses conseils.

Mme Azouaoui Ouahiba, Directrice de Recherche au Centre de Développement des Technologies Avancées (CDTA) 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 mérite d'être ardemment remerciée. Je lui suis considérablement reconnaissante et je manque d'expressions de remerciements digne de tout ce qu'elle m'a donné durant ma thèse.

Mes vifs remerciements s'adressent aussi à Mme Ghanem Khalida, Directrice de Recherche au CDTA pour son suivi minutieux, ses conseils constructifs et ses relectures de mes travaux qui m'ont permis de renforcer la rigueur de ce travail. Les discussions fructueuses et les réunions portables qui nous ont unis m'ont fort servi.

Mes remerciements s'adressent à Mme Ait Sadi Rachida Docteur en télécommunication à l'université d'oxford, pour sa disponibilité, ses conseils et pour les lectures et corrections de mes travaux.

Je remercie également toute la famille "CDTA" pour tous les bons moments qu'on passe ensemble. Un grand Merci à Amel Azzouz, Fouzia Guessab, Yasmine Belaroussi, Samia Ourari, Amel Chenouf, et Djamila Meriche.

Je ne dois pas aussi oublier mon amie de L'U.S.T.H.B. Baya Fellag.

TABLE DES MATIERES

LISTE DES FIGURES ET TABLEAUX

INTRODUCTION GÉNÉRALE

Le développement des technologies d'acquisition et de transmission d'images et de vidéos numériques a ouvert la porte à de grandes perspectives et possibilités de création et de manipulation des contenus visuels à la fois sur le plan scientifique et artistique. Or, il faut le dire, étant une arme à double tranchant, la popularisation des outils de traitement et de transmission d'images, à travers les open sources, a également ouvert le champ à la copie et à la distribution illégale. Les premiers, mais non les seuls, à en souffrir sont les artistes, l'économie et tout secteur d'emploi de façon générale. On estime au niveau mondial à plus d'une vingtaine de milliards de dollars les pertes en matière de droits d'auteur [1]. Face à cette situation alarmante, il est clair qu'il est devenu urgent de mettre en œuvre des systèmes permettant de protéger les intérêts de tout un chacun, à savoir, assurer les droits d'auteur, contrôler l'authenticité et la distribution des copies d'images et de vidéos et protéger l'intégrité du contenu multimédia.

Plusieurs algorithmes de protection du contenu multimédia, plus particulièrement ceux qui sont pertinents pour l'image et la vidéo, ont été proposés dans la littérature, chacun étant approprié à une ou plusieurs applications. Parmi les technologies émergentes et qui augure des retombées très prometteuses, on distingue le tatouage numérique (watermarking). Cette technologie est en fait apparue au cours des vingt dernières années où elle a suscité l'intérêt de la communauté scientifique dès le début des années 1990 pour se présenter comme une alternative au cryptage dans le créneau de la sécurité du contenu numérique. L'idée de base du tatouage est d'insérer directement des informations subliminales (i.e. invisibles ou inaudibles suivant la nature du document), appelées aussi marque ou signature, dans le support multimédia afin d'empêcher le piratage des œuvres numériques.

Bien que de grands efforts aient été consentis pour permettre aux techniques de tatouage numérique du contenu, plus particulièrement celui se rapportant à la vidéo, d'élargir leur champ d'applications pour des besoins de sécurité, ces efforts restent insuffisants et l'étendue de leur application encore restreinte. Comme toute nouvelle technologie qui voit le jour, le tatouage numérique passe par le classique pallier du scepticisme de la part des industriels, qui hésitent à investir dans une technologie susceptible de connaitre des rebonds et sujette au développement du matériel technologique. En effet, les chercheurs eux-mêmes reconnaissent que le tatouage vivait son âge d'enfance et qu'il vient de gravir un échelon en entamant sa période d'adolescence [2]. Cela se reflète aussi dans les investissements récents des industriels dans des projets de recherche mais qui n'aboutissent pas encore à un déploiement définitif des solutions proposées puisque celles-ci sont jugées enclines à être améliorées. L'évolution des techniques de traitement de signal ainsi que la microélectronique et les dispositifs DSP (Digital Signal Processing) impliqués dans les appareils vidéo ont fait que des algorithmes qui étaient jugés prohibitifs en termes de complexité de calcul/traitement, le sont moins, car des versions plus élaborées et exploitant ces progrès voient le jour. D'autre part l'amélioration de l'optimisation d'utilisation des bandes passantes dans les réseaux de télécommunications et des techniques de fiabilisation des transmissions s'y rapportant, ainsi que l'augmentation incessante des capacités de stockage permise grâce aux nouvelles technologies de fabrication de supports, ont conduit à l'apparition de plusieurs standards de compression vidéo. En effet, une multitude de méthodes de tatouage ont vu le jour pour les anciens standards de compression tels que MPE-G2 [3] et MPEG-4 [4], mais pour le plus récent, le H.264/AVC, standardisé en 2003, bien que largement adopté, peu de travaux ont été consacrés [5]. Fort heureusement, vu les performances confirmées en termes de la grande qualité offerte par ce standard pour la sécurité du contenu vidéo, un nombre croissant de travaux de recherches dans des laboratoires de renom a été lancé. Ces travaux visent un développement plus poussé de ce standard qui s'adapterait avec les besoins actuels des utilisateurs et des parties impliquées dans la distribution/acquisition des images et vidéos (vidéo en mobilité, la TV flexible, etc.). Ce manuscrit s'inscrit dans ce contexte de développement de

techniques de la protection du contenu vidéo compressé par la norme H.264/AVC et présente nos travaux effectués en tatouage numérique pour la protection des droits d'auteur et l'authentification du contenu. Ceux-ci ont été effectués au sein de la division Information, Télécom et Multimédia du Centre de Développement des Technologies Avancées (CDTA).

Dans le cadre de cette thèse, nous avons proposé deux systèmes de tatouage numérique opérant durant le processus de codage H.264/AVC pour des fins de protection des droits d'auteur, de vérification de l'intégrité du contenu et son authentification. Dans le premier système, nous avons opté pour l'insertion d'une marque robuste avec une forte capacité d'insertion et appliquant un compromis contrôlé entre la qualité vidéo exprimée en PSNR (Peak Signal to Noise Ratio) et le débit. Par ailleurs, dans le deuxième système, nous avons proposé une méthode de tatouage fragile pour vérifier l'intégrité du contenu H.264/AVC et son authentification. La marque fragile insérée dépend du contenu vidéo H.264/AVC et est générée à partir des caractéristiques intrinsèques robustes du contenu. Nous avons abouti à une approche de tatouage fragile assurant un maximum de critères d'efficacité d'authentification tels que la sensibilité aux attaques, l'invisibilité de la marque, et la conservation de la qualité visuelle et du flux binaire après insertion.

Le manuscrit est organisé en quatre chapitres : les deux premiers chapitres établissent un état de l'art des différentes approches adoptées jusqu'à ce jour et introduisent les différentes terminologies dans lesquelles s'inscrivent nos travaux de thèse. Plus particulièrement, les concepts d'importance pour la technologie du tatouage numérique et le principe d'opération du standard H.264/AVC, y sont présentés. Nous avons consacré le premier chapitre aux éléments de base d'un système de tatouage numérique, aux contraintes à considérer, aux attaques possibles et à l'évaluation de la qualité perceptuelle de la vidéo. Le second chapitre quant à lui a été dédié au principe général de fonctionnement du codeur H.264/AVC et les principaux modules le constituant. Les chapitres 3 et 4 présentent les deux contributions majeures que sont la protection des droits d'auteurs et l'authentification des vidéos au cours de la compression H.264/AVC. Chaque chapitre commence par une énumération et une brève description des différentes approches de tatouage

numérique, et ce afin d'avoir une idée générale sur les points forts et les faiblesses des travaux réalisés et exploiter les premiers tout en palliant aux seconds. Dans le chapitre 3, une méthode de tatouage robuste est proposée pour la protection des droits d'auteur du contenu vidéo H.264/AVC et tire justement partie des points forts des méthodes existantes et remédie à leurs faiblesses en termes de capacité d'insertion et de robustesse face aux attaques de traitements d'images usuels. Pour cela l'approche développée consiste à traiter la marque avant de l'insérer en moyennant la transformation de Wash Hadamard. Les bits de la marque sont insérés dans le domaine DCT (Discrete Cosine Transform) de la norme, en changeant les valeurs des coefficients AC quantifiés appartenant au mode de prédiction Intra_4×4 de la composante de luminance des trames Intra de la séquence vidéo.

Le chapitre 4 décrit en détail le Système d'Authentification Stricte du Contenu H.264/AVC (SASC) que nous avons proposé. Deux versions du système basées sur deux fonctions de Hachage différentes sont présentées : la première version utilise la fonction de hachage MD5 (SASC-MD5) tandis que la seconde est basée sur la fonction de hachage sécurisée SASC-HMAC-SHA-256 (keyed-hash message authentication code). Les approches de tatouage réalisées sont basées sur l'insertion d'une marque fragile générée à partir d'un ensemble de coefficients robustes représentatifs du contenu H.264/AVC. La première version du système SASC-MD5 regroupe deux variantes, dont l'objectif visé est d'aboutir à un système d'authentification strict remplissant un nombre maximal de critères d'authentification, plus particulièrement le critère de conservation du taux de bits. La première variante SASC-MD5-1 assure l'authentification stricte des vidéos ayant une grande activité temporelle, la vidéo est soit authentifiée soit non. La seconde variante SASC-MD5-2 représente une amélioration de la première en termes de sensibilité et localisation des trames manipulées.

Bien que cette version du système réponde aux principales contraintes d'un système d'authentification, elle souffre néanmoins d'une faible capacité d'insertion pour les séquences vidéo à faible activité temporelle et n'a pas visé la sécurité de l'insertion. Ces deux points faibles de la première version ont été pris en compte dans l'élaboration de la deuxième version SASC-HMAC-SHA-256. L'augmentation de la

capacité d'insertion est réalisée en substituant deux bits de la marque aux deux bits de poids faible de chaque composante des vecteurs de mouvement, et ce afin de dissimuler les quatre bits de la marque dans un seul vecteur de mouvement (MV), au lieu de n'en utiliser que deux bits, comme il est d'usage de le faire. La sécurité est renforcée en opérant à deux niveaux. Le premier niveau consiste à employer la fonction HMAC-SHA-256 qui est plus sécurisée que son alternative MD5. Le deuxième niveau de sécurité ajouté repose sur l'utilisation d'une séquence pseudo-aléatoire pour la sélection des macroblocs (MBs) concernés par le tatouage de leurs MVs. Afin d'aboutir à une sécurisation accrue, le système SASC-HMAC-SHA-256, dans sa sélection des MVs, adapte dynamiquement le seuil en fonction de l'activité de mouvement la plus élevée de la trame. Cette deuxième version assure un maximum de critères d'efficacité, sensibilité, imperceptibilité, sécurité et conservation de la qualité vidéo après insertion, tout en maintenant le débit de la vidéo inchangé, comparé aux méthodes présentées dans la littérature. En plus des critères cités, le système proposé est en mesure de localiser les régions manipulées dans le contenu H.264/AVC.

Ce manuscrit se termine par une conclusion générale ainsi que la citation de quelques perspectives potentielles pour l'extension ultérieure de ce travail.

CHAPITRE 1

TATOUAGE NUMERIQUE ET COMPRESSION VIDEO

1.1. Introduction

La production des images de télévision sous forme numérique s'est répandue depuis le début des années 90, grâce aux avantages qu'elle apporte aux producteurs de programmes tels que la fidélité des sources après enregistrement ou duplication; les traitements de post-production aisés et sans détérioration de la qualité des images ; et enfin la transmission sans erreurs sur des liaisons numériques de contribution entre studios [1].

La numérisation, en raison des coûts élevés du matériel qu'elle induit, serait vraisemblablement restée cantonnée dans les sphères de la production sans la double impulsion apportée par les techniques de compression de l'image et les progrès en densité et rapidité des circuits intégrés. Initialement destinés à la réduction du débit des programmes échangés entre studios de télévision sur des liaisons de contribution, de puissants algorithmes de compression de débit ont été développés vers le milieu de la décennie 1980 dont les principes essentiels ont été retenus par le groupe de standardisation MPEG. Depuis, les codecs vidéo n'ont jamais cessé d'évoluer. Ainsi, après le MPEG (utilisé notamment pour les Vidéos CD) [2], le MPEG-2 (utilisé pour le DVD) [3] et le MPEG-4 (utilisé comme base sur le DivX ou le Xvid) [4], un nouveau codec H.264/AVC (Advanced Video Coding), lancé en 1998 a été publié en juillet 2003 [5].

Le codec de compression vidéo H.264/AVC est une évolution logique de la norme MPEG-4. Il a amélioré le taux de compression tout en proposant une meilleure qualité d'affichage. En effet, Il offre un taux de compression de 2 à 3 fois plus élevé que le MPEG-2 et de 1.5 à 2 fois plus élevé que le MPEG-4. La qualité

DVD est atteinte en utilisant un débit binaire (bitrate en anglais) de 2Mbps (250 Ko/sec) alors que la qualité VHS est accessible dès le Mbps (125 Ko/sec) [6].

La norme H.264/AVC se destine à de multiples types d'applications, allant de la vidéotéléphonie et la vidéoconférence aux applications de diffusion de media, en passant par les communications sans fil, la télévision et le stockage. Avec de telles performances, une telle flexibilité et une licence parmi les moins chères du marché, cette norme est certainement promise à un bel avenir autant dans l'industrie (DVD, vidéoconférence, télévision) que sur Internet [7].

Ce rapprochement des disciplines, industries de l'information, télécommunication et radiodiffusion a ouvert la porte à de nouvelles possibilités de programme, de service et de personnalisation, mais il est devenu plus facile de violer la propriété intellectuelle et l'intégrité des contenus véhiculés.

En effet, les films vidéo, s'ils sont légalement disponibles à la vente ou en téléchargement gratuit, sont néanmoins sujets à l'application du droit d'auteur et d'authentification du contenu. Le fait qu'on puisse avoir accès à certaines œuvres sur Internet ne suppose pas que les titulaires des droits d'auteur pour ces œuvres en autorisent la reproduction ou toute autre utilisation, et c'est bien dans cette optique que les auteurs, les éditeurs et les fournisseurs de contenus sont préoccupés par la protection de leurs œuvres. Il devient donc nécessaire de créer et de mettre en œuvre des dispositifs permettant de faire respecter les droits d'auteur, de contrôler les copies et de protéger l'intégrité des contenus.

Pour résoudre ce problème, le tatouage numérique est très rapidement apparu comme la solution « alternative » et supplémentaire au cryptage et à l'embrouillage pour renforcer la sécurité du contenu multimédia (texte, audio, image, vidéo) face au piratage et à la contrefaçon.

En effet, les approches d'embrouillage et de cryptage permettent une protection a priori d'un contenu multimédia quelconque [8] [9] [10] [11] [12]. Le problème de ces techniques est que lorsque le contenu est désembrouillé ou décrypté, il n'existe plus de protection, et ce dernier peut être rediffusé en toute impunité. Pour faire face à ce problème, à la fin des années 1990 s'est manifestée

une effervescence médiatique, mais également scientifique autour d'une nouvelle discipline, le plus souvent associée à des préoccupations de sécurité.

1.2. Définition du tatouage

Le principe du tatouage numérique consiste en l'insertion d'un message par modification imperceptible d'un ensemble de données support, encore appelé signal de couverture; en anglais cover-data [13] [14] [15]. Ces données de couverture peuvent être de nature diverse comme un message sonore, un texte, des images fixes, de la vidéo, des partitions musicales, etc.

Le tatouage est donc une technique de communication, de la même manière qu'en transmission radio le champ électromagnétique est modulé par le signal à transmettre (modulation AM d'amplitude ou FM de fréquence - par exemple), l'information de tatouage ou la marque présente la particularité d'être étroitement liée aux données de support. La modification s'effectue dans les composantes perceptibles (comme la luminance des pixels d'une image), et non dans l'en-tête d'un fichier par exemple. Ce tatouage doit pouvoir être détecté et décodé, mais doit être imperceptible, c'est-à-dire que la déformation doit être suffisamment faible pour que l'utilisateur ne puisse pas différencier le document tatoué de l'original. Cette notion d'imperceptibilité et d'insertion dans la trame même du document rejoint la traduction littérale du terme digital watermark : "filigrane électronique" [16].

On peut trouver les premiers filigranes sur des papiers du treizième siècle, dans le but de garantir leur qualité. Sur un billet de banque, les fibres sont marquées au moment de la sortie du bain d'eau, ce qui est à l'origine du terme anglais water mark. De la même manière que sur un billet de banque, le filigrane électronique est d'abord invisible et n'est révélé que par une transformation spécifique. L'intérêt d'une telle opération est que le tatouage est indépendant du format de stockage des données, puisqu'il est intrinsèque au document. C'est donc une solution simple au vieux problème du "trou analogique" des systèmes de DRM (Digital Rights Management) : comment conserver un DRM si l'utilisateur numérise le contenu analogique du document, par exemple par impression/numérisation ?

Le tatouage est une technique de dissimulation d'information (data hiding), principe qui englobe également la transmission d'une information secrète dans un réseau ou encore la stéganographie.

1.3. Position du tatouage par rapport à la stéganograhie et à la cryptographie

Les premiers travaux sur le tatouage, qui datent du milieu des années 1990, étaient motivés par les problèmes de défense du droit d'auteur dans un environnement numérique ouvert. La duplication sans perte de qualité et la rapidité de diffusion dans un environnement tel internet faisaient que toute œuvre numérique (image, film, musique, logiciel, etc.) pouvait être copiée et distribuée extrêmement facilement sans contrôle des ayant-droits. Une des premières idées pour assurer la protection des œuvres a été d'utiliser les techniques de cryptographie: Une œuvre est proposée chiffrée, et les utilisateurs peuvent acheter une clé de déchiffrement pour visualiser l'œuvre originale. Cette idée est à la base de la diffusion des chaines cryptées par exemple. Cependant, cette méthode montre clairement ses limites : une fois que l'utilisateur dispose de l'œuvre en clair, rien ne l'empêche de la copier et de la redistribuer ou de la revendre. Un mécanisme de protection intrinsèque pour l'œuvre en clair est donc rapidement apparu indispensable. Le tatouage permet d'étendre la protection des œuvres: en la dotant d'une « signature» invisible et persistante, il devient possible de tracer automatiquement son utilisation dans un réseau. Alternativement, on peut insérer par marquage, un identifiant de l'acquéreur afin de le responsabiliser et de le dissuader de laisser le piratage s'effectuer par négligence ou avec son consentement tacite.

La définition du tatouage le rapproche en fait beaucoup plus de la stéganographie que de la cryptographie. Cependant, il faut garder à l'esprit les différences essentielles entre les deux. En particulier, il est indispensable en stéganographie que l'existence même du message soit dissimulée; au contraire, en tatouage, la connaissance publique de l'existence d'une marque dans un document peut être un moyen de dissuasion contre le piratage. D'autre part, en stéganographie, le message de couverture n'est pas important en soi, alors qu'en tatouage, il est

primordial que ce message ne soit pas dénaturé, à la fois lors de l'insertion du tatouage et lors d'une attaque visant à détruire la marque. Le lien entre le message caché et les données support est donc beaucoup plus fort.

1.4. Applications

Si le terme digital watermarking a été introduit en 1990, l'explosion du nombre de publications à ce sujet date de 1995, ce qui s'est concrétisé par la création de l'atelier IHW (Information Hiding Workshop) en 1996, d'une conférence spécifique au sein de SPIE en 1999 et de l'atelier IWDW (International Workshop on Digital Watermarking) en 2002 [16]. Quatre journaux dédiés aux problématiques de sécurité de l'information ont été créés: IEEE Transactions on Information Forensics and Security [17], IEE Proceeding Information Security [18], LNCS Transactions on Data Hiding and Multimedia Security [19] et European Association for Signal Processing Journal (EURASIP) on Information Security, ce qui souligne le dynamisme du domaine.

Les promesses du tatouage ont conduit à la prolifération d'entreprises dans le domaine, même si l'enthousiasme initial semble retombé. Digimarc [20], firme pionnière, rassemble des brevets de base sur le tatouage dont elle vend la licence. Elle est également auteur du module de tatouage du logiciel de traitement d'images Photoshop [21]. Les systèmes proposés par Epson [22] et Kodak [23] intègrent au niveau de leurs appareils photo numériques des processus de protection par tatouage afin de protéger les images dès leurs acquisitions.

Le tatouage numérique a de nombreuses applications, l'une de ses premières applications importantes est la protection de la propriété intellectuelle (ou droit d'auteur). Dans un premier temps, le tatouage concentrait la majorité des efforts de recherche dans le domaine du traitement du signal, pour la protection des droits d'auteurs en particulier. Il s'agit d'identifier le propriétaire d'un média, en insérant un identifiant dans le média. Dans ce cas, la marque insérée doit pouvoir résister au mieux aux modifications qu'est susceptible de subir le média. Il s'agit de préserver au maximum ce qui va permettre aux ayants droits de revendiquer la possession du

média. Par exemple, redimensionner une photographie tatouée ne doit pas faire disparaître son tatouage. L'auteur de la photo originale doit également être l'auteur de la photo redimensionnée.

La protection contre la copie (copy control) consiste en revanche à prévenir la copie du média. C'est notamment le cas avec les DVDs, pour lesquels on insère une marque destinée à reconnaitre la source des images si elles venaient à être copiées [24]. De la même manière, un « beta-test » de jeu vidéo peut intégrer une marque tatouée sur l'écran. Cela permet d'identifier l'auteur d'une fuite d'information si des copies d'écrans sont diffusées.

D'autres champs d'intérêt englobent le tatouage sans perte [25], ou tatouage réversible [26]. Ce tatouage sert à protéger le contenu, c'est-à-dire à s'assurer que le média n'a pas été modifié. Il ne s'agit ni d'empêcher la copie, ni d'identifier l'auteur, mais de garantir par un contrôle d'intégrité la conformité du média d'origine. Dans ce cas, le tatouage est volontairement vulnérable aux attaques dans le but de détecter une manipulation éventuelle du document. Un cas particulier de ce tatouage est le tatouage légiste (forensic watermarking). Il regroupe les applications qui peuvent directement entraîner l'intervention des tribunaux. Le but est d'attaquer le pirate en justice, à titre dissuasif pour les autres utilisateurs. Ce scénario est souvent évoqué pour un contenu musical ou pour le cinéma en ligne. Différents travaux ont porté sur les niveaux de finesse de ce contrôle d'intégrité [24]. Le niveau le plus bas consiste simplement à détecter la modification. À un degré plus élevé, il est possible de localiser ce qui a été modifié. Enfin, le niveau le plus fin consiste à déterminer la nature de la modification.

D'autres applications englobent le tatouage semi-fragile [27] [28]. Dans ce cas, le tatouage vise à résister à certains traitements du document, tant que son contenu sémantique n'est pas altéré. On distingue les attaques légitimes (ex : compression du contenu), auxquelles la méthode est robuste, des attaques illégitimes auxquelles elle est fragile.

Enfin, Une autre application du tatouage est l'enrichissement de document ou l'indexation. Une information supplémentaire est insérée dans le document, sans

contrainte de sécurité ou de robustesse dans le but d'améliorer le contenu. Cette information peut être de nature diverse (lien, données personnelles, ou autre).

1.5. Contraintes liées à la conception d'algorithmes de tatouage numérique

Les processus de tatouage numérique sont assez variés, et donnent par conséquent des résultats assez différents, adaptés à des contextes applicatifs propres. En effet, dans certaines circonstances comme la protection des droits d'auteur, on souhaiterait que le tatouage soit robuste, c'est-à-dire qu'il ne s'efface pas (résiste) même si le document a subi des transformations, ou plus exactement on souhaiterait que tant que le document conserve une valeur marchande, il contienne l'information de tatouage. Dans d'autres cas, comme l'authentification, on peut souhaiter que la moindre transformation du document tatoué fasse "sauter" la marque, et ce afin de détecter la perte d'intégrité du document. Enfin, on peut souhaiter une situation intermédiaire, dans laquelle la marque résiste pour certaines transformations, et non pour d'autres. Toutes les applications utilisant le tatouage numérique impliquent des contraintes qui sont plus ou moins générales dans tous les cas mais qui peuvent être différemment hiérarchisées selon l'utilisation envisagée. Dans ce qui suit, nous explicitons les contraintes vis à vis des applications prévues dans cette thèse.

1.5.1. Contraintes liées à un système de protection des droits d'auteur

Pour chaque type d'application envisagée, le système de tatouage doit satisfaire un certain nombre de contraintes. Les contraintes canoniques sont l'imperceptibilité, la robustesse et le débit utile, qui correspondent aux contraintes classiques des systèmes de communication. S'ajoutent certaines contraintes spécifiques que l'on peut qualifier de cryptographiques.

- **Imperceptibilité :** Elle désigne la similarité perceptuelle entre la donnée d'origine et celle tatouée. La procédure d'insertion doit assurer que la marque

est imperceptible pour toutes les personnes. Il est indispensable que le tatouage soit invisible pour ne pas dégrader la qualité du document support. Cette contrainte implique notamment que l'énergie du tatouage soit suffisamment faible comparée à l'énergie des données hôtes. Sachant qu'une imperceptibilité totale est impossible (ce qui reviendrait à ne pas modifier l'hôte, donc à ne rien insérer), il convient alors de définir un seuil à partir duquel on peut considérer la modification comme imperceptible. De plus, un tatouage peut être imperceptible de diverses manières. Soit parce que l'œil humain ne le voit pas (dans le cas d'une image), soit parce qu'une machine ne peut pas deviner sa présence. Dans la plupart des algorithmes proposés, l'imperceptibilité du tatouage s'obtient en utilisant diverses propriétés du Système Visuel Humain (SVH). Ces propriétés, souvent trouvées à partir d'heuristiques, proposent des modélisations du comportement psycho-visuel humain. L'utilisation de ces propriétés tend de plus en plus à se généraliser pour insérer une quantité d'information importante tout en gardant la marque invisible [29] [30].

- **Robustesse :** C'est la capacité que possède un algorithme de tatouage à résister aux attaques extérieures, qu'elles soient bienveillantes ou malveillantes. Ce critère se décline généralement en trois degrés [24] : robuste, fragile et semi-fragile. Un tatouage robuste correspond au degré le plus élevé. 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 (il s'agit ici de traitements bienveillants qui ne visent pas forcément à retirer la marque). On peut aussi avoir des attaques plus élaborées, qui ont pour seul but de retirer la marque, comme des attaques statistiques aveugles ou des attaques basées sur la connaissance de l'algorithme utilisé [31] [32]. En revanche, un tatouage fragile correspond au degré de robustesse le plus faible. La moindre modification de l'hôte efface la marque. Le tatouage semi-fragile est un degré intermédiaire entre ces deux extrêmes. Ce type de tatouage est robuste à un ensemble défini de manipulations et est fragile à d'autres.

- **Taux d'insertion :** C'est la quantité maximale d'information pouvant être insérée dans un document source sans dégrader sa qualité visuelle. On parle aussi de capacité.
- **Localisation :** La localisation est la possibilité que peut offrir une marque extraite, de déterminer quelles sont les zones du contenu dont les modifications ont endommagé cette marque [24].
- **Complexité :** La complexité d'un tatouage s'exprime en fonction de la complexité de l'algorithme d'insertion et/ou celui de la détection de la marque. Cette contrainte constitue un paramètre déterminant pour l'efficacité d'un système. Pour les applications de preuve de propriété ou de traçabilité des documents, la détection peut être effectuée en temps différé. Par contre, il est primordial que la lecture soit effectuée en temps réel pour les applications de contrôle automatique. Généralement, la complexité en écriture est moins cruciale que la complexité en lecture. Ceci n'est cependant pas le cas pour les applications de traçabilité, où l'on souhaite au contraire avoir une insertion rapide puisqu'un grand nombre de documents doit être marqué, et où l'on peut tolérer une plus grande complexité en lecture.

1.5.2. Contraintes liées à la conception d'un système d'authentification

La notion d'intégrité est un concept bien connu en sécurité. Sa définition repose sur une décision binaire qui garantit que les données reçues sont rigoureusement identiques à celles émises. Cette définition est applicable à tout type de documents numériques, néanmoins, dans la pratique elle s'avère être beaucoup trop stricte et inadaptée pour les documents multimédia. En effet, l'interprétation que l'on a d'une image (vidéo) dépend principalement des éléments la constituant, plutôt que des valeurs numériques des pixels ou de sa résolution. En d'autres termes, le problème de l'intégrité des images se pose principalement en termes de contenu sémantique ; c'est-à-dire la détection des modifications du document pouvant engendrer une gêne dans sa visualisation et/ou une erreur dans son interprétation (modification de la légende, disparition d'un visage, etc.) [33].

Différents systèmes d'authentification vidéo ont été proposés [28]. Pour être efficace, ces derniers doivent satisfaire les critères suivants :

- **Sensibilité aux modifications:** Le système doit être capable de déceler des manipulations pouvant modifier l'interprétation que l'on a d'une vidéo, telles que des recadrages (cropping) ou des retouches locales;
- **Visibilité** : Les données d'authentification doivent être invisibles (dans les conditions normales de visualisation). Il s'agit de faire en sorte que l'impact visuel du tatouage (i.e. distorsion) soit le plus faible possible afin que le document numérique marqué reste fidèle à l'original;
- **Tolérance** : La tolérance d'un tatouage définit sa conservation face à des modifications souhaitées comme la compression. le système doit être tolérant vis-à-vis des algorithmes de compression avec pertes tels que MPEG pour la vidéo et JPEG ou JPEG2000 pour les images fixes ;
- **Localisation des régions altérées** : Le système doit être en mesure de donner à l'utilisateur une information permettant d'identifier les régions qui ont été manipulées ;
- **Reconstruction des régions altérées** : Le système doit éventuellement permettre une restauration partielle des zones de l'image qui ont été manipulées ou détruites, afin de donner à l'utilisateur la possibilité de se faire une idée sur le contenu original de ces régions ;
- **La capacité : L**a capacité du tatouage doit être suffisante pour embarquer l'ensemble des informations de la preuve.

En plus des critères précédents, d'autres contraintes techniques sont à prendre en considération [34] :

- **Mode de stockage** : Il est préférable de cacher les données d'authentification dans le média (image, vidéo) lui-même, sous la forme d'un tatouage, plutôt que dans un fichier séparé comme dans le cas d'une signature externe;
- **Mode d'extraction** : Suivant que les données d'authentification sont dépendantes ou non du document numérique, on optera pour un mode

d'extraction du tatouage aveugle ou semi-aveugle. En mode d'extraction aveugle, la marque représentant les données d'authentification est récupérée à partir du média marqué seul (éventuellement manipulé), alors qu'en semi-aveugle il s'agit principalement de vérifier la présence d'une telle marque dans le média (via un score de corrélation);

- **Algorithme asymétrique** : Contrairement aux services de sécurité plus classiques comme le « copyright » où l'on peut se contenter d'une même clé (privée) pour l'insertion et l'extraction de la marque, un service d'intégrité nécessite de préférence l'utilisation d'un algorithme de tatouage asymétrique (ou de chiffrement, selon le cas) dans la mesure où tout un chacun doit pouvoir s'assurer de l'intégrité d'une image ;

1.6. Sécurité en tatouage

La sécurité concerne le comportement du système de tatouage face à des attaques intentionnelles qui visent à rendre le tatouage inutilisable. Elle s'appuie sur deux facteurs essentiels [24] : l'indétectabilité et la confidentialité. Le premier facteur, l'indétectabilité, consiste à faire dépendre la règle d'insertion de paramètres secrets, déduits d'une clé cryptographique liée à l'utilisateur. Le second facteur, la confidentialité, consiste à rendre les informations tatouées indéchiffrables sans la connaissance de la clé. Dans le cas où la marque est frauduleusement détectée et extraite, le chiffrement des informations rend leur exploitation impossible.

Le type d'attaques pertinentes, et donc la définition de la sécurité d'un système de tatouage, sera par conséquent différent selon les applications. Ainsi, en contrôle automatique de diffusion, seul le brouillage du tatouage sera considéré. Au contraire, dans une application de preuve de propriété, il existe des attaques plus subtiles qui laissent intact le tatouage d'origine mais jettent le doute sur son authenticité. Il est donc important de toujours avoir à l'esprit une application précise lorsque l'on juge de la sécurité d'un système de tatouage. Vassaux et al. [35] présentent un survol des différentes attaques en image fixe et en vidéo. La vidéo étant une succession d'images fixes, on peut appliquer la plupart des attaques de l'image fixe à la vidéo.

Cependant, certaines attaques couramment utilisées en image fixe ne sont pas applicables à la vidéo, c'est le cas par exemple de l'attaque stirmark [36], qui consiste en une succession de distorsions géométriques aléatoires appliquées localement à plusieurs endroits dans l'image. Les schémas d'attaque peuvent être classés en deux grandes familles, les bienveillantes et les malveillantes [37].

1.6.1. Attaques bienveillantes

Il s'agit de 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 (MPEG-1, MPEG-2, MPEG-4, MPEG4-AVC, etc.), à des filtrages (réduction de bruit), à un changement de résolution, au type de codage (progressif ou entrelacé). Un autre traitement couramment utilisé en vidéo est la conversion analogique/numérique, et inversement. Enfin, certaines distorsions géométriques peuvent être utilisées : inversion verticale ou flip vertical (couramment utilisé pour rendre les publicités non reconnaissables dans une séquence), recadrement, perte d'une ligne ou d'une colonne, etc.; Parmi celles-ci, nous trouvons :

- **Symétrie horizontale :** Certaines images peuvent être "flippées" sans perdre de leur sens (par exemple un paysage). Bien qu'il ne s'applique qu'à peu d'images, lorsqu'il se produit, très peu de marquages lui survivent.
- **Rotation :** C'est une transformation qui est très utilisée après avoir scanné une image. Elle sert à réaligner des images (avec des petits angles) et peut être fatale à certains types de marquages.
- **Recadrement :** appelé aussi attaque par cropping. Il consiste à extraire un morceau non tatoué d'un flux média pour le réutiliser. Pour être résistant à ce type d'attaque, le tatouage doit être présent sur tout le média. La même situation se produit dans le domaine fréquentiel du média où la marque doit être partout présente afin d'éviter une destruction par filtrage.
- **Changement d'échelle :** Ce genre de transformations peut être séparé en deux groupes : les transformations uniformes (pour lesquelles on conserve les proportions, l'échelle en X varie comme l'échelle en Y) et bien sûr les

transformations non uniformes (où l'échelle en X ne varie pas comme l'échelle en Y).

- **Transformations géométriques:** On se contente de faire un mélange de rotations, changements d'échelles non uniformes.
- **Filtrage passe-bas:** Ce filtrage est un outil de base de traitement d'image, il est utilisé généralement pour la suppression du bruit. Ce type de filtrage a généralement pour effet d'atténuer les composantes hautes fréquences de l'image et par conséquent dégrader les composantes de la marque insérées dans ces fréquences.
- **Accentuation des contours :** Ou encore appelé filtre "passe-haut" (car il supprime les basses fréquences), ou "Sharpen". Il s'agit de l'inverse du filtre passe-bas (encore appelé "Blur"). L'intérêt d'une telle attaque est assez faible, sachant que l'on conserve le bruit (et les forts gradients de l'image), et que c'est souvent à ce niveau là que se situe le tatouage (car c'est dans ces zones où l'on cache de préférence de l'information).

1.6.2. Attaques malveillantes

Ce type d'attaque vise explicitement à rendre le tatouage inopérant. Ces attaques, comme souvent dans le domaine numérique, sont difficiles à prouver d'un point de vue juridique. Toutefois, une attaque malveillante qui a réussi devra produire un contenu à la fois lavé de son tatouage et encore exploitable. Parmi celles-ci, on peut citer :

- **L'attaque par sur marquage** : consiste à tatouer à nouveau un média déjà tatoué. Pour certains schémas, en particulier si les lieux de tatouages sont fixés, cette attaque peut être très dangereuse. Certains algorithmes de tatouage se protègent en vérifiant, avant de distribuer une clé que le média d'origine proposé n'est pas tatoué. Cette protection n'est utile que si le schéma de tatouage demeure inconnu. En effet, s'il est connu, un pirate peut ajouter une marque de sa fabrication qui invalidera la détection.

- **L'ajout d'un bruit** : L'ajout involontaire d'un bruit avec des proportions importantes peut avoir un effet de masquage de la marque et par conséquent gêner son extraction/détection.
- **L'attaque par recopie** : consiste à recopier une marque obtenue préalablement (par exemple par estimation) sur un média non marqué. Le détecteur validera alors le nouveau média comme étant tatoué.
- **Attaque par collusion** : plusieurs utilisateurs se rassemblent pour accumuler différents documents tatoués. Ils les combinent ensuite pour obtenir des documents qui ne contiennent plus aucun signal de tatouage. Il existe principalement deux grandes familles d'attaques par collusion [38] [39]. D'une part, quand différentes versions tatouées de la même image sont disponibles, il suffit souvent de les moyenner pour obtenir une estimation de l'image d'origine non-tatouée. D'autre part, les utilisateurs peuvent amasser différentes images contenant le même tatouage. Dans ce cas, le but est d'estimer ce signal tatouage et de l'enlever par la suite de chaque image.

1.7. Exemples de manipulations malveillantes

Dans notre société, les messages véhiculés par les images ont un impact considérable. En effet, le réalisme d'une photographie est tel que nous avons tendance à prendre pour réelles des scènes qui ne le sont pas (toutes les images, y compris celles réalisées en toute innocence, ont la capacité d'être détournées de leur sens). Les manipulations, qui avant, nécessitaient des moyens coûteux sont désormais à la portée de tout le monde et les progrès de la technique et du tout numérique les rendent quasi indécelables. Dans ce contexte, un service d'intégrité d'image n'a bien évidemment pas la prétention de vérifier la véracité des événements, mais de déceler des manipulations qui auraient pu y être apportées a posteriori (i.e. entre la prise de la photographie et sa diffusion) dans le but de détourner le contenu de l'image ou de rendre impossible toute interprétation. Nous donnons ci-après un exemple célèbre de manipulations intentionnelles d'images [28]. La couverture du magazine « Time » du 27 juin 1994 est un bel exemple de falsification d'image (figure

1.1). Les éléments ajoutés à la photographie d'origine de Simpson sont le flou, le noircissement du visage et le halo obscur, créés de toutes pièces au moyen d'un logiciel de retouche à des fins de manipulation. La figure 1.2 illustre un exemple d'attaque par collusion par estimation du tatouage et re-modulation pour enlever le signal de tatouage dans chaque trame vidéo.

Figure 1.1 : Exemple de falsification d'image de l'affaire O.J. Simpson [28].

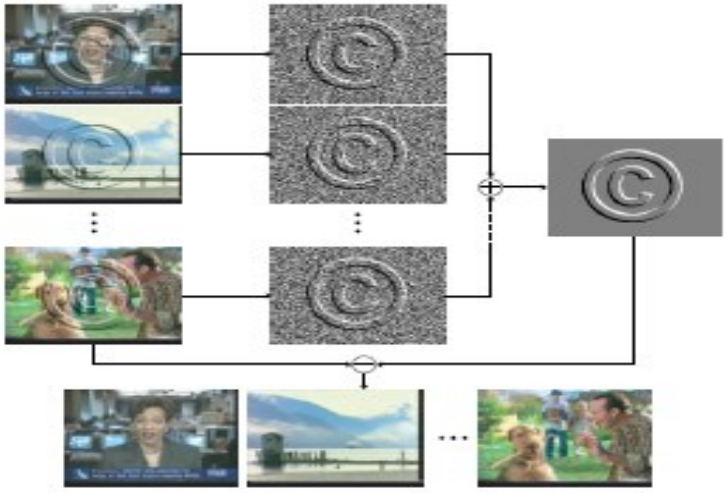


Figure 1.2 : Exemple d'attaque par collusion [39].

1.8. Classification des systèmes de tatouage numérique

Les algorithmes de tatouage numérique sont classifiés selon plusieurs critères à savoir (figure 1.3), le mode d'insertion de la marque (schéma additif ou schéma substitutif), la façon dont est inséré le tatouage : directement dans le document (domaine spatial), dans une transformé du document (domaine fréquentiel), selon le type de tatouage (aveugle ou non aveugle, perceptible ou imperceptible et

symétrique ou asymétrique) et aussi selon le type de la marque insérée (fragile ou robuste). Chaque espace de travail utilisé en tatouage possède ses propres avantages et inconvénients. Pour mieux étudier les méthodes de tatouage numérique, ces critères vont être traités séparément.

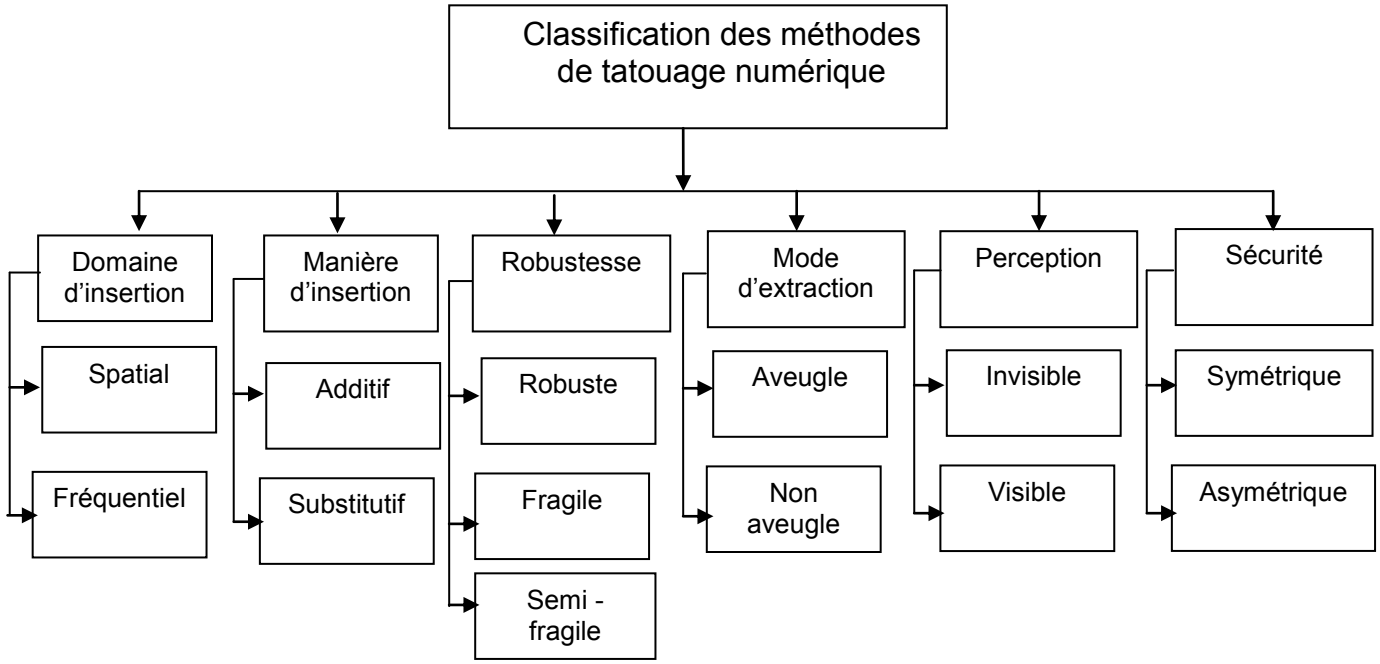


Figure 1.3 : Classification des méthodes de tatouage numérique.

1.8.1. Classification selon le domaine d'insertion

Selon le domaine d'insertion, les techniques du tatouage proposées dans la littérature peuvent être groupées en trois classes : celles qui opèrent dans le domaine spatial, dans le domaine transformé, ou dans le domaine compressé. Cependant, il existe des techniques qui utilisent la combinaison de plusieurs domaines, appelées techniques hybrides [43]. Ces algorithmes sont peu rencontrés dans la littérature.

Le choix du domaine influe directement sur la fiabilité de la technique en termes de robustesse et de capacité. C'est aussi un élément de contre-attaque. Ce

choix dépend de l'application visée, vu que chaque domaine possède ses propres caractéristiques.

- **Le domaine spatial** : Dans les techniques basées sur le domaine spatial comme domaine d'insertion, la marque est insérée par une modification directe des pixels. Ces pixels sont généralement sélectionnés par une clé secrète ou en se basant sur un modèle psycho-visuel. Les premières approches de tatouage numérique ont été conçues pour travailler dans ce domaine à cause de sa simplicité. Leurs principes consistaient à modifier directement la luminance (luma) des pixels ou insérer chaque élément de la marque dans le bit le moins significatif de chaque pixel L'avantage principal de ce domaine est son faible coût de calcul, ce qui favorise son utilisation dans les applications du tatouage en temps réel [44]. Leur inconvénient est la sensibilité aux attaques géométriques telles que la compression avec perte et le filtrage. C'est la raison pour laquelle beaucoup de méthodes de tatouage utilisent l'insertion dans d'autres domaines basés sur des transformations.
- **Le domaine fréquentiel :** Domaine transformé ou domaine fréquentiel obtenu du domaine spatial par une transformation en une dimension ou deux dimensions. La transformation peut se réaliser sur tout le document ou sur des blocs obtenus par une subdivision de celui-ci. L'avantage principal de ce domaine par rapport au domaine spatial et que l'insertion de la marque se fait dans les coefficients de la transformée, et ainsi, elle assure que les modifications appliquées sur un sous ensemble de ces coefficients seront propagées à tous les pixels dans le domaine spatial. Ce qui rend ces modifications imperceptibles. Les transformées les plus utilisées dans le domaine du tatouage numérique sont : DCT (Discrete Cosine Transform) ou TCD Transformée en Cosinus Discrète en français) [45], DFT (Discrete Fourier Transform ou TFD Transformée de Fourier discrète) [46] et DWT (Discret Wavelet Transform ou TOD Transformée en ondelettes) [47] [48] [49] [50].
- **Domaine compressé :** l'insertion peut être effectuée directement dans le bitstream (flux) compressé du signal vidéo.

1.8.2. Classification selon la manière d'insertion

La construction du signal de marque w lors de l'insertion et son extraction depuis le document reçu caractérise les techniques de tatouage. On classe ces techniques en deux catégories : d'abord les techniques additives, où le signal ajouté *w* n'est pas corrélé au signal hôte X, puis les techniques substitutives, où les données hôtes Xi sont modifiées afin de correspondre à un message codé [40].

- **Schéma additif :** Le tatouage additif consiste à ajouter un signal w à X (Y = X + w), sans que le codage amenant à w soit déterminé par X, même si les échantillons wi peuvent être modulés par un facteur perceptuel dépendant de X. Classiquement, on pose E [X, w] = 0, où E[X] est l'espérance de la variable aléatoire X. L'extraction se fait en décodant le signal *Y'* reçu, c'est-à-dire en décodant *w* bruité par l'attaque et par le signal hôte *X*. De plus, afin de respecter la contrainte d'imperceptibilité, l'énergie de *w* est très inférieure à celle de *X*.

- **Schéma substitutif :** Plutôt que de construire un signal w n'ayant que peu de rapport avec les données hôtes, le tatouage substitutif se propose de modifier ces données (la donnée Xi est remplacée par une donnée Yi très proche) afin de les faire correspondre au message que l'on souhaite transmettre. Ce schéma de tatouage correspond principalement à deux comportements : le premier est un tatouage substitutif avec contraintes. Il consiste à imposer un ensemble de contraintes aux données marquées. La méthode la plus répandue de cette variante consiste à remplacer les bits les moins significatifs ou les bits de poids faibles des pixels d'une image par les bits de la marque. Le second est un tatouage quantitatif qui regroupe les méthodes de tatouage substitutif avec dictionnaire. On peut classer dans ces méthodes le tatouage par quantification tel que la quantification par la modulation d'index [41] [42]. Le principe de cette technique consiste à quantifier des paramètres de l'image selon un ensemble de quantifieurs partitionné en autant de sous-ensembles que de symboles de la marque par paramètre.

1.8.3. Classification selon le type de la marque insérée

- **Tatouage robuste :** Il a pour objectif de transmettre une information malgré la modification du document. Lors de la lecture de la marque, certains algorithmes permettent d'extraire un message complet (une suite de symboles), tandis que d'autres indiquent simplement si le document a été marqué ou pas (on parle de détection de marque). Le tatouage robuste est particulièrement adapté au suivi et à la gestion de droits. Même si un fraudeur modifie le document, il est possible de retrouver l'auteur initial en insérant un numéro d'identification par tatouage robuste [28].
- **Tatouage fragile :** Il permet de prouver l'authenticité et l'intégrité du document marqué. Dans ce schéma de tatouage, la marque est très sensible aux modifications du contenu tatoué. Une technique de tatouage fragile doit détecter (avec une forte probabilité) toute altération du document tatoué. Une comparaison de la marque extraite et de la marque d'origine est effectuée afin d'identifier si le document est manipulé ou pas.
- **Tatouage semi-fragile :** Il combine les caractéristiques du tatouage robuste et fragile pour avoir une situation intermédiaire, dans laquelle la marque est robuste pour un ensemble défini de dégradations, et fragile à d'autres.

1.8.4. Classification selon le type de tatouage

- **Tatouage imperceptible :** Dans ce type, on n'observe pas l'existence de la marque. En conséquence, elle n'affecte pas la qualité du document et ce dernier garde sa qualité commerciale.
- **Tatouage perceptible :** Par contre, dans ce type de tatouage, la marque est bien visible dans le document. Il est utilisé plus dans des applications non commerciales.
- **Tatouage aveugle et non aveugle** : les schémas de tatouages peuvent être classés suivant les éléments nécessaires pour l'extraction (lecture du message depuis le document) de la marque. Un schéma **aveugle** n'a pas besoin du

document d'origine pour extraire la marque. Au contraire, un schéma **non aveugle** nécessite le media d'origine pour pouvoir lire correctement le message. Ces types de schémas sont de moins en moins étudiés, les applications concrètes étant assez rares.

- **Tatouage symétrique et asymétrique :** Un dernier point discriminant est l'utilisation des clés. La marque insérée est issue du codage du message à transmettre. Il est dépendant d'une clé. Si cette même clé est nécessaire au décodage (c'est-à-dire à l'extraction du message), le schéma est **symétrique** et dans le cas contraire, il est **asymétrique** (systèmes à clé privée et clé publique). On retrouve cette classification dans les algorithmes de cryptographie.

1.9. Evaluation des algorithmes de tatouage

Les techniques de tatouage se multiplient, mais aucune ne parvient encore à s'imposer. En effet, les utilisateurs potentiels de schémas de tatouage ne savent à quel algorithme se fier, car il n'existe toujours pas de programme permettant une évaluation fine, ni un cahier des charges qui fixe la longueur de la marque, ou la qualité de tatouage en terme d'imperceptibilité ou même l'ensemble des transformations auxquelles le tatouage doit être robuste dans le cas de la protection du droit d'auteurs ou sensible dans le cas de l'authentification du contenu. Mais en général, les chercheurs évaluent leur approche en se basant sur le meilleur compromis entre qualité, capacité et robustesse (figure 1.4).

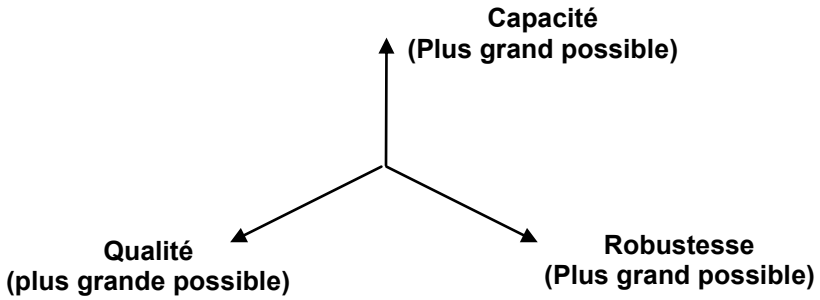


Figure 1.4 : Illustration graphique du compromis entre les caractéristiques du tatouage numérique.

1.9.1. Qualité

La notion de qualité intervient deux fois dans le cahier des charges d'un processus de tatouage. Il faut, d'une part, que la vidéo tatouée soit de la même qualité que la vidéo d'origine, c'est la contrainte d'imperceptibilité que doit satisfaire un algorithme de tatouage. D'autre part, les attaques auxquelles doit être robuste le tatouage, doivent conserver la qualité de la vidéo. Cette notion de qualité permet donc de caractériser les attaques et de restreindre leur ensemble afin d'étudier la robustesse de la marque.

1.9.2. Mesure de la qualité d'une vidéo

La première constatation est qu'il n'existe aucune méthode automatique pour mesurer la qualité absolue d'une vidéo. Aucun algorithme n'est capable, sans la vidéo de référence, de dire qu'une vidéo est de bonne ou mauvaise qualité. Afin de comparer la qualité des séquences vidéo obtenues, de nombreuses mesures, aussi bien subjectives qu'objectives, ont vu le jour. La méthode subjective donnant généralement les meilleurs résultats est basée sur la recommandation ITU-R BT.500 intitulée « Methodolgy for the subjective assessment of the quality of television pictures» [51]. Le principe de cette recommandation consiste à présenter à un groupe d'observateurs composé d'experts et de non-experts une paire de séquences vidéo: une séquence de référence et une séquence traitée. La position de la séquence de référence varie d'une manière pseudo aléatoire. A la fin, les observateurs expriment leurs jugements. Pour permettre de se repérer, la qualité a été divisée en cinq intervalles égaux correspondant aux mêmes qualificatifs utilisés dans le cas des images fixes : excellent, bon, assez bon, médiocre et mauvais (tableau 1.1).

Tableau 1.1 : Notes de qualité des images [51].

Note	Qualité
5	Excellente
4	Bonne
3	Assez bonne
2	Médiocre
1	Mauvaise

Toutefois les tests menés dans le cadre de la mesure subjective sont lourds à mettre en œuvre, chers et surtout très longs et ne constituent donc pas une solution pratique pour les différents opérateurs. Pour éviter un tel inconvénient, les métriques perceptuelles, qui représentent la deuxième alternative, ont pour objectif de définir des mesures de qualité qui soient fortement corrélées aux notes de qualité qu'auraient donné un ensemble d'observateurs. Jusqu'à un passé récent, ces métriques étaient simples et leurs performances restaient limitées. Le développement technologique de ces dernières années a permis la mise en place d'outils psychophysiques qui ont aidé à mieux comprendre le comportement du SVH et affiner les modèles associés. L'intégration de ces modèles dans les métriques perceptuelles est devenue alors un domaine de recherche et d'application très actif. En général, les plus utilisées dans les applications vidéo sont le rapport signal à bruit crête PSNR (Peak Signal to Noise Ratio), métrique de qualité vidéo VQM (Video quality metric) [52] [53] et l'index de similarité structurelle SSIM (Structural similarity Index Metric) [54].

Le PSNR est la métrique d'évaluation de la qualité de l'image la plus répandue. Il exprime le rapport entre la puissance maximale d'un signal et la puissance du bruit de corruption affectant la fidélité de la représentation. L'usage du PSNR est devenu presque exclusif vu sa simplicité et sa rapidité d'exécution, qui rend son utilisation très aisée. Le PSNR est mesuré en décibel (dB) à partir des relations suivantes :

$$MSE = \frac{\sum_{x=0}^{M}\sum_{y=0}^{N}(I_{x,y} - \overline{I_{x,y}})^2}{MN} \tag{1.1}$$

$$PSNR = 10\log_{10}\frac{(\max(I_{x,y}))^2}{MSE} \tag{1.2}$$

$I_{x,y}$ est la valeur du pixel à la position *(x,y)* de l'image référence et $\overline{I_{x,y}}$ celle de l'image à tester, les deux images étant de taille *MxN.* MSE représente l'erreur quadratique moyenne entre $I_{x,y}$ et $\overline{I_{x,y}}$.

Le critère VQM est proposé par Wolf et Pinson du NTIA (National Telecommunications and Information Administration) [52] [53]. Il représente une solution précise et souple pour mesurer et surveiller la qualité vidéo perçue par les utilisateurs finaux (aussi appelée QoE: Quality of Experience). VQM produit des notes de qualité vidéo perçue exprimées sur une échelle MOS (Mean Opinion Score). Les métriques de qualité vidéo perçue intégrées dans VQM sont dédiées aux formats d'encodage, MPEG-4/AVC (H.264) et MPEG-2. Ces métriques de qualité vidéo ont été optimisées pour produire des notes de qualité hautement corrélées avec des jugements humains recueillis lors de tests subjectifs d'évaluation de qualité réalisés dans des conditions d'observation normalisées.

VQM est basé sur le calcul des caractéristiques locales extraites de la séquence de référence et de celle à évaluer avant de les comparer. L'extraction des caractéristiques se fait par régions spatio-temporelles élémentaires. VQM utilise six caractéristiques. Deux caractérisent l'activité spatiale, une troisième les distorsions dans les composantes chromatiques, une quatrième le contraste local, une cinquième la quantité d'information temporelle et enfin la sixième est le produit de la caractéristique du contraste local et de celle de l'information temporelle. Des fonctions sont utilisées pour mesurer les distorsions de la vidéo dégradée par rapport à la vidéo d'origine en fonction de l'évolution des caractéristiques. Il s'agit de différentes fonctions de comparaison, de calcul de distance, de cumul et de seuillage. La combinaison des caractéristiques et des fonctions fournit des paramètres de gain ou de perte de qualité en grande quantité. Cela permet de couvrir une large gamme de contextes et explique la diversité des modèles proposés. Chaque modèle consiste ensuite en une combinaison linéaire d'une sélection de fonctions appliquées aux caractéristiques. Les pondérations de ces combinaisons sont obtenues par optimisation. La sélection des fonctions est faire de sorte à retenir les plus pertinentes. Enfin, chaque modèle produit une valeur comprise entre 0 et 1. La valeur nulle correspond à la perception d'aucune distorsion, alors que la valeur un correspond à la perception maximale de distorsions.

La troisième métrique de qualité vidéo utilisée est l'index de similarité structurelle SSIM [54]. Elle appartient aux approches basées sur la fidélité structurelle. Son principe se base non pas sur des propriétés bas niveau de la vision, mais sur des propriétés supposées haut niveau portant sur la réaction du SVH à une image dégradée. La principale hypothèse est que la perception humaine est particulièrement adaptée à l'extraction de l'information structurelle d'une image. L'idée est donc de mesurer les dégradations de cette information structurelle. Le principe de SSIM utilise l'index de qualité d'image (UQI : Universal image quality index) des mêmes auteurs [54]. Cet index UQI définit des mesures de comparaison de luma l (x, y), de contraste c(x, y) et de structure s(x, y) entre deux signaux x et y de luma :

$$l(x,y)=\frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2},\quad c(x,y)=\frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2},\quad s(x,y)=\frac{cov_{xy}}{\sigma_x\sigma_y}, \tag{1.3}$$

avec μ_x la moyenne de x, μ_y la moyenne de y, σ_x^2 la variance de x, σ_y^2 la variance de y et cov_{xy} la covariance entre x et y. L'index de similarité UQI entre x et y correspond alors à :

$$UQI(x,y)=l(x,y).c(x,y).s(x,y)=\frac{4\mu_x\mu_y\, cov_{xy}}{(\mu_x^2+\mu_y^2)(\sigma_x^2+\sigma_y^2)} \tag{1.4}$$

Le passage à SSIM résulte de la prise en compte des cas où $\mu_x^2+\mu_y^2$ ou $\sigma_x^2+\sigma_y^2$ peuvent être proches de zéro. La formule est alors transformée de la manière suivante [54]:

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\,cov_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{1.5}$$

avec $c_1=(k_1L)^2$, $c_2=(k_2L)^2$, L la dynamique des valeurs des pixels, soit 255 pour des images codées sur 8 bits, $k_1 = 0.01$ et $k_2= 0.03$ par défaut.

Vu que notre travail consiste à réaliser un système de protection du contenu vidéo compressé, il est nécessaire de donner quelques notions de bases du signal vidéo et sa compression.

1.10. Notion de base du signal vidéo et sa compression

Une séquence vidéo est une suite d'images fixes, qui peut être caractérisée par trois paramètres : sa résolution en luma, sa résolution spatiale et sa résolution temporelle. La résolution en luma détermine le nombre de couleurs possibles pour un pixel. Celle-ci est généralement de 8 bits pour les niveaux de gris et de 24 bits pour les séquences en couleurs. La résolution spatiale définit le nombre de lignes et de colonnes de la matrice de pixels. Enfin, la résolution temporelle est le nombre d'images par seconde. La valeur de ces trois paramètres détermine l'espace mémoire nécessaire pour stocker chaque image de la séquence. Cet espace mémoire est caractérisé par le débit, qui est le coût de stockage pour une seconde (capacité mémoire nécessaire pour stocker une seconde de vidéo).

Une image animée est une suite d'images décrivant un mouvement. Le nombre d'images par seconde doit être suffisant pour donner à l'œil une sensation de fluidité. La fréquence idéale est de 25 images par seconde. A cette fréquence, l'œil perçoit le mouvement de façon claire.

1.10.1. Formats de compression

Le développement de la compression et des équipements de traitements vidéo ont permis à l'ère digitale de s'épanouir, et de remplacer progressivement l'ère analogique. Le but de développer les formats de compression et les traitements de la vidéo en général, est d'en optimiser le contenu, afin d'en réduire l'espace de stockage, tout en maintenant une excellente qualité. De ce fait, le but d'un système de compression est d'éliminer les redondances spatio-temporelles d'un média afin d'en diminuer la taille (nous n'évoquerons ici que le cas de la vidéo).

Dans le monde analogique, ces redondances sont exploitées via le codage de la couleur, basé sur la vision et les techniques d'entrelacement. Le monde numérique permet quant à lui d'utiliser de nouvelles méthodes de codage qui seront présentées dans le paragraphe suivant. Le codage de la couleur consiste à déterminer un espace qui se rapproche au mieux des caractéristiques de la vision humaine. De nombreux standards vidéo, tels que le standard PAL, NTSC, ou MPEG introduisent un modèle SVH pour traiter la couleur. Ces standards prennent en effet en compte la perception non linéaire de luma, l'organisation des canaux de la couleur et les lois de l'acuité visuelle vis à vis de la chrominance.

1.10.2. Principes élémentaires du codage de la couleur

La théorie des couleurs opposées établit que le système visuel humain décorrèle ses entrées entre des signaux noir-blanc, rouge-vert et bleu-jaune, qui sont traités dans des canaux séparés. De plus, l'acuité visuelle pour la chrominance est inférieure à l'acuité pour la luma. Afin d'exploiter cet aspect de la vision humaine, les couleurs primaires Rouge, Vert et Bleu (RVB) sont rarement utilisées directement pour le codage ; à la place, on utilise couramment des systèmes de couleurs où les signaux correspondent à des différences. En vidéo, l'espace résultant de ces considérations est souvent l'espace YUV (ou YC_BC_R), où *Y* dénote la luma, *U* (ou C_B) la différence entre la couleur primaire bleue et la luma, et *V* (ou C_R) la différence entre la couleur primaire rouge et la luma. En vidéo, il est courant d'employer un sous-échantillonnage pour coder la couleur, les termes couramment utilisés sont les suivants :

- **4:4:4** correspond à l'absence de sous-échantillonnage ;
- **4:2:2** correspond à un sous échantillonnage de la chrominance par un facteur 2, horizontalement;
- **4:2:0** correspond à un sous-échantillonnage de la chrominance par un facteur 2, à la fois horizontalement et verticalement. Ce format est celui qui se rapproche le plus de l'acuité visuelle pour les couleurs, lorsque l'on considère la seule opération de sous-échantillonnage des couleurs;

- **4:1:1** correspond à un sous-échantillonnage horizontal de la chrominance par un facteur 4.

1.10.3. Formats vidéo

Il existe de nombreux formats de trame vidéo. En pratique, on utilise des formats intermédiaires tels que le CIF (Common Intermediate Format) et ses dérivés. Le format 4CIF est le standard de définition pour la télévision et le DVD (Digital Video Disc). CIF et QCIF sont utilisés en vidéoconférence et les formats QCIF et SQCIF pour les applications multimédia mobiles. Le tableau 1.2 présente la taille, exprimée en bits, nécessaire à la représentation d'une trame (non compressée) pour chaque format. On considère le modèle 4:2:0 pour des échantillons de luma et de chroma de 8 bits.

Tableau 1.2 : Formats de trame vidéo.

Format	Résolution de la luma	Bits par trames (4 :2 :0)
Sub-QCIF	128×96	147456
Quarter-CIF	176×144	304128
CIF	352×288	1216512
4CIF	704×576	4866048

1.10.4. Compression vidéo

La bande passante disponible pour la diffusion de la télévision numérique et ses applications est très limitée. C'est pourquoi, une réelle motivation est née afin de pallier ce besoin : développer de nouvelles technologies de compression pour la diffusion de la télévision numérique et ses applications [55]. Jusqu'à l'arrivée de la norme H.264/AVC, le standard de compression MPEG-2 était le système le plus utilisé au monde. Si un codec permet de diminuer le coût de codage, et donc la bande passante utile, il doit répondre à un certain nombre d'autres exigences :

- accès aléatoire dans la séquence décodée ;

- possibilité de compresser l'information dans plusieurs formats d'image;
- possibilité d'avoir un débit allant jusqu'à 80 Mbit/s.

La plupart des codecs de compression vidéo sont basés sur des transformées telles que la DCT, ou la DWT. L'architecture d'un codec est généralement composée de deux systèmes complémentaires, un codeur et un décodeur [55]. Le codeur convertit la source de données en une forme compressée occupant un nombre de bits inférieur, avant transmission ou stockage. Le décodeur convertit la forme compressée en une représentation de la source d'origine. La paire codeur/décodeur est appelée codec.

Un codeur comprend trois fonctionnalités principales [55] : un modèle temporel, un modèle spatial et un codeur entropique (Figure 1.5). L'entrée du système est une séquence vidéo non compressée. Le modèle temporel réduit la redondance temporelle en exploitant les similarités entre les trames autour d'un même instant. Il construit ainsi une prédiction. Celle-ci est faite à partir d'une ou plusieurs trames futures ou passées et est améliorée en compensant les différences entre les trames (prédiction par compensation de mouvement). La sortie du modèle temporel est constituée d'une trame résiduelle (la trame courante moins la prédiction) et d'un jeu de paramètres tels que des vecteurs de mouvement (MVs) décrivant la compensation.

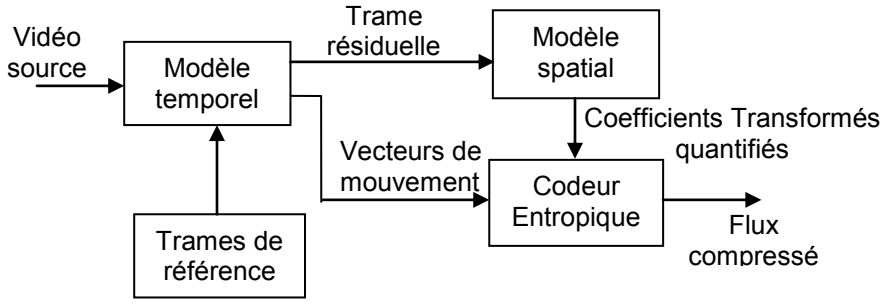


Figure 1.5 : Bloc diagramme d'un codeur vidéo [2].

La trame résiduelle devient l'entrée du modèle spatial qui utilise les similarités entre les pixels d'un même voisinage pour réduire la redondance spatiale. Ceci est

généralement effectué par transformation du résidu et quantification du résultat. La transformation convertit l'image résiduelle dans un autre domaine dans lequel elle est représentée par des coefficients de transformation. Ceux-ci sont ensuite quantifiés afin d'éliminer les valeurs non significatives, laissant un nombre réduit de coefficients significatifs. On obtient ainsi une représentation plus compacte de la trame résiduelle.

La sortie du modèle spatial est un ensemble de coefficients de transformation quantifiés. Les paramètres du modèle temporel (typiquement des MVs) et du modèle spatial (coefficients) sont codés par le codage entropique (comme le codage de Huffman ou le codage arithmétique). Le codage entropique se base sur la probabilité d'apparition des symboles. Celui-ci retire la redondance statistique des données et produit un flux binaire compressé qui peut être transmis ou stocké. Ainsi, une séquence compressée consiste en des paramètres de MVs, des coefficients résiduels et un en-tête d'information. Le décodeur vidéo reconstruit la séquence à partir du flux compressé. Les coefficients et les MVs sont décodés par un décodeur entropique, les trames résiduelles sont reconstruites par le modèle spatial. Il ne reste plus qu'à recréer les prédictions auxquelles on ajoute les trames résiduelles [62]. Il existe de nombreux codecs vidéo, nous pouvons, sans être exhaustif, citer les suivants : MPEG-1 [2], MPEG-2 [3], MPEG-4, MPEG4-AVC (H.264) [5] [55].

1.11. Conclusion

Dans ce premier chapitre, nous avons abordé les différentes définitions jugées nécessaires pour se familiariser avec les notions de tatouage numérique et la compression que nous allons utiliser dans le contexte de notre étude. La nature du média est la vidéo compressée par le standard H.264/AVC. Dans les prochains chapitres, nous commençons par rappeler les principales étapes de fonctionnement de ce standard, puis nous donnons les algorithmes développés pour les applications de protection des droits d'auteur et d'authentification du contenu dans cette norme.

CHAPITRE 2

LE STANDARD DE COMPRESSION VIDEO H.264/AVC ET LA PROTECTION DE SON CONTENU

2.1. Introduction

Différentes techniques de tatouage ont été proposées pour la protection du contenu dans les standards de compression vidéo antérieurs tels que MPEG-1 [2] et MPEG-2 [3], mais peu de travaux ont été effectués pour la norme H.264/AVC actuelle. Dans la première partie de ce chapitre, nous décrivons le principe de fonctionnement du codec (codeur/décodeur) H.264/AVC et nous détaillons les différents modules le constituant. Nous présentons ensuite dans la seconde partie les différentes applications utilisant le tatouage numérique réalisées dans la norme H.264/AVC pour protéger le contenu vidéo.

2.2. Fonctionnement du codeur H.264/AVC

En comparaison avec les anciens standards de compression vidéo, le H.264/AVC est basé sur une véritable révolution algorithmique qui permet d'atteindre un seuil de codage qui n'était pas prévisible huit ans auparavant [5] [6] [55] [56] [57] [58]. Ces progrès ont été rendus possibles par l'union des experts vidéo de l'ITU-T et de MPEG (Moving Picture Experts Group) qui ont établi la JVT (Joint Video Team) en décembre 2001 afin de finaliser la norme. H.264/AVC fut finalisée en mars 2003 et approuvée par l'ITU-T en mai 2003 (Figure 2.1). Il s'agit du codec vidéo le plus performant en termes de débit-distorsion dans la mesure où il apporte un gain de 60% de débit pour une même qualité par rapport à MPEG-2 [55] [56] [58].

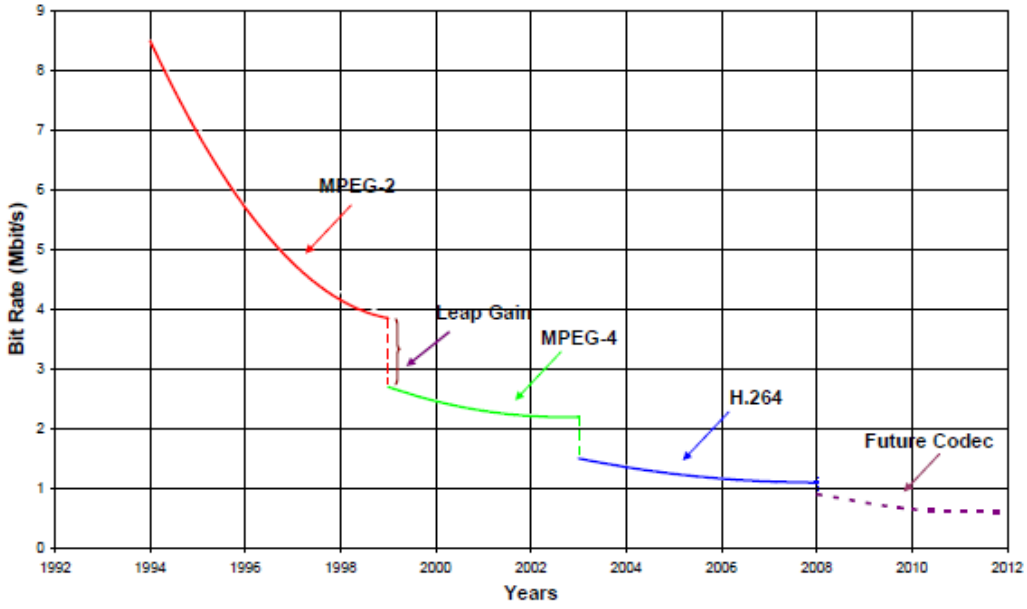


Figure 2.1 : Chronologie des standards vidéo [55].

La norme H.264/AVC représente l'aboutissement des techniques de compression hybrides du fait que deux techniques de réduction de redondances sont utilisées : d'une part, une prédiction temporelle, d'autre part, une transformation des résidus de prédiction.

Le codeur inclut deux chemins de données [57] [58]: le chemin avant (de gauche à droite) et le chemin de reconstruction (de droite à gauche). C'est pourquoi le décodeur est présenté de droite à gauche, le décodeur étant ainsi inclus dans le codeur.

Le schéma bloc général d'un codeur de type H.264/AVC est illustré sur la Figure 2.2. L'image d'entrée (trame) est partitionnée en blocs de pixels de taille 16×16 appelés macroblocs (MBs). Chaque MB est composé de trois composantes: Y, U et V. Du fait que la vision humaine est moins sensible à la chroma qu'à la luma, les MBs de chroma sont sous-échantillonnés d'un facteur 2 dans les directions horizontales et verticales. Par conséquent, chaque portion élémentaire de l'image est composée d'un MB de luma de 16×16 pixels et de deux MBs de chroma de 8×8 pixels. Chaque MB est prédit en mode Intra ou Inter. Ces deux outils caractérisent l'étage de prédiction du codeur. Le mode Intra exploite les redondances spatiales des

images, il permet de construire une estimation d'un MB en utilisant exclusivement les informations contenues dans l'image courante. Le mode Inter tire parti des redondances temporelles entre les images, il permet de prédire le MB courant en utilisant les informations contenues dans des images de référence, qui ont déjà été codées, décodées puis stockées dans une mémoire (images décodées). Ce principe de compensation en mouvement (CM) repose sur l'estimation d'un vecteur de déplacement associé à chaque bloc. Ce vecteur caractérise la position du bloc le plus vraisemblable dans l'image de référence. Il est évident que la prédiction Inter est beaucoup plus efficace que la prédiction Intra car les redondances temporelles représentent une forte proportion de l'énergie du signal. Le mode Inter est donc utilisé en priorité dans les codeurs vidéo, excepté dans la situation où la mémoire ne contient aucune image de référence (première image d'une séquence vidéo par exemple). Les images de référence peuvent aussi bien faire parti du passé que du futur de l'image courante.

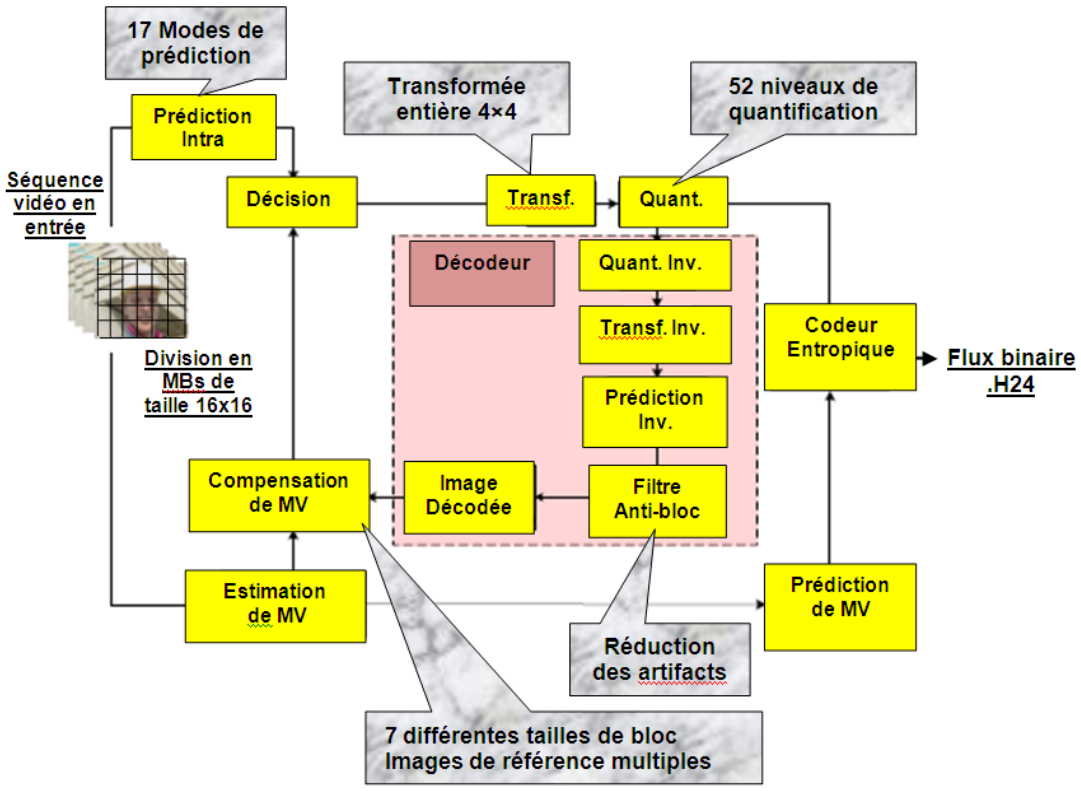


Figure 2.2 : Structure basique de codage de la norme H.264/AVC [6] [7].

L'erreur de prédiction, qui correspond à la différence entre le bloc d'origine et le bloc prédit, est ensuite transformée par la DCT 4×4 entière, quantifiée puis codée par le biais d'un codage entropique. Les informations de contrôle ainsi que les MVs sont également codés avant d'être transmis. Afin de reconstruire les mêmes images au codeur et au décodeur, les coefficients quantifiés de chaque bloc sont inversés et ajoutés au signal de prédiction pour reconstruire chaque MB codé, qui pourra ensuite servir de référence pour la prédiction des autres MBs. Ces mêmes coefficients ainsi que d'autres informations nécessaires au décodage forment un flux compressé qui passe à la couche d'abstraction réseau NAL (Network Abstraction Layer) pour transmission ou stockage.

Afin de réduire les effets de blocs (artefacts) générés par le décodage interne au codeur, un filtre débloquant a été intégré dans la boucle de compression. Ce dernier permet de lisser les informations visuelles avant de les stocker dans la mémoire de référence.

La structure de traitement du décodeur est plus simple que celle du codeur (Figure 2.3). Il commence par décoder les différents types d'information : paramètres de contrôle, MVs et coefficients quantifiés. Les MBs sont ensuite successivement prédits en utilisant le mode approprié (Intra ou Inter). En parallèle, les résidus de prédiction sont reconstruits grâce aux coefficients quantifiés, puis ajoutés au signal de prédiction. Suite à l'opération de filtrage, le MB est complètement décodé et peut être stocké en mémoire pour les prédictions futures.

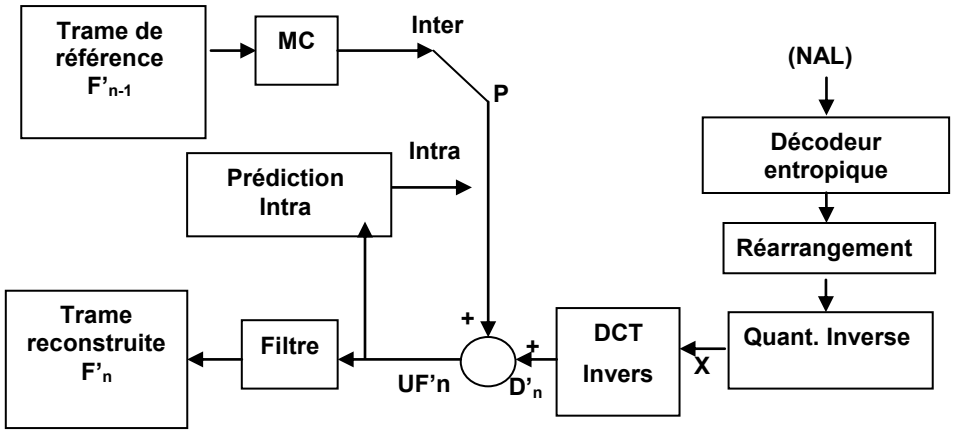


Figure 2.3 : Diagramme du décodeur H.264/AVC [58 [59].

2.2.1. Partitionnement en tranches (slices)

Cinq types d'images (slices) sont supportés par la norme H.264/AVC et sont nommées : I, P, B, SI et SP. Ils sont classés selon les méthodes de prédiction utilisées.

- **Les images I ou les « Intra » images** : Une image Intra décrit une image statique dans le GOP. Elle ne tient compte que des pixels se trouvant dans la même image que le bloc à prédire (elle fait référence à elle-même). Tous les MBs sont codés sans référence à d'autres images de la séquence vidéo. De manière idéale, une image I devrait intervenir lors d'un changement de scène, c'est-à-dire lorsque les redondances temporelles entre les images sont faibles.

- **Les images P ou les images « prédictives »** : Dans ce type d'images, chaque MB est prédit en utilisant soit le mode Intra, soit le mode Inter. Lorsque le mode Inter est activé, chaque MB est associé à une seule image de référence.

- **Les images B ou les images « bidirectionnelles prédictives »** : fonctionnent comme des slices P sauf que pour la prédiction, elles nécessitent les anciennes et les futures images I ou P, comme images de référence.

- **Les images SP et SI ou les tranches de « commutation »** : peuvent être employées pour des transitions entre deux flux différents de la vidéo H.264/AVC. Le but des images SP et SI est de faciliter l'accès aléatoire aux trames en phase de décompression d'une vidéo [55].

2.2.2. Partitionnement de la séquence vidéo

Une séquence vidéo est décomposée en groupes d'images GOP (Group Of Pictures) (Figure 2.4), dont chaque image de chaque groupe est décomposée en bandes. La bande est un groupe de MBs organisés en 16 blocs dont chacun représente la plus petite entité de cette hiérarchie permettant de réduire les redondances spatiales. Les MBs sont organisés en tranches (Slices) qui représentent

en général des sous-ensembles d'une image donnée pouvant être codés indépendamment.

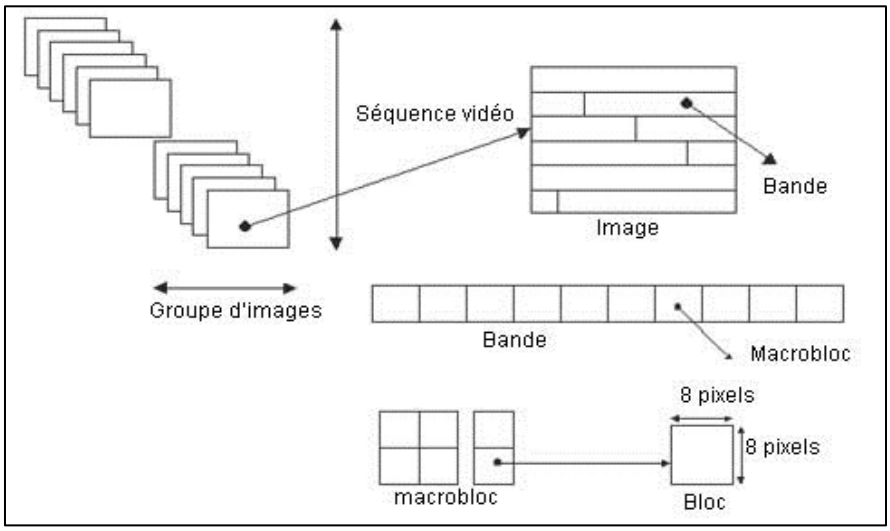


Figure 2.4 : Partitionnement de la séquence vidéo [6] [7] [58] [59].

Un GOP débute par une image I et contient ensuite une succession d'images P et B. La structure classique d'un GOP est illustrée sur la figure 2.5. Généralement, la première image d'un GOP vide la mémoire de référence (Instantaneous Decoding Refresh ou IDR). Par conséquent, les GOPs sont indépendants entre eux. Cette technique permet au décodeur de se resynchroniser sur le flux dans le cas d'une transmission avec pertes.

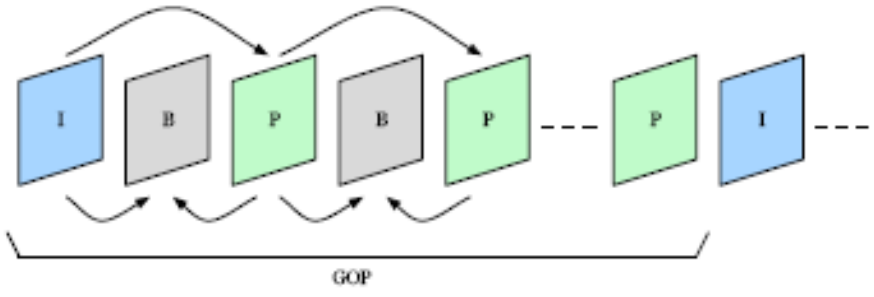


Figure 2.5 : Structure du GOP.

2.2.3. Prédiction

Le but de la prédiction est de réduire la redondance spatiale présente dans l'image courante et la redondante temporelle produite entre les images transmises. Avant de décrire les deux types de prédiction (prédiction Intra et Inter) prévus dans la norme H.264/AVC, il est nécessaire de définir ce qu'est un résiduel et comment il est codé puis comment il est calculé.

Le résiduel est la différence entre le prédicteur et le MB à coder. Le prédicteur, noté $\hat{p}$ est calculé à partir de blocs déjà codés se trouvant au-dessus et à gauche du bloc à prédire. On note *e* le résiduel et *p* le pixel à coder, l'expression du résiduel est alors donnée par la formule suivante:

$$e(x,y) = p(x,y) - \hat{p}(x,y) \tag{2.1}$$

C'est ce résiduel *e* qui est transmis au décodeur qui le décode puis recalcule le prédicteur $\hat{p}$. Le pixel initial *p* est alors retrouvé comme suit:

$$p(x,y) = \hat{p}(x,y) + e(x,y) \tag{2.2}$$

2.2.3.1. Prédiction Intra

La prédiction Intra n'utilise pas les images de référence mais uniquement l'image courante. Ce mode de prédiction consiste à estimer les échantillons d'un MB en utilisant les informations contenues dans les blocs contigus appartenant au passé spatial de l'image courante. Ces blocs de référence doivent déjà avoir été codés puis décodés par le codeur (de manière à retrouver des résultats identiques au codeur et au décodeur). La norme H.264/AVC propose deux types de traitement Intra pour prédire le signal de luma. Le premier mode est appelé Intra_4×4 et le second Intra_16×16 [59] [60].

- **Prédiction Intra_4×4** : Pour la prédiction Intra, les standards précédents tels que le codeur MPEG-1, MPEG-2 [3] travaillaient généralement avec des blocs de taille 8×8. C'est donc une nouveauté de descendre jusqu'à des blocs de

taille 4×4, ce qui engendre bien évidemment, une augmentation de la quantité de calculs [60] [61] [62]. Dans le type Intra_4×4, le MB est divisé en seize blocs de 4×4 pixels et chaque bloc est codé individuellement. Neuf modèles de prédiction sont fournis par la norme et l'objectif du codeur est de sélectionner le mode le plus adapté au bloc courant (Figure 2.6) [56] [57] [58] [59] [60] [63]. Huit de ces modèles caractérisent des modèles de prédiction directionnels. Chaque valeur prédite du bloc courant est déterminée à partir d'une combinaison linéaire entre les pixels situés dans les blocs contigus. Ces combinaisons linéaires sont définies dans la norme en fonction d'une direction spécifique.

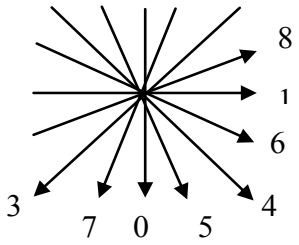


Figure 2.6 : Modes de prédiction Intra [63].

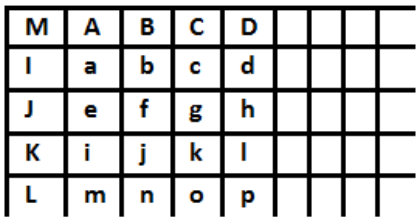


Figure 2.7 : Labellisation des échantillons de prédiction (4×4) [63].

La figure 2.7 présente la manière dont les pixels sont labellisés. Avec a, b, c,.., p sont des pixels du bloc courant, et M, A, B, …, L sont des pixels des blocs voisins. Le mode 0 (prédiction verticale) et le mode 1 (prédiction horizontale) sont montrés explicitement sur la figure 2.8. Par exemple, si le mode de prédiction verticale est appliqué, tous les échantillons au-dessous de l'échantillon A sont prédits par l'échantillon A. Tous les échantillons au-dessous de l'échantillon B sont prédits par B et ainsi de suite. Les échantillons des autres modes de prédiction sont calculés à partir des échantillons de A à M comme montré dans le tableau 2.1. Si les pixels voisins manquent (cas des blocs 4×4 dans les premiers MBs en haut et à gauche de l'image), le mode DC avec les pixels disponibles est utilisé [56] [57] [58]. Les flèches illustrées sur la figure 2.8 indiquent la direction de la prédiction de chaque mode.

Pour les modes de 3 à 8, les pixels prédits sont calculés par une formule appliquée aux pixels [A-M].

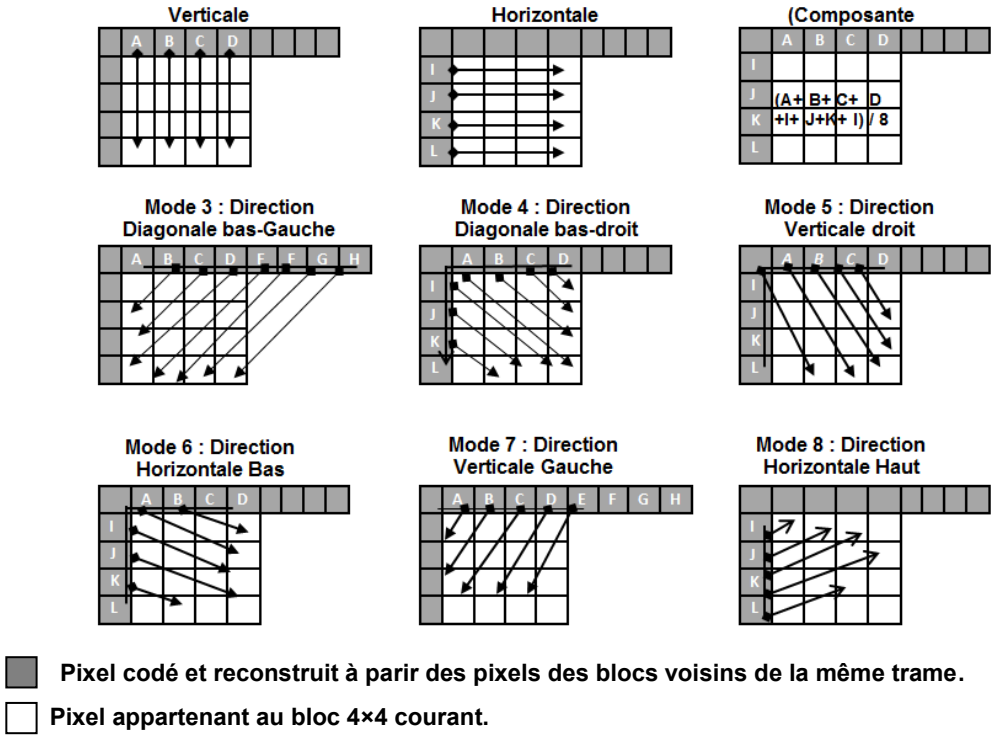


Figure 2.8 : Les neuf modes de prédiction Intra_4×4 [60] [63].

Tableau 2.1 : Les neufs modes de prédiction Intra_4×4 [60] [63].

Mode	Direction
0 (verticale)	[A-D] sont extrapolés verticalement
1 (horizontale)	[I-L] sont extrapolés horizontalement
2 (composante continue)	[a-p] sont prédits par la moyenne de [A-D] et [I-L]
3 (diagonale bas-gauche)	[a-p] sont interpolés à un angle de 45° bas-gauche
4 (diagonale bas-droit)	[a-p] sont interpolés à un angle de 45° bas-droit
5 (verticale-droit)	[a-p] sont interpolés à un angle de 22.5° vertical-droit
6 (horizontale-droit)	[a-p] sont interpolés à un angle de 22.5° horizontal-droit
7 (verticale-gauche)	[a-p] sont interpolés à un angle de 22.5° vertical-gauche
8 (horizontale-haute)	[a-p] sont interpolés à un angle de 22.5° horizontal-haut

Une fois calculés, les neuf modes sont évalués par une Somme de Différence Absolue (SAD) donnée par l'équation 2.3 [63]. Le mode fournissant la valeur minimale du SAD est retenu pour la prédiction du bloc.

$$SAD(x,y) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \left| C_{i,j} - R_{i+x,j+y} \right| \tag{2.3}$$

Où $C_{i,j}$ est la valeur du pixel de la trame courante et $R_{i+x,j+y}$ celle du bloc de référence. Les composantes du vecteur de déplacement sont (x,y).

- **La prédiction Intra 16×16 :** La prédiction par blocs de taille 4×4 demande des calculs importants et n'est pas toujours justifiée pour des régions de faible variation. C'est pourquoi la prédiction intra image peut se faire par blocs de taille 16×16. Cette alternative a l'avantage d'être évidemment plus rapide et moins coûteuse. Quatre modèles de prédiction sont définis dans la norme dont les trois premiers (mode 0, mode 1 et mode 2) sont semblables aux modes de prédiction Intra_4×4 et un mode plan (mode 3). La prédiction est une fonction linéaire entre les échantillons voisins se trouvant à gauche et au-dessus afin de prédire les échantillons courants (les échantillons sont calculés par moyenne des valeurs obliques dans les deux sens de direction). Les quatre modes sont représentés sur la figure 2.9 et le tableau 2.2 les énonce [56] [57] [58].

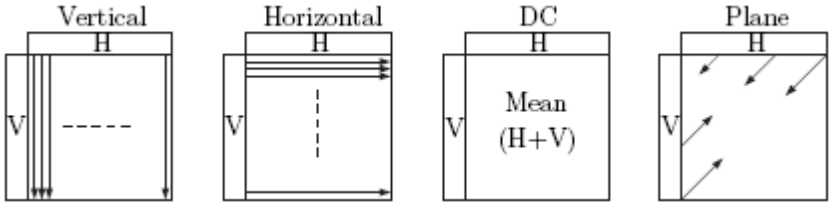


Figure 2.9 : Les quatre modes de prédiction Intra_16×16 [58] [59].

Tableau 2.2 : Modes de prédiction des blocs 16×16 de luma [58] [59].

Modes	Directions
0 (vertical)	Extrapolation à partir de H
1 (horizontal)	Extrapolation à partir de V
2 (composante continue)	Moyenne de H et V
3 (plan)	Une fonction linéaire plane est ajustée à partir de H et V

La prédiction des blocs de chroma : Pour le codage des MBs de chroma *Cb* et *Cr*, seule la prédiction Intra_8×8 est utilisée. Ce schéma est suffisant car les variations des signaux de chroma sont très faibles. Le mode Intra_8×8 permet d'estimer les 8×8 pixels d'un MB de chroma en proposant quatre modèles de prédiction (tableau 2.3) : DC, vertical, horizontal et plan. Les types Intra_8×8 et Intra_16×16 sont donc identiques à un facteur d'échelle près.

Tableau 2.3 : Modes de prédiction des blocs 8×8 de chroma [58] [59].

Modes	Directions
0 (composante continue)	Moyenne de H et V
1 (horizontal)	Extrapolation à partir de V
2 (vertical)	Extrapolation à partir de H
3 (plan)	Une fonction linéaire est ajustée à partir de H et V

Toutes ces informations (choix de prédiction, mode de prédiction, vecteur de mouvement) sont transmises au flux compressé.

2.2.3.2. Prédiction Inter

En mode Inter, la prédiction crée un modèle prédictif à partir d'une ou plusieurs images préalablement codées (images de référence) qui peuvent être antérieures ou postérieures à l'image à coder. La norme H.264/AVC comprend de nombreuses fonctionnalités nouvelles qui lui permettent de compresser beaucoup plus efficacement les vidéos que les normes précédentes. Parmi celles-ci, notons un plus large support de taille de blocs et un échantillonnage plus fin des MVs, d'où une

complexité calculatoire accrue. Une partition de luma peut être composée de 16×16, 16×8, 8×16 ou 8×8 échantillons [63] [65]. Un sous-MB de 8×8 pixels peut être à nouveau redécoupé en partitions de taille 8×4, 4×8 ou 4×4. Cette décomposition pyramidale appelée aussi méthode de compensation de mouvement (CM) à structure d'arbre permet d'isoler les différents objets composant une image et de s'adapter à leurs caractéristiques (sens de déplacement, vitesse). Un MV est associé à chaque partition d'un MB. Ce vecteur de déplacement spécifie l'écart spatial entre la partition courante de l'image actuelle et sa meilleure représentation dans l'image de référence. Les partitions d'un MB et d'un sous-MB sont illustrées sur la figure 2.10.

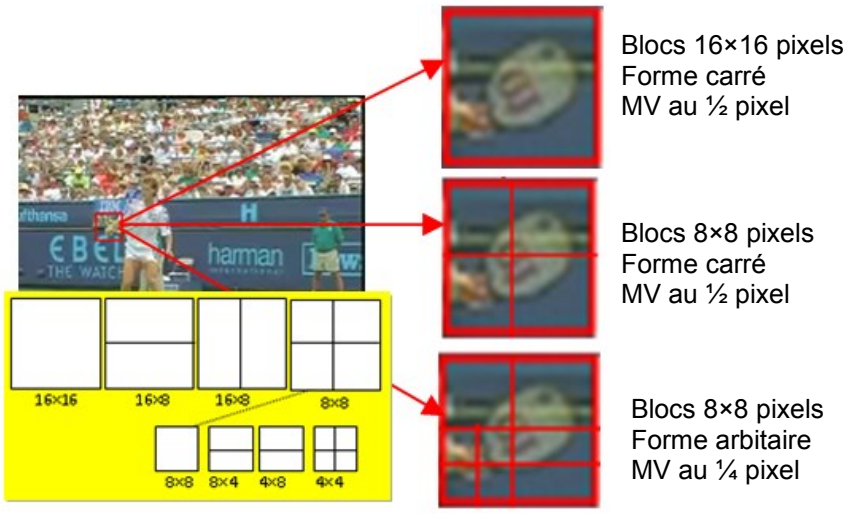


Figure 2.10 : Découpage d'un macrobloc.

2.2.4. Estimation et compensation de mouvement

Le but de l'estimation/compensation de mouvement est de réduire la redondance temporelle entre les images transmises, en formant une image prédite depuis une ou plusieurs images de référence et en la soustrayant à l'image courante [68] [69] [70]. Plus la prédiction est performante et moins l'image résiduelle résultante contiendra d'énergie. L'image résiduelle est par la suite codée, transmise au décodeur qui recrée l'image de prédiction, ajoute les données résiduelles décodées et reconstruit l'image courante (prédiction inter-images). La norme H.264/AVC

s’appuie sur l'estimation de mouvement basée bloc [71]. Les blocs sont rectangulaires de taille NxM avec N, M {4, 8, 16} pixels. Pour chaque bloc, l'algorithme de (CM) basé sur les blocs (block matching) est appliqué (figure 2.11). Cet algorithme consiste à comparer chaque bloc dans la trame d’origine avec les blocs de la même taille dans les trames de référence (passée ou future) dans le but de trouver le meilleur bloc qui satisfasse un critère d’erreur basé sur une mesure particulière. Le vecteur pointant vers le bloc sélectionné est choisi comme vecteur de déplacement. Les techniques basées sur les blocs sont à l’heure actuelle très efficace en termes de qualité et de débit. Elles restent les techniques les plus adoptées dans les standards de compression y compris la norme H.264/AVC.

La valeur du SAD est codée et transmise avec les MVs à l’étage suivant du codeur. Le MV n’est pas nécessairement composé d’entiers. La précision des MVs peut atteindre le quart-de pixel (¼ pel) dans la norme H.264/AVC. Un tel déplacement (résolution fractionnelle) peut pointer sur des positions qui sont spatialement situées entre les échantillons du signal image. Le signal de l’image de référence doit donc être interpolé entre les pixels.

Dans la norme H.264/AVC, le sur-échantillonnage d’un facteur de 2 du signal de luma est généré en appliquant un filtre RIF uni-dimensionnel d’ordre 6 sur les pixels de l’image. Les éléments correspondant au ¼ pel sont obtenus en moyennant les échantillons intermédiaires du signal.

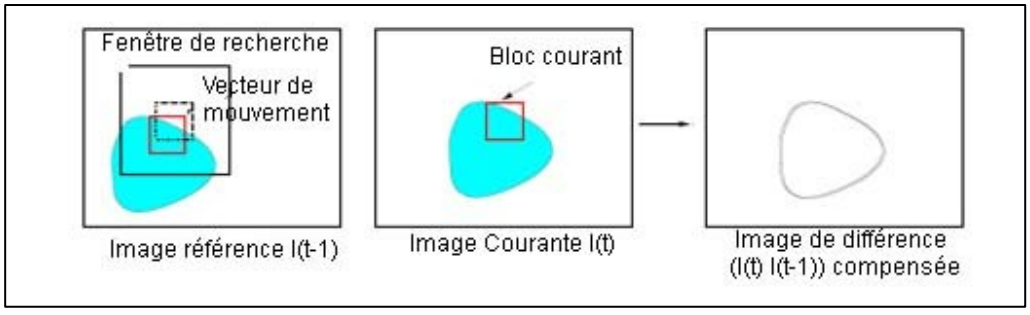
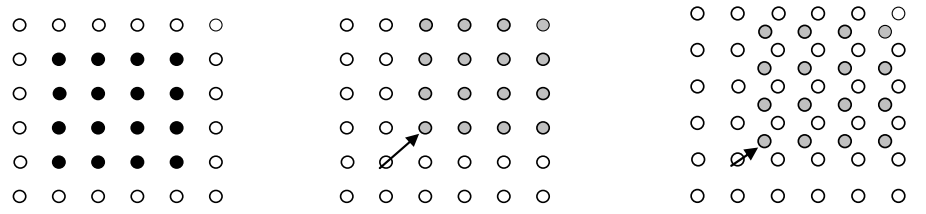


Figure 2.11 : Estimation et compensation du mouvement [65].

Le principe de l'estimation de mouvement consiste à prédire chaque sous partition 4×4 dans la trame courante par une région voisine de l'image de référence (Figure 2.12.a). Si les composantes horizontales MVx et verticales MVy du MV sont des nombres entiers, les échantillons appropriés dans le bloc de référence existent réellement (les points gris) (Figure 2.12.b). Si une ou les deux composantes du vecteur sont des valeurs fractionnaires, les échantillons de prédiction (points gris) sont produits par interpolation entre échantillons adjacents dans la trame de référence (points blancs) (Figure 2.12.c).

Dans la composante de luma, les échantillons aux positions d'un ½ pel sont produits et interpolés des échantillons des pixels voisins à l'aide d'un filtre. Ceci signifie que chaque échantillon de ½ pel est une somme de six échantillons voisins de pixels. Lorsque tous les échantillons de ½ pel sont disponibles, chaque échantillon de ¼ pel est produit en utilisant l'interpolation bilinéaire entre les échantillons voisins de un ou ½ pel.

Les MVs pour les partitions voisines sont souvent fortement corrélés; pour cela, chaque MV est prédit par des vecteurs voisins des partitions précédemment codées. Le vecteur prédit est formé à l'aide des MVs précédemment calculés. La différence entre le vecteur courant et le vecteur prédit est codé et transmis. Lors du décodage, le MV prédit est formé de la même manière et ajouté au vecteur de différence décodé.



(a) bloc 4×4 dans la trame courante

(b) bloc de référence Vecteur (1, -1)

(c) bloc de référence Vecteur (0.75, -0.5)

Figure 2.12 : Prédiction d'un pixel et d'un sous-pixel [6] [58] [59].

2.2.5. Transformation

Les résiduels des pixels prennent des valeurs équiprobables sur [-255; 256], la transformée permet alors de décorréler le signal afin de rendre le codage entropique plus efficace et de réduire les redondances spatiales de l'erreur de prédiction. La norme H.264/AVC dispose d'une transformée basée sur des entiers. La matrice de transformation est généralement composée de 4×4 éléments, mais peut être réduite à 2×2 éléments pour le codage de certaines informations de chroma. La diminution de la taille de la fenêtre d'analyse permet au codeur de mieux adapter le codage de l'erreur de prédiction aux frontières des objets en mouvement. En effet, la taille du bloc est similaire aux dimensions de la plus petite zone d'analyse de l'estimation Inter 4×4 ou Intra_4×4 et la transformée s'ajuste donc mieux aux erreurs de prédiction locales. La norme H.264/AVC utilise trois transformations selon le type de données résiduelles qui doivent être codées :

- Les MBs prédits en mode 16×16, les coefficients DC (coefficients de plus basse fréquence) sont placés dans une matrice 4×4 et une transformation de Hadamard lui est appliquée.
- Tous les MBs, les coefficients continus de chroma sont placés dans une matrice 2x2 et une transformation de Hadamard lui est également appliquée.
- Enfin, pour le reste des blocs 4×4, une transformation DCT 4×4 entière est appliquée afin de décorreler le signal en séparant les hautes fréquences des basses.

La transformée DCT classique est donnée par la formule (2.4) [72]. Dans la norme H.264/AVC, la transformée utilisée est une transformée entière [73]. Son implémentation ne comporte que des additions et des décalages ce qui donne l'avantage de stocker des résiduels entiers et non plus flottants comme dans ses prédécesseurs. Notons que cette transformée conserve les mêmes propriétés qu'une DCT classique donnée par :

$$F(i,j) = \frac{1}{\sqrt{2N}} C(i)\,C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} P(x,y) \cos\left(\frac{(2x+1)i\pi}{2N}\right).\cos\left(\frac{(2y+1)j\pi}{2N}\right) \tag{2.4}$$

avec $$C(i) = C(j) = \begin{cases} \frac{1}{\sqrt{2}} & si\ u = 0 \\ 1 & si\ u > 0 \end{cases}$$

où p(x,y) représente l'intensité du pixel à la position (x,y).

La formule de la DCT entière de taille 4×4 est donnée par [73]:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{pmatrix} \begin{pmatrix} x_{00} & x_{01} & x_{02} & x_{03} \\ x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ x_{30} & x_{31} & x_{32} & x_{33} \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{pmatrix} \tag{2.5}$$

où les $x_{i,j}$ sont les 16 pixels d'un bloc 4×4.

2.2.6. Quantification

La quantification scalaire a pour but de réduire l'espace des valeurs des résiduels transformés pour réduire l'entropie du signal. Cette opération consiste à diviser chaque coefficient transformé par son coefficient de quantification provenant d'une matrice de quantification et à ne garder que la partie entière. La quantification est donnée par la formule suivante :

$$q_{i,j} = arrondi\left(\frac{F_{i,j}}{QP}\right) \tag{2.6}$$

La quantification dans H.264/AVC fonctionne d'une manière différente de celle employée dans MPEG-1/2/4 [74]. Elle est sur une échelle logarithmique. Le pas de quantification *QP* est donné par :

$$QP(H.264/AVC) = 12 + 6\,Log_2\big(QP(MPEG)\big) \tag{2.7}$$

Par exemple, QP de MPEG égal à 4 est équivalent à QP de H.264/AVC égal approximativement à 20. Le paramètre *QP* correspond au pas de quantification choisi parmi 52 valeurs prédéfinies dans le standard H.264/AVC ce qui permet d'ajuster le critère débit/distorsion. Cette opération introduit naturellement des pertes d'information dans le signal.

2.2.7. Codage entropique

Le codage entropique est un procédé dit "sans-perte", qui représente l'information sous forme binaire et structurée afin d'être compréhensible par le décodeur. Dans le standard H.264/AVC, deux méthodes de codage entropique sont proposées:

- Le Context Adaptative Variable Length Coding (CAVLC) qui est une méthode de faible complexité basée sur l'utilisation d'un codage à longueur variable VLC suivant le contexte. Chaque composante (résiduel, mode, MV,..) à coder dispose ainsi de son propre contexte.
- Le Context Adaptative Binary Arithmetic Coding (CABAC) est un codage arithmétique plus complexe que CAVLC.

Ces deux méthodes représentent l'amélioration principale en termes d'efficacité de codage comparées aux normes précédentes.

2.2.7.1. Codage à longueur variable CAVLC

Le codage CAVLC est utilisé pour coder les résidus des blocs 4×4 parcourus en zig-zag. Le codage est effectué de telle sorte que les différents mots de code n'ont pas nécessairement la même longueur en bits. Ce codage utilise une table de mots de code définie pour tous les éléments syntaxiques. Les éléments de syntaxe incluent :

- l'élément de syntaxe de l'image, de la tranche et son en-tête.
- les indicateurs des MBs sautés.
- le type des MBs.
- les paramètres de quantification.

- l'index de trames de références.
- les MVs.
- les coefficients transformés quantifiés.

Ces éléments sont codés par le codage exponentiel de Golomb [75].

2.2.7.2. Codage arithmétique CABAC

L'efficacité du codage entropique peut être plus améliorée si le codage arithmétique CABAC est utilisé [7] [76]. Par rapport au CAVLC, le CABAC garantit en général une réduction du débit binaire de 10 à 15% lors du codage de signaux TV d'une même qualité. Le codage CABAC est plus efficace que CAVLC pour des probabilités de symbole supérieures à 50%, puisqu'il permet à un symbole d'être représenté avec moins d'un bit.

Les codes adaptatifs réduisent l'inefficacité des statistiques non stationnaires de symbole provoquées par la disparité entre les longueurs statiques de mot-code et les probabilités du symbole qui changent en raison du débit binaire, du type de mouvement à la source et d'autres facteurs.

2.2.8. Filtre de déblocage

Un défaut du codage axé sur le bloc est la visibilité de la structure en blocs. Les bords sont en général reconstitués avec moins de précision que les pixels intérieurs : la pixellisation est l'un des artefacts les plus visibles des méthodes actuelles de compression. Pour cette raison, la norme H.264/AVC définit un filtre de «déblocage» adaptatif en boucle où la puissance du filtrage est contrôlée par les valeurs de plusieurs éléments syntaxiques [77] [78]. La pixellisation est réduite sans affecter outre mesure la clarté du contenu et la qualité subjective est considérablement améliorée. En même temps, le filtre réduit le débit binaire de 5 à 10% tout en produisant la même qualité objective que la vidéo non filtrée.

2.3. Profils de la norme

Tous les outils de la norme ne sont pas indispensables pour une application donnée. Pour limiter la complexité des décodeurs, des groupes d'outils appelés profils ont été définis. Pour être conforme à la norme, un décodeur peut ne supporter que les outils d'un profil donné (Figure 2.13).

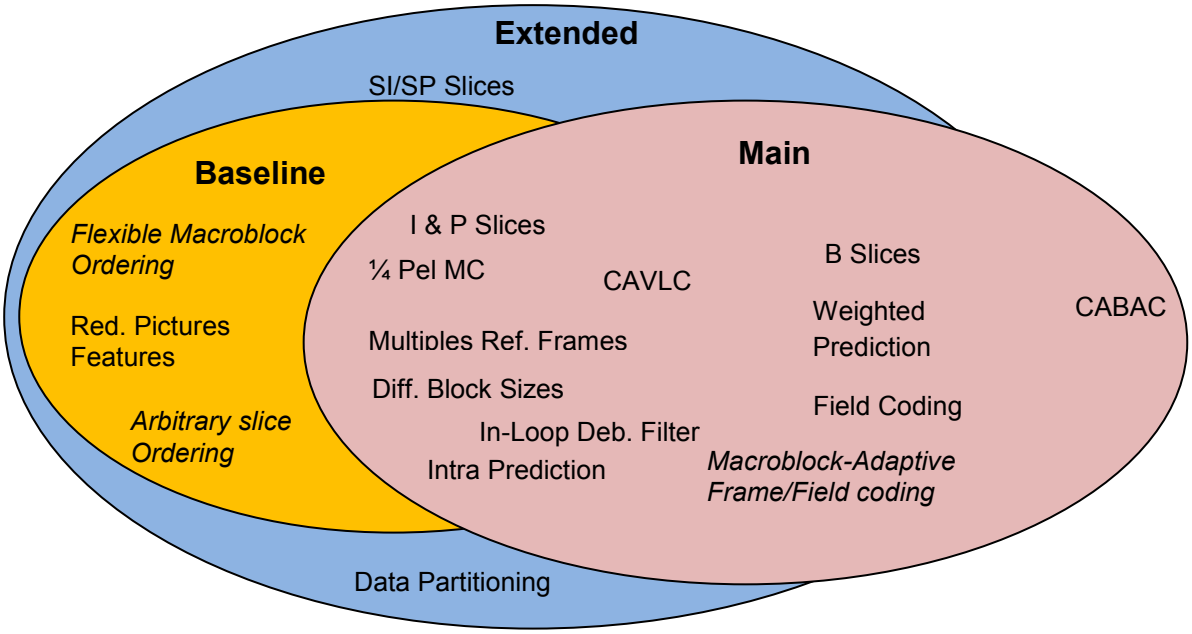


Figure 2.13 : Les profils de la norme H.264/AVC [7].

Trois profils ont été introduits initialement dans la norme [7]. « Baseline profile » et « main profile » ont été destinés respectivement aux applications à bas-coût (mobiles, vidéo-conférence) et aux applications grand public de diffusion et de stockage. Ce dernier a perdu de l'importance quand le « High profile » a été ajouté avec le même objectif. « Extended Profile » a été introduit pour la diffusion en flux (streaming). Les autres profils sont destinés à la production et aux applications professionnelles. En 2004, l'amendement FREXt (Fidelity Range Extensions initialement destiné aux environnements studio, définit de nouveaux outils avec quatre profils additionnels (High, High 10, High 4 :2 :2 et High 4 :4 :4) [79]. En plus des profils qui restreignent les outils, des niveaux (levels) ont été définis afin de

borner la puissance de calcul et la mémoire nécessaire pour les décodeurs. Un niveau spécifie des bornes en termes de taille d'image, de fréquence d'image et de débit compressé.

Nous avons vu dans le paragraphe précédent qu'il est devenu très aisé de gérer les données volumineuses en termes de stockage et de transmission grâce à l'évolution des techniques de codage vidéo. La norme H.264/AVC introduit des outils de codage outrepassant amplement les performances des standards précédents comme MPEG-2 ou MPEG-4 Partie 2. Néanmoins, cette évolution technologique a accéléré la piraterie sur la propriété intellectuelle, l'intégrité et la traçabilité du contenu. En effet, différentes techniques de tatouage ont été proposées pour les codecs vidéo antérieurs, mais peu de travaux ont été effectués pour la dernière norme de compression vidéo H.264/AVC. Le but du paragraphe suivant est de présenter dans les grandes lignes les différentes applications de tatouage numérique dédiées pour protéger le contenu H.264/AVC.

2.4. Application du tatouage numérique pour le contenu vidéo H.264/AVC

Les applications les plus répandues pour lesquelles le tatouage numérique a apporté une solution ou une partie de solution pour protéger le contenu H.264/AVC peuvent être classées selon trois domaines d'insertion: domaine spatial (pixel) où l'insertion est effectuée avant la compression, domaine transformé qui consiste à effectuer l'insertion au cours de la compression et domaine compressé où l'opération d'insertion est opérée dans le flux binaire (*.264) à la sortie du décodeur H.264/AVC [80] [81].

Dans le domaine pixel la marque est insérée dans la source vidéo par une simple addition/substitution des bits de certaines positions des pixels sélectionnés. Le principal avantage de l'utilisation de ces techniques est qu'elles sont conceptuellement simples à comprendre et la complexité de leur implémentation est faible. Ce qui favorise leur implémentation en temps réel. Mais ces techniques font généralement défaut dans la sensibilité aux attaques géométriques et à la

compression sans perte. C'est pourquoi, les méthodes de tatouage vidéo avant la compression sont peu développées car elles doivent être elles-mêmes robustes à la compression. Dans la littérature, la méthode de Pröfrock et al. [82] fut l'un des premiers travaux proposés pour insérer la marque d'une façon imperceptible dans les parties pertinentes et visibles de la vidéo. Ceci est réalisé par la quantification du centre de gravité normalisé NCG (Normed Centre of Gravity) des frontières des objets. Pour les deux autres domaines d'insertion à savoir l'insertion au cours (domaine transformé) et après la compression, les techniques de tatouage développées dans la norme H.264/AVC peuvent être classées selon l'application visée (figure 2.14) :

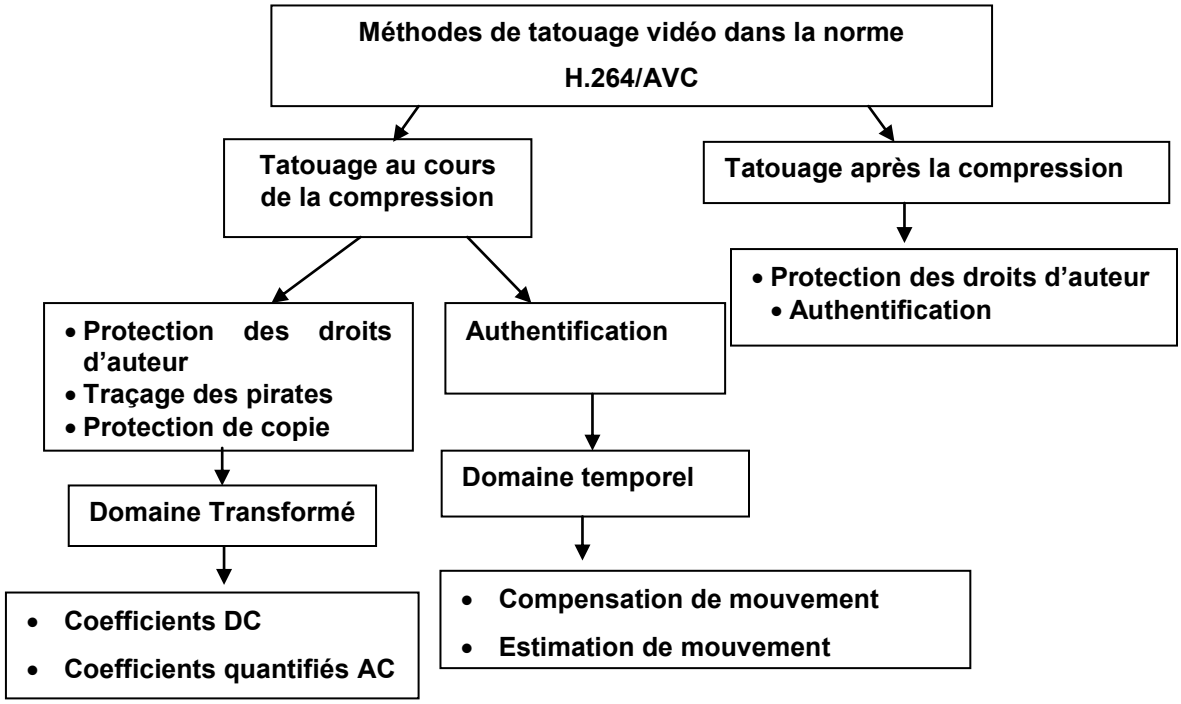


Figure 2. 14 : Classification des méthodes de tatouage dans la norme H.264/AVC.

- **Protection des droits d'auteur** : classiquement, le tatouage utilisé est le tatouage robuste qui permet de protéger la vidéo, même lorsque celle-ci est diffusée. L'objectif est d'incruster une information dans la vidéo source, typiquement le copyright du propriétaire, afin de prévenir toute revendication frauduleuse de propriété. Cette signature ne doit être connue que de la

personne ou de l'organisme de tatouage. Elle dépend donc d'une clé secrète, qui permet son insertion et sa détection.

- **Authentification de données** : L'objectif est de détecter toutes modifications éventuelles des données, afin de pouvoir certifier si celles-ci ont été modifiées ou non. On rencontre ici une problématique de contrôle de documents. Ce qui peut être obtenu avec un tatouage fragile dont le principe consiste à insérer une marque ou logo binaire (généralement prédéfinie et indépendante des données à protéger) dans la vidéo d'origine de telle sorte que les moindres modifications apportées à la vidéo se reflètent également sur la marque insérée. Pour vérifier l'intégrité de la vidéo, il suffit alors de vérifier localement la présence de cette marque.
- **Erreurs de transmission du flux H.264/AVC** : Une autre application de tatouage visée dans la norme H.264/AVC est celle qui correspond à la détection d'erreurs dans la transmission de flux vidéo compressés [83] [84] [85]. En effet, les vidéos compressées sont moins robustes aux erreurs de transmission, qui peuvent se propager sur les images prédites, donnant lieu à des artefacts gênants (par exemple la structure des frames I B P des schémas de compression du type H.264/AVC). Pour compenser ces imperfections, les protocoles de contrôles basés sur les codes correcteurs d'erreurs (ECC) dont le principe repose sur la composition de messages redondants à l'aide d'un certain nombre de symboles de contrôle sont employés [86]. Le principal problème des systèmes basés sur les codes ECC réside dans l'augmentation considérable de la taille des données. En contrepartie, le système de tatouage dégrade légèrement la qualité de la vidéo, sans que cela soit perceptible. Ces systèmes sont donc utilisés pour la détection d'erreurs associés à des systèmes de correction d'erreurs.
- **Traçage de traîtres (fingerprinting ou traitor tracing)** est une application destinée à tracer les copies légales d'un contenu en insérant un identifiant propre au possesseur d'une copie [87] [88] [89]. A la suite d'une transaction entre un vendeur et un/des acheteur(s), le traçage de traîtres permet au vendeur de déterminer le/les acheteur(s) qui redistribue(nt) illégalement le

produit. Lors de la vente d'un film, le vendeur intègre un mot de code qui lui permettra d'identifier l'acheteur à partir de la vidéo. Si le vendeur découvre que son film est redistribué illégalement, la technique de traçage de traîtres doit lui permettre d'identifier le/les acheteur(s) qui sont responsables de cette redistribution. L'approche de traçage de traîtres doit fournir un mécanisme d'accusation sûre qui permet au vendeur d'engager des sanctions ou des poursuites judiciaires à l'encontre du/des fraudeur(s).

- **Protection de la copie :** dans cette application, le tatouage consiste à insérer une marque afin d'interdire la copie du contenu. Avant toute opération de copie, le dispositif permettant cette action analysera le contenu ; si une telle marque est détectée, la copie est refusée [90] [91].

Les méthodes de tatouage appliquées pour les applications citées au-dessus sont réalisées dans la norme H.264/AVC au cours et après la compression selon le type de la marque à insérer : insertion d'une marque robuste ou semi fragile dans le modèle transformé de la norme pour l'application des droits d'auteur, traçages des pirates et protection de copie vidéo (modèle regroupant respectivement les opérations de la transformation DCT 4×4 et l'opération de quantification) et l'insertion d'une marque fragile dans les MVs pour l'application d'authentification.

2.5. Conclusion

Nous venons de présenter dans ce chapitre le principe général de fonctionnement du codeur H.264/AVC et les principaux modules le constituant. Dans le dernier paragraphe du chapitre, nous avons cité les applications les plus considérées pour protéger le contenu vidéo compressé par la norme. Les recherches effectuées dans le cadre de la protection du contenu vidéo H.264/AVC visent essentiellement à améliorer les performances en robustesse, capacité d'insertion et sécurité des méthodes développées jusqu'à présent. Pour cela, nous nous sommes inspirés de certains procédés issus du traitement d'image et les avantages des approches réalisées afin d'aboutir à un système de protection des droits d'auteur

assurant un maximum de robustesse avec une plus grande quantité d'information insérée et un système d'authentification du contenu H.264/AVC efficace satisfaisant le maximum de critères d'authentification tels que la sensibilité l'invisibilité et la préservation des bits à transmettre après l'insertion. Dans les chapitres suivants, nous allons présenter en détail les systèmes proposés.

CHAPITRE 3

PROTECTION DES DROITS D'AUTEUR DANS LA NORME H.264/AVC

3.1. Introduction

Le développement des technologies de capture et de transmission d'images et de vidéos numériques a ouvert de grandes perspectives et possibilités de création et de manipulation des contenus visuels à la fois sur le plan scientifique et artistique. De là, le besoin d'un échange d'information sécuritaire entre les personnes a fait naître la cryptologie, qui désigne les techniques de chiffrement pour rendre un message incompréhensible lors de sa transmission. Néanmoins, une fois décryptés, le contenu numérique ne possède aucune protection. A cet effet, le tatouage numérique a été introduit comme une technique permettant de protéger le contenu numérique. Pour la protection du droit d'auteur du contenu vidéo compressé par la norme H.264/AVC, l'étude est portée sur les algorithmes de tatouage vidéo robuste. Depuis la standardisation de la norme de compression H.264/AVC en 2003, les recherches de procédés de tatouage numérique pour protéger le contenu H.264/AVC n'ont pas cessé. Jusqu'à présent il n'existe pas de méthodes suffisamment robustes pour protéger les droits d'auteur ou copyright dans la norme H.264/AVC. Tirant partie des avantages des méthodes développées dans la littérature, une méthode est proposée dans cette thèse afin d'augmenter les performances des résultats du point de vue capacité d'insertion et de remédier à la faiblesse de ces méthodes en termes de robustesse face aux traitements d'images usuels.

3.2. Etat de l'art des méthodes de protection des droits d'auteur dans la norme H.264/AVC

Comme nous l'avons déjà évoqué, la plupart des contenus multimédia (images, son, vidéos, etc.) sont stockés et échangés sous une forme compressée et le plus souvent la compression se fait avec perte. Si le tatouage est effectué avant l'étape de compression, le signal de tatouage subit une première dégradation (attaque) alors que le média n'a pas encore été distribué. C'est la raison pour laquelle la majorité des techniques de tatouage numérique pour la protection des droits d'auteur dans la norme H.264/AVC se font conjointement avec la compression. L'insertion s'effectue dans le domaine fréquentiel au cours de la compression. Généralement, les techniques où l'insertion s'opère dans le domaine temporel sont dédiées à l'authentification du contenu. Le domaine fréquentiel dans la norme H.264/AVC a recours aux modules de la transformation DCT 4×4 et de la quantification. L'insertion est effectuée sur les coefficients transformés quantifiés AC ou au niveau des coefficients continus quantifiés DC.

Avant la phase de quantification, l'opération d'insertion consiste à modifier les coefficients DCT de luminances en prenant soin d'avoir un paramètre de puissance d'insertion calculé en fonction du paramètre de quantification choisi par l'utilisateur pour effectuer la compression. Les auteurs Golikeri et al. [93] ont proposé d'insérer 1 bit par MB, en utilisant un masque psycho-visuel et une approche mixant l'étalement de spectres et la quantification afin de minimiser les distorsions entre les MBs. Cette approche est parmi les plus robustes approches adaptées au flux vidéo H.264/AVC. La première méthode de tatouage dans la norme H.264/AVC opérant après l'opération de quantification est proposée par Qiu et al. [94]. L'insertion est effectuée au cours du processus de codage et la détection au cours du processus du décodage. Deux types de marques sont insérées (figure 3.1): une marque robuste W_R est insérée dans les blocs DCT 4×4 et la seconde marque fragile W_F est insérée dans les MVs. La méthode proposée peut réaliser conjointement deux applications, à savoir, la protection de copyright et l'authentification du contenu H.264/AVC. Pour le tatouage robuste, la marque est insérée en changeant les coefficients AC quantifiés

des blocs de taille 4×4 de la luma des images de type I. Afin de surmonter l'attaque de re-compression, la marque est quantifiée. Un seul coefficient AC quantifié appartenant aux hautes fréquences le long des positions de la diagonale est choisi pour l'insertion. Ainsi, le coefficient AC est remplacé par le coefficient tatoué. L'algorithme ainsi présenté n'est pas robuste face aux attaques de traitements usuelles.

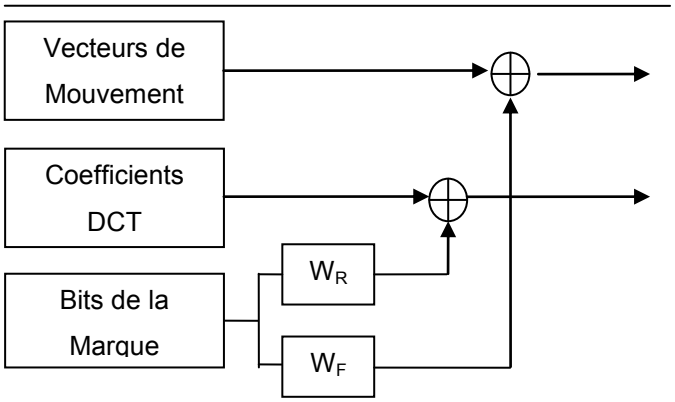


Figure 3.1 : Insertion hybride dans les coefficients DCT et Vecteurs de mouvement de Qui et al. [94].

Noorkami et al. [95] ont proposé une méthode de faible complexité, basée sur la sélection aléatoire des coefficients quantifiés AC. La sélection est sous le contrôle d'une clé générée à partir d'une combinaison d'une clé publique extraite de certaines fonctionnalités du MB et une clé secrète détenue par le propriétaire du copyright (figure 3.2).

Le schéma a prouvé sa robustesse face aux attaques d'auto-collusion, cependant la capacité d'insertion est limitée. Augmenter la capacité d'insertion était l'objectif de l'approche proposée par Tian et al. [96]. Les auteurs ont amélioré la méthode en sélectionnant uniquement les coefficients AC quantifiés appartenant au mode Intra_4×4, puis ces coefficients sont modifiés selon le codage CAVLC. Pour améliorer leur approche proposée dans le premier schéma [95], Noorkami et al. proposent un second schéma [97] qui consiste à insérer les bits de la marque dans les trames I en mesurant l'intensité de mouvement afin d'éviter l'insertion dans les

régions ayant beaucoup de mouvement. Leur but est d'augmenter la capacité d'insertion tout en limitant la distorsion de la qualité visuelle de la vidéo tatouée. Dans les deux procédés d'amélioration, la capacité est améliorée en entrainant une augmentation du débit et une dégradation de la qualité visuelle.

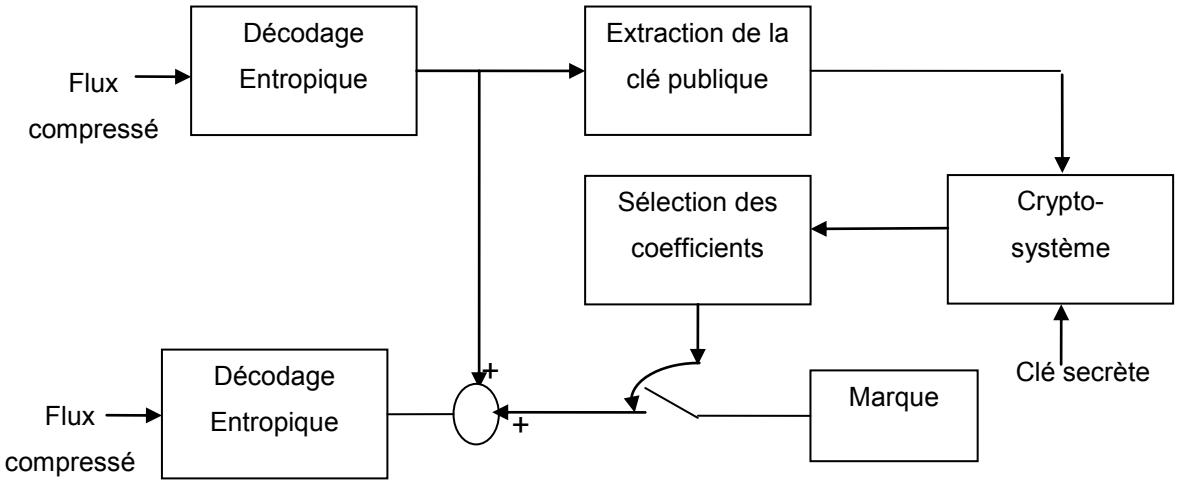


Figure 3.2 : Shéma d'Insertion robuste face d'auto-collusion de Noorkami et al. [95].

D'autres auteurs ont proposé de prétraiter la marque en niveaux de gris avant son insertion dans le contenu H.264/AVC. Le but du prétraitement est de diminuer la redondance dans l'insertion de la marque. Zhang et al. [98] [99] ont proposé deux méthodes de protection des droits d'auteur basées sur le prétraitement de la marque avant son insertion. Les deux méthodes sont similaires dans le processus d'insertion, mais diffèrent dans le traitement de la marque. L'insertion des bits de la marque est effectuée dans les coefficients AC des blocs 4×4 des MBs appartenant aux trames I. Dans le schéma proposé dans [98], la marque à insérer est transformée dans le domaine fréquentiel. La marque est subdivisée en blocs de taille 4×4 puis pour chaque bloc la DCT 4×4 est appliquée. Les coefficients ainsi obtenus subissent un balayage en zig-zag et sont rangés dans un vecteur. Pour atteindre une plus grande efficacité, un masque leur est appliqué pour sélectionner les 6 premiers coefficients de hautes fréquences et écarter les autres. Les six coefficients sont ensuite binarisés et transformés en séquence bipolaire pour augmenter la robustesse. La méthode a

prouvé sa robustesse face à certains traitements tels que le transcodage, les filtrages gaussien et circulaire, mais elle a conduit à des distorsions dans la qualité visuelle de la vidéo tatouée. Pour améliorer la première technique, Zhang et al. [99] ont proposé un autre schéma de prétraitement de la marque. Cette amélioration consiste à introduire une classification des différents alphabets (26 caractères) en plus des 10 chiffres numériques constituant la marque. Pour les deux techniques présentées, il est extrêmement difficile d'effectuer l'insertion et l'extraction en temps réel.

Une autre catégorie d'insertion consiste à cacher les bits de la marque dans les coefficients DC. Mohammad Ali et al. [100] ont proposé un schéma basé sur la stabilité de la valeur des coefficients DC dans les blocs 4×4 transformés. La marque cryptée est insérée dans les coefficients de hautes fréquences de quelques coefficients DC consécutifs appartenant aux trames aléatoirement sélectionnées. L'ensemble de valeurs aléatoires utilisées pour sélectionner les valeurs DC est considéré comme une clé. Les résultats expérimentaux de l'approche montrent que la technique est invisible avec un effet minimal de distorsion sur la qualité. En outre, étant donné que seuls les coefficients DC sont utilisés pour contenir la marque, la marque insérée est résistante seulement à la compression.

Lu et al. [101] ont proposé un tatouage aveugle pour protéger le contenu H.264/AVC en se basant sur la polarité du bloc et l'indice de modulation. La polarité du bloc est employée pour sélectionner les blocs DCT 4×4 appropriés à l'insertion afin d'assurer le critère d'invisibilité. Elle est déterminée sur la base du coefficient DC quantifié non nul dans chaque bloc 4×4. L'indice du bloc est l'activité du bloc représentée par la somme de l'amplitude des coefficients DC quantifiés. L'indice de modulation du bloc opère selon la polarité des blocs et la nature de la marque (niveaux de gris ou binaire) à insérer. L'insertion est effectuée par l'indice de modulation, de telle façon à modifier légèrement les valeurs des coefficients AC quantifiés et par conséquent, forcer l'activité du bloc à être quantifiée dans une région spécifique. Leur schéma est robuste à la compression H.264/AVC avec un taux de

compression supérieur à 40: 1 dans les trames I. Cependant, la technique nécessite la décompression du flux pour insérer la marque.

En plus de l'insertion de la marque dans la caractéristique DCT 4x4 entière offerte par la norme, d'autres caractéristiques telles que le codage CABAC et CAVLC sont prises en compte par certains auteurs pour le tatouage du contenu vidéo au cours de la compression H.264/AVC. Hu et al. [102] ont présenté une approche d'insertion de données basée sur la modification des modes de prédiction intra 4×4 en cadrant ces modes avec les bits de la marque. Afin d'améliorer la capacité d'insertion et contrôler le flux binaire à la sortie du codeur Yang et al. [103] ont proposé d'insérer chaque deux bits de la marque dans un groupe contenant trois blocs intra 4×4 vérifiant les exigences d'insertion. Chaque deux bits de la marque sont modulés aux modes de prédiction des blocs du groupe et un seul bloc des trois est nécessaire pour changer son mode de prédiction. Dans le processus d'insertion, la sécurité des données secrètes est assurée par chiffrement d'une part, et par la sélection des blocs d'insertion qui est commandée par une clé privée, d'autre part. L'algorithme a assuré une bonne qualité visuelle (PSNR élevé) avec une légère augmentation du flux binaire après compression. Cependant, ces méthodes nécessitent des connaissances préalables des modes de prédiction de la norme et ne peuvent pas être utilisées dans les applications en temps réel. Kapotas et al. [104] ont proposé un schéma d'insertion basé sur la sélection de la taille des blocs dans les modes de prédiction Inter. Kim et al. [105] ont inséré un bit de la marque dans le bit signe du train binaire des 1 (séquences binaires de 1) du contexte de codage adaptatif à longueur variable (CAVLC) du flux H.264/AVC. L'inconvénient de cette dernière approche est la connaissance au préalable du codeur entropique utilisé au cours de la compression. Dans le cas où le codeur CABAC est utilisé pour le codage entropique, l'algorithme de tatouage ne peut être appliqué.

Toutes les méthodes susmentionnées montrent clairement que les algorithmes cités dans la littérature n'ont pas atteint toutes les performances nécessaires du système de protection des droits d'auteur du contenu vidéo H.264/AVC en termes de robustesse, invisibilité et sécurité. Notre objectif dans ce chapitre est de développer

un système de protection des droits d'auteur du contenu vidéo H.264/AVC assurant un maximum de critères de robustesse, invisibilité, sécurité et capacité d'insertion élevée. Dans cette optique, une méthode d'insertion d'un logo du propriétaire du contenu vidéo opérant au cours de la compression H.264/AVC est développée.

3.3. Méthode de tatouage robuste proposée

Un schéma d'insertion de logo du propriétaire du contenu vidéo opérant dans le domaine fréquentiel de la norme est proposé [106]. L'approche développée consiste à traiter la marque avant de l'insérer. Les bits de la marque sont insérés dans le domaine DCT de la norme, en changeant les valeurs des coefficients AC quantifiés appartenant au mode de prédiction Intra_4×4 de la composante de luma des trames I. La méthode proposée suit le même principe d'insertion présenté par Qiu [94], la différence réside dans le traitement de la marque et la localisation des blocs d'insertion. Les trames I sont choisies pour l'insertion car ce type de trames est crucial pour la reconstruction du signal vidéo. Aussi, les trames P et B sont fortement compressées par la compensation de mouvement, ce qui conduit à une faible capacité d'insertion dans ces trames.

Bien que la taille 4×4 des blocs rende l'insertion de la marque plus sensible aux attaques ou au transcodage [98], l'insertion de la marque dans notre système de protection de copyright est effectuée en prenant en compte le mode de prédiction Intra_4×4. La marque n'est pas insérée dans les MBs codés en mode de prédiction Intra_16×16 pour deux raisons [98]. La première est que le mode de prédiction Intra_16×16 est utilisé pour les régions homogènes de la trame et l'insertion dans ces régions cause la dégradation de la qualité visuelle de la vidéo. La seconde raison est due à l'application de la transformée de Hadamard qui décorrèle fortement les coefficients DC de telle façon à rendre la majorité de ces coefficients nuls.

La marque avant d'être insérée est transformée en valeurs réelles en moyennant la transformation de Wash Hadamard puis binarisée selon une table de conversion. Le choix de la transformation de Hadamard pour le prétraitement de la

marque est motivé par le fait que Hadamard-2D a été utilisée avec beaucoup de succès pour le traitement d'images en général et pour la compression d'images et vidéo en particulier. D'autre part, nous voulons montrer l'efficacité de l'application de la transformation de Hadamard dans le prétraitement de la marque pour l'insertion afin de comparer les résultats expérimentaux obtenus aux résultats présentés dans la littérature.

Le processus de tatouage est considéré comme une tâche de communication composée de trois étapes principales (figure 3.3):

- le prétraitement de la marque,
- l'insertion,
- l'extraction et le post-traitement de la marque.

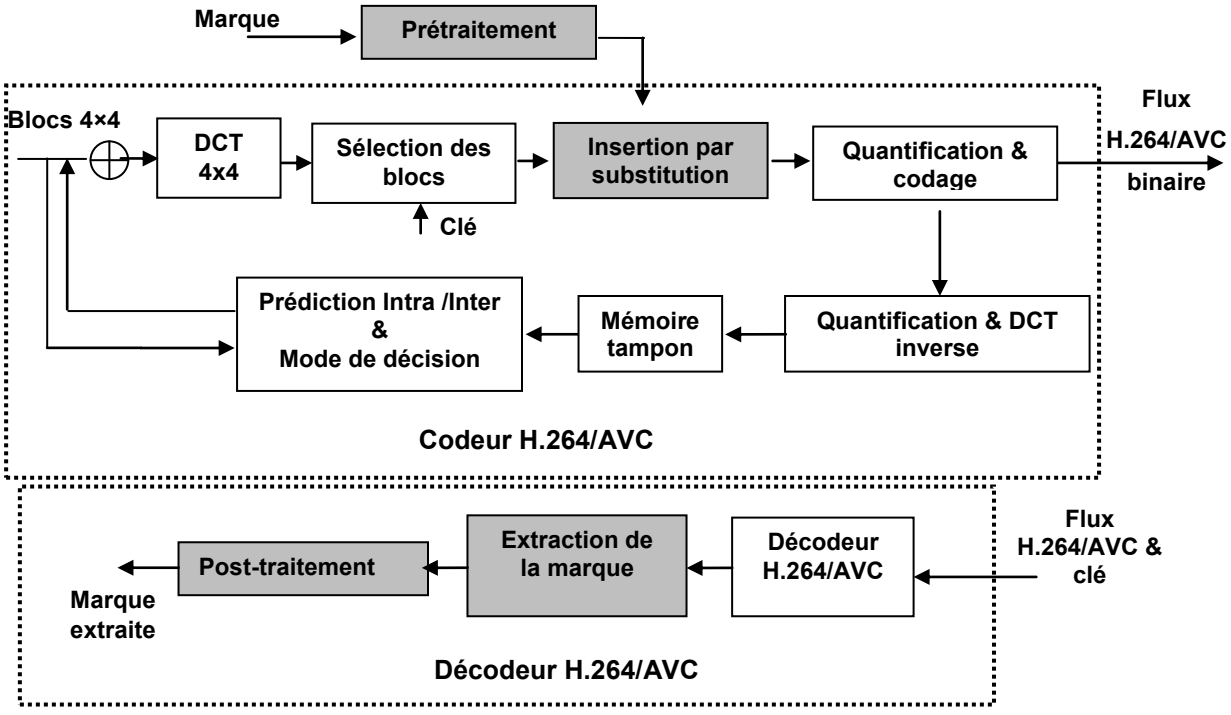


Figure 3.3 : Schéma proposé du système de protection de copyright dans la norme H.264/AVC.

3.3.1. Processus de prétraitement de la marque basé sur Hadamard

Le prétraitement de la marque dans la méthode proposée est effectué en deux étapes :

- Transformation de Wash-Hadamard: la marque est tout d'abord décomposée en blocs non chevauchants de taille 4×4, désignés par *M(x,y)* (0≤x, y≤3). Puis chaque bloc *M(x,y)* est transformé par l'équation suivante:

$$M'(x,y) = HM(x,y)H^{-1} \tag{3.1}$$

Où H représente la transformation de Wash- Hadamard 4×4 utilisée dans la norme H.264/AVC [68].

Les valeurs résultantes de la transformée de Hadamard sont des valeurs réelles. Pour les convertir en valeur binaire 0 ou 1, nous avons utilisé une table de conversion.

Le choix de la transformation de Hadamard pour le prétraitement de la marque peut être justifié par le fait que Hadamard-2D a été utilisée avec beaucoup de succès pour la compression d'images et vidéo [107] [108] et le tatouage d'images [109] [110] [111]. Contrairement aux autres transformées bien connues telles que la transformée de Fourier discrète (TFD) et la DCT, les éléments des vecteurs de base de la transformation de Hadamard ne prennent que les valeurs +1 et -1. Par conséquent, la transformation de Hadamard est bien adaptée pour les applications de traitements numériques d'images où la simplicité de calcul est nécessaire. D'autre part, nous voulons démontrer l'efficacité de l'application de la transformation de Hadamard dans le prétraitement de la marque pour l'insertion afin de comparer ses performances par rapport aux résultats obtenus par l'utilisation de la DCT pour le prétraitement de la marque [98] [99].

3.3.2. Processus d'insertion

La marque insérée contient des informations liées à l'utilisateur (droits d'auteur) afin de repérer d'où provient la fuite si le contenu est changé. Nous abordons ainsi une des principales contraintes liées à l'ajout d'un tatouage : sa robustesse face à toute transformation du contenu (compression, filtrage, etc.). En effet, la marque ne doit pas pouvoir être effacée. De plus, cette marque ne doit pas détériorer le contenu et elle ne doit pas être perceptible par l'utilisateur. La dernière contrainte qui a pris de plus en plus d'importance est la sécurité et le choix d'une clé secrète, pour l'insertion et l'extraction afin que la localisation du message caché soit uniquement accessible aux possesseurs de la clé. Pour cela, l'approche du tatouage robuste développée doit satisfaire les exigences fondamentales, à savoir : la sécurité d'insertion, l'invisibilité, la capacité d'insertion et la robustesse.

Par rapport à la méthode développée dans [94] où la capacité d'insertion est de 99 bits dans une trame de type I, dans l'approche proposée, notre objectif est d'augmenter cette capacité vu que la vidéo est un media contenant beaucoup de texture donc une grande quantité d'information peut être insérée sans altérer la qualité visuelle de la vidéo. Dans l'approche adoptée, l'insertion se fait dans les valeurs des coefficients résiduels AC de la luminance des blocs de type Intra_4×4. Le problème qui se pose alors est comment insérer une grande capacité de bits (la marque) dans les composantes significatives du spectre d'une façon sécurisée et invisible ?

Pour sécuriser le processus d'insertion, la position du coefficient d'insertion *G(i)* dans le bloc de type Intra_4×4 dans un MB donné est sélectionnée aléatoirement via un générateur à congruence linéaire (GCL), en anglais Linear Congruential Generator (LCG) [112]. C'est l'algorithme le plus utilisé pour produire des nombres aléatoires depuis qu'il a été inventé en 1948 par D. H. Lehmer. Il est déterminé par quatre valeurs entières définies dans le tableau 3.1 [113].

Tableau 3.1 : Les quatre paramètres entiers déterminant le générateur à congruence linéaire (GCL).

Paramètres	Signification	valeurs
M	le modulo	M > 0
m	le multiplieur	0≤ a <m
cr	l'incrément	0≤ c <m
G(i-1)	la valeur initiale	0≤ G(0)< m

La séquence est donnée par :

$$G(i) = (m * G(i-1) + cr) \bmod M \quad (3.2)$$

Où *mod* est la fonction « reste entière » (le reste entier de la valeur entre parenthèses divisé par M). Ce générateur est désigné par *GCL (m, cr, M, G (0)).*

La formule est simple mais le choix des valeurs des trois paramètres *m*, *cr* et *M* ne doit pas être fait à la légère. Ils sont choisis afin de maximiser la période qui ne peut excéder *M* et qui est égale à 16, la taille du MB dans notre cas. *G(0)* représente la clé d'insertion et d'extraction de la marque.

Pourquoi le GCL est–il utilisé dans notre système?

D'un point de vue pratique, les générateurs GCL sont bons s'ils produisent des résultats corrects dans autant d'applications que possible. Les générateurs de nombres aléatoires sont rien de plus que des algorithmes déterministes qui produisent des nombres avec une distribution uniforme. Ces séquences qu'on appelle des nombres «Aléatoires» sont périodiques. Leur tâche n'est pas de simuler «aléatoirement », qui est une notion difficile à définir en termes de pertinence pratique, mais pour donner de bons résultats d'une simulation, et c'est pourquoi le choix des valeurs de *m* et *cr* est critique. Le générateur GCL n'est pas le meilleur générateur de nombres aléatoires, mais il reste toujours bon à condition que les paramètres soient choisis avec soin. Il est donc nécessaire de se demander comment choisir *m*, *cr* et *M* convenablement.

- **Le choix du modulo *M* :** Les GLC font intervenir un calcul modulo *M*, et donc a priori une division euclidienne, ce qui peut avoir un coût de calcul important

dans le cadre d'une utilisation fréquente du générateur. La solution la plus simple est d'utiliser un module de type $M = 2^n$.

En effet, les ordinateurs calculant naturellement en base binaire, un tel choix est tout à fait transparent pour eux, ce qui rend inutile une division euclidienne. Cependant, un tel choix présente une limite importante: les bits dits de poids faible (LSB) sont beaucoup moins aléatoires que ceux de la partie de poids fort (MSB). En effet, si *d* est un diviseur de *M*, alors la suite *Y(i)*, telle que :

$$Y(i) = G(i) \bmod d \tag{3.3}$$

Satisfait à la suite congruentielle linéaire :

$$G(i) = (m * G(i-1) + cr) \bmod d \tag{3.4}$$

En particulier, avec *d = 2k*, pour *k* fixé, compris entre 1 et *n*, on voit que les *k* chiffres de LSB ont une période maximale de *2k*, évidemment inférieure à *M*. Pour remédier à ce problème, on peut ne garder que MSB, c'est-à-dire garder les bits les plus à gauche du nombre obtenu. Si l'on tronque les *k* derniers bits, on aura alors un générateur pseudo aléatoire de nombres compris entre 0 et 2n-k-1. Dans notre cas *n = 16*, qui représente la taille du MB.

- **Choix du multiplicateur m et de l'incrément cr** : Afin de pouvoir choisir une clé G(0) sans contraintes entre 0 et (M-1), il faut chercher à maximiser la période du générateur. Or, il se trouve que les valeurs de *m* et *cr* sont connues, ce qui permet d'obtenir une période maximale (égale à *M*). La période d'un GCL est maximale si et seulement si :*cr* est premier avec *M*, *PGCD(cr, M) = 1*.

La marque insérée doit être invisible pour ne pas abaisser la valeur commerciale de la vidéo et l'utilisateur ne doit pas sentir l'existence de marque dans la vidéo. Pour cela, dans notre approche, la sélection de la position d'insertion est faite au niveau bloc en sélectionnant la position du bloc Intra_4×4 dans un MB donné et au niveau pixel en distinguant le pixel à tatouer. Au niveau bloc, l'insertion est

effectuée seulement dans des blocs de forte entropie sélectionnés parmi les modes de prédiction (3-8) du mode Intra_4×4 de la norme H.264/AVC. Ce choix découle des résultats obtenus dans la littérature [114], où les chercheurs ont montré expérimentalement que dans le MB de type Intra_4×4, le codeur sélectionne les modes (3-8) comme les meilleurs modes de prédiction lorsque les blocs ont une texture complexe et écarte les modes (0-2) qui sont choisis pour les blocs à texture uniforme.

En outre, au niveau pixel, les coefficients quantifiés AC du bloc sont balayés en zig-zag de haut en bas (figure 3.4) et le processus d'insertion est effectué sur tous les coefficients quantifiés AC à l'intérieur du bloc afin de trouver le meilleur coefficient quantifié AC qui vérifie l'invisibilité de la marque. Toujours au niveau pixel, l'amélioration de l'imperceptibilité de la marque est opérée en testant la substitution du bit de la marque avec tous les bits du coefficient AC quantifié sélectionné.

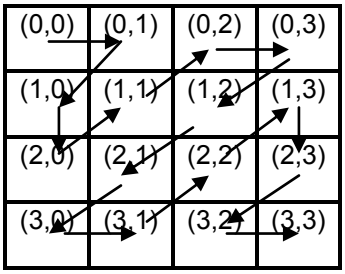


Figure 3.4 : Balayage en zig-zag des blocs de type Intra_4×4.

Note : les coefficients DCT sont représentés sur 16 bits.

Avant d'être marqué, le coefficient AC quantifié sélectionné *Xq(i,i*) est masqué afin d'effacer le bit à substituer. Cette opération est comme suit :

$$Q(Xq(i,i)) = \begin{cases} Xq(i,i)\,\&\,(0xEFFF) & si\ Xq(i,i) \geq 0 \\ -(Xq(i,i))\,\&\,(0xEFFF) & si\ Xq(i,i) < 0 \end{cases} \quad (3.5)$$

où la porte ET binaire (&) est utilisée pour effacer le bit du coefficient *Xq(i,i)*.

Le bit de la marque est inséré dans le coefficient *Xq(i,i)* de la manière suivante :

$$\overline{Xq(i,i)} = \begin{cases} Q(Xq(i,i)) \mid (0x1000) & si\ w_i = 1 \\ Xq(i,i) & ailleurs \end{cases} \quad (3.6)$$

où w_i est le bit de la marque à insérer et la barre verticale (|) représente l'opérateur OU exclusif.

3.3.3. Processus d'extraction

La détermination du droit de possession est basée sur le résultat de l'extraction de la marque qui est effectuée au cours du processus de décodage d'une manière aveugle. Le processus d'extraction est réalisé au niveau du décodeur H.264/AVC. Le flux compressé est partiellement décodé afin d'obtenir les coefficients AC quantifiés. Pour cela, la marque est extraite après le décodage entropique des coefficients AC quantifiés de la trame de type I. Les blocs Intra_4×4 dont les coefficients sont tatoués sont sélectionnés en utilisant le GCL avec la même clé *G(0)* employée au cours du processus d'insertion. Une fois les blocs sont distingués, le bit de la marque est extrait du pixel tatoué *X'(i,i)*. La reconstruction de la marque est le processus inverse de l'opération du prétraitement, les étapes de binarisation inverse et la transformation de Hadamard inverse sont appliquées respectivement. Cette étape de traitement est appelée opération de post-traitement de la marque.

3.4. Analyse et résultats expérimentaux

Le schéma de tatouage robuste est intégré dans le logiciel de référence H.264 JM-7.6 [115]. Les plus importants paramètres de la configuration du codeur sont donnés dans le tableau 3.2. Le reste des paramètres ont conservé leurs valeurs par défaut. Les tests sont effectués sur les séquences vidéo benchmark sous le format QCIF(176×144) à 372 kbits/s: telles que Claire, Table, Coastguard, Flower, Bridge-close et Container. La marque utilisée est un logo de taille 16×16 (figure 3.5).

Figure 3.5 : Logo utilisé comme une signature du propriétaire

Tableau 3.2 : Paramètres de configurations du codeur JM-7.6.

Profile	Baseline (Profile IDC=66)
Number of trames	150 pour l'ensemble des séquences tests, sauf pour la séquence Table comprenant 99 images.
Frame rate	30 fps
Source Bit Depth luma	8
Source Bit Depth chroma	8
Motion estimation	Full Search
Entropy coding	CAVLC
Search range	16
Quantization parameter (Qp)	28
Intra period	15 only the first frame is intra

Pour mesurer l'intégrité de la marque, la corrélation normalisée *NC* est utilisée. Elle mesure la ressemblance entre la marque extraite *w*' et la marque d'origine *w*. Autrement, dit, elle représente le nombre de données erronées produites entre les données binaires de w et celles de *w'*. Elle est définie comme suit [116]:

$$NC = \frac{\sum_{i,j} w_{u,v} w'_{u,v}}{\sum_{i,j} w_{u,v}^2} \tag{3.7}$$

Pour le choix de la position d'insertion, nous avons effectué l'insertion sur tous les coefficients quantifiés le long de la première diagonale et la meilleure position trouvée est la position (3.3) du bloc Intra_4×4. Le schéma d'insertion est un schéma substitutif, un bit de la marque est substitué avec un bit du coefficient *Xq(3,3)* choisi pour l'insertion. En substituant tous les bits de Xq(3,3) un à un avec le bit de la marque, nous avons constaté que le meilleur bit assurant le critère d'imperceptibilité est trouvé au cinquième bit du coefficient quantifié *Xq(3,3*). D'où la meilleure position d'insertion trouvée est au cinquième bit du coefficient AC quantifié localisé à la position (3,3) du bloc Intra_4×4. La figure 3.6 illustre la qualité visuelle de la vidéo «Claire» qui résulte des tests d'insertion effectués sur certaines positions dans le bloc

Intra_4×4 sélectionné et le tableau 3.3 regroupe les résultats du PSNR (dB) et la corrélation normalisée *NC* de l'insertion effectuée sur les différentes positions sélectionnées du bloc Intra_4×4 de la séquence Claire.

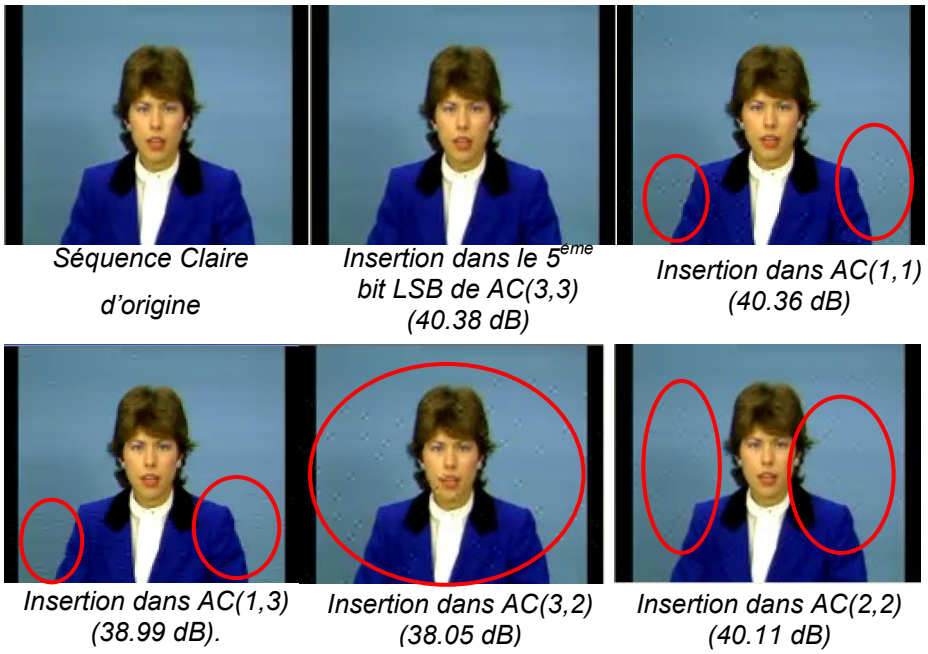


Figure 3.6 : Qualité visuelle résultante de l'insertion dans les différentes positions dans un bloc Intra_4×4 sélectionné.

La première analyse réalisée a porté sur la mesure de la qualité perceptuelle afin d'évaluer la distorsion introduite par l'opération d'insertion. L'évaluation subjective a été conduite à partir d'un panel de 10 observateurs. Celui-ci a établi que la méthode développée satisfait le critère d'invisibilité: aucune différence visuelle significative n'a pu être identifiée entre les vidéos d'origine et les vidéos tatouées. Pour une évaluation de la qualité perceptuelle objective, le PSNR qui est la métrique classique exploitée pour comparer deux images est utilisée. Celui-ci tente de déterminer le niveau de distorsion d'une image compressée par rapport à sa source. Il est considéré comme une mesure très indicative qui dépend grandement du format de compression choisi ou des particularités du codeur. Dans la norme H.264/AVC, la

métrique PSNR est directement délivrée par le standard dans le fichier *log.dat* créé au cours du processus de codage.

Tableau 3.3 : PSNR (dB) et corrélation normalisée obtenus à partir de l'insertion effectuée sur les différentes positions dans un bloc Intra_4×4 sélectionné.

Position (x,y)	PSNR (dB)	*NC*
0,1	40.82	0.75
1,0	40.51	0.67
2,0	40.36	0.5
1,1	40.36	0.7
0,2	40.55	0.97
0,3	37.08	0.33
1,2	37.27	0.41
2,1	36.99	0.54
3,0	37.56	0.7
3,1	37.88	0.8
2,2	40.11	0.79
1,3	38.99	0.69
2,3	37.88	0.59
3,2	38.05	0.88
3,3	40.38	1

La figure 3.7 illustre le résultat des tests d'insertion effectués sur la vidéo « Claire » pour les différents bits du coefficient *Xq(3,3)* ainsi que les PSNR (dB) correspondants. La dégradation de la qualité visuelle est clairement montrée sur la figure.

Claire d'origine | Insertion dans le 6ème bit (PSNR = 38.24 dB) | Insertion dans le 8ème bit (PSNR = 34.93 dB) | Insertion dans le 5ème bit (PSNR = 40.38 dB)

Figure 3.7 : Vidéo « Claire » tatouée à différentes positions des bits du coefficient *Xq(3,3)* sélectionné pour l'insertion.

La figure 3.8 illustre le PSNR moyen résultant de l'insertion effectuée sur l'ensemble des images tests. L'algorithme de tatouage robuste proposé provoque une légère dégradation de la qualité visuelle de la vidéo tatouée par rapport à l'originale. Cette dégradation de qualité qui n'est pas perceptible à l'œil nu (qualité subjective) est mesurée entre 0.23 dB et 0.88 dB pour toutes les séquences tests compressées à un débit de 372 kbits/s. La figure 3.9 montre la qualité subjective des deux séquences "Foreman" et "Container" tatouées.

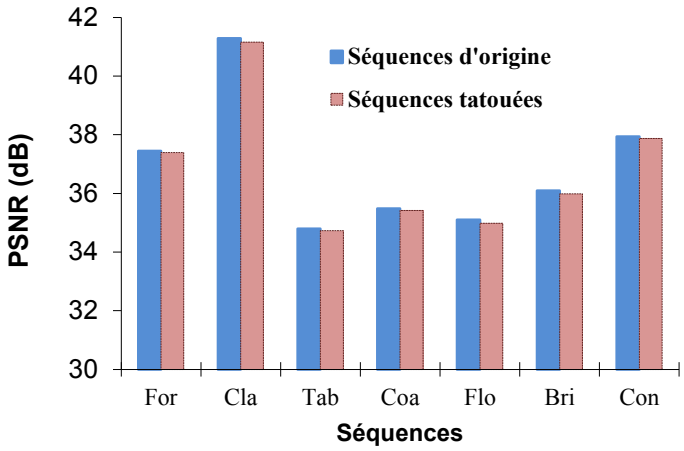


Figure 3.8 : PSNR (dB) des séquences tests tatouées et non tatouées : Foreman, Claire, Table, Coastguard, Flower, Bridge-close, Container (notées For, Cla, Tab, Coa, Flo, Bri, Con dans l'axe horizontal) à un débit de 372 kbits/s.

Trame d'origine de la vidéo Container (37.94 dB) | Trame tatouée de la vidéo Container (37.52 dB) | Trame d'origine de la vidéo Foreman (37.45 dB) | Trame tatouée de la vidéo Foreman (37.03 dB)

Figure 3.9 : Les 10ème trames d'origine et tatouées des séquences Container et Foreman (à 372 kbits/s).

La seconde analyse conduite a porté sur l'évaluation de la robustesse. La robustesse de la technique proposée est évaluée en effectuant des manipulations usuelles comme le changement de format, la compression avec pertes, le transcodage et le filtrage numérique. Deux codecs vidéo H.264/AVC et H.263 ont été appliqués pour re-compresser les séquences vidéo tatouées. Le transcodage utilisé est l'opération qui transforme la séquence du format YUV (4:2: 0) au format RVB (4:4:4). La troisième manipulation appliquée évalue la robustesse par rapport au changement de format. Cette attaque redimensionne la résolution de la vidéo du QCIF (176×144) à la résolution CIF (352×288). Les attaques communes de traitement des images appliquées sont: le filtre gaussien 5×5, le filtre passe-bas, les filtres moyen et circulaire ainsi que le bruit gaussien additif (variance = 0.001).

La re-compression par le codeur H.264/AVC a été appliquée avec différents pas de quantification *QP*. D'après les tests effectués, seuls les *QP* = [28, 32, 36], correspondant aux *QP* typiques aux applications à bas débit, conviennent pour obtenir la robustesse face à la re-compression du contenu vidéo. La figure 3.10 illustre les différents résultats obtenus des essais effectués avec différentes valeurs de *QP*. Pour la valeur de *QP* inférieure à 28, la marque n'est pas complètement retrouvée, au-delà de *QP* égal à 36, la marque est visible et elle est fortement corrélée à la marque d'origine.

Figure 3.10 : Marque extraite avec différentes valeurs de QP.

La figure 3.11 illustre respectivement la qualité visuelle des essais effectués avec différentes valeurs de *QP* sur les vidéos "Claire" et " Container " ainsi que leur PSNR. Les valeurs moyennes des PSNR sont calculées en comparant les séquences vidéo tatouées et décompressées aux séquences d'origine. La dégradation de la qualité est

clairement montrée sur la figure pour l'insertion réalisée avec la valeur de *QP* supérieure à 36.

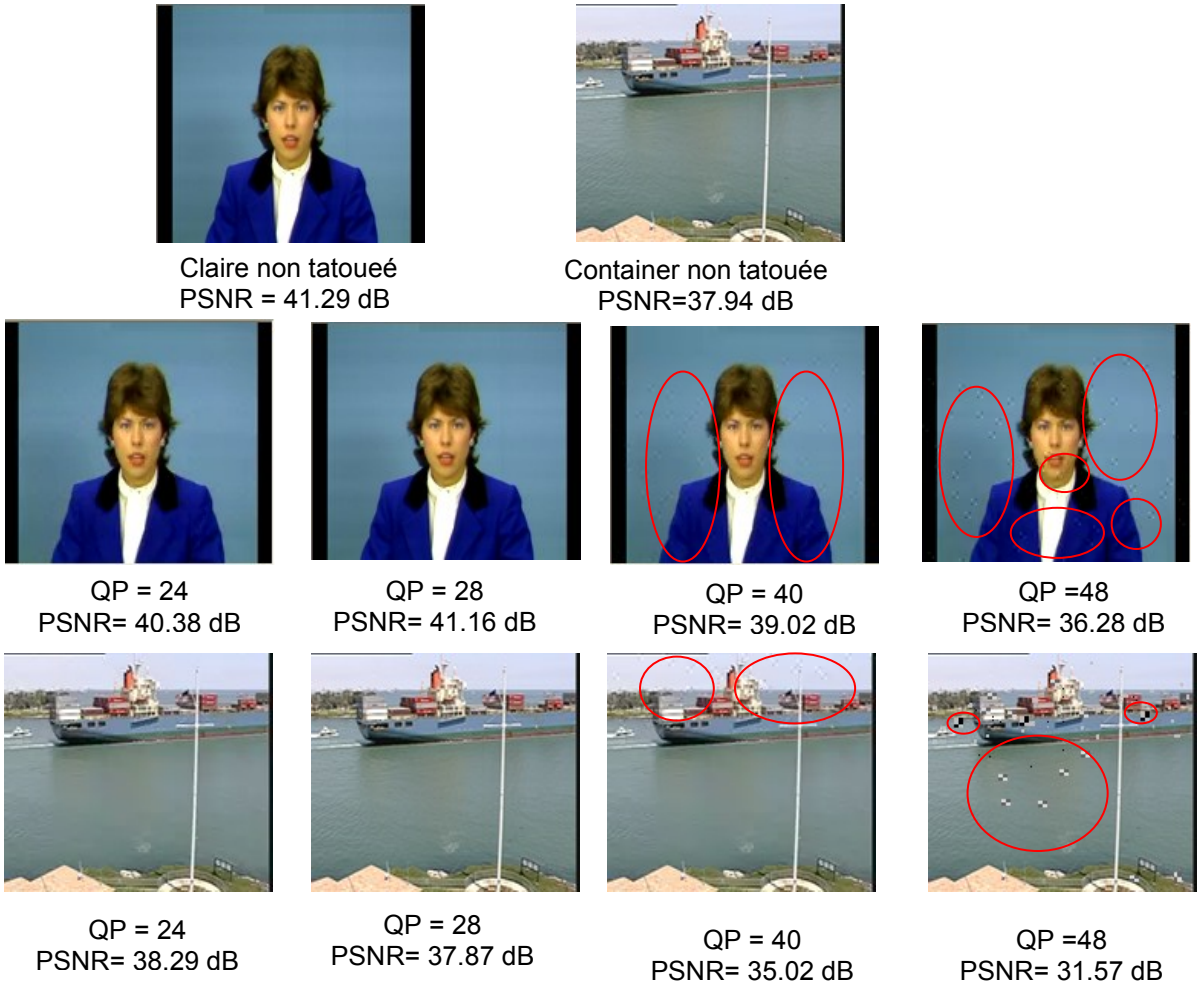


Figure 3. 11 : Les séquences Claire et Container tatouées avec différentes valeurs de QP.

Le tableau 3.4 regroupe les résultats obtenus de la qualité visuelle objective (PSNR) et le taux de corrélation de la marque extraite après l'application des traitements cités sur les séquences «Claire » et « Container ». Dans toutes les attaques analysées, la marque reconstruite après le tatouage est fortement corrélée à la marque d'origine. Les valeurs de la corrélation normalisée *NC* entre la marque d'origine et la marque extraite sont supérieures à 0.75 avec une conservation de la

qualité visuelle sauf pour l'attaque de compression par le codeur H.263 qui est de 0.22.

Tableau 3. 4 : Qualité et taux de reconstruction de la marque après les différentes attaques appliquées

Séquences tests	**Claire** **PSNR = 41.29 dB**	**Container** **PSNR = 37.94 dB**
Extraction à partir de la vidéo tatouée	 PSNR = 40.38 dB *NC*= 1	 PSNR = 37.33 dB *NC*= 1
Extraction après re-compression par le codeur H.264/AVC	 PSNR = 39.94 dB *NC*= 1	 PSNR = 37.33 dB *NC* = 1
Extraction après compression par le codeur H.263	 PSNR = 35.44 dB *NC* = 0.22	 PSNR = 35.44 dB *NC* = 0.22
Extraction après transcodage vers le format RGB (4:4:4)	 PSNR = 38.53 dB *NC* = 1	 PSNR = 36.45 dB *NC* = 1
Extraction après le filtre gaussien 5x5	 PSNR = 38.05 dB *NC* = 0.75	 PSNR = 37.05 dB *NC* = 0.8
Extraction après redimensionnement	 PSNR=38.62 dB *NC* =0.97	 PSNR=37.36dB *NC* =0.94
Extraction après filtrage circulaire	 PSNR = 38.28 dB *NC* = 0.79	 PSNR = 37.54 dB *NC* = 0.9
Extraction après l'application du bruit gaussien	 PSNR=37.36 dB *NC* = 0.75	 PSNR = 36.00 dB *NC* = 0.79

La fragilité de l'algorithme développé face à l'attaque par compression H.263 pourrait s'expliquer par le fait que, dans cette attaque, les séquences tests avant insertion sont compressées avec QP égal à 28. Dans le codeur H.263, la fourchette du paramètre de quantification est entre 1 et 31. Plus la valeur de QP est élevée, plus le débit est faible et, par conséquent, plus la qualité de la vidéo compressée est faible. Pour QP égal à 28, la marque insérée dans la vidéo est perdue. Afin de rendre l'algorithme robuste à cette attaque, il est nécessaire de trouver QP adéquat du codeur H.263 qui s'adapte avec le débit désiré pour compresser la séquence tatouée et pouvoir extraire la marque.

Un autre critère important visé par cette méthode est l'augmentation de la capacité d'insertion. Dans cette approche, une capacité d'insertion minimale de 495 (99×5) bits est obtenue en fixant le nombre minimal de blocs appropriés à l'insertion à cinq. Ainsi cinq (05) bits sont insérés dans un MB de taille 16×16. Le choix des blocs d'insertion est sélectionné aléatoirement en utilisant la fonction GCL. Si le bloc choisi n'est pas approprié pour l'insertion, à savoir, les données insérées provoquent une détérioration de la qualité de la vidéo perceptible par un observateur humain, le bloc est rejeté et un autre bloc sélectionné est testé.

Le schéma de tatouage robuste est comparé à celui présenté dans la méthode [94]. L'insertion proposée a fait augmenter la capacité d'insertion de 396 (99x5) bits au minimum tout en conservant la qualité visuelle par rapport au schéma de Qui et al. [94] qui permet d'insérer 99 bits dans une trame Intra appartenant à une séquence de résolution QCIF. Dans le cas où la vidéo à tatouer est texturée (avec beaucoup de détails spatiaux) telles que les séquences «Foreman» ou «Table», le nombre de blocs de type Intra_4×4 est considérable. Le nombre de blocs sélectionnés pour l'insertion augmente à condition de respecter le critère d'invisibilité.

D'autre part, l'approche proposée est robuste à la compression par H.264/AVC, au transcodage et traitements usuels d'images en la comparant à l'approche proposée dans [98] où la méthode n'est pas robuste face aux traitements usuels d'images. La méthode mise au point est adaptable à tout type d'images; images texturées et peu texturées tout en gardant une bonne qualité visuelle. Du point de

vue qualité visuelle, nous déduisons que notre méthode provoque une légère dégradation visuelle face aux traitements usuels par rapport à la méthode présentée dans [98] où la qualité visuelle est diminuée de façon significative. Cette dégradation est mesurée à 6.7 dB et 6.3 dB pour toutes les séquences tests texturées et codées à 396 kbits/s. En revanche, la méthode mise au point ne fait diminuer la qualité que de 0.5 dB à 2 dB, pour les mêmes attaques et pour les mêmes séquences tests. Pour les séquences tests non texturées ayant beaucoup de régions homogènes telle que «Akiyo» (séquences peu texturées), la diminution de la qualité est mesurée dans l'intervalle 2dB - 4dB sans observation de bruit visible sur les séquences.

Avec la méthode proposée, la marque extraite de la vidéo tatouée est très corrélée à la marque d'origine. La corrélation normalisée *NC* est égale à 1, alors que dans [98] [99], elle est égale à 0.9. Le tableau 3.5 présente une comparaison de *NC* entre la méthode proposée appliquée sur la vidéo "Container" et les méthodes [98] [99] utilisant un prétraitement de la marque avant l'insertion. L'algorithme développé surpasse nettement les deux méthodes proposées dans la littérature [98] [99], en termes d'intégrité de la marque. Aussi, il conduit à une capacité d'insertion considérable tout en étant robuste face aux attaques usuelles par rapport à la méthode développée dans [94].

Tableau 3. 5 : Tableau comparatif de la corrélation normalisée entre la méthode proposée et celle décrite par Zhang et al. [98] et [99].

	Corrélation Normalisée (*NC*)		
Marque extraite après les attaques testées	**Méthode proposée**	**Zhang et al. [98]**	**Zhang et al. [99]**
Sans attaques	1	0.93	0.93
Filtre gaussien 5x5	0.8	0.72	0.72
Filtre circulaire	0.9	0.70	0.87
Bruit gaussien	0.79	0.69	0.75

3.5. Conclusion

Dans ce chapitre, nous avons abordé la protection vidéo pour assurer le droit d'auteur du contenu vidéo compressé par la norme H.264/AVC. Un état de l'art des méthodes de tatouage robuste développées dans cette optique est élaboré. Cependant, toutes les méthodes évoquées ne satisfont pas tous les critères d'un schéma de tatouage robuste et sécurisé. Nous avons proposé alors une technique de tatouage robuste avec prétraitement de la marque en moyennant la transformation de Hadamard. L'approche développée assure une bonne robustesse face aux attaques de traitement d'images usuels tels que la compression par le codeur H.264/AVC, le transcodage, les filtres gaussien 5×5 et circulaire et le bruit gaussien. Elle conduit aussi à une capacité d'insertion considérable tout en maintenant une bonne qualité visuelle des vidéos tatouées.

CHAPITRE 4

AUTHENTIFICATION DU CONTENU VIDEO H.264/AVC

4.1. Introduction

La cryptographie a été une première proposition pour sécuriser des transferts du contenu vidéo. Aujourd'hui les algorithmes de chiffrement modernes, avec des clés de longueur importante, permettent d'assurer la confidentialité. Néanmoins, une fois décrypté, le contenu n'est plus protégé et il peut être distribué ou modifié malhonnêtement. Le tatouage numérique peut être une réponse à ce problème. L'insertion d'une marque dans un contenu vidéo permet de l'authentifier et de garantir son intégrité. L'authentification du contenu vidéo est de vérifier l'intégrité des images (trames) constituant la vidéo, c'est à dire de vérifier si la vidéo a subi une altération depuis sa création. Le tatouage numérique est devenu une technique prometteuse pour l'authentification du contenu vidéo en raison de ses performances exceptionnelles et la capacité de détection des altérations au niveau de la vidéo. Toutefois, de nombreux défis pour l'authentification du contenu vidéo H.264/AVC basée sur le tatouage numérique demeurent non résolus ou doivent être améliorés comme la localisation des altérations, le maintien de la qualité vidéo, l'augmentation de la taille du flux compressé, la sécurité, etc.

La deuxième problématique abordée dans cette thèse est l'authentification du contenu vidéo compressé par le standard H.264/AVC. L'état de l'art des systèmes d'authentification dédiés à la protection du contenu H.264/AVC classe les techniques élaborées en deux grandes catégories: (1) les méthodes d'authentification basées sur le contenu appelées également l'authentification par signature numérique où la marque peut être fragile ou semi-fragile selon les critères d'intégrité [117] et (2) les approches indépendantes du contenu dans lesquelles la marque est uniquement fragile [118]. L'authentification basée sur le tatouage fragile, appelée aussi

authentification stricte [119] [120] [121], est en mesure de détecter les régions où le contenu a été modifié. Le but de cette application est d'insérer dans le contenu vidéo une marque qui puisse authentifier le contenu ou apporter la preuve que le contenu de cette vidéo n'a pas été modifié depuis cette insertion. Cependant, dans ces techniques les manipulations malveillantes et classiques (bienveillantes) ne sont pas distinguées. En revanche, l'authentification basée sur le tatouage semi-fragile, appelée authentification douce [120], est robuste face aux modifications accidentelles telles que la compression, mais fragile à d'autres modifications.

Pour l'authentification stricte, les conditions générales du système d'authentification doivent être remplies lors de l'exécution. Ces exigences sont (1) sensibilité - ce qui signifie que l'approche doit être en mesure de détecter toute modification ou manipulation du contenu. Pour les algorithmes d'authentification fragiles, non seulement la modification du contenu est recherchée, mais aussi la détection de toute manipulation; (2) localisation locale – implique que le système doit être en mesure de localiser les régions modifiées dans les images de la séquence vidéo; (3) conservation du taux de bits - ce qui implique que le débit binaire doit être inchangé avant et après le tatouage; (4) imperceptibilité - impliquant que la méthode de tatouage doit maintenir la qualité de la vidéo d'origine; (5) sécurité – signifie que la marque d'authentification insérée doit résister à toute tentative de falsifications et (6) capacité d'insertion (appelée aussi charge utile de données) correspond à la quantité d'informations qui peut être insérée dans la séquence vidéo, plus la capacité est élevée, plus le système d'authentification est meilleur.

L'authentification stricte est utilisée dans les applications critiques où les modifications peuvent avoir des effets drastiques et coûteux telles que les images médicales, vidéosurveillance, etc. Pour de telles applications, afin d'assurer l'intégrité du contenu, on utilise des marques fragiles qui deviennent non détectables dès qu'une valeur des données change dans le contenu. Par exemple, dans le cadre de vérification de l'intégrité de la vidéosurveillance, les vidéos enregistrées sont tatouées et stockées. Par ailleurs, lorsque ces vidéos sont sollicitées pour aider à élucider certains actes criminels, la marque insérée est détectée pour en vérifier l'intégrité.

Dans la présente thèse, les techniques de tatouage fragiles sont considérées, pour cela, un état de l'art des techniques de tatouage fragile développées dans la littérature pour protéger le contenu H.264/AVC est abordé.

4.2. Etat de l'art des techniques d'authentification du contenu H.264/AVC

Les techniques d'authentification du contenu H.264/AVC proposées dans la littérature sont classées selon le domaine d'insertion de la marque. La marque peut être soit insérée dans le domaine compressé c'est-à-dire au cours de la compression [122] [123] [90] [124] [125], soit directement dans le flux binaire compressé de la norme H.264/AVC [126] [127] ou tout simplement insérée comme informations supplémentaires à l'entête du flux H.264 dans l'espace « amélioration d'informations supplémentaires » SEI (**S**upplemental **E**nhancement **I**nformation) alloué par le codeur [128] (figure 4.1).

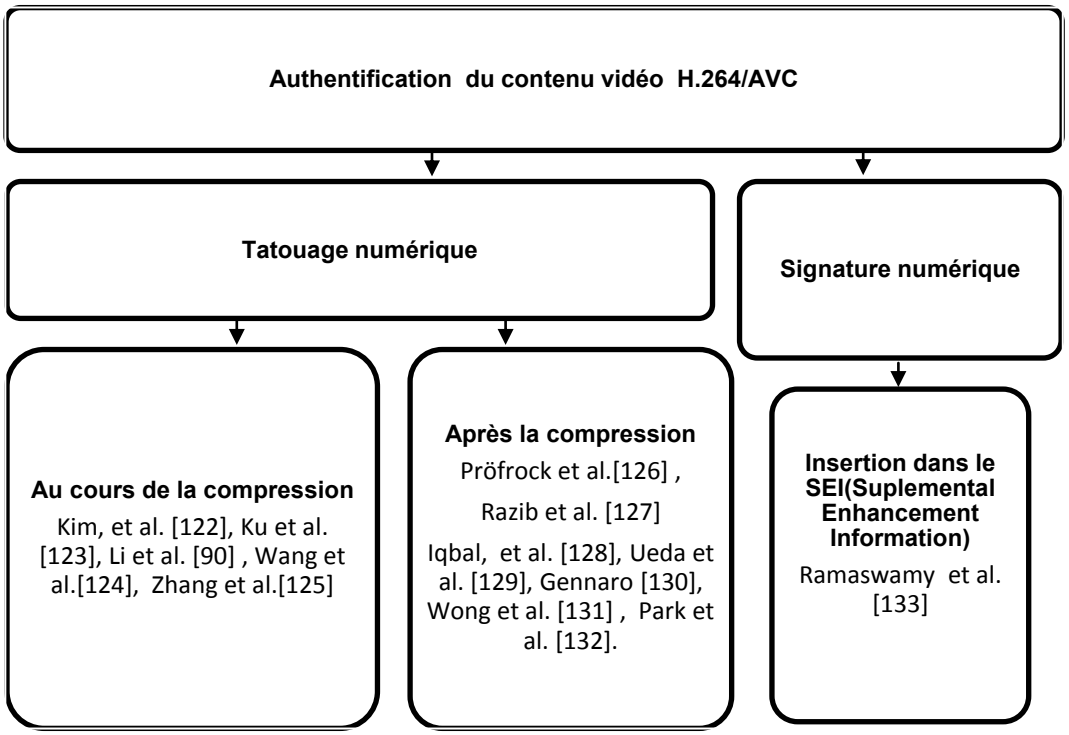


Figure 4.1 : Classification des méthodes d'authentification du contenu vidéo H.264/AVC.

Au cours du processus de compression, la plupart des systèmes d'authentification du contenu vidéo H.264/AVC prennent en compte la dimension temporelle pour insérer l'information au niveau des MVs. Qiu et al. [94] ont présenté une méthode basée sur l'utilisation des coefficients DCT pour le tatouage robuste et les MVs pour le tatouage fragile. L'authentification est indépendante du contenu H.264/AVC, ce qui signifie que la marque insérée est constituée d'une séquence binaire externe. Le choix des MVs pour l'insertion est basé sur la sélection des MVs optimaux déterminés par le coût minimal de la fonction de Lagrange. L'insertion est effectuée dans les MVs en changeant une des composantes de l'ensemble des MVs sélectionnés. Cependant, comme l'erreur de prédiction des MVs doit être codée dans le flux vidéo, les auteurs dans leur méthode, changent le MV courant de telle façon à rendre l'erreur de prédiction paire ou impaire au lieu de modifier directement les MVs. D'où pour chaque trame P, un bit seulement est inséré dans la composante horizontale du MV. L'avantage de la méthode réside dans la sélection du mode de prédiction et les MVs optimaux dans la norme H.264/AVC afin de ne modifier que légèrement les fonctions d'optimisation de Lagrange pour freiner l'augmentation du taux de bits dans le flux.

L'inconvénient de ce type de méthodes est que la sécurité ne peut être garantie. Les utilisateurs malveillants peuvent modifier les coefficients non marqués pour rendre l'authentification non opérationnelle, ou estimer une autre marque de la vidéo, puis l'insérer dans d'autres vidéos. En effet, les auteurs ont mentionné que la robustesse de leur méthode est limitée à l'opération de transcodage. Zhang et Ho [125] ont proposé un nouveau schéma de tatouage fragile qui utilise les opérations de compensation et d'estimation de mouvement de la norme H.264/AVC. Cet algorithme est capable d'effectuer une authentification stricte dans laquelle les altérations du contenu sont fortement détectées par la sensibilité de changement du mode de décision au niveau des opérations d'estimation et de compensation de mouvement. Cependant, l'algorithme n'est pas capable de localiser les zones attaquées. Ce point faible de la technique se trouve également dans les approches présentées par Wang et Hsu [124] et Nguyen et al. [118]. Wang et Hsu ont proposé un tatouage fragile basé sur une marque de taille fixe indépendante du contenu [124].

L'analyse de leurs résultats a montré que la sensibilité d'authentification de la méthode est assurée uniquement pour les attaques de re-compression et de suppression de GOPs. Nguyen et al. [118] ont proposé un algorithme fragile indépendant du contenu basé les MVs codés. Le principe de la technique consiste à insérer une information dans les deux bits les moins significatifs des deux composantes horizontale *MVx* et verticale *MVy* des MVs délivrés par le codage de Golomb de la norme. L'insertion est effectuée selon certaines restrictions qui sont (1) les MVs calculés avec une précision inférieure à ½ pel sont écartés, car si tous les MVs sont candidats sans tenir compte de la précision, le calcul des MVs peut aller jusqu'à ¼ pel et cela engendre une perte significative dans la qualité visuelle; (2) les MBs en mode SKIP (les MBs sautés) sont aussi exclus car ceux-ci sont représentés seulement par les vecteurs de mouvement prédits (MVP) dans le flux binaire (aucun vecteur de mouvement de différence MVD n'est transmis dans le flux binaire). Un MV issu d'un MB en mode SKIP ainsi que ses MBs voisins ne sont donc pas pris en compte. Une autre technique d'authentification du contenu vidéo H.264/AVC a été proposée par Kim et al. [122]. Son principe consiste à insérer les bits de la marque dans les MVs des MBs inter-codés ou dans le nombre de mode des MBs intra-codés. Cette technique a atteint une capacité d'insertion élevée, tout en maintenant la même taille du flux compressé avec une faible dégradation de la qualité perceptible. Kuo et al. [123] ont présenté une approche de tatouage fragile utilisant les MVs comme zones d'insertion. Le principe se base sur l'extraction aléatoire des caractéristiques des blocs DCT 4×4 des trames précédemment codées et leur insertion dans les MVs de la trame courante. La position d'insertion est sélectionnée selon le mode présentant le coût du taux de distorsion minimal. L'algorithme est sensible à l'attaque de transcodage tout en conservant la qualité visuelle de la vidéo, cependant, les attaques spatiales et temporelles n'ont pas étaient testées.

Pour l'authentification du flux compressé, Pröfrock et al. [126] ont proposé une approche de tatouage fragile, effaçable et aveugle pour assurer l'intégrité du flux compressé H.264/AVC. L'authentification est de type strict, La marque est insérée dans les MBs de type sauté (skipped) de la norme. La méthode a conduit à une faible dégradation de la qualité vidéo avec une faible augmentation du taux de bits,

cependant, la capacité d'insertion est très faible. Pour le même but, Razib et al. [127] ont utilisé le processus d'adaptation MPEG-21 pour chiffrer et authentifier le flux compressé H.264/AVC. Ce processus consiste à produire une description XML (e**X**tensible **M**arkup **L**anguage) au format gBSD (**g**eneric **Bi**tsream **S**yntax **D**escription) [128] du flux H.264/AVC. Cette description peut être produite au moment du codage ou juste avant la diffusion, par un module connaissant le format binaire. Une description gBSD décrit la structure d'un flux en indiquant, pour chaque élément de la syntaxe binaire, sa position dans le flux et sa longueur. Un marqueur est également associé à chaque élément de syntaxe. Ces informations sont regroupées dans un élément XML nommé « gBSD Unit ». Cette description est ensuite adaptée par une transformation du document XML pour produire une nouvelle description gBSD. Cette nouvelle description correspond à un flux adapté. Elle est produite en supprimant (ou en modifiant) des éléments de la description gBSD initiale sur la base des marqueurs. L'utilisation du processus gBSD exige la préservation et la connaissance de la structure du flux ainsi que la structure du contenu qui sont stockées sous la forme de métadonnées. Cependant, la conservation des métadonnées, le décryptage, l'authentification et l'adaptation conduisent à des informations supplémentaires dans l'entête du flux. Ce processus est très lent, d'autant plus lent qu'il rend les applications classiques lentes et inefficaces.

Une autre méthode d'authentification développée dans le flux compressé de la norme H.264/AVC est proposée par Ueda et al. [129]. Son principe général est d'effectuer une série de traitements pour générer une signature numérique qui est ensuite insérée dans les paquets selon un choix bien précis. L'authentification du contenu vidéo dans cette méthode est réalisée par l'authentification de chaque paquet constituant le flux compressé. Dans cette méthode, l'auteur s'est basé sur trois approches de tatouage : (1) un schéma d'authentification proposé par Gennaro [130] qui consiste à associer une signature numérique sur plusieurs paquets. Le flux vidéo est divisé en blocs, chacun est composé de plusieurs paquets. En utilisant une fonction de hachage, chaque paquet a l'adresse du paquet suivant. Le premier paquet de chaque bloc est signé. Cette fonction de hachage construit une chaine entre les paquets et donc entre les blocs. L'inconvénient de ce schéma réside dans la

perte de paquets de données binaires, donc du contenu vidéo. En effet, la perte d'un seul paquet entraine la perte de la chaine et donc la perte de la vidéo. Pour remédier à ce problème, Wong et al. [131] ont proposé un autre schéma qui consiste à vérifier chaque paquet séparément. C'est le même principe que celui de Gennaro et al. [130], sauf que dans cet algorithme chaque paquet conserve l'adresse du paquet racine donc en cas de perte d'un paquet, la chaine est retrouvée dans un autre paquet. Cet algorithme malgré qu'il soit robuste face à la perte de paquets, il a l'inconvénient d'accroitre l'entête de chaque paquet. Afin de résoudre les problèmes rencontrés dans les schémas proposés [129] [131], Park et al. [132] ont proposé une approche nommée SAIDA (Signature Amortization using IDA) basée sur l'algorithme IDA (Information Dispersal Algorithm). Une signature est associée à chaque paquet. Les signatures générées sont concaténées selon le processus IDA pour sortir avec une seule signature qui est la signature globale. Un rapport de correction FEC (Forward Error Correction) est transmis à chaque paquet pour éviter la perte des signatures. Le point commun entre ces trois schémas est que l'authentification se fait au niveau des paquets ce qui perturbe la souplesse au niveau des couches de transport.

L'utilisation de la signature cryptographique pour authentifier le contenu H.264/AVC après compression est proposée en premier par N. Ramaswamy [133]. La signature est générée à partir du contenu. Les caractéristiques des MBs en mode Intra et Inter de chaque GOP constituant la séquence sont hachées par la fonction de hachage cryptographique MD5 (Message Digest 5) pour produire une signature de 128 bits pour chaque GOP. Cette signature est ensuite envoyée comme une information supplémentaire (SEI) dans l'entête de chaque GOP du flux compressé. La méthode peut détecter la cause de l'échec de l'authentification et peut localiser les trames altérées, cependant, elle augmente le taux de bits de la vidéo, ce qui conduit à l'augmentation de la bande passante pour la transmission du flux. Un autre schéma d'authentification stricte de faible complexité a été proposé par Horng et al. [134]. Son principe repose sur l'extraction des caractéristiques de type fragile au sein des MBs intra et inter. Ces dernières constituent le code d'authentification, et sont cachées dans le GOP au sein de la couche d'adaptation réseau appelé NAL (Network

Abstraction Layer). Dans leur approche, une clé générée du contenu a été utilisée pour assurer la sécurité de l'authentification et la sélection des coefficients AC quantifiés non nuls pour obtenir la fragilité. L'algorithme a conduit à une faible dégradation de la qualité visuelle de la vidéo tatouée avec une légère augmentation du taux de bits, cependant, son inconvénient réside dans la connaissance de la structure du flux compressé binaire afin d'extraire les informations cachées.

De nouvelles fonctionnalités de la norme H.264/AVC sont étudiées dans d'autres algorithmes pour authentifier le contenu, telles que le codage CAVLC, le mode intra prédiction et l'index de référence pour le tatouage numérique dans la norme H.264/AVC [135] [136]. Néanmoins ces algorithmes sont fragiles à certaines attaques communes.

Toutes les méthodes mentionnées ci-dessus montrent clairement que chaque méthode proposée permet à chaque contrainte particulière (transparence, sensibilité, charge utile de données, sécurité) d'être atteinte individuellement pour des applications d'authentification stricte du contenu vidéo H.264/AVC. Cependant, aucune de ces méthodes n'est en mesure d'atteindre conjointement toutes les exigences. D'où, aboutir à un système d'authentification stricte répondant à la majorité des exigences nécessite plus de recherche. Développer un système d'authentification stricte et performant du contenu vidéo compressé par la norme de compression H.264/AVC est le second objectif de cette thèse. Cette performance relève de contraintes et d'enjeux spécifiques [118] [119] [120]: sensibilité, invisibilité, localisation des régions altérées, sécurité et préservation de la qualité visuelle et taux de bits.

Deux versions du système d'authentification stricte du contenu (SASC) basées sur deux fonctions de Hachage différentes sont proposées dans cette thèse: la version SASC-MD5 utilise la fonction de hachage cryptographique MD5 présentée dans [137] et améliorée par la méthode publiée dans [138] et la version SASC-HMAC-SHA-256 basée sur la fonction de hachage sécurisée HMAC-SHA-256 parue dans [139].

4.3. Méthode proposée : Version SASC-MD5

La première version SASC-MD5 regroupe deux variantes [137] [138]. L'objectif visé dans le développement de la première version est d'aboutir à un système d'authentification strict remplissant un nombre maximal de critères d'authentification plus particulièrement le critère de conservation du taux de bits, critère que la technique de Ramaswamy et al. ne remplit pas [133]. L'apport de notre technique réside dans le processus d'insertion, où nous optons pour les MVs comme régions d'insertion. En effet, Ramaswamy et al. ont proposé de cacher la signature numérique dans la zone SEI. Cette zone est un espace alloué par le codeur H.264/AVC. Il est dédié pour l'insertion des informations supplémentaires dans le flux H.264/AVC dans le but d'améliorer son usage pour un grand nombre d'applications. L'inconvénient qui en découle est que l'insertion engendre une augmentation de la bande nécessaire pour transmettre le flux compressé. L'idée de base de notre méthode élaborée consiste à extraire certaines caractéristiques intrinsèques dans le domaine transformé (fréquentiel) de la vidéo d'origine et à les insérer ensuite dans les MVs des trames prédéfinies sous la forme d'un tatouage fragile et invisible. La marque fragile est obtenue en appliquant la fonction de Hachage MD5 sur les caractéristiques extraites de la vidéo. Le rôle de cette fonction de hachage et de produire un condensé unique, représentatif du contenu vidéo d'origine à partir des caractéristiques extraites. Lorsque l'on souhaite vérifier l'intégrité d'une vidéo, on compare simplement la signature numérique générée à partir des caractéristiques intrinsèques de cette vidéo tatouée avec la signature générée à partir de la vidéo d'origine (processus d'extraction). Si les signatures sont identiques, cela signifie que la vidéo n'a pas été manipulée, sinon les différences indiquent les régions qui ont été altérées. Le choix des caractéristiques de la vidéo et la position d'insertion sont primordiaux dans la mesure où ils vont conditionner les manipulations que l'on pourra détecter et celles qu'on laissera passer. D'une manière générale, on sélectionne les traits de la vidéo en fonction de leur stabilité face aux différentes attaques. Typiquement, pour un système d'authentification strict, on recherche des caractéristiques qui sont sensibles à des retouches locales de la vidéo. Le tatouage

utilisé est un schéma aveugle, l'extraction de la marque ne nécessite pas la vidéo d'origine, elle est effectuée au cours du décodage du flux H.264/AVC.

4.3.1. Variante SASC-MD5-1

La première variante du système d'authentification SASC-MD5 nommée SASC-MD5-1 présentée dans [137] consiste à subdiviser chaque séquence vidéo en entrée en plusieurs groupes d'images (GOPs) et à générer ensuite une signature numérique unique pour chaque GOP séparément (figure 4.2). L'idée de base de cette approche consiste à extraire certaines caractéristiques intrinsèques et robustes à partir des coefficients DCT 4×4 de chaque GOP de la séquence d'origine, les hacher en appliquant une fonction de hachage cryptographique MD5 pour générer une signature numérique de 128 bits pour chaque GOP. Ensuite, les bits de la signature (marque fragile) sont insérés dans les MVs sélectionnés dans les GOPs correspondants.

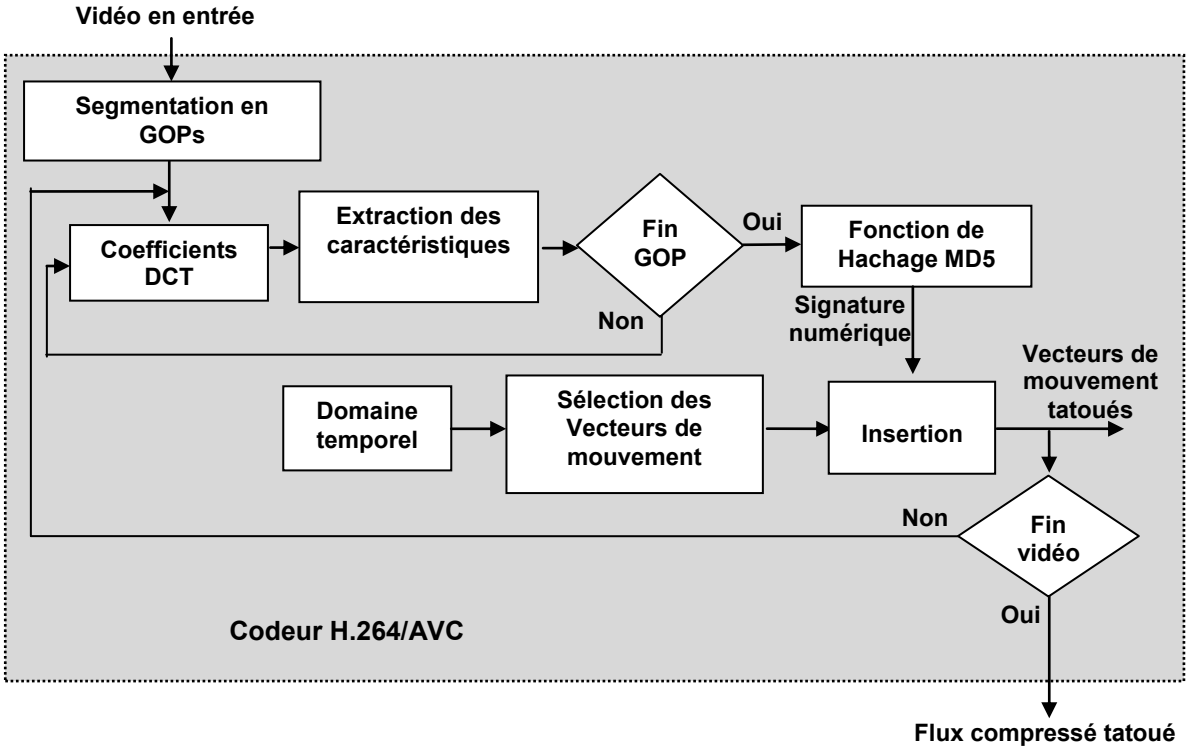


Figure 4.2 : Schéma d'insertion de la signature.

Au niveau du décodeur H.264/AVC l'authentification du contenu vidéo est effectuée. Si les signatures numériques des GOPs sont identiques, cela signifie que la séquence n'a pas été manipulée, sinon les différences indiquent que le contenu a été altéré, la vidéo n'est donc pas authentique. Les différentes étapes du processus d'insertion de la signature sont:

- Segmentation de la séquence vidéo en GOPs;
- Extraction des caractéristiques et génération de la signature;
- Sélection des positions d'insertion;
- Insertion.

4.3.1.1. Segmentation de la séquence vidéo en GOPs

La structure et la taille d'un GOP ne sont pas spécifiées dans la norme H.264/AVC, elles sont définies selon l'application. Dans nos expériences, le GOP se compose de la trame I et toutes les autres trames P (figure 4.3), qui sont temporellement enfermées entre des images IDR (Instantaneous Decoder Refresh). Par conséquent, une sous-séquence vidéo codée commence par une image IDR et se termine lorsqu'un nouveau IDR est reçu, signalant ainsi la disponibilité d'une nouvelle sous-séquence à coder ou la fin de transmission de la vidéo. La présence d'une image IDR indique qu'il n'y a pas d'autres images suivantes dans le flux binaire nécessitant des références dans l'ordre de décodage juste avant la trame I. Par conséquent, les GOPs sont indépendants entre eux. Cette technique permet au décodeur de se resynchroniser sur le flux dans le cas d'une transmission avec pertes.

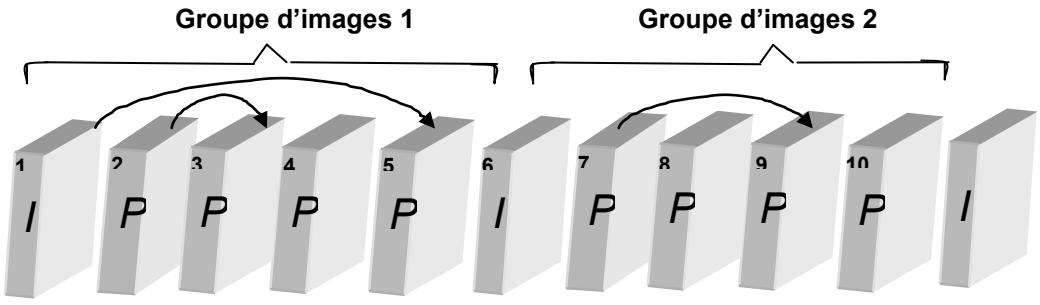


Figure 4.3 : Structure d'un groupe d'images GOP.

4.3.1.2. Extraction des caractéristiques et génération de la signature

L'utilisation d'un tatouage fragile est de détecter tout changement du contenu. Ceci est réalisé en utilisant une fonction de hachage cryptographique MD5 ayant en entrée des données robustes de la vidéo d'origine. Les données à l'entrée de la fonction MD5 sont des caractéristiques visuelles robustes pour lesquelles l'œil humain est sensible. Il a été démontré dans la littérature que les caractéristiques visuelles robustes dans le domaine DCT [140] [117] [134] [141] sont représentées par un nombre restreint de coefficients fréquentiels appartenant aux basses fréquences. En se basant sur cette constatation, la génération de la signature est basée sur l'extraction des caractéristiques intrinsèques spatiales (Intra_16×16 et Intra_4×4) et temporelles Inter_4×4 dans un macrobloc de taille 16×16 de la composante de luma (figure 4.4).

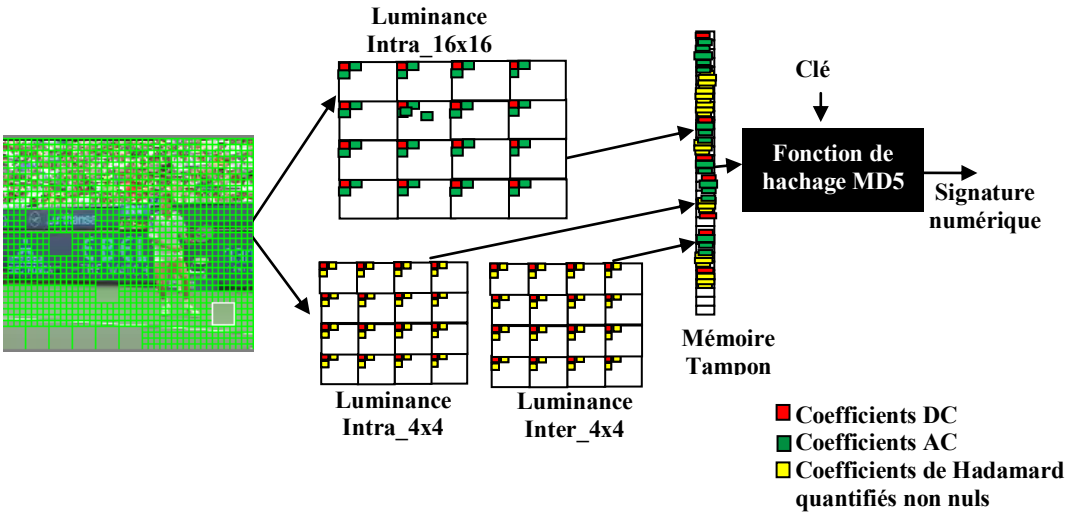


Figure 4.4 : Extraction des caractéristiques et génération de la signature numérique.

Pour les blocs de type Intra_4×4 et Inter_4×4, les coefficients pris en compte pour la génération de la signature sont les coefficients quantifiés DC et les deux premiers coefficients transformés et quantifiés AC appartenant aux coefficients de basse fréquence dans l'ordre de balayage en zig-zag. Pour les MBs de type Intra_16×16, les coefficients pris en compte sont tous les coefficients DC de chaque

bloc et les coefficients transformés et quantifiés AC non nuls. L'ensemble des coefficients sont collectés dans une mémoire tampon jusqu'à atteindre l'indicateur IDR. Ce dernier permet de mentionner la fin d'un GOP tout en rafraîchissant instantanément la mémoire de référence du décodeur interne au codeur H.264/AVC. À la fin de chaque GOP, Les valeurs présentes dans la mémoire tampon sont hachées par la fonction MD5 pour produire un message de 128 bits. Ce message constitue la marque fragile qui est insérée dans les MVs sélectionnés.

La fonction de hachage sert à produire un condensé (ou une empreinte) unique, représentatif du contenu d'origine. Son rôle principal est de vérifier l'intégrité de la séquence sans avoir recours à la séquence d'origine [31]. Une fonction de hachage opère généralement sur un message (caractéristiques de la vidéo) de longueur arbitraire pour fournir une séquence de valeurs de taille fixe (pour MD5 128 bits). Pour qu'une telle fonction soit considérée comme sûre, elle doit vérifier les propriétés suivantes :

- il est ″facile″ de calculer la séquence binaire connaissant le message,
- il est ″difficile″ de retrouver le message connaissant la séquence de valeurs,
- il est ″difficile″de trouver un message autre que le message d'origine ayant le même condensé de valeurs.

4.3.1.3. Sélection des positions d'insertion

Le processus d'insertion est un schéma substitutif, qui se base sur les premières techniques d'insertion utilisées pour vérifier l'intégrité de l'image [31]. Il consiste à insérer les bits de la marque fragile dans les bits les moins significatifs des MVs. L'idée de base de l'opération d'insertion consiste à sélectionner l'ensemble des MVs appartenant à des trames ayant de grandes valeurs de la quantité de mouvement moyenne (séquence avec beaucoup de mouvement). Afin d'assurer l'imperceptibilité de la marque, des restrictions sur les régions d'insertion sont établies:

- les MBs sautés (MBs non traités) et leurs blocs voisins sont écartés car leurs vecteurs de mouvement de différence MVD ne sont pas transmis au flux compressé. Les MVs des MBs sautés sont produits uniquement par des vecteurs de mouvement prédits MVP et l'insertion dans les blocs voisins provoque des erreurs de mouvement au niveau de la reconstruction des MBs sautés donc ils ne peuvent pas être reconstruits correctement.
- La seconde restriction appliquée découle de l'estimation du mouvement et le mode de décision délivré par le standard H.264/AVC [68] [69]. L'insertion est effectuée uniquement dans les MVs de la partition 8×8 et ses quatre sous partitions 8×8, 8×4, 4×8 et 4×4 des trames de la séquence ayant une grande activité moyenne de mouvement. En sélectionnant ces derniers modes de partition qui représentent les régions ayant une grande activité, il serait difficile pour l'œil humain de détecter les distorsions introduites par l'insertion de la marque.

4.3.1.4. Opération d'insertion

Après l'application des restrictions ci-dessus, les *K* MVs restants des trames *P* ayant une activité de mouvement élevée sont sélectionnés en calculant l'intensité de la quantité de mouvement moyenne donnée par :

$$MV_{aver} = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} \left| MV_p(x,y) \right| \tag{4.4}$$

Où *M* et *N* sont respectivement la largeur et la hauteur de la trame P mesurées en bloc de taille 4×4. L'expression $|MV_p(x,y|$ est l'amplitude du MV d'un bloc 4×4 à la position *(x,y)* dans la trame P, calculée à partir de ses deux composantes *MVx* et *MVy*.

L'insertion est réalisée en modifiant le dernier bit du poids faible (LSB) des deux composantes *MVx* et *MVy* des MVs sélectionnés. Dans le schéma proposé, la précision des MVs est de un quart de pixel (¼ pel). Un tel déplacement (résolution

fractionnelle) peut pointer sur des positions qui sont spatialement situées entre les pixels de la trame. Avant d'être tatouée, chaque composante du MV sélectionné est quantifiée à la position la plus proche du pixel en entier. En appliquant cette opération, on limitera la distorsion au maximum à ½ pel pour chaque composante du MV. L'opération de quantification est définie comme suit :

$$Q(MVx) = \begin{cases} (1 + MVx)\,\&\,(0xFFFE) & MVx \geq 0 \\ -(1 - MVx)\,\&\,(0xFFFE) & MVx < 0 \end{cases} \tag{4.5}$$

$$Q(MVy) = \begin{cases} (1 + MVy)\,\&\,(0xFFFE) & MVy \geq 0 \\ -(1 - MVy)\,\&\,(0xFFFE) & MVy < 0 \end{cases} \tag{4.6}$$

Où & est l'opérateur ET binaire.

Les bits de la marque wi sont insérés en substituant le dernier bit LSB des composantes de MV comme suit :

$$\overline{MVx} = \begin{cases} Q(MVx) \mid (0x0001) & si\ wi = 1 \\ Q(MVx) & ailleurs \end{cases} \tag{4.7}$$

$$\overline{MVy} = \begin{cases} Q(MVy) \mid (0x0001) & si\ wi = 1 \\ Q(MVy) & ailleurs \end{cases} \tag{4.8}$$

où | est l'opérateur OU exclusif.

Afin de minimiser la distorsion causée par l'insertion de la marque, le processus d'insertion doit assurer la condition de synchronisation suivante :

$$Q(\overline{MVx}) = Q(MVx) \quad et \quad Q(\overline{MVy}) = Q(MVy) \tag{4.9}$$

Enfin, la trame *P* est composée des blocs de MVs tatoués ($\{\overline{MVx}, \overline{MVy}\}$) et des blocs de MVs d'origine. Ceci conduit à la trame *P* de MVs composée de :

$$\overline{MV} = \begin{cases} \{\overline{MVx}, \overline{MVy}\} & si\ MV \in MBs\ sélectionnés \\ \{MVx, MVy\} & ailleurs \end{cases} \tag{4.10}$$

4.3.1.5. Processus d'extraction et de vérification

L'opération d'extraction et de vérification est cruciale pour vérifier l'intégrité de la vidéo. Elle s'opère à la réception du flux tatoué au niveau du décodeur H.264/AVC (figure 4.5). La marque est extraite en procédant comme suit:

- décodage des MVs,
- application des restrictions semblables à celles de l'opération d'insertion pour déterminer les trames P ayant une activité de mouvement élevée,
- détermination des trames *P* taouées moyennant la clé *K*,
- obtention des bits de la marque par extraction du dernier bit des deux composantes des MVs décodés.

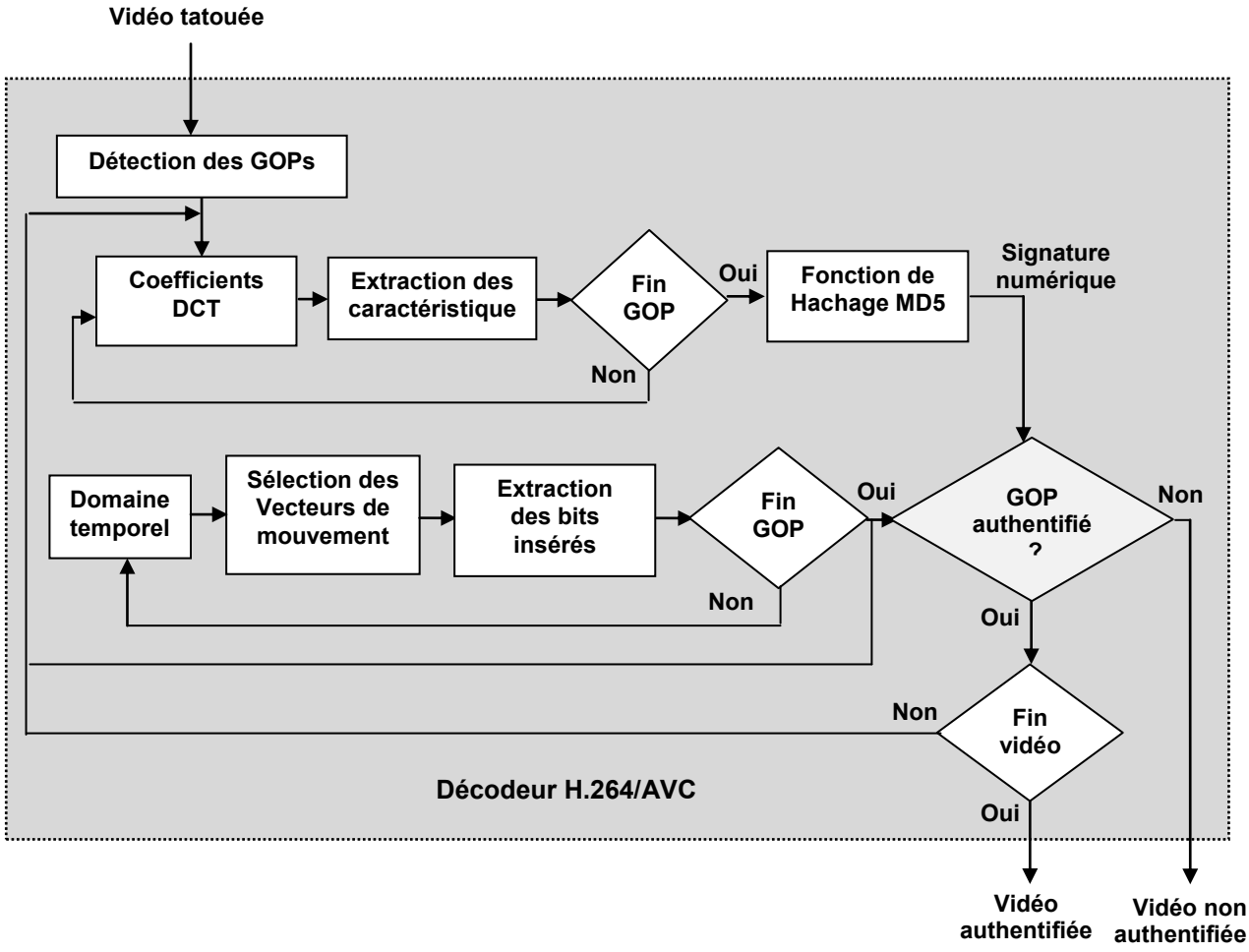


Figure 4. 5 : Schéma du processus de vérification de signature.

L'opération de vérification regroupe deux processus: la génération de la signature au niveau du décodeur et la comparaison. La génération de la signature est effectuée au niveau du décodeur à partir des coefficients DCT quantifiés. Un décodage partiel du flux compressé est effectué pour extraire les mêmes caractéristiques intrinsèques que celles adoptées pour générer la signature au niveau du codeur, à savoir :

- le coefficient quantifié DC plus les deux premiers coefficients transformés et quantifiés AC des blocs de type Intra_ 4×4,
- le coefficient quantifié DC plus les deux premiers coefficients transformés et quantifiés AC des blocs de type Inter_4×4,
- le coefficient quantifié DC plus tous les coefficients quantifiés de Hadamard non nuls pour les blocs de type Intra_16x16.

Lorsque la fin de la séquence est détectée, les données sont ensuite hachées par le MD5 pour produire au niveau du décodeur une signature numérique unique pour chaque GOP séparément.

L'approche développée appartient à la classe des techniques de tatouage aveugle, la séquence vidéo d'origine n'est pas nécessaire pour vérifier l'intégrité du contenu. Ceci car la marque représentant les données d'authentification est récupérée à partir du flux H.264/AVC tatoué. L'intégrité est vérifiée en comparant pour chaque GOP décodé, la signature extraite des MVs décodés et la signature générée à partir des caractéristiques du contenu tatoué. Si toute les signatures générées des GOPs sont semblables aux signatures extraites alors la vidéo est considérée comme authentique et donc elle n'a pas été altérée par une source externe, dans le cas contraire, la vidéo a subi des manipulations.

4.3.1.6. Résultats expérimentaux

La méthode conçue est intégrée dans le profil "restreint de base" (constrained baseline profile) du logiciel de référence Joint Model (JM) H.264 JM-10.1 [142]. Ce

profil constitue le noyau des extensions développées par le JVT (Joint Video Team) dans la norme [143]. Les résultats obtenus sont indépendants de la version du logiciel de référence JM, puisque chaque version du standard H.264/AVC est une mise à jour de la précédente [143]. Par conséquent, l'extension vers les autres versions du logiciel H.264/AVC nécessite uniquement la modification du fichier de configuration dans le profil de base (Profil IDC = 66) selon les paramètres illustrés sur le Tableau 4.1. Le logiciel est configuré pour produire une image I une fois toutes les 15 trames, rendant ainsi le nombre total de GOPs égal à 10 pour une séquence de 150 images. Les tests sont effectués sur les séquences vidéo benchmark au format YUV 4:2:0 de résolution QCIF (176x144 pixels) à une fréquence de 30 images/seconde [144] et classifiées en deux groupes (figure 4.6): (1) Vidéos avec beaucoup de régions homogènes et peu d'activité temporelle (Groupe A) et (2) Vidéos présentant beaucoup de mouvements et moyennement ou fortement texturées (Groupe B).

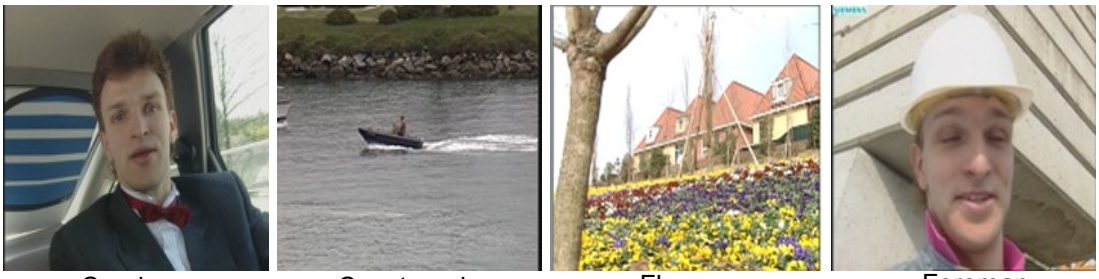
Tableau 4.1 : Paramètres de configurations du codeur JM-10.1.

Profile	Baseline (Profile IDC=66)
Number of frames	150 for all test sequences except for Table which includes 88 frames
Frame rate	30 fps
Source Bit Depth luma	8
Source Bit Depth chroma	8
Motion estimation	Full Search
Entropy coding	CAVLC
Search range	16
Quantization parameter	28
Intra period	15, only the first frame is intra

La capacité d'insertion par GOP, la taille de l'espace mémoire occupé par la vidéo et la qualité perceptuelle exprimée en termes de PSNR (dB), sont utilisés comme métriques d'évaluation de la méthode implémentée. Les deux dernières métriques sont lues à partir du fichier *log.date* créé par le JM 10.1 au cours du processus compression-insertion. Par contre la capacité d'insertion est déterminée par l'activité de mouvement et la distorsion introduite par l'insertion.

Miss America Claire Bridge-Close Akiyo

Groupe A



Carphone Coastguad Flowers Foreman

Suzie Table

Groupe B

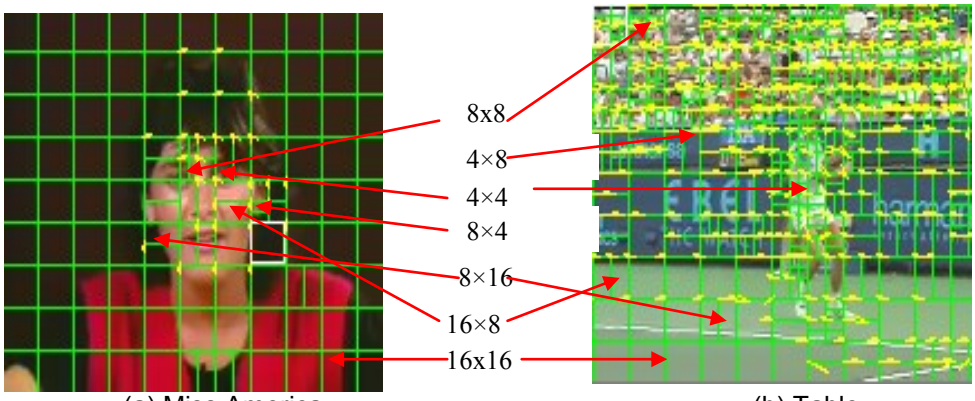
Figure 4.6 : Différentes séquences vidéo utilisées pour les tests.

Le tableau 4.2 illustre les résultats de simulation en termes de capacité maximale d'insertion pour toutes les séquences tests appartenant aux deux groupes. Dans le groupe A, l'arrière-plan est plus souvent statique où les trames présentent un mouvement limité dans certaines trames, par conséquent, il y a moins de MVs appartenant au mode de partition 8x8 et ses sous partitions 4×4 , 4×8 , 8×4, et 8x8 pour insérer tous les bits de la marque fragile (128 bits délivrés par le MD5) (figure 4.7(a)). La séquences Table est une séquence qui contient beaucoup de mouvements (l'arrière-plan et l'avant plan sont tous les deux en mouvement), contient donc beaucoup de partitions de taille 8x8 et ses sous partitions (8×4, 4×8, 4×4)

(Figure 4.7(a)). Ceci permet d'insérer un nombre important de bits dans un GOP qui peut aller au maximum jusqu'à 5182 bits. Par contre, la séquence Miss America permet l'insertion d'un nombre faible de bits (88 bits) dans un seul GOP. Ceci est justifié par le fait que la séquence est une séquence plus ou moins statique contenant peu de mouvements (figure 4.7(a)). Elle comprend beaucoup de partitions de taille 16×16 et ses sous partitions (16×8, 8×16), qui correspondent aux régions homogènes. Elle contient peu de partitions 8x8 pour insérer toute la signature. L'insertion est donc tronquée et l'authentification échoue. Les figures 4.8 et 4.9 montrent respectivement la dégradation de la qualité visuelle causée par l'insertion dans les MVs appartenant au mode partition 16x16 et l'insertion sans tenir compte des conditions d'insertion. Toujours, en se référant au tableau 4.2, nous pouvons conclure que le PSNR qui exprime la performance de la qualité perceptuelle est préservé et la taille du flux tatoué est restée inchangée. La figure 4.10 illustre la qualité visuelle obtenue après l'application des restrictions établies.

Tableau 4.2 : Résultats de simulation pour les séquences vidéo appartenant aux deux groupes.

	Séquence vidéo	Capacité d'insertion (bits) par GOP	PSNR (dB)		Taille du contenu vidéo (Mo)	
		Méthode proposée [137]	Vidéo d'origine	Méthode proposée [137]	Vidéo d'origine	Méthode proposée [137]
Groupe A	**Miss America**	88	40.056	40.04	5.43	5.43
	Claire	135	39.681	39.67	5.43	5.43
	Bridge-close	264	34.847	34.85	5.43	5.43
	Akiyo	133	38.205	38.17	5.43	5.43
Groupe B	**Carphone**	1495	37.315	37.29	5.43	5.43
	Coastguard	2806	34.026	34.02	5.43	5.43
	Flower	4304	34.308	34.31	5.43	5.43
	Foreman	2569	36.45	35.75	5.43	5.43
	Suzie	1349	37.141	37.12	5.43	5.43
	Table	5182	34.915	34.91	3.22	3.22



(a) Miss America (b) Table

Figure 4.7 : Illustration des différents modes de partition et les vecteurs de mouvement des séquences (a) Miss America et (b) Table.

Image d'origine Image tatouée

Figure 4.8 : Insertion dans les partitions 16x16 de la 36ème image de Table.

Trame d'origine Trame tatouée

Figure 4.9 : Insertion sans prise en compte des conditions d'insertion.

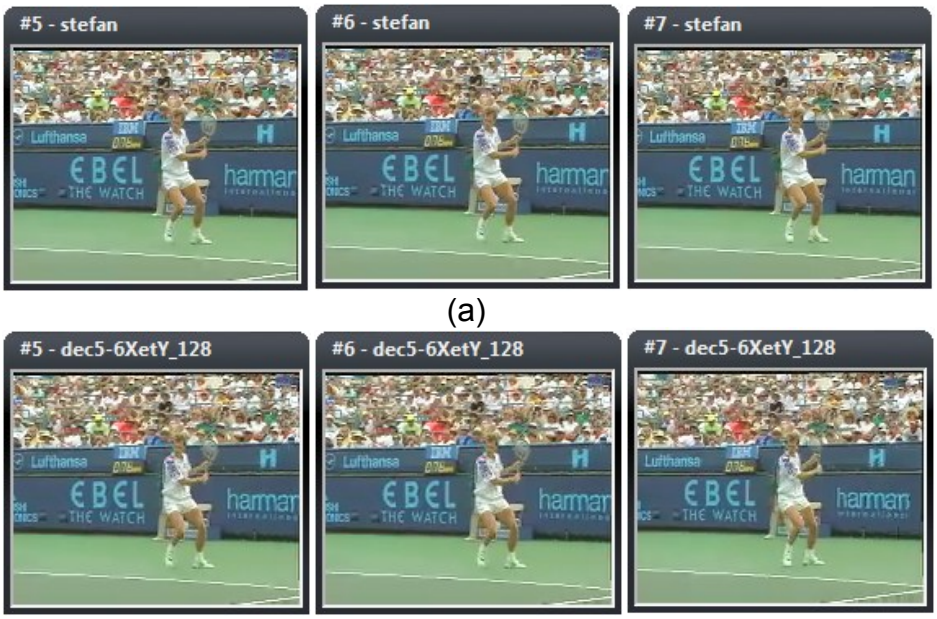


Figure 4.10 : Qualité visuelle des trames 5, 6 et 7 de la séquence Table après le processus d'insertion:(a) trames d'origine (b) trames tatouées.

Le système SASC-MD5-1 ainsi développé assure l'authentification stricte du contenu vidéo présentant une forte activité temporelle (séquences du groupe B), vu qu' il y a assez de positions pour insérer les 128 bits de la marque, à l'inverse, pour les vidéos ayant une faible activité temporelle (séquence du groupe A), la marque est interrompue car il y a peu de positions d'insertion, d'où l'échec de l'authentification de ce type de vidéo.

Le système proposé assure l'authentification stricte du contenu vidéo caractérisé par une forte activité temporelle. Il permet de résoudre l'augmentation de l'espace mémoire engendré par l'insertion de la méthode de Ramaswamy et al. [133] tout en garantissant les critères d'invisibilité, de sensibilité et d'altérations spatiales et temporelles. Par contre, les critères sécurité et localisation des régions altérées ne sont pas satisfaits. La méthode n'est pas en mesure de donner à l'utilisateur une information visuelle permettant d'identifier les régions qui ont été manipulées.

Ces défaillances ont été prises en compte dans la seconde variante publiée dans [138]. La sécurité est assurée par un brouillage selon une permutation pseudo-

aléatoire de la marque avant son insertion. Les performances du critère de sensibilité sont améliorées en introduisant les caractéristiques de la chrominance (chroma) en plus de celles de luma prises en compte dans la première variante SASC-MD5-1.

4.3.2. Variante SASC-MD5-2

4.3.2.1. Principe général de la méthode

La sécurité, la sensibilité aux changements de couleurs et la localisation des régions altérées constituent les principales améliorations apportées à la première variante SASC-MD5-1 [137]. En plus des caractéristiques de luma utilisées dans SASC-MD5-1 pour générer les signatures numériques, celles de chroma sont aussi prises en compte pour détecter les manipulations de couleur dans la vidéo. Les signatures numériques de chaque GOP sont donc générées à partir des caractéristiques spatiales (Intra_16x16 et Intra_4×4) et temporelles Inter_4×4 dans un MB de taille 16x16 ainsi que les caractéristiques de chroma de chaque bloc 4×4 dans un MB de taille 8x8 (Figure. 4.11).

Les échantillons de chroma de taille 8x8 du MB sont prédits en utilisant une technique de prédiction similaire à celle de luma Intra_16x16 car la composante chroma est habituellement uniforme sur de grandes surfaces. Pour les MBs de taille 8x8 pour chacune des composantes de chroma, les coefficients pris en compte sont tous les coefficients DC de chaque bloc et les coefficients transformés et quantifiés AC non nuls. Les caractéristiques de chaque trame de la séquence vidéo sont collectées dans une mémoire tampon jusqu'à la fin du GOP soit indiquée par l'IDR. A la fin du GOP, les valeurs présentes dans la mémoire tampon sont hachées par la fonction MD5, puis brouillées par une fonction de permutation pseudo aléatoire commandée par une clé *K*. Les bits de la marque (128 bits) ainsi obtenus sont insérés ensuite dans les derniers bits LSB des MVs sélectionnés dans les GOPs correspondants. Le même processus d'insertion que celui utilisé dans SASC-MD5-1 [137] est adopté [138].

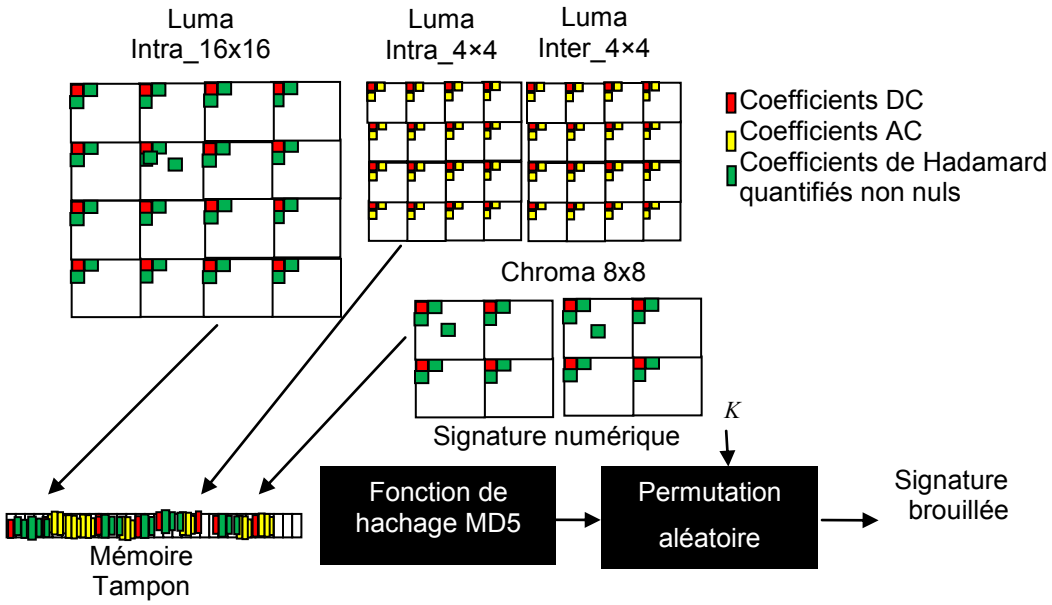


Figure 4.11 : Génération de la marque d'un GOP.

La différence des étapes d'authentification entre SASC-MD5-1 [137] et SASC-MD5-2 [138]) réside dans l'introduction de : (1) la composante couleur pour la génération de la signature et ceci dans le but d'augmenter les performances de sensibilité de la méthode, (2) la fonction de brouillage (au niveau du codeur) et la désembrouillage (au niveau du décodeur) pour augmenter la sécurité d'insertion et enfin (3) l'opération de détection des trames altérées (Figure 4.12). Si les signatures extraite et générée de chaque GOP décodé sont identiques, la séquence n'a donc pas été manipulée, sinon les différences indiquent les GOPs modifiés.

Le système offre une parfaite sécurité contre les attaques spatiales, temporelles et de changement de couleur. En cas d'échec de l'authentification dans un GOP, et afin de trouver les raisons de l'échec, nous suggérons dans cette variante du système de produire le condensé de la fonction MD5 de chaque trame du GOP localisé altéré et d'envoyer ensuite les signatures résultantes des trames au codeur afin de les comparer aux signatures d'origine. Si les signatures du codeur et du décodeur d'une trame donnée dans le GOP ne correspondent pas, la dite trame est

signalée modifiée et la localisation de l'altération est localisée au niveau de la trame et non au niveau du GOP (Figure 4.13).

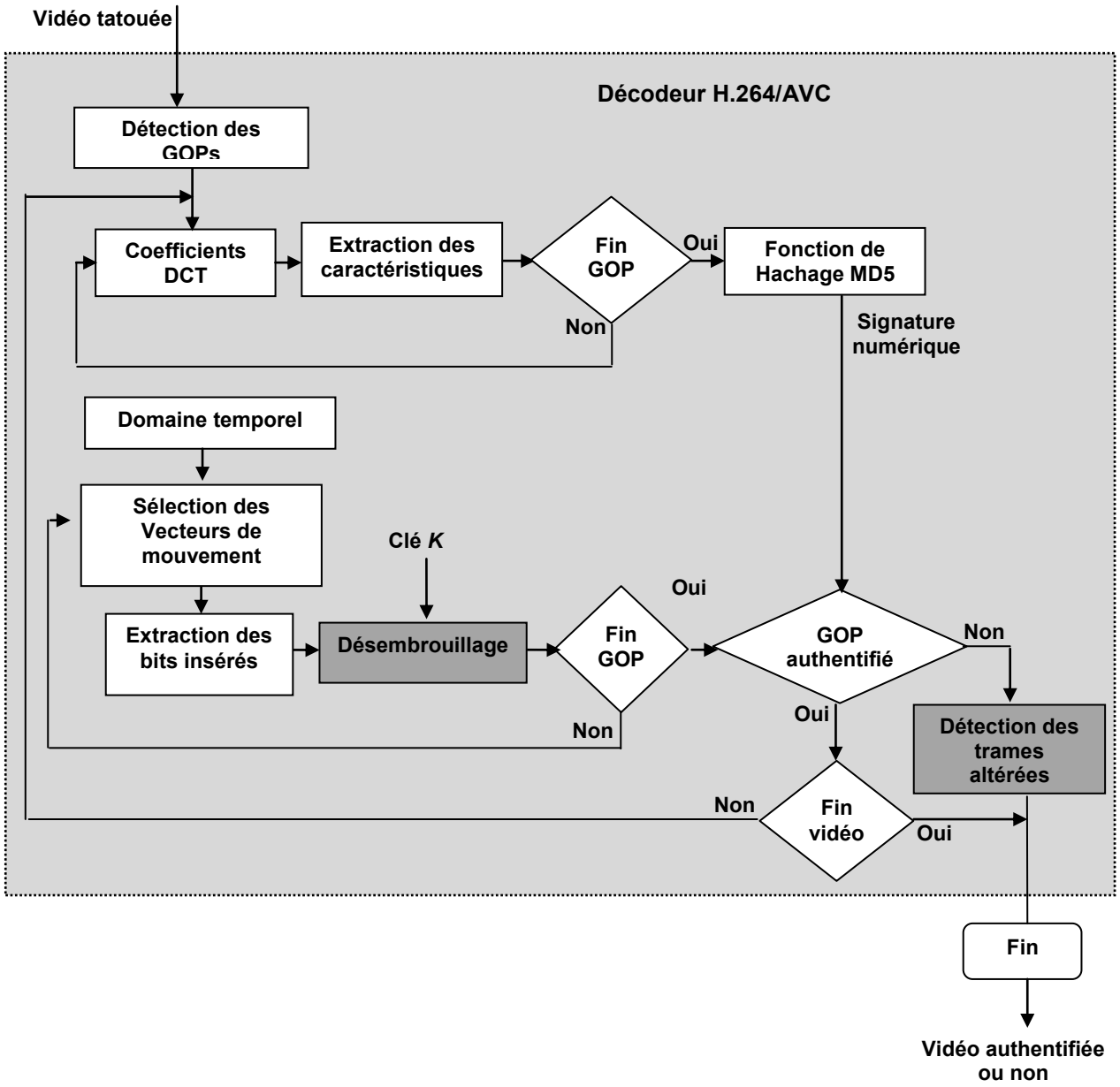


Figure 4.12 : Schéma du processus de vérification de signature.

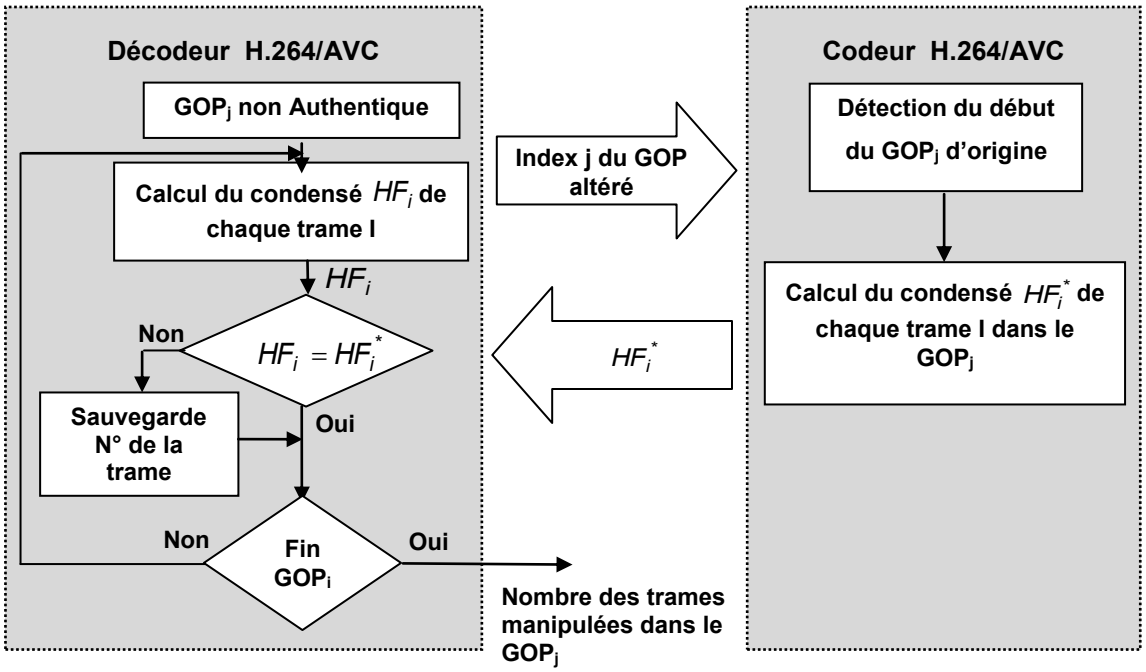


Figure 4.13 : Détection des trames altérées.

4.3.2.2. Analyse et résultats expérimentaux

L'algorithme est intégré dans le même logiciel de référence H.264 JM-10.1 [142] avec les mêmes séquences vidéo et la même configuration (paramètres du codeur et segmentation des vidéos) utilisées dans la variante SASC-MD5-1 [137]. Les résultats d'évaluation de la qualité perceptuelle de la vidéo tatouée exprimée en termes de PSNR, la capacité d'insertion et l'espace mémoire occupé par la vidéo après l'opération d'insertion restent inchangés car le même processus d'insertion utilisé dans SASC-MD5-1 est adopté. Pour la variante SASC-MD5-2 [138], les expériences sont menées surtout pour évaluer la sensibilité de l'algorithme face aux altérations spatiales, temporelles et les manipulations de couleurs. Pour cela, deux différentes attaques bienveillantes ont été appliquées : attaque par DC et attaque par rotation verticale. L'attaque par DC consiste à modifier les coefficients AC du bloc Intra_4×4 sans changer la valeur moyenne (DC) du bloc. Un grand changement de la valeur de DC affecte grossièrement la qualité perceptuelle de la trame et aussi la signature correspondant au GOP (figure 4.14). Un petit changement de la valeur de

DC préserve la qualité visuelle de la vidéo, mais change le condensé du MD5 du GOP correspondant. Un exemple d'attaque par DC est effectuée sur le premier bloc Intra_4×4 du 77ème MB de la 7ème trame (1er GOP) de la séquence Table est donné ci-dessous.

Bloc Intra_4×4 à la position (0,0) du 77ème MB de la 7ème trame

185	212	203	197
173	201	204	172
173	171	180	167
149	187	186	153

$$\sum_{i=1}^{15} AC(i) = 2913 .$$

Altération du bloc Intra_4×4 à la position (0,0) du 77ème MB de la 7ème trame

200	180	205	214
200	155	152	210
233	175	125	120
182	173	173	216

$$\sum_{i=1}^{15} AC(i) = 2913 .$$

Nous remarquons que toutes les signatures extraites des GOPs sont semblables aux signatures générées sauf celle du premier GOP qui est différente. D'où la zone modifiée est détectée, et est localisée dans le premier GOP.

L'attaque par rotation verticale consiste à effectuer une rotation de 90 degrés d'une ou plusieurs images de la séquence vidéo. Les résultats obtenus après l'attaque de rotation sont illustrés sur la figure 4.15. Nous remarquons une différence entre la signature générée et la signature extraite dans le deuxième GOP. D'où la détection de manipulation au niveau du second GOP de la séquence.

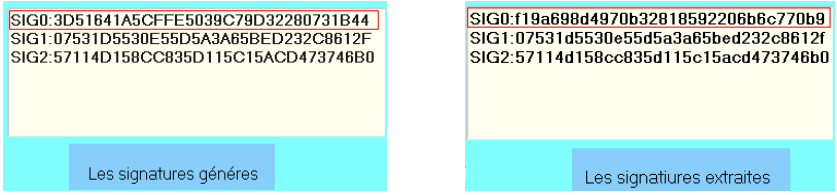


Figure 4.14 : Signatures générées et extraites après l'attaque par DC de la séquence Table.

Figure 4.15 : Signatures générées et extraites après l'attaque par rotation verticale de la 27ème trame de la séquence Table.

Une autre caractéristique importante de l'algorithme proposé est que pour la génération de la signature numérique, la dimension temporelle est prise en compte par l'ensemble des coefficients extraits des blocs de type Inter_4×4 prédits à partir des blocs des trames précédemment codées. Tout changement, dans l'une des trames de la séquence, se reflète sur la signature générée. Toute modification comme le recadrage (cropping) et le réarrangement se traduit par la modification de la signature numérique. En plus de la sensibilité aux attaques spatiales et temporelles, la méthode proposée est sensible aux attaques de changement de couleurs par la prise en considération des caractéristiques extraites des deux composantes de chroma du MB de taille 8x8 au sein du GOP. Toute modification de couleur apportée dans une des régions d'un GOP, se répercute sur la signature numérique, et par conséquent sur l'authentification du GOP. La figure 4.16(b) illustre la qualité visuelle résultante des vidéos tatouées. Les figures 4.16(c) à 4.16(f) montrent les résultats qualitatifs obtenus des différentes altérations spatiales et temporelles (attaque par DC, le recadrage, la rotation et le réarrangement des images) effectuées sur la séquence Table.

Pour les attaques de remplacement, le mot "Lufthansa" sur le panneau d'affichage dans le stade est remplacé par le mot "Harman" comme illustré dans la figure 4.16(d). La méthode est également sensible à la rotation et le réarrangement des trames (figure 4.16(e) et 4.16(f)). La figure 4.17 illustre l'avantage d'inclure les composantes de chroma comme caractéristiques pour générer la signature numérique de chaque GOP. L'attaque est réalisée en changeant la couleur du chapeau porté par Foreman dans la 20ème trame de la séquence. Ainsi, la signature résultante du GOP altéré est différente de celle produite à partir du GOP correspondant de la séquence tatouée.

26ème trame 27ème trame 28ème trame

(a) Trames d'origine

26ème trame 27ème trame 28ème trame

(b) Trames tatouées

26ème trame 27ème trame 28ème trame

(c) Trames après attaque par DC

26ème trame 27ème trame 28ème trame

(d) Trames après recadrage des régions

26ème trame 27ème trame 28ème trame

(e) Trames après attaque par rotation

26ème trame 27ème trame 28ème trame

(f) Trames après réarrangement

Figure 4.16 : Trames d'origine, tatouées et altérées de la séquence Table.

19ème trame 20ème trame 21ème trame

Figure 4.17 : Foreman après modification de couleurs.

4.3.2.3. Conclusion

Deux variantes SASC-MD5-1 et SASC-MD5-2 d'authentification ont été développées pour obtenir la première version. SASC-MD5-1 assure l'authentification stricte des vidéos ayant une grande activité temporelle, la vidéo est soit authentifiée soit non. L'échec de la méthode réside dans la localisation des régions manipulées et la sécurité. Alors que la SASC-MD5-2 permet d'assurer un maximum de critères tels que (1) la sensibilité : le système est capable de déceler des manipulations pouvant modifier l'interprétation que l'on a de la vidéo, tels que des recadrages (cropping) ou des retouches locales (rotation), (2) l'invisibilité de l'insertion, (3) la conservation de l'espace occupé par la vidéo, afin de ne pas augmenter la bande nécessaire pour transmettre la signature pour authentifier le contenu vidéo et (4) la localisation des trames manipulées afin de donner à l'utilisateur une information permettant d'identifier rapidement les régions manipulées. L'inconvénient majeur qui reste à résoudre réside dans la capacité d'insertion qui est faible pour les séquences vidéo à faible activité temporelle et la sécurité de l'insertion qui reste insatisfaite. En effet, dans les deux variantes développées, les seuils introduits pour borner les MVs à tatouer sont considérés comme clé d'insertion. Mais, si la vidéo est manipulée par un attaquant malveillant qui connait la structure du codec, il pourra comparer les séquences vidéo d'origine et tatouées pour détecter les MVs modifiés et par conséquent détecter la signature insérée et la changer.

4.4. Méthode proposée : Version SASC- HMAC-SHA-256

L'augmentation de la capacité d'insertion, la sécurité, et la sensibilité d'insertion constituent les principales améliorations apportées à SASC-MD5-2 [138]. La modification apportée au système d'authentification afin d'augmenter la capacité d'insertion consiste à insérer plus de bits dans les MVs [139]. Celle-ci est réalisée en substituant deux bits de la marque aux deux bits de poids faible de chaque composante des MVs afin de cacher quatre bits de la marque dans un seul MV au lieu de deux bits. La sécurité est renforcée en opérant à deux niveaux. Le premier niveau consiste à employer la fonction de hachage HMAC-SHA-256 (keyed-hash message authentication code) plus sécurisée au lieu de la fonction MD5 [138]. Le deuxième niveau de sécurité ajouté repose sur l'utilisation d'une séquence pseudo-aléatoire pour sélectionner les MBs pour tatouer leurs MVs. Dans ce travail, la sélection des MVs est réalisée d'une manière différente que celle utilisée dans [138]. Le seuil est sélectionné dynamiquement en fonction de l'activité de mouvement la plus élevée de la trame. Le schéma synoptique du système SASC-HMAC-SHA-256 est illustré sur la figure 4.18. Il regroupe les mêmes opérations de traitements que dans le système SASC-MD5-2, à savoir : la génération de la signature numérique, le processus d'insertion et l'extraction et la vérification.

4.4.1. Génération de la signature

Pour la génération de la signature numérique, les mêmes caractéristiques spatiales, temporelles et de couleurs sont extraites du contenu vidéo en cours de compression pour générer la signature numérique. La différence réside au niveau de la fonction de hachage utilisée. Dans cette version, à la fin de chaque GOP, les données présentes dans la mémoire tampon, sont brouillées par la fonction de hachage sécurisée avec la clé HMAC-SHA-256 [145]. Cette dernière est créée sur la base de la fonction de hachage SHA-256. Le processus HMAC mélange d'abord le message présent dans le tampon avec la clé secrète choisie *K*. Ensuite, il hache la séquence résultante avec le SHA-256 et la mélange à nouveau avec la clé secrète *K*, puis il applique le

traitement SHA-256. La séquence de hachage de 256 bits de longueur résultante est utilisée comme marque fragile à insérer dans les MVs du contenu H.264/AVC.

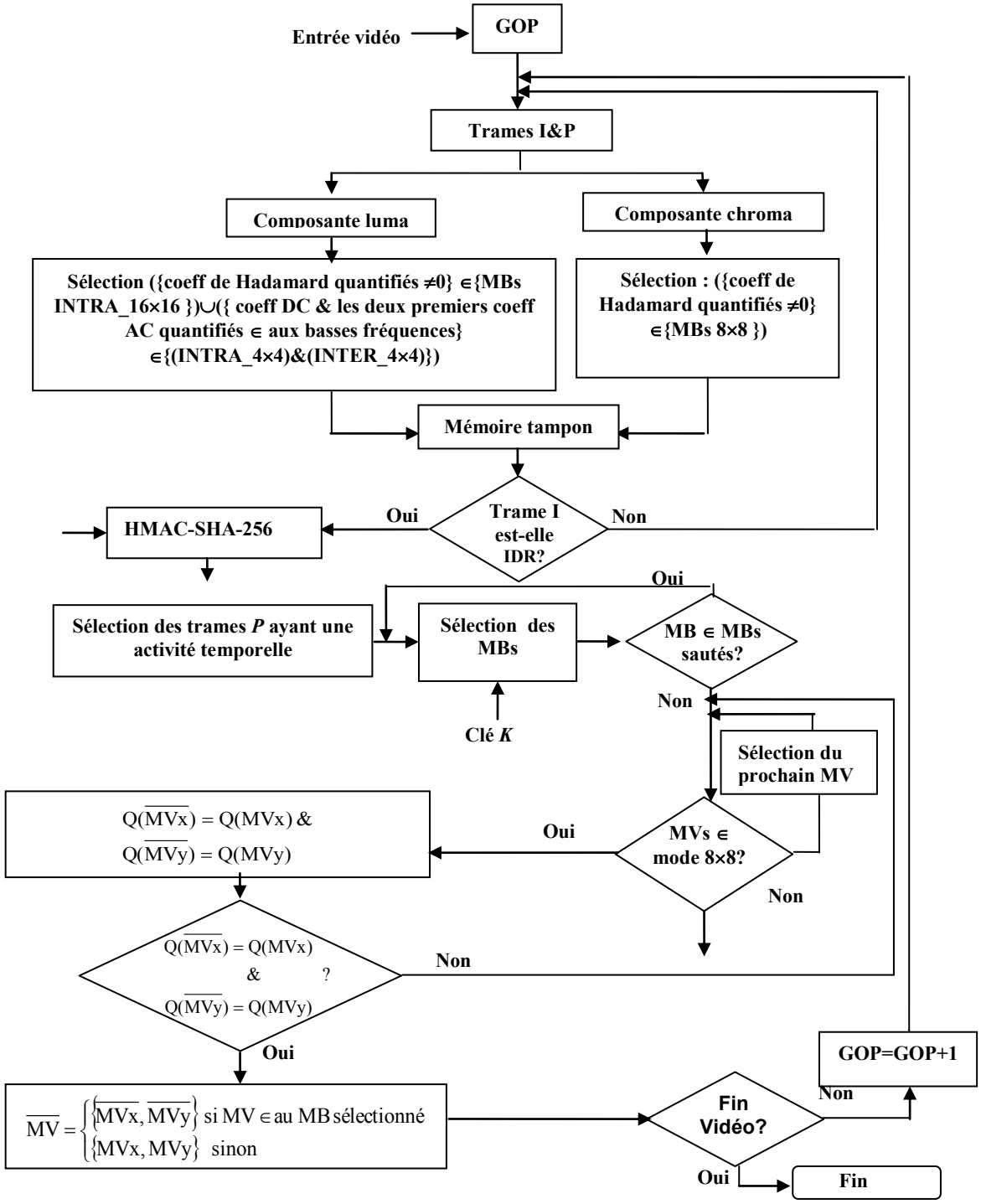


Figure 4.18 : Schéma bloc de la seconde version du système d'authentification stricte du contenu H.264/AVC.

4.4.2. Processus d'insertion

Les bits de la marque sont insérés dans les MVs des trames P caractérisées par une activité temporelle élevée. Une fois les trames P reconnues, les restrictions employées dans la première version à savoir les MBs sautés, leurs blocs voisins sont écartés et les MVs appartenant aux partitions 8×8 et leurs sous partitions 8×8, 8×4, 4×8 et 4×4 des trames P sélectionnées sont tatoués. Les trames P d'activité temporelle élevée sont déterminées selon leur intensité de mouvement. Pour une trame P donnée, la matrice d'activité spatiale est calculée comme suit [146]:

$$C = \{MV(i,j)\} \tag{4.11}$$

et

$$MV(x,y) = \sqrt{(MV_x(x,y))^2 + (MV_y(x,y))^2} \tag{4.12}$$

où *(i,j)* est l'indice du bloc dans un MB.

Pour chaque trame P, la matrice de l'activité temporelle moyenne est donnée par:

$$C_{avg} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} C(x,y) \tag{4.13}$$

où *M* et *N* sont la longueur et la largeur du MB.

L'activité temporelle de la trame, définie comme étant l'écart type de l'amplitude de mouvement, est calculée comme suit :

$$\sigma_{Fi} = \sqrt{\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (C(x,y) - C_{avg}(x,y))^2} \tag{4.14}$$

Les trames ayant une activité temporelle élevée sont sélectionnées si l'écart type satisfait à la condition suivante:

$$\sigma_{Fi} \geq T \tag{4.15}$$

où *T* est un seuil défini expérimentalement.

En appliquant les trois conditions d'insertion, il est difficile à l'œil humain de percevoir les distorsions introduites par le tatouage. Similaire à la version SASC-MD5, l'unité de longueur des composantes de MV (*MVx* et *MVy*) correspond à ¼ de pel. Avant d'être tatouée, chacune des composantes est quantifiée à la position du pixel entier le plus proche. L'opération de quantification est appliquée comme suit:

$$Q(MVx)=\begin{cases}(2+MVx)\,\&\,(0xFFFC) & MVx\geq 0\\ -(2-MVx)\,\&\,(0xFFFC) & MVx<0\end{cases} \tag{4.16}$$

$$Q(MVy)=\begin{cases}(2+MVy)\,\&\,(0xFFFC) & MVy\geq 0\\ -(2-MVy)\,\&\,(0xFFFC) & MVy<0\end{cases} \tag{4.17}$$

Où & est l'opérateur ET binaire.

Les bits de la signature sont ensuite substitués aux deux derniers bits LSB du vecteur *MVx* et *MVy* d'origine selon l'équation suivante :

$$\overline{MVx}=\begin{cases}Q(MVx)-\alpha & si\ Q(MVx)\geq 0\ \&\ \alpha=2\\ Q(MVx)+\alpha & sinon\end{cases} \tag{4.18}$$

$$\overline{MVy}=\begin{cases}Q(MVy)-\alpha & si\ Q(MVy)\geq 0\ \&\ \alpha=2\\ Q(MVy)+\alpha & sinon\end{cases} \tag{4.19}$$

où α représente les bits de la signature, ses valeurs sont -1, 0, 1 et 2 correspondant respectivement à une des paires de bits 11, 00, 01, 10.

Pour s'affranchir de la contrainte de synchronisation temporelle et minimiser les distorsions causées par insertion, la condition de synchronisation suivante doit être vérifiée :

$$Q(\overline{MVx})=Q(MVx)\quad et\quad Q(\overline{MVy})=Q(MVy) \tag{4.20}$$

Le MV à la sortie du codeur est composé des MVs tatoués $\left\{\overline{MVx},\overline{MVy}\right\}$ et des MVs non tatoués comme suit :

$$\overline{MV} = \begin{cases} \{\overline{MVx}, \overline{MVy}\} & si\ MV \in MBs\ sélectionnés \\ \{MVx, MVy\} & Sinon \end{cases} \quad (4.21)$$

4.4.3. Processus d'extraction et de vérification

Pratiquement, le même processus d'extraction utilisé dans la variante SASC-MD5-2 est appliqué dans cette seconde version. La différence réside dans les points suivants :

- L'application de la condition de sélection des MBs contenus dans les trames ayant une activité temporelle élevée;
- Les bits insérés sont extraits des deux derniers bits LSB des deux composantes *MVx*' et *MVy*' du MV reconstruit;
- L'utilisation de la fonction de hachage sécurisée HMAC-SHA-256 avec la clé *K* au lieu de la fonction MD5.

Les étapes utilisées pour la vérification sont les mêmes que celles utilisées dans les deux variantes de la version SASC-MD5.

4.5. Analyse et résultats expérimentaux du système SASC-HMAC-SHA-256

La version SASC-HMAC-SHA-256 développée est intégrée dans le même logiciel de référence H.264 JM-10.1 [142] avec les mêmes paramètres du codeur et les mêmes vidéo tests. La performance du schéma proposé est évaluée en fonction de (1) la capacité d'insertion maximale ; (2) la qualité perceptuelle de la vidéo tatouée exprimée en termes de rapport signal/bruit (PSNR) ; la qualité vidéo (VQM) et l'indice de SIMilarité structurelle (SSIM); (3) la conservation de la taille de l'espace mémoire de la vidéo tatouée; (4) la localisation des altérations et (5) la fragilité des manipulations du contenu.

Pour les séquences vidéo avec des activités de mouvement significatives (l'activité de mouvement de la trame σ_{Fi}≥3.870), de nombreux MBs sont attribués au mode 8×8. Par conséquent, un nombre important de MVs sont tatoués. Cependant, si

la vidéo contient une faible activité de mouvement, le nombre de MVs approprié diminue donc moins de bits peuvent être cachés dans le contenu vidéo. Le tableau 4.3 illustre les résultats comparatifs entre l'insertion effectuée dans SASC-MD5-1 [137], SASC-MD5-2 [138] et SASC-HMAC-SHA-256 [139]. Le schéma proposé de la version SASC-HMAC-SHA-256 montre une nette amélioration en termes de capacité d'insertion moyenne par GOP comparé aux deux variantes de la version SASC-MD5. Plus encore, cette augmentation n'a changé ni la qualité visuelle subjective ni la taille des fichiers vidéo tatoués. La capacité d'insertion moyenne par GOP, donnée dans le tableau 4.3 est presque trois fois plus élevée pour la plupart des séquences appartenant au groupe B par rapport au processus d'insertion proposé dans SASC-MD5-2. Pour les séquences vidéo du groupe A, malgré que la capacité d'insertion moyenne par GOP a pratiquement doublé, le nombre de MVs pour insérer les 256 bits de la marque reste insuffisant. Afin de surmonter ce problème et insérer les 256 bits de la marque dans chaque GOP des séquences du groupe A, nous avons augmenté le nombre de trames dans un GOP à 25 comme c'est illustré sur le tableau 4.4.

En ce qui concerne la qualité visuelle des séquences vidéo, nous remarquons que le PSNR reste pratiquement inchangé (tableau 4.3) pour les séquences du groupe A qui comprend des clips vidéo à faible mouvement et de nombreuses régions homogènes. Cependant, une légère diminution a été observée dans les valeurs du PSNR (0.025 à 0.315 dB) pour les séquences vidéo du groupe B. Pour une évaluation objective, les résultats de simulation de la qualité perceptuelle trame par trame sur la composante de luma (YPSNR) des deux séquences vidéo Table et Foreman sont illustrés sur la figure 4.19. Nous remarquons clairement sur cette figure que le PSNR moyen est inférieur à 0.268 dB pour toutes les trames Intra, des échantillons de luma (Y) des séquences vidéo d'origine et tatouées. Nous concluons donc que la qualité visuelle est préservée.

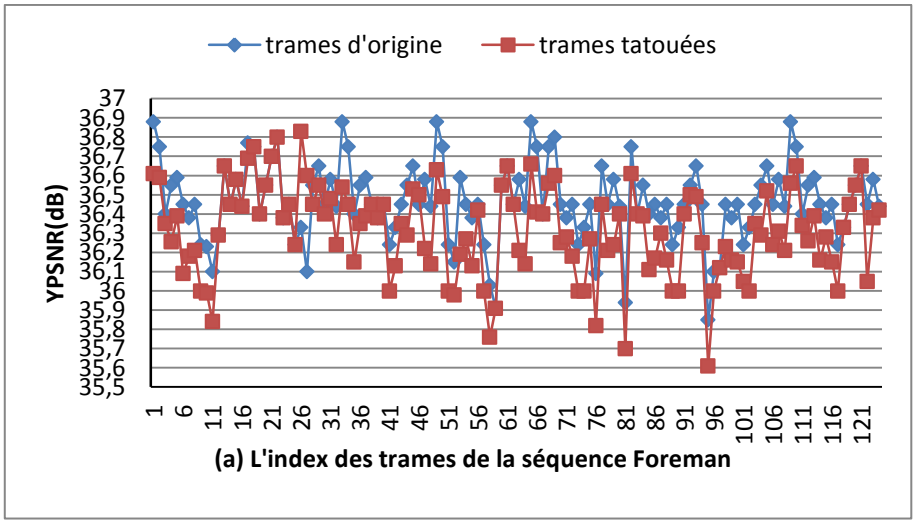
Tableau 4.3 : Capacité d'insertion et PSNR des séquences vidéo appartenant aux deux groupes avec $\sigma_{Fi} \geq 3.870$ et comparaison avec les résultats de SASC-MD5-[137].

	Séquences vidéo	Capacité moyenne par GOP (bits)		PSNR (dB)					Taille du contenu vidéo (Mo)	
		Version2 [139]	Version1 [137]	Vidéo d'origine	Version2 [139]	Diff-1	Version1 [137]	Diff-2	Vidéo d'origine	Version2 [139]
Groupe A	Miss America	162	88	40.056	40.056	0	40.04	0.016	5.43	5.43
	Claire	238	135	39.681	39.681	0	39.67	0.011	5.43	5.43
	Akiyo	248	133	38.205	38.205	0	38.17	0.035	5.43	5.43
	Bridge-close	518	264	34.847	34.847	0	34.85	0.03	5.43	5.43
Groupe B	Carphone	2960	1495	37.340	37.315	0.025	37.29	0.025	5.43	5.43
	Coastguard	5334	2806	34.181	34.026	0.155	34.02	0.006	5.43	5.43
	Flower	8578	4304	34.336	34.308	0.028	34.31	0.02	5.43	5.43
	Foreman	7688	2569	36.687	36.45	0.237	35.75	0.3	5.43	5.43
	Suzie	2760	1349	37.357	37.141	0.216	37.12	0.021	5.43	5.43
	Table	9992	5182	35.23	34.915	0.315	34.91	0.05	3.22	3.22

Version1 : SASC-MD5-1.
Version2 : SASC-HMAC-SHA-256.
Diff-1 : Différence entre les vidéos d'origine et les vidéos tatouées dans la version SASC-HMAC-2 [138].
Diff-2 : Différence entre La variante SASC-MD5-1 [137] et la version SASC-HMAC-SHA-256 [139].

Tableau 4.4 : Capacité d'insertion et PSNR des séquences vidéo du groupe A avec un GOP de 25 trames ($\sigma_{Fi} \geq 3{,}870$).

Séquences vidéo du groupe A	Capacité moyenne par GOP (bits)	PSNR (dB)	Taille du contenu vidéo (Mo)
Miss America	265	40.056	5.43
Claire	283	39.681	5.43
Akiyo	301	38.205	5.43
Bridge-close	555	34.847	5.43



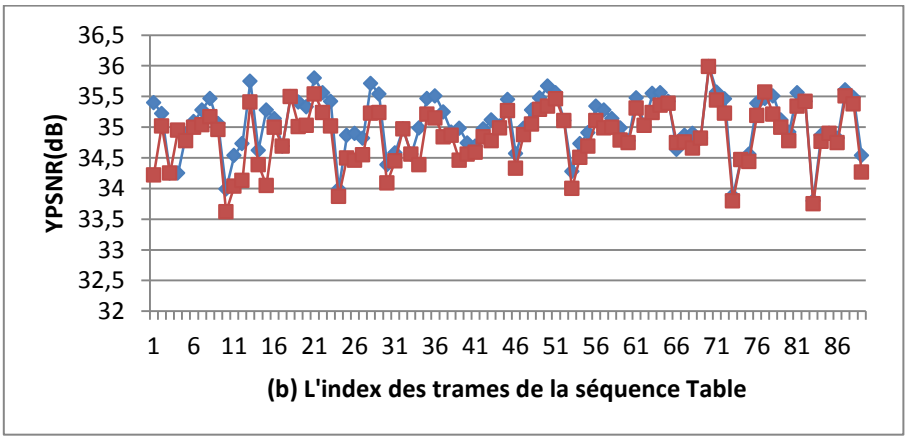


Figure 4.19 : YPSNR trame par trame des séquences d'origine et tatouées de Foreman et Table avec *QP=28*.

Les logiciels VQM (Video Quality Metric) et SSIM (Structural SIMilarity) sont utilisés pour mesurer les qualités perceptuelles temporelles et structurelles. Une valeur de VQM qui est proche de 0 signifie la présence d'une distorsion insignifiante. Dans nos expériences, l'évaluation est effectuée uniquement sur les séquences du groupe B. Comme c'est illustré sur le tableau 4.5, les valeurs de VQM sont comprises dans la plage 0.292 à 0.398 et sont toutes inférieures à 0.4, ce qui se traduit par une différence insignifiante entre les séquences vidéo d'origine et les séquences vidéo tatouées. Il est clair que notre approche, en maintenant de faibles valeurs de VQM, conduit à une bonne imperceptibilité. L'indice SSIM est généralement utilisé dans l'évaluation des images traitées par rapport aux images d'origine en termes de

similitudes en se basant sur trois composantes : la luminance, le contraste et la structure de l'image. SSIM combine ces similitudes pour retourner une valeur unique. Une valeur de SSIM proche de 1 indique une grande similitude des deux vidéos par contre une valeur proche de 0 signifie un écart total. Le tableau 4.5 montre clairement que les valeurs obtenues se situent entre 0.968 et 0.998 avec la plupart des valeurs situées au-dessus de 0.97. Nous concluons qu'il n'y a pas de dégradation de la qualité visuelle sur la vidéo après insertion de la marque car toutes les valeurs sont très proches de 1.

Tableau 4.5 : Qualité visuelle calculée par les métriques VQM et SSIM.

	Vidéos	VQM	SSIM		Vidéos	VQM	SSIM
Groupe A	Miss America	0.192	0.998	Groupe B	Carphone	0.292	0.987
	Claire	0.206	0.978		Coastguard	0.234	0.998
	Akiyo	0.187	0.995		Flower	0.290	0.995
	Bridge-close	0.245	0.981		Foreman	0.351	0.976
					Suzie	0.289	0.983
					Table	0.398	0.968

L'aptitude de localiser le tatouage fragile proposé est évaluée sur la séquence Foreman (Figure 4.20 (a)). Dans la première expérience, nous appliquons des altérations spatiales comprenant l'attaque par DC, le recadrage et la rotation des trames dans la séquence tatouée. La figure 4.20(b) montre la séquence Foreman tatouée en tenant compte des conditions d'insertion. Les figures 4.20(c) et 4.20(d) montrent les résultats obtenus à partir de différents tests d'altérations spatiales. La figure 4.20(c) illustre la distorsion traduite par l'application de l'attaque DC sur le $52^{ème}$ MB de la $10^{ème}$ trame du premier GOP. L'attaque DC est appliquée en modifiant les coefficients d'origine du bloc 4×4 tout en maintenant sa valeur moyenne d'origine. Le résultat révèle que l'altération a conduit à une dégradation apparente de la qualité visuelle aussi bien qu'au contenu de la signature du GOP falsifié. Toutes les signatures insérées sont semblables à celles extraites des GOPs de la séquence tatouée à l'exception de la première signature du GOP qui est différente. La région modifiée est donc détectée et localisée dans le premier GOP. La manipulation de cadrage est illustrée sur la figure 4.20(d). Elle consiste à supprimer le mot 'SIEMENS'

affiché sur l'arrière-plan de la 8ème trame du premier GOP. La méthode a aussi la capacité de détecter les manipulations illégales telles que la rotation de la 9ème trame (Figure 4.20(e)). La signature numérique dans le premier GOP (manipulé) diffère de la signature d'origine.

Figure 4.20 : Trames du 1er GOP de la séquence Foreman: (a) trames d'origine, (b) trames tatouées, (c) attaque par DC, (d) attaque de cadrage et (e) attaque de rotation.

Dans la deuxième expérience, la sensibilité aux attaques temporelles est étudiée par re-compression, transcodage, ré-ordonnancement et suppression de trames. Pour l'attaque de compression, la séquence Table tatouée est compressée par la norme MPEG-2. En conséquence, les signatures des GOPs sont complètement supprimées du fait de la modification de la structure des trames et les MVs dans les GOPs. L'attaque de transcodage est appliquée sur la séquence Table, la vidéo est re-compressée avec les mêmes paramètres de compression utilisés lors du processus compression-insertion, sauf la valeur du pas de quantification *QP* qui est fixé à 32 au lieu de 28 dans la version d'origine. De même, les marques sont brouillées comme dans l'expérience précédente. Par conséquent, l'algorithme proposé est capable de détecter efficacement l'attaque de transcodage tout en gardant la perception vidéo inchangée (Figure 4.21). Le ré-ordonnancement et suppression de trames sont des attaques intentionnelles qui exploitent la redondance temporelle des trames de la séquence pour détruire la marque insérée sans provoquer de dégradation visuelle. Dans le schéma proposé, chaque GOP contient une marque indépendante des autres pour s'authentifier. L'attaque de ré-ordonnancement est produite en échangeant l'ordre de la 9ème par la 10ème trame dans le premier GOP de la séquence Foreman (Figure 4.22(a)). L'attaque de suppression est appliquée sur les 56ème et 58ème trames dans le quatrième GOP de la même séquence (figure 4.22(b)).

6ème trame 7ème trame 8ème trame

Figure 4.21 : Trames du 1er GOP de la séquence Table après l'attaque de transcodage.

Figure 4.22 : Attaque de ré-ordonnancement appliquée sur le 1er GOP et attaque de suppression de trames appliquée sur le 4ème GOP de la séquence Foreman.

La corrélation entre la marque insérée dans le GOP et celle extraite du GOP correspondant est mesurée par la corrélation normalisée *NC* (Normalised correlation). La valeur 0 du *NC* indique une destruction complète de la marque insérée et la valeur 1 signifie que la marque est identique à l'originale. La figure 4.23 illustre les valeurs de *NC* résultant des attaques temporelles. Les résultats obtenus révèlent que les valeurs de *NC* sont proches de zéro dans les GOPs attaqués temporellement, ce qui se traduit par la destruction complète des marques insérées.

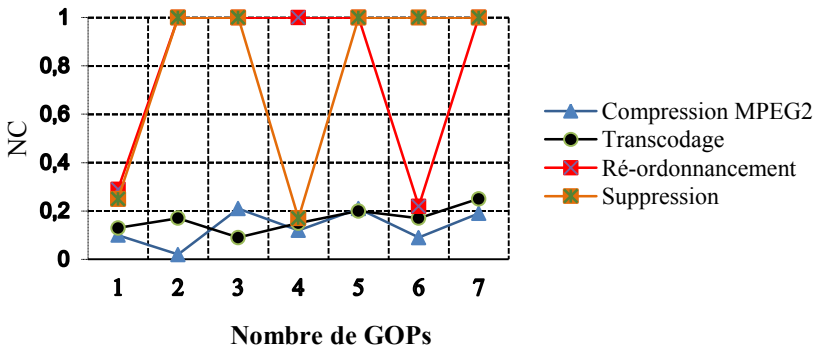


Figure 4.23 : Sensitivité des attaques temporelles appliquées sur la séquence Table.

La figure 4.24 illustre l'avantage d'inclure les composantes de chroma comme caractéristiques dans la génération de la marque. L'attaque est réalisée en changeant la couleur du mur qui se trouve derrière Foreman dans la 7ème trame du 1er GOP de la séquence. Cette attaque a fourni une signature du GOP modifié différente de la signature d'origine.

6ème trame 7ème trame 8ème trame

Figure 4.24 : Foreman après modification de couleurs.

La complexité d'insertion est liée au temps de traitement requis par les différentes étapes du système SASC- HMAC-SHA-256. Pour le processus de codage, le temps de traitement englobe les temps nécessaires pour générer la marque de chaque GOP et son insertion dans les MVs. Le temps de traitement du tatouage correspond à la différence des temps pris par le codeur pour compresser la vidéo avec et sans tatouage. La simulation a été effectuée sur un ordinateur personnel équipé d'un processeur dual-core de 3,2 GHz et de 2 Go de RAM. Le tableau 4.6 donne les temps d'exécution moyens correspondants. Cette évaluation est effectuée sur les séquences vidéo du groupe B uniquement. Selon les résultats, nous constatons que le temps du traitement compression-insertion augmente de 3.3 à 11.4 secondes par rapport au temps de compression seul. Cette légère augmentation provient non seulement de la génération de la marque, mais aussi de l'identification du meilleur mode 8×8 et la sélection des MVs à tatouer. D'un autre côté, les résultats obtenus (tableau 4.6) montrent que l'extraction de la marque et la détection des trames altérées des GOPs se font de façon transparente puisque le temps d'exécution supplémentaire dans ce cas est insignifiant. Il est intéressant de noter que le temps d'exécution supplémentaire est le temps relatif au traitement du

GOP. Comme mentionné précédemment, l'augmentation du temps est rencontrée lors de la localisation des trames manipulées dans le GOP non authentifié.

Tableau 4.6 : Temps de codage et de décodage des séquences vidéo appartenant au groupe B avec et sans insertion.

Vidéo Groupe B	Temps de codage (s)			Temps de décodage (s)		
	Sans insertion	Avec insertion	Diff	Sans insertion	Avec insertion	Diff
Carphone	21.60	24.86	3.26	15.1	15.31	0.21
Coastguard	22.21	26.31	4.10	16.2	16.44	0.22
Flower	23.39	30.33	6.94	16.47	16.73	0.26
Foreman	30.48	35.54	5.06	17.6	17.78	0.18
Table	32.80	44.20	11.4	17.81	18.05	0.24

Diff : Différence

4.6. Comparaison avec des travaux antérieurs.

Le tableau 4.7 illustre la comparaison du SASC-HMAC-SHA-256 proposé avec d'autres techniques d'authentification fragiles utilisant soit un tatouage dépendant du contenu, soit un tatouage indépendant du contenu pour le codec H.264/AVC. En premier lieu, la comparaison est faite avec les techniques de tatouage dépendantes du contenu. Le schéma de Ramaswamy et al. [133] nécessite plus de mémoire pour la vidéo tatouée alors que la technique proposée n'entraîne pas d'augmentation de la taille du fichier, l'espace mémoire de la vidéo tatouée est le même que celui de la vidéo d'origine avec d'excellentes performances en termes de qualité vidéo. En effet, les résultats montrent (tableau 4.7) que le PSNR a diminué de 0.315 dB seulement, ce qui se traduit par une dégradation perceptible insignifiante. En termes de détection de manipulations, le schéma proposé dévoile une sensibilité élevée aux manipulations spatiales, temporelles et colorimétriques par rapport à la méthode de Kuo et al. [123] qui n'inscrivait pas les effets des altérations temporelles et colorimétriques.

La seconde comparaison est faite avec les méthodes de tatouage indépendantes du contenu. Les résultats montrent (tableau 4.7) que même si la plupart des algorithmes sont sensibles à certaines manipulations spatiales telles que

le transcodage ou la compression, ils ne fournissent aucune information nécessaire pour caractériser les attaques, ou simplement négligent d'inclure certaines attaques telles que l'altération spatiale [90], les altérations temporelles et de couleur [123]. En revanche, l'approche développée est capable de localiser l'échec d'authentification au niveau du GOP, tout en conservant la qualité vidéo et la taille du fichier du flux binaire compressé inchangés. Le temps supplémentaire induit par la localisation des trames manipulées dans le ou les GOPs identifié(s) non authentique(s) est le seul inconvénient induit par la méthode développée. En ce qui concerne la qualité visuelle, bien que Horng et al. [134] ont obtenu de meilleures performances par rapport à notre approche tant en termes de YPSNR que de VQM, cela implique une distorsion perceptuelle non significative car l'insertion dans leur approche était effectuée sur les trames de type I seulement. D'autre part, notre approche basée sur l'insertion dans les MVs assure une meilleure qualité perceptive des séquences vidéo. En outre, les valeurs de SSIM de notre méthode couvrent approximativement le même intervalle que dans la méthode de Horng et al. [134] et leurs valeurs moyennes rapportées sont très semblables aux nôtres avec une capacité d'insertion plus élevée révélée par notre méthode. En effet, en dépit de sa capacité à préserver l'efficacité d'insertion, dans le travail de Horng et al., la capacité d'insertion est moindre, plus particulièrement avec les séquences vidéo du groupe B.

En termes de complexité de calcul, la méthode développée dans [134] a atteint une meilleure performance. Ceci vient du fait que les différentes étapes des processus d'insertion et de détection sont réalisées au niveau du flux binaire en utilisant les éléments syntaxiques du NAL.

Tableau 4.7 : Comparaison de la méthode proposée avec les techniques d'authentification stricte du contenu H.264/AVC.

Paramètres	Qiu [94]	Zhang [125]	Wang [124]	Kim [122]	Kuo [123]	Ramaswamy [133]	Horng [134]	SASC-HMAC-SHA-256 [139]
Type de tatouage	Indépendant du contenu				Dépendant du contenu (extractions des caractéristiques dans le domaine transformé)			
Type de caractéristiques					luma	luma	luma	luma & chroma
Type de données					Coeffs AC	Coeffs DC & AC	Coeffs AC	Coeffs DC & AC de luma & chroma
Espace d'insertion	Estimation de mouvement	Coeffs AC. (Trame I)	Coeffs AC. (Trames P & B)	MVs des MBs des trames inter-codées	MVs	SEI	Blocs Intra_4×4	MVs
Capacité (bits)	1 bit /MB	Faible (150 bits)	Faible (400 bits)	Elevée	Non évoquée	Signatures pour toute la vidéo (y*160) bits y : Nombre total de GOPs dans la vidéo	Elevée	Elevée: Signatures pour toute la vidéo y*256) bits y : Nombre total de GOPs dans la vidéo
Augmentation de la taille du fichier	Non	Non	+ 1%	+1% ~2%	+ 4%	Large	Non	Non
Détection des manipulations spatiales	Non	oui	Suppression de GOPs & recompression	Suppression de trames & recompression	transcodage	oui	oui	oui
Détection des manipulations temporelles	Non	Non	Non	Non	Non évoquée	Non	oui	oui
Détection des manipulations de la couleur	Non	Non	Non	Non	Non évoquée	Non	non	oui
Localisation des altérations	Non	Oui (Au niveau trame)	Oui (Au niveau trame)	Non	Oui (Au niveau trame)	Oui (Au niveau trame)	oui	Oui (Au niveau trame)
PSNR (dB)	Dégradation insignifiante	Maintenue	- 0.12	-0.06	- 0.27	Maintenue	-0.005	- 0.315
Complexité	Faible	Faible	Faible	Elevée	Non évoquée	Faible	Faible	Elevée (localisations des trames manipulées)

4.7. Conclusion

Deux versions du système d'authentification des séquences vidéo compressées par la norme H.264/AVC opérant au cours de la compression ont été développées. Le premier système assure l'authentification stricte, la vidéo est soit authentifiée soit non. L'inconvénient dans ce système est la non localisation des régions manipulées, par contre le second système assure un maximum de critères d'efficacité. En effet, en termes de sensibilité, le système est capable de déceler des manipulations pouvant modifier l'interprétation que l'on a de la vidéo, telles que des recadrages ou des retouches locales (rotation). En termes d'imperceptibilité, l'insertion de la marque est invisible et en termes de qualité, la vidéo tatouée a conservé sa qualité après insertion tout en maintenant le taux de bits de la vidéo après insertion inchangé. Le système proposé est en mesure de donner à l'utilisateur une information permettant d'identifier les régions manipulées. L'inconvénient majeur qui reste à résoudre réside dans la complexité de traitement au niveau de la détection des trames manipulées qui se fait en temps différé.

CONCLUSION GENERALE ET PERSPECTIVES

Dans cette thèse il a été question de la protection du contenu vidéo compressé par la norme H.264/AVC, une problématique qui a pris de plus en plus d'importance depuis le développement d'Internet et des réseaux d'échange. Le tatouage numérique, technologie permettant d'insérer de façon invisible une information dans le contenu multimédia, a été proposé comme une solution, qui combiné au cryptage, assurerait une protection supérieure de la propriété intellectuelle et de l'authenticité du contenu vidéo numérique.

Notre premier objectif dans le cadre de cette thèse était de contribuer à l'amélioration des techniques actuelles de tatouage robuste et fragile en proposant de nouvelles approches. Une amélioration effective de la robustesse du tatouage « robuste » a été atteinte en partie par l'augmentation de la capacité à résister aux différentes attaques que peut subir le contenu tatoué. A l'opposé, l'augmentation de la robustesse du tatouage fragile est que la marque insérée ne résiste à aucune attaque. Avant de montrer que nos objectifs ont été atteints, les deux premiers chapitres ont été dédiés à l'introduction des différents concepts relatifs au tatouage numérique d'image et la classification des différents schémas rencontrés dans la littérature, ainsi qu'au principe général de fonctionnement du codeur H.264/AVC et les principaux modules le constituant.

La deuxième partie introduit notre contribution dans le domaine du tatouage numérique et valide nos résultats. Ces travaux sont aussi divisés en deux volets. Le premier volet se rapporte à nos travaux sur la protection des droits d'auteur du contenu vidéo H.264/AVC où les contraintes sécurité, robustesse et capacité sont prises en charge. Nous avons proposé une nouvelle approche de tatouage robuste opérant dans le domaine fréquentiel du codeur de la norme H.264/AVC. L'approche développée consiste à traiter la marque avant de l'insérer dans le contenu H.264/AVC.

Le second volet quant à lui s'est penché sur la problématique d'authentification du contenu H.264/AVC, discipline permettant de vérifier l'intégrité du contenu. Nous avons proposé deux nouvelles versions du système d'authentification stricte du contenu (SASC) H.264/AVC basées sur deux fonctions de Hachage différentes : la version SASC reposant sur le MD5 (SASC-MD5), et la version SASC-HMAC-SHA-256 se basant sur la fonction de hachage sécurisée HMAC-SHA-256 (keyed-hash message authentication code). Les approches de tatouage réalisées sont sensibles à toute altération, et sont basées sur l'insertion d'une marque fragile dans le domaine temporel de la norme H.264/AVC. Cette marque, insérée de manière non robuste, est le résultat d'une fonction de hachage cryptographique particulière calculée sur un ensemble de coefficients représentatifs du contenu. Lors de l'extraction, si les coefficients représentatifs ont été modifiés entre temps, il en résulterait que la signature qui a été originalement insérée et la signature calculée à l'extraction seraient différentes, conduisant à la conclusion que le contenu est non authentique.

L'un des points forts du premier système SASC-MD5 est que son authentification du contenu H.264/AVC est stricte. Par contre, sa limitation est la non localisation des régions manipulées. D'autre part, le second système HMAC-SHA-256 proposé vérifie un maximum de critères d'efficacité d'authentification tels que la sensibilité (le système est capable de donner à l'utilisateur une information permettant d'identifier les images manipulées), l'invisibilité de la marque, la conservation de la qualité visuelle et du flux binaire après insertion.

Il est évident que les applications liées à la sécurité, dans le cas du tatouage numérique du contenu multimédia, sont celles qui ont drainé le plus d'intérêt pour la recherche et l'industrie. En effet, de nombreux industriels ont vu dans le tatouage numérique une solution venant en soutien à la cryptographie pour sécuriser le contenu vidéo afin de préserver les droits d'auteur, et vérifier l'intégrité du contenu et son authentification. Bien que notre souci dans cette thèse ait été d'améliorer les performances des techniques existantes, il est évident que des efforts supplémentaires doivent être consentis pour assurer un produit compétitif sur le

marché. Pour cela, certaines perspectives peuvent être citées afin d'améliorer le présent travail.

Une première perspective concerne l'amélioration de la méthode de tatouage robuste proposée afin d'atteindre une robustesse plus élevée tout en garantissant un meilleur compromis invisibilité/robustesse à toutes les attaques. L'idée est d'insérer la marque dans les régions saillantes de la vidéo. En effet, le tatouage dans ces zones permet de fournir certaines informations relatives au contenu et qui sont utiles pour le tatouage. Par exemple, localiser les différents points physiques dans la scène et tatouer par la suite, par la même marque, tous les pixels représentant le même point physique dans les différentes positions de la scène.

Comme deuxième perspective, nous visons l'exploitation des résultats obtenus par la technique de tatouage robuste pour concevoir une approche de tatouage dédiée pour l'application de traçage des traîtres dans le contexte de la Video On Demand (VOD). Dans cette application, chaque utilisateur de la vidéo acquiert une version personnalisée du contenu, contenant un identifiant personnel inséré grâce à une technique de tatouage robuste. Ainsi si une copie est rediffusée illégalement, il sera alors possible de remonter à l'utilisateur malhonnête, puis calculer un score pour chaque utilisateur afin de déterminer les personnes qui ont participé dans la génération de la copie illégale.

Une dernière perspective concerne l'exploitation des résultats obtenus par la technique d'authentification stricte pour concevoir une approche de tatouage semi fragile dédiée à la détection des manipulations malicieuses ayant trait à la fraude tout en tolérant un nombre limité de traitements tels que la compression, le changement de format, l'ajout du tatouage pour couvrir un large spectre d'applications. L'utilisation de telles méthodes est principalement motivée par le fait que les vidéos sont généralement transmises et stockées sous forme compressée et que pour la majorité des applications, les pertes liées au processus de compression n'affectent pas l'intégrité du contenu au sens de son interprétation.

APENDICE A :

LISTE DES SYMBOLES ET DES ABREVIATIONS

AVC	Advanced Video Coding
CABAC	Context-based Adaptive Binary arithmetic coding
CAVLC	Context-Adaptive Variable-Length Coding
CIF	Common Intermediate Format
CM	Compensation de Mouvement
dB	Décibel
DCT	Discrete Cosine Transform
DFT	Discrete Fourier transform
DWT	Discrete wavelet Transform
DVD	Digital Video disc
EM	Estimation de Mouvement
FRExt	Fidelity Range Extensions:
GOP	Group Of Pictures
H.26H4, H.263	Normes de compression vidéo numérique définies par ITU. H.264 est issue d'un travail conjoint entre ITU et ISO
HMAC-SHA-256	Keyed-Hash Message Authentication Code : fonction de hachage sécurisée
IUT	Union Internationale des Télécommunications
ISO	International Organization for Standardization
IDR	Instantaneous Decoding Refresh
MB	Macrobloc
MD5	Message Digest 5 : fonction de hachage cryptographique
MPEG	Motion Picture Expert Group
MV	Motion Vector : Vecteur de mouvement
MVD	Motion Vector Difference
MVP	Vecteur de mouvement prédit
NAL	Network Abstaction Layer
PSNR	Peak Signal to Noise Ratio : Mesure objective de qualité d'image et de vidéo
QCIF	Quart de CIF : Résolution de séquence vidéo 176×144 pixels
SSIM	Structural Similarity Index Metric : l'index de similarité

	structurelle
SVH	Système Visuel Humain
Trame Intra	Trame prédite à partir de la trame courante. Codée à partir d'un algorithme de prédiction spatiale
Trame Inter	Trame prédite à partir des trames de références précédentes codées ou futures.
VQM	Video quality metric : Mesure objective de qualité d'image et de vidéo

APENDICE B :

FONCTIONS DE HACHAGE CRYPTOGRAPHIQUE

B.1. Introduction

Une fonction de hachage est une méthode permettant de caractériser une information, une donnée. En faisant subir une suite de traitements reproductibles à une entrée, elle génère une empreinte servant à identifier la donnée initiale. De telles fonctions datent de la fin des années 1980 (algorithme MD2) mais l'idée est plus ancienne, et a germé dès l'apparition des codes correcteurs d'erreurs (théorie de l'information).

Une fonction de hachage prend en entrée un message de taille quelconque, applique une série de transformations et réduit ces données. A la sortie, une chaîne de caractères hexadécimaux (le condensé) est générée, Celle-ci résume en quelque sorte le fichier. Cette chaîne a une taille fixe qui varie selon les algorithmes (128 bits pour MD5 et 160 bits pour SHA-1).

B.2 Utilité des Fonctions de hachage

Les fonctions de hachage sont très utilisées en informatique et en cryptographie. L'une des utilisations est de pouvoir vérifier que le transfert d'un fichier s'est bien passé : avant d'envoyer le fichier, l'émetteur calcule en local l'empreinte du fichier et envoie ces deux informations (le fichier et l'empreinte) vers le destinataire. Une fois les informations reçues par le destinataire, celui-ci va recalculer l'empreinte sur la base du fichier reçu et en utilisant le même algorithme de hachage et comparer le résultat de son calcul avec l'empreinte qui lui a été envoyée par l'émetteur : si les empreintes correspondent cela indique qu'il n'y a pas eu d'erreur durant le transfert.

B. 3 Propriétés des fonctions de Hachage

Les fonctions de hachage possèdent de nombreuses propriétés :

- Elles peuvent s'appliquer à n'importe quelle longueur de message M ;
- Elles produisent un résultat de longueur constante ;
- Il doit être facile de calculer h = H(M) pour n'importe quel message M ;
- Résistance aux préimages : étant donné y il est difficile de trouver x tel que y=H(x) On parle de propriété à sens unique.
- Résistance aux collisions : il est difficile de trouver x et x' différents tels que H(x)=H(x')
- – Résistance aux secondes préimages : étant donnés x et y tels que y=H(x), il est difficile de trouver x' ≠ x tel que H(x') = y

B.4 Les Fonctions de hachage pour l'authentification et l'intégrité des données

Semblable aux algorithmes de chiffrements symétriques, les fonctions de hachage évoluent dans le temps par la puissance croissante des attaques par force brute. Elles sont passées du MD5 à SHA et à Ripemd-160. Il est aujourd'hui conseillé d'utiliser le SHA-256 (SHA-2), suite aux récentes attaques développées pour SHA-1.

B.4.1 MD5

Conçu par Ronald Rivest (le R dans RSA) en 1991, c'est le dernier d'une série (MD2, MD4). Cet algorithme produit un condensé de 128 bits. Il était il y a encore quelques temps l'algorithme de hachage le plus largement répandu. La cryptanalyse et l'attaque par force brute (2004) l'ont affaibli. Ses spécifications sont disponibles sur Internet dans le RFC 1321 [147].

B.4.1.1 Fonctionnement Les étapes principales de l'algorithme MD5

Le déroulement général de l'algorithme est illustré à la figure B.1.

- Complétion : Le message est constitué de b bits m1...mb. le message est complété par un 1, et suffisamment de 0 pour qu'il soit étendu à une longueur congruente à 448, modulo 512.
- Ajout de la longueur : l'ajout à ce message la valeur de b, codée en binaire sur 64 bits (on a donc b qui peut valoir jusque 264... ce qui est énorme). En conséquence, la taille totale du bloc atteint 512 bits. Si la longueur nécessite plus de 64 bits, Seulement les 64 bits de poids faible sont pris en compte.
- Initialisation : initialiser 4 buffers de 32 bits chacun (A,B,C, D), qui constitue la valeur initiale (IV)
- Calcul itératif : traiter le message par blocs de 512 bits. Il y a 4 rondes de 16 opérations. Pour chaque bloc de 512 bits du texte, les opérations suivantes sont appliquées (figure B.1) ;
- Ecriture du résumé : Le résumé final sur 128 bits est obtenu en concaténant les résultats des additions des buffers A,B,C,D de 32 bits avec la variable chainée obtenue par la manipulation du $q^{éme}$ bloc.

B.4.2 Sécurité du MD5

Le condensé MD5 dépend de tous les bits du message, ce qui apporte un effet d'avalanche : chaque bit du haché est une fonction de chaque bit d'entrée.

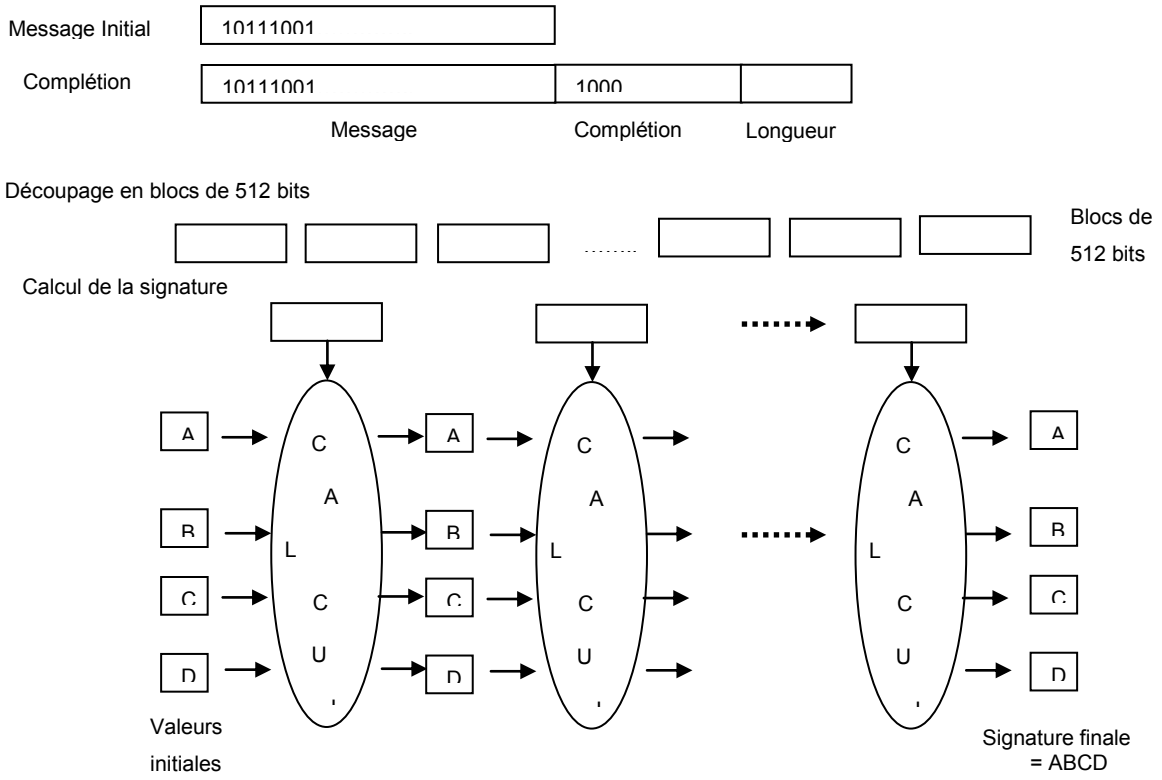


Figure B1. Principe de fonctionnement du MD5

B.5 Les algorithmes SHA

SHA (Secure Hash Algorithm) a été conçu par NIST et NSA en 1993, et révisé 1995 pour étendre ses capacités en matière de sécurité. Ses spécifications sont publiées dans le RFC 3174 [148]. L'algorithme est le SHA, la norme est SHS (Secure Hash Standard). Contrairement au MD5 qui produit des condensés de 128 bits, le SHA produit des valeurs condensées de 160 bits. Jusqu'à 2005, il était l'algorithme généralement préféré pour le hachage, mais des rumeurs de cassage le font peu à peu évoluer vers des versions plus sophistiquées. Il convient aujourd'hui (depuis 2009) d'utiliser le SHA-2, fonction de hachage qui produit 256 bits en sortie.

B.5.1 Fonction de hachage pour les MAC

Un HMAC, de l'anglais keyed-hash message authentication code (code d'authentification d'une empreinte cryptographique de message avec clé), est un type de code d'authentification de message (CAM), ou MAC en anglais (Message Authentication Code), calculé en utilisant une fonction de hachage cryptographique en combinaison avec une clé secrète [149]. Le code MAC peut être utilisé pour vérifier simultanément l'intégrité de données et l'authenticité d'un message. N'importe quelle fonction itérative de hachage, comme MD5, SHA-1 ou SH-2, peut être utilisée dans le calcul d'un HMAC ; le nom de l'algorithme résultant est HMAC-MD5, HMAC-SHA-1 ou HMAC-SHA-2. La qualité cryptographique du HMAC dépend de la qualité cryptographique de la fonction de hachage et de la taille et la qualité de la clé.

Il est souhaitable de créer des MACs à partir de fonctions de hachage plutôt qu'à partir de fonction de chiffrement par bloc. Plusieurs raisons sont avancées telles que la rapidité des fonctions de hachage, et l'absence de contrôle à l'exportation. La différence majeure entre les MAC et les Hachés étant la gestion d'une clé secrète, il faut intégrer cette clé dans l'algorithme de hachage. Une idée simple fut avancée, et connue sous le nom de *HashSur*. Le principe était de concaténer directement le message à la clé (i.e. Hash(clé|Message)). Cependant quelques faiblesses ont été trouvées, ce qui a amené au développement de HMAC.

La fonction HMAC est donnée par l'équation suivante [150] :

$$HHMACK(M) = H(K' \oplus opad, H(K' \oplus ipad, M)) \quad \text{(B.1)}$$

où *ipad* et *opad* sont des constantes

Les étapes de construction de HMAC est comme suit :

- Ajouter des 0 à K pour avoir K' de 512 bits ;
- Calculer K' $\oplus$ ipad ;
- Concaténer K' $\oplus$ ipad et M ;
- Appliquer H au résultat pour obtenir T ;
- Calculer K' $\oplus$ opad ;

- Concaténer K' $\oplus$ opad et T ;
- Appliquer H au résultat pour obtenir le MAC.

B.5.2 Sécurité de HMAC

Si la fonction de hachage H est résistante aux collisions avec probabilité ε, alors on ne peut pas forger de MACs avec une attaque à messages choisis avec probabilité supérieure à 2ε. Attaque encore possible ici dès que $2^{n/2}$ MACs ont été calculés (car $\varepsilon = 1/ 2^{n/2}$ au mieux) [151].

B.5.2 HMAC-SHA-256

HMAC-SHA-256 est un type d'algorithme de hachage à clé qui est construit à partir de la fonction de hachage SHA-256 et utilisé en tant qu'un Hash-based Message Authentication Code (HMAC). Le processus HMAC mélange une clé secrète aux données du message, hache le résultat avec la fonction de hachage, mélange de cette valeur de hachage avec la clé secrète à nouveau, puis applique la fonction de hachage une deuxième fois. Le hachage de sortie est la longueur de 256 bits.

Toute modification apportée aux données ou à la valeur de hachage entraîne une incompatibilité, car la connaissance de la clé secrète est requise pour modifier le message et reproduire la valeur de hachage correcte. Par conséquent, si les valeurs de hachage d'origine et calculées correspondent, le message est authentifié.

HMAC-SHA-256 accepte des clés de toute taille et produit une séquence de hachage longueur de 256 bits.

BIBLIOGRAPHIE

1. Pirat, P., “Les bases techniques de la télévision numériques”, Technical report, (2000).

2. Gall, D. L., “MPEG: A video compression standard for multimedia applications”, Communications of the ACM, V. 34, n° 4, (1991), 46 - 58.

3. ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG2), “Generic Coding of Moving Pictures and Associated Audio Information” Part 2: Video”, (November 1994).

4. ISO/IEC JTC 1, “Coding of Audio-Visual Objects — Part 2: Visual,” ISO/IEC 14496-2, MPEG4 Visual Version 1, (April 1999); Amendment 1, Version 2, (February 2000); Amendment 4, streaming profile, (January 2001).

5. ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-AVC), “Advanced video coding for generic audiovisual services,” V. 1(May, 2003); V. 2, (January 2004), V. 3 (with FRExt), (September. 2004), V. 4, (July 2005).

6. Luthra, A., Sullivan, G. J. et Wiegand, T., “H.264/AVC video coding standard”, IEEE Transactions Circuits and Systems for Video Technology, V. 13, n°. 7, (July 2003), 688 - 703.

7. Ostermann, j., Bormans, J., List, P., Marpe, D. et Wedi, T., “Video coding with H.264/AVC: Tools, Performance, and Complexity”, IEEE Circuits and Systems Magazine, V. 4, n°1, (September 2004), 7 - 28.

8. Li, Y., Chen, Z., Tan, S. M. et Campbell, R.H., “Security enhanced MPEG player”, IEEE 1st International Workshop on Multimedia Software Development (MMSD), Berlin, Germany, (March 1996), 169 - 175.

9. Tang, L., “Methods for encrypting and decrypting MPEG video data efficiently”, 4th ACM International Multimedia Conference, Boston, MA, (November 1996), 219 - 229.

10. Roche, S., Dugelay, J. L. et Molva. R., “Multi-resolution Access Control Algorithm Based on Fractal Coding”, International Conference on Image Processing, Lausanne Switzerland, (September 1996), 235 - 238.

11. Pazarci, M. et Dipc, V., “A MPEG2-transparent scrambling technique”, IEEE Transactions on Consumer Electronics, V. 48, n° 2 (August 2002), 345 - 355.

12. Tewfik, A. H. et Swanson, M.D., “Data hiding for multimedia personalization, interaction and protection”, IEEE Signal Processing Magazine, (July 1997), 41 - 44.

13. Nguyen, P. et Baudry, S., “Le tatouage de données audiovisuelles”, LCN, V. 4, n° 3-4, (2003), 135 - 165.

14. Cox, I., Miller, M. et Bloom, J., “Digital watermarking digital steganography”, The Morgan Kaufmann Series in Multimedia Information and Systems, (2001).

15. Martin, V. “Contribution des filtres LPTV et des techniques d’interpolation au tatouage numérique”, Thèse de doctorat en Signal image acoustique et optimisation, Ecole doctorale en Informatique et Télécommunications, (novembre 2006).

16. Kim, H. J., “Digital Watermarking”, First International Workshop, Seoul, Korea, (November 2002), http://www.iwdw.org/.

17. Fei; C., Kundur, D., Kwong R. H., “Analysis and design of secure watermark-based authentication systems”, IEEE transactions on Information Forensics and Security, (2006), V. 1, n°1, 43-55.

18. Lu; Z. m., Li, Y. -n., Wang, H. –x., Sun, S. -h., “Multipurpose video watermarking algorithm in the hybrid compressed domain”, IEE Proceeding Information Security, V. 153, n° 4, (2006), 173 - 182.

19. Luis, P. F., “Watermarking Security: A Survey”, Transaction on data Hiding and Multimedia security, (2006). 41-72.

20. Digimarc Corporation, http://www.digimarc.com/, (May 2004).

21. Adobe systems inc, http://www.adobe.com/, (November 2006).

22. Seiko Epson corporation, at: http://www.epson.comamericanorth.html/, (November 2006).

23. Kodak Australia, http://www.kodak.com/AU/en/consumer.jhtml/, (November 2006).

24. Stadler, Y., “Tatouage d’image semi-fragile pour appareil mobile intégré dans une chaîne de certification”, Thèse de PhD, Université de Lorraine, (Novembre 2012).

25. Fridrich, J. et Goljan, M., “Lossless data embedding - new paradigm in digital watermarking”, EURASIP Journal on Applied Signal Processing, Emerging applications of multimedia data hiding, V. 2, n° 2, (February 2002), 185 - 196.

26. Fridrich, J. et Goljan, M., “Invertible authentication watermark for JPEG images”, ITCC, Las Vegas, Nevada, (April 2001), 223 - 227.

27. Barni, M. et Bartolini, F., “Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications," Signal Processing and Communications Series, (February 2004), ISBN:0824748069.

28. Rey, C. et Dugelay, J., “Un panorama des méthodes de tatouage permettant d’assurer un service d’intégrité pour les images”, Traitement du Signal, n° spécial Tatouage, V. 18, n° 4, (2001).

29. Autrusseau, F., Saadane, A. et Barba, D., “Psychovisual approach for watermarking”, SPIE Electronic Imaging Security and watermarking of Multimedia, V. 3, san Jose United States, (January 2001), 495 - 504.

30. Winkler, S. et Kutter, S., “Vers un tatouage à étalement de spectre optimal utilisant le système visuel humain”, COmpression et REprésentation des Signaux Audiovisuels (Coresa), Sophia Antipolis, France, (June 1999), 25 - 33.

31. Deguillaume, F., Csurka, G., O’Ruanaidh, J. J. K. et Pun, T., “Robust 3D DFT video watermarking”, Electronic Imaging, Session: Security and Watermarking, V. 3657, (April 1999), 113 - 124.

32. Su, K., Kundur, D. et Hatzinakos, D., “A content-dependent spatially localized video watermark for resistance to collusion and interpolation attacks”, IEEE International Conference on Image Processing, V. 1, (October 2001), 818 - 821.

33. Barni, M. et Bartolini, F., “Watermarking systems engineering: enabling digital assets security and other applications," CRC Press, Boca Raton, FL, USA, (2004), ISBN: 0824748069.

34. Ling, C. et Ur-Rehman, O., “Watermarking for image authentication”, Chapter 2 in Robust Image Authentication in the Presence of Noise, Springer International Publishing Switzerland, (April 2015), 43 - 52.

35. Vassaux, B., Nguyen, P., Baudry, S, Bas, P. et Chassery, J. M., “A survey on attacks in image and video watermarking”, International Symposium on Optical Science and Technology, International Society for Optics and Photonics, Seattle, USA, (July 2002), 169 - 179.

36. Petitcolas, F. A. P., Anderson, R. J. et Kuhn, M. G., “Attacks on copyright marking systems”, International workshop on information hiding, V. 1529 of lecture notes in computer science, (April 1998), 218 - 238.

37. Voloshynovskiy, S., Pereira, S., Pun, T., Eggers, J. J. et Su, J. K., “Attacks on digital watermarks: classification, estimation based attacks, and benchmarks”, IEEE Communications Magazine, V. 39, n° 8 (August 2001), 118 - 126.

38. Su, K., Kundur, D. et Hatzinakos, D., “A novel approach to collusion resistant video watermarking”, Proceeding of Electronic Imaging, Security and Watermarking of Multimedia, Contents IV, San Jose, CA, (January 2002), 491 - 502.

39. Doërr, G. et Dugelay, J. L., “Collusion issue in video watermarking”, Proceeding of Electronic Imaging, Security and Watermarking of Multimedia, Contents VII, V. 5681, (January 2005), 685 - 696.

40. Fadoua, D., “Tatouage d'image par techniques multidirectionnelles et multirésolution”, Mémoire d’études approfondies, Laboratoire d’Informatique en Images et Systèmes Information, Université Claude Bernard Lyon, (Juillet 2003).

41. Chen, B. et Wornell, G. W., “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding”, IEEE Transactions on Information Theory, V. 47, n° 4, (May 2001), 1423 - 1443.

42. Furon, T. et Bas, P., “A new measure of watermarking security applied on QIM”, International Workshop on Information Hiding, Berkeley, USA, (May 2013), 207 - 223.

43. Agarwal, C., Mishra, A. et Sharma, A., “Gray-scale image watermarking using GA-BPN hybrid network”, Journal of Visual Communication and Image Representation, V. 24, n° 7, (October 2013), 1135 - 1146.

44. Emami, M. S., Omar, K., Sahran, S. et Abdullah, S. N. H. S., “Spatial domain approaches for real-time ownership identification”, Journal of Advances in Information Technology, V. 5, n° 1, (February 2014), 1 - 4.

45. Piva, A., Barni, M., Bartolini, F. et Cappellini, V., “DCT-based watermark recovering without resorting to the uncorrupted original image”, International Conference on Image Processing, V. 1, Santa Barbara, USA, (October 1997), 520 - 523.

46. Hartung. F., “Digital watermarking and fingerprinting of uncompressed and compressed video”, Berichteaus der Kommunikations-und Informationstechnik, V. 13, Shaker Verlag, Aachen, Germany, (February 2000).

47. O’Ruanaith, J. J. et Pun, T., “Rotation, scale and translation invariant spread spectrum digital image watermarking”, Signal Processing, V. 66, n° 3, (May 1998), 303 - 317.

48. Kundur, D. et Hatzinakos, D., “Digital watermarking using multiresolution wavelet decomposition”, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), V. 5, (May 1998), 2659 - 2662.

49. Wang, S. H. et Lin, Y. P., “Wavelet tree quantization for copyright protection watermarking”, IEEE Transactions on Image Processing, V. 13, n° 2, (February 2004), 154 - 165.

50. Meerwald, P., “Quantization watermarking in the JPEG2000 coding pipeline”, Fifth Joint Working Conference on Communications and Multimedia Security (IFIP TC6/TC11), Issues of the New Century, Eds. Kluwer Academic Publishing, (May 2001), 69 - 79.

51. CCIR. Projet de révision de la recommandation 500-4, “Méthode d'évaluation subjective de la qualité des images de télévision”, Document commissions d'études du CCIR, 11/BL/51-F, (1992).

52. Watson, A. B., et Malo, J., “Video quality measurement based on the standard spatial observer”, International Conference on Image Processing, V. 3, Rochester, New York, USA, (September 2002), 41 - 44.

53. Wang, Z., Lu, L. et Bovik, A.C., “Video quality assessment Based on structural distortion measurement”, Signal Processing: Image Communication, V. 19, n° 2, (February 2004), 121 - 132.

54. Zhou, W. A., Bovik, C., Sheikh, H. R. et Simoncelli, E. P., “Image quality assessment: from error measurement to structural similarity”, IEEE Transactions on Image Processing, V. 13, n° 1, (April 2004), 600 - 612.

55. Wong, P.H.W., Yeung, G.Y.M. et Au, O.C., “Capacity for JPEG2000- to-JPEG2000 images watermarking”, International Conference on Multimedia and Expo, V. II, Baltimore, Maryland, USA, (July 2003), 485 - 488.

56. Bodo, Y. “Elaboration d’une technique d’accès conditionnel par tatouage et embrouillage vidéo basée sur la perturbation des vecteurs de mouvement”, Thèse de doctorat de l’école national supérieur des télécommunications, (2004).

57. Ghanbari, M., “Standard Codec: Image compression to advanced video coding”, Institution Electrical Engineers, (2003), ISBN:0852967101978085296710 19780852967102.

58. Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC). ITU-T and ISO/IEC JTC 1, Version 1 : May 2003, Version 2 : May 2004, Version 3 : March 2005, Version 4 : September. 2005, Version 5 and Version 6, June 2006, Version 7: April 2007.

59. Wiegand, T., Sullivan, G. J., Bjntegaard, G. et Luthra, A., “Overview of the H.264/AVC video coding standard,” IEEE Transactions on Circuits and Systems for Video Technology, V. 13, n° 7, (July 2003), 560 - 576.

60. Organisation Internationale de Normalisation, ISO/IEC JTC1/SC29/WG11 N2196. Coding of Moving Pictures and Audio, “Overview of the MPEG-4 Version 1 Standard”, Tokyo, (March1998).

61. Wien, M., “Intra coding using variable block sizes”, Proposal VCEG-O31, ITU-T VCEG, Pattaya, Thailand, (December 2001).

62. Wiegand, T., “New intra prediction using intra macroblock motion compensation”, Document JVT-B118r2 of Joint Video Team of ISO/IECMPEG and ITU-T VCEG, (February 2002).

63. Meng, B. J., Au, O. C., Wong, C.W. et Lam, H. K., “Efficient intra-prediction mode selection for 4x4 blocks in H.264”, International Conference on Multimedia and Expo, V. III, Baltimore, USA, (July 2003), 521 - 524.

64. Meng, B.J., Au, O.C., “Fast mode selection method for 4x4 blocks intra-predication in H.264”, IEEE International conference On Image Processing, (September 2003), 14 - 17.

65. Chen, Z., Zhou, P. et He, Y., “Fast integer pel and fractional pel motion estimation for JVT,” Joint Video Team (JVT), Docs, JVT-F017, (December 2002).

66. Yang L., Yu K., Li J., Li S. “An effective variable block size early termination algorithm for H.264 video coding,” IEEE Transactions Circuits and Systems Video Technology., V. 15, n° 6, (June 2005), 784 - 788.

67. Gang, Z., Li, G. et He, Y., “The intra prediction based on sub block”, 7th International Conference on Signal Processing Proceedings (ICSP), V. 1, Beijing, China, (August-September 2004), 467 - 469.

68. Yeh, J., Vetterli, M. et Khansari, M., “Motion compensation of motion vectors”, IEEE International Conference on Image Processing, V. 1, Washington, District de Columbia, USA, (October 1995), 574 - 577.

69. Yoon Yung, L. et Woods, J. W., “Motion vector quantization for video coding”, IEEE Transactions on Image Processing, V. 4, n° 3, (March 1995), 378 - 382.

70. Darshna, D. et Shah, S. N., “Analysis of block matching algorithms for motion estimation in H.264 video codec”, International Journal of Engineering Research and Applications, V. 2, n° 6, (November 2012), 1396 - 1401.

71. Wiegand, T., Zhang, X. et Girod, B., “Long-term memory motion-compensated prediction”, IEEE Transactions Circuits and Systems for Video Technology, V. 9, n° 1, (February 1999), 70 - 84.

72. Ahmed, N., Natarajan, T. et Rao, K.R., “Discrete cosine transform”, IEEE Transactions on Computers, V. 23, n° 1, (January 1994), 90 - 94.

73. Malvar, H., Hallapuro, A., Karczewicz, M. et Kerofsky, L., “Low complexity transform and quantization in H.264/AVC”, IEEE Transactions on Circuits and Systems for Video Technology, V. 13, n° 7, (July 2003), 598 - 603.

74. Shahid, Z., “Protection des Vidéos Hiérarchiques par Cryptage et Tatouage”, Thèse de PhD, Université de Montpellier II, (Octobre 2010).

75. Cédric, M., “ Vers une solution réaliste de décodage source-canal conjoint de contenus multimédia ”, Thèse de PhD, Université de Paris Sud, (Mars 2009).

76. Marpe, D., Schwarz, H. et Wiegand, T., “Context-Adaptive binary arithmetic coding in the H.264/AVC video compression standard”, IEEE Transactions

on Circuits and Systems for Video Technology, V. 13, n° 7, (July 2003), 620 - 636.

77. Cote, G. et Winger, L., “Recent advances in video compression standards”, IEEE Canadian Review, (Spring 2002), 21 - 24.

78. List, P. Joch, A. Lainema, J. Bjontegaard, G. et Karczewicz, M. “Adaptive deblocking filter”, IEEE Transactions on Circuits and Systems for Video. Technology, V. 13, n° 7, (July 2003), 614 - 619.

79. Sullivan, G.J., Topiwala, P. et Luthra, A., “The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions”, Applications of Digital Image Processing XXVII, SPIE, V. 5558, Denver, Colo, USA, (August 2004), 454 - 474.

80. Hedayath, B., Gangatharan, N. et Tamilchelvan, R., “A survey on video watermarking technologies based on copyright protection and authentication”, International Journal of Computer Applications Technology and Research, V. 5, n° 5, (May 2016), 295 - 303.

81. Su, P. C., Hsu, C. W., et Wu, C. Y., “A practical design of content protection for H.264/AVC compressed videos by selective encryption and fingerprinting”, Multimedia Tools and Applications, V. 52, n° 2 - 3, (January 2010), 529 - 549.

82. Pröfrock, D., Schlauweg, M. et Müller E., “A new uncompressed domain video watermarking approach robust to H.264/AVC compression”, Signal Processing, Pattern Recognition and Applications, Innsbruck, Austria, (February 2006), 99 - 104.

83. Rodriguez, E., Superiori, L., Nemethova, O. et Rupp, M., “Performance of watermarking as an error detection mechanism for corrupted H.264/AVC video sequences”, Talk, 17th European Signal Processing Conference (EUSIPCO), Glasgow, Scotland, (August 2009), 2206 - 2210.

84. Nemethova, O., Forte, G. C. et Rupp, M., “Robust error detection for H.264/AVC using relation based fragile watermarking”, 13th International Conference on Systems, Signals and Image Processing (IWSSIP), Budapest, Hungary, (September 2006).

85. Superiori, L., Nemethova, O. et Rupp, M., “Performance of a H.264/AVC error detection algorithm based on syntax analysis”, 4th International Conference on Mobile Computing and Multimedia (MoMM), Yogiakarta, Indonesia, (December 2006), 49 - 58.

86. Chen, X., chung, Y., Xu, F. Otoom, A .F. et Bae, C., “Combined copyright protection and error detection scheme for H.264/AVC”, 6th International Conference on Multimedia Systems and Signal Processing, Hangzhou, China, (April 2006), 7 - 12.

87. Shahid, Z., Chaumont, M. et Puech, W., “Spread spectrum based watermarking for Tardos code based fingerprinting of H.264/AVC video”, IEEE International Conference on Image Processing (ICIP), Hong-Kong, China, (September 2010), 2105 - 2108.

88. Ait sadi, K., Bouridane, A. et Guessoum, A., “H.264/AVC Digital fingerprinting based on content adaptive embedding”, 7th International Conference on Information Assurance and Security, Malacca, Malaysia (December 2011).

89. Shahid, Z., Chaumont M. et Puech, W., “H.264/AVC video watermarking for active fingerprinting based on Tardos code”, Signal, Image and Video Processing, Springer, Special Issue on Image and Video Processing for Security, n° 11760, (2013).

90. Li, Z. et Chen, J., “Efficient compressed domain video copy detection”, International Conference on Management and Service Science (MASS), Wuhan, China, Wuhan, China, (August 2010), 1 - 4.

91. Ali, A. et Edirisinghe, E. A., “Efficient spatio-temporal matching for video copy detection in H.264/AVC video”, International Journal of Computer Applications (IJCA), V. 41, n° 15, (March 2012), 1 - 7.

92. Beghdad, A., “Le tatouage numérique : une solution de protection du bien numérique”, Rapport du projet collaboratif HD3D-IIO labellisé par le pôle de compétitivité CAP DIGITAL”, (2009), 13 -11.

93. Golikeri, A., Nasiopoulos, P. et Wang, Z. J., “Robust digital video watermarking scheme for H.264 advanced video coding standard”, Journal of Electronic Imaging, V. 16, n° 4, (December 2007), doi: 10.1117/1.2816054.

94. Qiu, G., Marziliano, P., Ho, A. T. S., He, D. J. et Sun, Q. B., "A hybrid watermarking scheme for H.264/AVC video," 17th International Conference on Pattern Recognition, V. 4, Cambridge, UK, (August 2004), 865 - 868.

95. Noorkami, M. et Mersereau, R. M., “Compressed domain video watermarking scheme for H.264”, 12th IEEE International Conference on Image Processing, Genova, Italy, V. 2, (September 2005), 890 - 893.

96. Tian, L., Zheng, N., Xue, J. et Xu, T., “A CAVLC-based blind watermarking method for H.264/AVC compressed video”, IEEE Asia-Pacific Services Computing Conference, Yilan, Taiwan, (December 2008), 1295 - 1299.

97. Noorkami, M. et Mersereau, R. M., “Improving perceptual quality in video watermarking using motion estimation”, 13th IEEE International Conference on Image Processing, Atlanta, GA, USA, (October 2006), 1389 - 1392.

98. Zhang, J. et Ho, A. T. S., “Robust digital Image-in-video watermarking for the emerging H.264/AVC standard”, IEEE Workshop on Signal Processing Systems Design and Implementation, Athens, Greece, (November 2005), 657 - 662.

99. Zhang, J. et Ho, A. T. S., “Robust video watermarking of H.264/AVC”, IEEE Transactions on Circuits and Systems, V. 54, n° 2, (February 2007). 205 - 209.

100. Ali, M.A., et Edirisinghe, E. A., “Watermarking H.264/AVC by modifying DC coefficients”, International Conference on CyberWorlds, Bradford, England, UK, (October 2009), 241 - 245.

101. Lu, T. T., Hsu, W. L. et Chang, P. C. “Blind video watermarking for H.264”, Canadian Conference on Electrical and Computer Engineering (CCECE), Ottawa, ON, Canada, (May 2006), 2353 - 2356.

102. Hu, Y., Zhang, C. et Su, Y., “Information hiding based on intra prediction modes for H.264/AVC,” IEEE International Conference on Multimedia and Expo (ICME), Beijing, China, (July 2007), 1231 - 1234.

103. Yang, G., Li, J., He, Y. et Kang, Z., “An information hiding algorithm based on intra-prediction modes and matrix coding for H.264/AVC video stream”, International Journal of Electronics and Communications, V. 65, n° 4, (April 2010), 331 - 337.

104. Kapotas, S. K. et Skodras, A. N., “Real time data hiding by exploiting the IPCM macroblocks in H. 264/AVC streams”, Journal of Real-Time Image Processing, V. 4, n° 1, (2009), 33 - 41.

105. Kim, S. M., Kim, S. B., Hong, Y. et Won, C. S., “Data hiding on H.264/AVC compressed video”, International Conference on Image Analysis and Recognition (ICIAR), Montreal, Canada, V. 4633, (August 2007), 698 - 707.

106. Ait sadi, K., Guessoum, A. et Bouridane, A., “Efficient pre-processing watermark for robust video watermarking of H.264/AVC”, International Journal of Advanced Media and Communication, V. 4, n° 3, (2010), 219 - 234.

107. Fourati, W. et Bouhlel, M. S., “A novel approach to improve the performance of JPEG 2000”, Journal of Graphics Vision and Image Processing , V. 5, n° 5, (2005), 1 - 9.

108. Tasdizen, O. et Hamzaoglu, I., “A high performance and low cost hardware architecture for H.264 transform and quantization algorithms”, 13th European Signal Processing Conference, Turkey, (September 2005), 4 - 8.

109. Santhi, V. et Arulmozhivarvarman, P., “Hadamard transform based adaptive visible/invisible watermarking scheme for digital images”, Information Security and Applications, V. 18, n° 4, (December 2013), 167 - 179.

110. Mohanty, S. P. et Bhargava, B. K., “Invisible watermarking based on creation and robust insertion extraction of image adaptive watermarks”, ACM Transactions on Multimedia Computing Communications and Applications (TOMCCAP), V. 5, n° 2, (November 2008), 12:1 - 2:22.

111. Ho, A. T. S., Shen, J., Tan, S. H. et Kot, A. C., “Digital image-in-image watermarking for copyright protection of satellite images using the fast Hadamard transform”, 24th IEEE International Geoscience and Remote Sensing Symposium, Toronto, Canada (June 2002), 3311 - 3313.

112. Lee Raymond, S. T. et Lam Henry, W. S., “A chaotic real-time cryptosystem using a switching algorithmic-based linear congruential generator (SLCG)”, International Journal of Computer Science and Network Security, V. 6, n° 8, (August 2006),116 - 124.

113. Leeb, H. et Wegenkittl, S., “Inversive and linear congruential pseudorandom number generators in empirical tests”, ACM Transactions on Modeling and Computer Simulation, V. 7, n° 2, (April 1997), 272 - 286.

114. Wang, Z. N., Yang, J., Peng, Q., Ma, Z. et Zhu, C. Q., “A fast transform domain based algorithm for H.264/AVC intra prediction”, IEEE International

Conference on Multimedia and Expo, Beijing, China, (July 2007), 1563 - 1566.

115. H.264/AVC Joint Model 7.6 (JM-7.6).

116. Jung, S., Lee, D., Lee, S. et Paik, J., “Fingerprint watermarking for H.264 streaming media”, International Conference Frontiers in the Convergence of Bioscience and Information Technologies, Jeju, Island, Korea (October 2007), 671 - 675.

117. Farfoura, M. E., Horng, S. J., Guo, J. et Al-Haj, A., “Low complexity semi-fragile watermarking scheme for H. 264/AVC authentication”, Multimedia Tools and applications, V. 75, n° 13, (June 2015), 7465 - 7493.

118. Le, B., Nguyen, H. et Tran, D., “A robust fingerprint watermark-based authentication scheme in H.264/AVC video“, Vietnam Journal of Computer Science, V. 1, n° 3, (August 2014), 193 - 206.

119. Bovik, A. C., “Handbook of image and video processing”, Elsevier Academic Press, Boston, MA, (2010).

120. Feng, D., Siu, W. C. et Zhang, H. J. “Multimedia information retrieval and management: Technological fundamentals and applications”, Signals and Communication Technology, Springer Science & Business Media, (January 2003).

121. Zhu, B. B., Swanson, M. D. et Tewfik, A. H., “When seeing isn’t believing multimedia authentication technologies”, IEEE Signal Processing Magazine, V. 21, n° 2, (March 2004), 40 - 49.

122. Kim, T., Park, K. et Hong, Y., “Video watermarking technique for H.264/AVC”, Optical Engineering, V. 51, n° 4, (April 2012), 047402 - 47412.

123. Kuo, T. Y., Lo, Y. C. et Lin, C. I., “Fragile video watermarking technique by motion field embedding with rate distortion minimization”, 4th International

Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP), Harbin, China, (August 2008), 853 - 856.

124. Wang, C. C. et Hsu, Y. C., “Fragile watermarking for H.264 video stream authentication”, 8th International Conference of Intelligent Systems Design and Applications, Kaohsiung, Taiwan, (November 2008), 77 - 80.

125. Zhang, J. et Ho, Ho, A. T. S., “Efficient video authentication for H. 264/AVC”, 1st International Conference on Innovative Computing Information and Control, V. 3, Beijing, China, (August 2006), 46 - 49.

126. Pröfrock, D., Richter, H., Schlauweg, M. et Müller, E., “H.264/AVC video authentication using skipped macroblocks for an erasable watermark”, Visual Communications and Image Processing, SPIE, V. 5960, Beijing, China, (July 2005), 1480 - 1489.

127. Razib, I., Shirmohammadi, S. et Zhao, J., “Compressed domain authentication of live video”, IEEE International Conference on Signal Processing and Communications (ICSPC), Dubai, United Arab Emirates (November 2007), 1443 - 1447.

128. Iqbal, R., Shirmohammadi, S. et Zhao, J., “Hard authentication of H.264 video applying MPEG-21 generic Bitstream Syntax Description (gBSD)”, IEEE International Conference on Multimedia and Expo, Beijing, China, (July 2007), 875 - 878.

129. Ueda, S., Shigeno, H. et Okada, K. I., “NAL level stream authentication for H.264/AVC”, IPSJ Transactions on Databases, V. 48, n° 2, (2007), 635 - 643.

130. Gennaro, R. et Rohatgi, P., “How to sign Digital Streams”, Kaliski, B.S. (eds.) Advances in Cryptology – CRYPTO’97, LNCS, Springer, Heidelberg, V. 1294, (1997), 180 - 197.

131. Wong, C. et Lam, S., “Digital signature for flows and multicasts”, IEEE/ACM Transactions on Networking, V. 7, n° 4, (1999), 502 - 513.

132. Park, J., Chong, E. et Sieggl, H., “Efficient multicast stream authentication using erasure codes”, ACM Transactions on Information and System Security, V. 6, n° 2, (May 2003), 258 - 285.

133. Ramaswamy, N. et Rao, K., “Video authentication for H.264/AVC using digital signature standard and secure hash algorithm”, International workshop on Network and operating systems support for digital audio and video, Newport, Rhode Island, (May 2006).

134. Horng, S. J., Farfoura, M. E., Fan, P., Wang, X., Li, T. et Guo, J. M., “A low cost fragile watermarking scheme in H.264/AVC compressed domain”, Multimedia Tools and Applications, V. 72, n° 3, (October 2014), 2469 - 2495.

135. Xu, D. et Wang, R., “Watermarking in H.264/AVC compressed domain using Exp-Golomb code words mapping”, Optical Engineering, V. 50, n° 9, (September 2011), 267 - 279.

136. Xu, D., Wang, R. et Shi, Y. Q., “Data hiding in encrypted H.264/AVC video streams by codeword substitution”, IEEE Transactions on Information Forensics and Security, V. 9, n° 4, (April 2014), 596 - 606.

137. Ait Sadi, K., Bouridane, A. et Guessoum, A., “Combined fragile watermark and digital signature for H.264/AVC video authentication”, 17th European Signal Processing Conference, Glasgow (August 2009), 1799 - 1803.

138. Ait Sadi, K., Bouridane, A. et Guessoum, A., “H.264/AVC video authentication based video content”, 5th International Symposium on I/V Communications and Mobile Network, Rabat, Morocco, (September 2010), 1 - 4.

139. Ait Sadi, K., Guessoum, A., Bouridane, A. et Khelifi, F., “Content fragile watermarking for H.264/AVC video authentication”, International Journal of Electronics (IJE), V. 104, n° 4, (January 2017), 673 - 691.

140. Chang, L., “Comparison of transformed-based visual features for automatic lip reading”, EURASIP Journal on Image and Video Processing, (2008), 1 - 9, doi:10.1155/2008/810362.

141. Weng, L. et Preneel, B., “On encryption and authentication of the DC DCT coefficient”, International Conference on Signal Processing and Multimedia Applications, Barcelona, Spain, (July 2007), 375 - 379.

142. Suehring, K., “JVT JM reference software”, http://iphome.hhi.de/suehring/tml/download/KTA/, (2007).

143. Ohm, J. R. et Sullivan, G. J., “High efficiency video coding: The next frontier in video compression”, IEEE Signal Processing Magazine, V. 30, n° 1, (January 2013), 152 - 158.

144. http://www.cipr.rpi.edu/resource/sequences/sif.html.

145. Kim, J., Biryukov, A., Preneel, B. et Hong, S., “On the security of HMAC and NMAC based on HAVAL, MD4, MD5, SHA-0 and SHA-1”, Chapter in Security and cryptography for networks, Springer Berlin Heidelberg, (2006), 242 - 256.

146. Sun, X., Divakaran, A. et Manjunath, B., “A motion activity descriptor and its extraction in compressed domain”, IEEE Multimedia Information processing, V. 2195, (November 2001), 450 - 457.

147. Rivest, R., “The MD5 Message-Digest Algorithm ”,Request for Comments: 1321, RFC 1321, MIT Laboratory for Computer Science and RSA Data Security, Inc, (April1992).

148. NIST, FIPS PUB 180-1: Secure Hash Standard, (April 1995).

149. Bellare, M., Canetti, R., and Krawczyk, H., “Message authentication using hash functions: The HMAC construction,” RSA Laboratories, CryptoBytes, V.2, n° 1, Spring, (1996).

150. Krawczyk, H., Bellare, M., and Canetti, R., “HMAC: Keyed-Hashing for Message Authentication,” Internet RFC 2104, (February 1997).

151. Rogaway, P., Shrimpton, T., “Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision Resistance”,: Fast Software Encryption: 11th International Workshop, Computer Science, Springe, V. 3017, (2004). 245–259.