

Université du Québec en Outaouais
Département d'Informatique et d'Ingénierie

La vidéo surveillance automatique: sécurisation du contenu et traitements coopératifs

Rapport de recherche

(RR 07/06-1, Juin 2007)

Par

Nadia Baaziz

nadia.baaziz@uqo.ca

Table des Matières

I-	Introduction	3
II-	La vidéo surveillance intelligente	3
III-	La sécurisation du contenu visuel	XX
	III-1 La cryptographie	
	III-2 Le marquage numérique pour authentification	
	III-3 L'incorporation du marquage	
IV-	La détection des changements	5
	IV-1 La détection du fonds de scène et analyse de l'avant de scène	
	IV-2 L'estimation du mouvement	
	IV-3 La sélection vidéo	XX
V-	La protection de l'information privée	XX
VI-	Conclusion.....	26
	Références bibliographiques.....	27
	Annexe	

I. INTRODUCTION

De nos jours, le monde est de plus en plus complexe dans ses infrastructures et ses interactions, ce qui a pour effet d'augmenter le nombre d'événements tragiques de nature accidentelle ou intentionnelle. En contrepartie, les sociétés sont aussi plus exigeantes en termes de sécurité et de prévention et exploitent les avancées technologiques afin de répondre à ces besoins de la meilleure façon possible. C'est ainsi que l'on assiste de nos jours à la prolifération de systèmes de vidéo surveillance numériques conçus et installés dans des lieux résidentiels, publics ou de travail (banques, centres commerciaux, usines, aéroports, écoles, tours résidentielles, gares routières, voies ferroviaires, autoroutes...). Ces systèmes sont pour la plupart utilisés dans le but d'assister les gardiens de sécurité dans leur travail. Dans leur forme de base, ces systèmes bien que numériques ont des fonctions qui se limitent à la capture, la transmission, le stockage et l'affichage de données visuelles au niveau du centre de surveillance. Ainsi les opérateurs humains peuvent effectuer des surveillances tout en minimisant leurs déplacements. Mais il leur incombe d'observer continuellement les milliers d'images qui défilent, de détecter d'éventuelles anomalies et de prendre les mesures adéquates aux situations. Ce travail est particulièrement intensif et fastidieux, exigeant un degré d'attention et de vigilance qui dépassent la capacité moyenne d'opérateurs humains. Mieux encore, si la détection d'événements est sujette aux erreurs d'inattention, la possibilité de faire des prédictions et de prévenir les accidents sont quasiment non faisables avec de telles infrastructures de base. La solution est dans l'intégration technologique (logicielles et matérielle) qui augmente la part d'assistance à l'opérateur humain et diminue la part d'intervention de ce dernier. Le système de vidéo surveillance doit évoluer de sa forme traditionnelle vers de nouvelles formes qui intègrent des fonctionnalités additionnelles de traitement de données, d'analyse et de décision. Le système de vidéo surveillance doit assurer une surveillance intelligente !

Dans ce qui va suivre nous allons d'abord décrire les objectifs de l'incorporation de l'intelligence dans les systèmes de vidéo surveillance et plus précisément dans les caméras. La section III est une étude sur la sécurisation des données et du contenu de séquences de vidéo surveillance pour permettre leur authentification. Puis, nous allons procéder, dans les sections IV et V, à une étude des traitements et fonctionnalités pouvant s'incorporer conjointement avec les procédés de sécurisation des données vidéo ou en coopération avec ceux-ci.

II. La vidéo surveillance intelligente

La vidéo surveillance intelligente repose sur des systèmes qui incorporent de manière automatique les technologies de pointe de plusieurs domaines, notamment en vision par ordinateur, en traitement du signal, en intelligence artificielle et en fouilles de données, afin d'élargir le spectre des fonctionnalités et des applications de la vidéo surveillance et les amener à :

1. Prédire les incidents en détectant des comportements suspects et à déclencher des alarmes en temps réel. Comme exemples de prédictions nous pouvons citer les cas de détection suivants : un intrus en mouvement, un sac abandonné, une pièce ou objet disparu d'un endroit, un véhicule dépassant une vitesse maximale, un véhicule mal stationné, une salle archi-comble, une chaîne trop longue, un rôdeur dans un parc de stationnement ainsi que la détection de tout comportement jugé anormal suite à une séquence d'apprentissage de ce qui est supposé "normal".
2. Aider aux opérations d'investigation et leur amélioration en effectuant des recherches basées sur le contenu, des suivis spatio-temporels, et des extractions vidéo automatiques.

Bien entendu, un tel système de vidéo surveillance n'est pas disponible à l'exploitation vu le coût de l'infrastructure (caméra, réseaux, serveurs, etc.) ainsi que la complexité des traitements requis. Mais tout porte à croire que l'on n'est pas loin de voir son introduction graduelle sur le marché si l'on se fie aux quelques prototypes en cours de réalisation [2].

Une partie des recherches et développements qui se font de nos jours s'oriente vers la construction de caméras dont les fonctionnalités dépassent le simple fait de capturer et envoyer des images. Il s'agit en effet de caméras intelligentes (Smart Cameras) qui incorporent en leur sein des unités d'analyse et de traitement d'images leur permettant de délivrer aux serveurs de fichiers des séquences vidéo dont le contenu est pertinent, soit des vidéos de 'grande valeur'. Des exemples de ces traitements incluent :

- La sécurisation du contenu visuel
- La détection des changements dans les séquences vidéo
- La sélection de séquences contenant des faits ou des changements potentiellement pertinents (présentant une signification importante par rapport aux objectifs de la vidéo surveillance)
- La compression des données
- La protection des informations à caractère privé

De toute évidence, ces traitements une fois incorporés doivent s'effectuer en temps réel. Autrement dit, leur exécution ne doit en aucun cas gêner ou retarder les cadences standards de capture et d'envoi d'images (exemple : 30images/seconde).

Dans le cadre de nos travaux de recherche et développement, nous nous intéressons en particulier aux traitements pouvant s'incorporer dans une caméra, et donc acceptant les contraintes de temps réel. Nous considérons en priorité les possibilités d'incorporation de la sécurisation pour authentification comme une étape essentielle et requise pour la crédibilité des données et des contenus collectés. Nous considérons aussi les possibilités de traitements connexes pouvant précéder la sécurisation ou bien coopérer avec celle-ci. Dans ce qui suit, nous dressons un état de l'art sur l'existant et nous donnons au même temps plus de détails sur nos orientations.

III. La sécurisation du contenu visuel

Les systèmes de vidéo surveillance sont déployés afin d'aider à assurer au mieux la sécurité des biens et des personnes en général. Dépendamment des taux d'automatisme et d'intelligence incorporés dans le système, cette aide peut prendre plusieurs formes, allant de la surveillance à distance qui réduit le déplacement des gardiens de sécurité, à l'émission d'alarmes suite à des détection de situations suspectes ou dangereuses, ou encore à l'apport de preuves tangibles aux procédures d'investigations policières ou judiciaires.

Afin de pouvoir utiliser les données visuelles collectées par un système de vidéo surveillance comme support pour l'investigation, il est très important d'avoir tout d'abord, une preuve sur l'authenticité et l'origine de ces données. Les méthodes de traitement numériques sont tellement disponibles et faciles d'utilisation qu'il est rendu aisé de falsifier et de manipuler les séquences vidéo. Il est donc nécessaire d'incorporer dans les systèmes de vidéo surveillance modernes des procédées qui sécurisent les données contre toutes sortes de manipulations. Plus précisément, cette sécurité incorporée doit permettre :

- l'authentification des données visuelles et/ou du contenu visuel contre toute sortes de manipulations,
- la divulgation de l'origine des données en termes de temps d'acquisition, du lieu d'acquisition, de l'équipement utilisé, etc..

La référence [11] consacre une grande partie à cette thématique et l'aborde sous plusieurs angles. Dans ce qui suit, nous tâcherons de répondre brièvement aux questions suivantes : comment incorporer la sécurité des données ? et quant l'incorporer ?

Afin de permettre une exploitation utile et correcte des données de vidéo surveillance, il est souhaitable que la sécurisation des données visuelles se fasse dans le respect des contraintes suivantes :

- 1- Les données additionnelles générées suite à la sécurisation doivent avoir un volume minimal.

- 2- Toute manipulation malicieuse doit être détectée, et distinguée d'une manipulation usuelle ou innocente.
- 3- Certaines manipulations dites innocentes (telles que la compression, la protection de l'information privée, le changement de format, etc.) doivent être tolérées. Ces manipulations doivent être définies au préalable pour chaque cadre d'application.
- 4- Les manipulations doivent pouvoir être localisées sur les images de la séquences.
- 5- Les données originales doivent pouvoir être retrouvées suite à une manipulation.

Deux approches d'incorporation de la sécurité sont généralement reconnues : la cryptographie, et le marquage numérique pour authentification, c'est ce que nous allons détailler dans les sections suivantes tout en répondant à notre première question.

III.1 L'approche cryptographique

L'utilisation de la cryptographie classique à base de clés secrètes pour sécuriser les séquences de vidéo surveillance peut se faire de plusieurs façons. Parmi celles-ci, nous pouvons citer les cas suivants:

- crypter chaque image de la séquence. Seuls les détenteurs de la bonne clé pourront décrypter les données, les visualiser et autres.
- Générer une signature de l'image (un résumé de son contenu, une empreinte), puis la crypter. L'authentification requiert alors la régénération de la même signature à partir de l'image, son cryptage à l'aide de la même clé, puis sa comparaison avec la signature cryptée fournie. Aucune différence n'est permise, sinon les données sont réputées non authentiques. La connaissance de la clé est requis.
- générer plusieurs signatures localisées de l'image et les crypter. Ce cas est similaire au précédent sauf qu'il introduit la localisation puisqu'il permet d'authentifier des régions de l'image.

Dans tous ces cas, le surplus de données générées est considérable, et la sécurité introduite est très restreinte à la vérification de l'intégrité des données plutôt qu'au contenu visuel. En d'autres mots, les contraintes 1), 2) ...5) citées ci-dessus ne sont quasiment pas respectées. De plus, toute visualisation, analyse ou fouille des données visuelles requiert le décryptage et le contrôle d'accès avec privilèges n'est plus possible. Par conséquent, les chercheurs se sont tournés vers d'autres approches qui permettent l'atteinte d'objectifs de sécurité en respect de ces contraintes. Ceci dit, ces nouvelles approches ne s'interdisent pas l'utilisation de concepts cryptographiques comme une étape partielle qui renforce le niveau de sécurité des procédés introduits (comme par exemple l'utilisation de clés secrètes).

III.2 L'approche par marquage numérique

La technologie du marquage numérique ou watermarking consiste à insérer une marque invisible dans l'image, afin de réaliser un objectif de sécurité bien défini. La marque en elle-même peut être une structuration d'une information pertinente. Sa détection ou son extraction suivie d'une analyse conduit à une décision quant à l'objectif de sécurité visé. Dans le cas de la vidéo surveillance, l'objectif de sécurité poursuivi est essentiellement l'authentification des données et du contenu. Il importe de définir le contenu de la marque et que sa détection ou son extraction nous renseigne de manière précise sur l'origine des données, leur authenticité et toute falsification subie. Bien que les recherches et développements dans le domaine du marquage numérique soient encore à leurs débuts, il existe déjà des avancées qui nous permettent de croire que l'authentification des données de vidéo surveillance par marquage est possible[11][25]. En effet, plusieurs travaux menés dans l'authentification semi fragile ou robuste sont prometteurs car ils permettent déjà la sécurisation des données tout en respectant la plupart des 5 contraintes citées ci-dessus. Ayant une connaissance de ce qui se fait, nous pouvons d'ores et déjà citer quelques caractéristiques et avantages de cette approche :

- Avec le marquage pour authentification, on peut faire mieux que ce qu'on fait avec la cryptographie classique. De plus, des concepts de cryptographie peuvent facilement être incorporés pour plus de sécurité et de fiabilité dans le processus d'authentification (crypter une donnée, utiliser des clés secrètes, etc.).
- Du fait que la marque est insérée dans l'image, le surplus de données que l'on doit gérer se limite généralement aux clés requises lors de l'insertion et/ou lors de la détection ou de l'extraction.
- Bien que l'opération de marquage altère les données de l'image, le critère d'imperceptibilité imposé lors de l'insertion de la marque fait que ces altérations n'ont aucune incidence grave sur l'analyse et l'interprétation du contenu des séquences vidéo (identification des personnes, reconnaissance et suivi de véhicules). En d'autres mots, ces altérations sont tout à fait acceptables dans le domaine de la vidéo surveillance et une contrainte de réversibilité du marquage n'est pas requise dans ce domaine d'application.
- Nous ajoutons à cela un autre avantage qui est la compatibilité possible avec la compression. Les données de vidéo surveillance sont volumineuses et leur compression peut être requise pour une meilleure gestion du stockage et de l'archivage. La compression efficace en taux de compression étant en général celle qui engendre des pertes, quelles conséquences cela a sur le processus d'authentification ? Il faut savoir qu'il existe, et heureusement, quelques méthodes d'authentification qui s'appliquent sur les données compressées (dans leur format de compression JPEG, MPEG, etc.), comme il existe aussi des méthodes dites semi fragiles qui s'appliquent avant la compression des données et qui sont conçues pour résister (tolérer) les effets de la compression (perçues comme des manipulations usuelles et innocentes de l'image).

- Il est à constater que la marque insérée peut avoir un contenu informatif relié au contenu sémantique de l'image. On peut aussi exploiter cette marque pour y cacher des données relatives à des traitements auxiliaires. Par exemple, pour une application de protection de l'information privée, la localisation des régions brouillées est une carte binaire qui peut être cachée parmi les bits de la marque. Ainsi, la marque servira deux objectifs, l'authentification et la divulgation des régions d'intérêts.

Parmi les nombreuses méthodes d'authentification existantes [11],[23-27], nous nous sommes intéressés en particulier à celle proposée par Fridrich et al. [23]. Il s'agit d'un marquage qui génère des approximations locales de régions ou blocs d'image, sous forme de versions compressées similaires au JPEG, et qui les structures en signatures binaires. La marque est construite comme une concaténation de ces signatures auxquelles peuvent s'ajouter des données auxiliaires. Son insertion spatiale dans les bits de poids faibles de pixels de l'image est contrôlée par une clé secrète qui active un générateur d'adresses pseudo aléatoire, indiquant l'emplacement des pixels où les bits de la marque peuvent être cachés. Lors du processus d'authentification, il faut d'abord être en possession de la clé secrète pour pouvoir extraire la marque de manière correcte (dans le bon ordre). La marque, ainsi extraite et reconstituée, est minutieusement analysée pour restituer les signatures locales qui s'y trouvent. D'un autre coté, le contenu de l'image est utilisé pour recalculer les signatures locales et les comparer à leurs équivalents restitués. Une égalité parfaite est une preuve d'authenticité des données alors que toute différence est un signe d'une manipulation possible. Un protocole, qui tient compte d'une analyse de l'information contextuelle (ou de voisinage) est alors exécuté afin de rendre la décision finale de non authenticité de la région considérée. Dans ce cas, il est possible de reconstituer une très bonne approximation du contenu original de la région (supposée manipulée) en décodant le contenu de sa signature. Donc, authentification, localisation des manipulations et reconstructions sont les atouts formidables de cette technique. Nous pensons aussi qu'il y aurait tout intérêt à étudier des améliorations de cette méthode surtout en ce qui a trait à la semi fragilité ou la tolérance des manipulations innocentes comme la compression.

En conclusion, nous pouvons dire que l'authentification des séquences de vidéo surveillance par marquage numérique est une avenue très attrayante. L'avancement de la recherche dans ce domaine est hautement désiré surtout si l'apport est concentré à relever les défis suivants : la semi fragilité ou la robustesse aux manipulations innocentes, et l'augmentation de la capacité du marquage (le nombre de bits inséré) qui se trouve en compromis avec la précision de l'extraction de la marque et de son invisibilité.

III.3 L'incorporation du marquage

La deuxième question à la quelle nous devons répondre concerne le choix du moment où il faut incorporer le traitement qui sécurise les données visuelles. Nous allons l'aborder avec l'hypothèse que l'approche suivie est basée sur le marquage numérique pour authentification. Idéalement, ce traitement doit intervenir immédiatement après l'acquisition des données et doit constituer un témoignage fiable de l'état des données et des conditions d'acquisition. L'implantation d'un tel processus peut se faire sous deux formes :

- Un composant matériel fonctionnant en temps réel et qui est intégré au hardware de la caméra. Ceci est tout à fait envisageable dans le cadre de la conception des caméras intelligentes (ou smart cameras). Une fois les données sécurisées, elles sont transmises à la station centrale pour analyse ou stockage.

- Un système de vision par ordinateur qui réceptionne les données visuelles envoyées par la caméra, procède à leur sécurisation par marquage puis les achemine à des unités d'analyse et/ou de traitements auxiliaires ou bien au système de stockage. Dans ce cas de figure, la transmission doit être fiable et sécurisée (ne doit pas permettre d'intercepter les données, les substituer ou les falsifier). Pour cela, une solution possible est le cryptage de toutes les données par la caméra. Une fois reçues, les données sont décryptées et un autre procédé de sécurisation (par marquage) est alors appliqué. Rappelons que, contrairement à la cryptographie, le marquage a les avantages de laisser les données dans un état visualisable, de minimiser le volume des données additionnelles, et de permettre l'analyse des données (fouilles, suivi, etc.) puisque les altérations dues au marquage sont mineures et acceptables.

Précisons quand même qu'un marquage opérant au sein de la caméra d'acquisition est plus intéressant si on ne peut pas assurer la transmission sécurisée. Cela évite aussi des traitements additionnels de cryptographie.

Il est possible que les deux opérations de capture et de marquage soit séparée ou intercalées par d'autres traitements. Dans ce cas, les propriétés suivantes doivent être garanties :

- la complexité des traitements doit obéir aux contraintes de temps réel si on est dans le cas d'une smart camera.

- le traitement doit être une conversion réversible ou bien une analyse, destiné uniquement à convertir la forme des données ou bien à extraire de l'information des séquences d'images et aucunement de les altérer. L'intégrité des données est ainsi garantie. Parmi ces traitements, nous pouvons citer tout changement de format sans perte et réversible (format de la couleur), ou bien toute analyse du mouvement effectuée afin de comprendre la dynamique de la scène ou détecter les changements ou les objets dans la scène.

- Le traitement peut être une transformation des données. Dans ce cas, l'intégrité des données est forcément perdue mais le contenu des données (au sens sémantique du terme) doit être préservé. En effet, la visualisation, l'analyse et l'observation des séquences vidéo originales ou transformées par les personnes autorisées doit aboutir aux mêmes interprétations et mener aux mêmes conclusions. Pour l'instant, ce que l'on reconnaît comme traitements valides de cette catégorie sont :

- la compression avec pertes (JPEG, MJPEG, MPEG, etc.) : encore faut-il rappeler qu'un marquage opéré conjointement à la compression est un thème d'intérêt.

- la protection de l'information privée : afin de préserver la sphère privée des individus impliqués dans les séquences de vidéo surveillance et de limiter l'étendue des abus possibles, on peut recourir à la dissimulation des régions d'intérêt aux yeux

d'opérateurs indirectement concernées. Cette dissimulation doit donc être réversible lorsque les données se trouvent dans les mains de personnes autorisées, représentant des instances légales. Notons que dans ce cas, nous supposons que le marquage pour authentification est opéré sur les séquences avec ROI dissimulées. La marque insérée peut contenir des informations relatives à la dissimulation.

Dans ce qui va suivre, nous allons aborder plus en détail ces différents traitements au travers d'analyses bibliographiques.

IV. La détection des changements

La détection des changements dans une séquence d'images constitue une source d'information très importante et voire même essentielle pour l'atteinte de plusieurs objectifs poursuivis par une vidéo surveillance automatique intelligente. Dépendamment de l'application visée, la détection de changement peut varier et aller d'une forme simple vers des formes de plus en plus complexes. En effet, effectuer une simple différence temporelle entre les images successives d'une scène nous permet de nous donner une indication sur le taux de changement spatio-temporel de la séquence vidéo. Une différence temporelle nulle (bien que rare) représente un cas évident d'une scène statique où il ne se passe absolument rien. Par contre, une différence temporelle non-nulle peut être une indication des situations suivantes :

- les changements de luminosité dans la scène (lumière on/off, lever du soleil...)
- les effets d'ajout de bruit lors de l'acquisition du signal
- les mouvements de la caméra (translation, balayage, zoom,...)
- les mouvements dans la scène mais sans importance pour la vidéo surveillance (la végétation sous l'effet du vent, chute d'eau, fontaines, vagues d'eau,...)
- les mouvements d'objets pertinents pour la vidéo surveillance (les personnes qui se déplacent, les véhicules en circulation...)

Les approches de détection de changement dans les séquences d'images peuvent être classées en deux catégories :

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

IV.1 La détection du fonds de scène et analyse de l'avant de scène

Une des approches classiques de détection des objets en mouvement dans une scène consiste dans une première étape, à identifier le fond de scène, puis, dans une seconde étape, de le soustraire de la scène entière afin d'isoler l'avant de scène. Cette approche, communément appelée background subtraction, est très efficace lorsqu'on est en présence d'objets en mouvement sur un fonds statique. Des extensions basées sur l'adaptation périodique du fond sont utilisées lorsque le fond de scène est dynamique (en cas de caméra en mouvement). Plusieurs techniques sont exploitées à cette fin allant de la modélisation gaussienne, aux méthodes de prédictions, aux différences temporelles et à l'estimation de mouvement. En général, les méthodes développées dans le cadre de cette approche sont capables de détecter les changements de lumière, les mouvements lents de la caméra, le bruit de la caméra, les objets à mouvements lents, les objets abandonnés ou disparus. Cependant, de grandes faiblesses sont constatées lorsque la scène contient des changements de lumière rapides ou brusques. À ce propos, nous pouvons nous référer aux travaux de Tian et al. [18] qui proposent une solution à ce type de faiblesses. Il s'agit d'une analyse de l'avant de scène effectuée une fois que le fond de scène est détecté. Le fond de scène est d'abord modélisé par une combinaison de gaussiennes. Des informations d'intensité et de texture sont intégrées afin d'éliminer les ombres et de s'adapter aux changements de lumières brusques ou rapides. Une fois la soustraction est effectuée, l'avant de scène ainsi obtenu est alors modélisé et analysé afin de déterminer des régions statiques (s'il y a lieu). Celles-ci sont alors poussées vers le fond. Grâce à une analyse basée sur une quantification de l'énergie des contours des régions de l'avant scène, il est possible de générer une classification de ces régions en objets disparus ou en objets abandonnés. Ce qui constitue une analyse d'avant scène à niveau sémantique élevé. De plus, cette méthode peut s'exécuter à une cadence de 130-150fps, ce qui la rend attrayante pour la vidéo surveillance temps réel. De leur côté, les chercheurs Pradeep et al. [22] ont essayé d'optimiser le temps de traitement d'une approche similaire en intégrant une technique de définition et de suivi de régions d'attention (experiential sampling), ce qui permet de concentrer le plus gros des traitements à ces régions plutôt que de l'appliquer à toute l'image de la séquence vidéo.

VI.2 L'estimation du mouvement

Dans les vidéos capturant des scènes complexes, le mouvement apparent des objets peut avoir des origines diverses. Car, en plus d'une origine due à un mouvement réel d'objets, les changements de l'illumination de la scène, le mouvement de la caméra, et les bruits introduits par les dispositifs électroniques d'acquisition du signal sont tous interprétés comme des changements apparents dans la séquence d'images. Par conséquent, un estimateur de mouvement classique aura tendance à détecter ces changements comme des mouvements apparents de la scène. Dans ce sens, la détection de mouvement correspondant à des mouvements réels d'objets constitue un défi faisant encore l'objet de plusieurs travaux de recherches. Pour répondre à l'objectif de détection de changements dans une séquence de vidéo surveillance, il est important de distinguer ou bien d'identifier les situations suivantes :

- Identifier les objets en mouvement dans la scène (personnes, véhicules, etc.).
- Retenir uniquement les mouvements pertinents et ignorer les mouvements inintéressants ou inutiles tels que le scintillement de l'eau, les vagues d'eau, les fontaines, les effets du vent sur les branches d'arbres, ou sur les rideaux de fenêtres.

Par conséquent, une séquence d'images est réputée utile ou informative si un mouvement pertinent y est détecté. Cette détection de changement peut être exploitée de plusieurs façons, dont les exemples suivants :

- Décider de retenir la séquence (la sélectionner) et de la diriger au stockage vue son utilité potentielle.
- Déclencher une alarme pour détection d'intrus.
- Diriger la séquence ainsi sélectionnée vers des traitements additionnels tels que le calcul de la vitesse, la reconnaissance de face et son identification, le brouillage de face, l'authentification...etc.

L'approche de détection des changements par détection de mouvements pertinents s'avère plus efficace que l'approche de soustraction du fond de scène vu les faiblesses constatées pour cette dernière :

- Nécessité d'une phase d'apprentissage avec de longues séquences vidéo.
- Difficulté de suivre les changements rapides.
- Non distinction entre les mouvements pertinents et les mouvements à caractère distrayant.

Dans leur publication [3] Tian et Hampapur proposent une méthode temps réel de détection de mouvements pertinents de la scène. L'algorithme correspondant est basé sur l'hypothèse de "mouvement à direction consistante" suivante :

- Tout objet à mouvement pertinent a, en général, tendance à se déplacer dans une direction approximativement consistante ou constante pendant un laps de temps.

À partir de cette hypothèse, la détection de changement dans la séquence d'image est basée sur un estimateur de mouvement différentiel (méthode du gradient) et récursif auquel s'ajoutent plusieurs post-traitements visant à distinguer les mouvements pertinents des mouvements distrayants. Ainsi, le processus de détection de changement, tel que proposé, s'effectue en 5 étapes :

- 1)- Calcul de la différence temporelle cumulative : identifie, dans une séquence, toutes les régions d'images ayant eu des changements à travers le temps. La paramètre de cumul permet de détecter les objets à l'arrêt ou à mouvement lent.
- 2)- Estimation de mouvement : s'effectue pour chaque paire d'images successives de la séquence. Un algorithme de type gradient direct et itératif est utilisé. Il minimise la différence temporelle déplacée sur un bloc de $N \times M$ pixels en se basant sur l'hypothèse

d'invariance de la luminance. Un champ de vecteurs de déplacement (dx,dy) dense (répétitif dans un bloc) est ainsi obtenu pour chaque paire d'images. Il est à noter que cet algorithme a fait l'objet de plusieurs études et améliorations à travers le temps [3]. Il est reconnu pour son degré de complexité relativement basse, l'exactitude et la consistance des mouvements estimés par rapport aux mouvements observés ainsi que sa bonne tolérance au bruit.

3)- Filtrage temporel : les champs de mouvement denses obtenus pour une séquence de T images sont filtrés temporellement. Pour chaque pixel de la première image de la séquence, il s'agit de faire le suivi, à travers le temps, pour déterminer si son mouvement jouit d'une direction constante et ceci de manière majoritaire (exemple : un pixel dont la composante de mouvement dx est toujours positive).

4)- Constitution de régions à direction de mouvement constante: les pixels sélectionnés à l'étape précédente (3) sont utilisés comme germes (seeds) pour initialiser une méthode de croissance de région. Le but est de former des régions compactes (fermées et bien délimitées) dont le mouvement est réputé à direction constante. Ceci permet de retrouver les objets de la scène et réduit les situations où un objet est subdivisé en plusieurs petites régions indépendantes. Il est à noter que les auteurs ne spécifient aucune méthode particulière.

5)- Détection d'objets à mouvement pertinent : combine le champ de mouvement, la différence temporelle et les régions constituées en 4). Leur mise en correspondance permet d'identifier les objets de la scène dont le mouvement est pertinent. Cette étape est sensée réduire les cas d'erreur. Le résultat est une map (masque binaire) dont les points mis à 1 indiquent les objets, accompagnée d'une décision (séquence pertinente O/N). Une map nulle correspond à une séquence statique ou une séquence à mouvements non pertinents.

Il est à noter que l'implémentation de toute la méthode de détection des changements, telle que c'est décrit dans [3], a révélé quelques faiblesses. Plus concrètement, nous avons constaté les faits suivants :

- La croissance des régions de l'étape 4, telle que nous l'avons appliquée, ne donne pas lieu à des régions bien délimitées et bien pleines. Leur correspondance à des objets de la scène n'est pas toujours fidèle. Par conséquent, d'autres méthodes plus sophistiquées doivent être considérées (des méthodes de morphologie mathématique).
- La mise en correspondances des résultats à l'étape 5) résulte en une map binaire où les objets en mouvement, bien que correctement localisés, sont fins et petits comparés à leur équivalent dans la séquence originale. L'objet en mouvement est détecté mais sa forme est perdue. Nous pensons que cette étape mérite l'introduction d'une technique plus sophistiquée pour la mise en correspondance. L'incorporation d'une détection de contours peut être envisagée.
- Les mouvements en zigzag seront interprétés comme des mouvements non pertinents.
- Un objet à mouvement lent ou qui s'arrête pendant un laps de temps ne sera pas détecté, sauf lorsqu'il aura repris son mouvement.

Ceci dit, la méthode telle que proposée permet en effet de détecter et rejeter une séquence statique ou bien une séquence dont le mouvement est à caractère distrayant tel

que le mouvement vibratoire de rideaux ou bien le changement brusque d'une lumière. Par contre, un changement lent de lumière n'est pas détecté. Ainsi, seules les séquences dont le mouvement est détecté comme pertinent sont retenues et envoyées pour des post-traitements ou stockage.

Concernant les performances temporelles, les auteurs soulignent la simplicité du processus de traitement, et annoncent un taux de détection de 50 images/seconde en moyenne, sur un Pentium III muni de 1GB de mémoire. Une telle performance fait que cette méthode est attrayante pour les applications temps réel, tel que les systèmes de vision par ordinateur sur PC en général, les systèmes de vidéo surveillance automatiques en particulier, ou encore l'intégration matérielle dans une caméra intelligente.

IV-3 La sélection vidéo

L'analyse de scène de vidéo surveillance visant à estimer le mouvement apparent et à caractériser les types de changements dans la scène peut s'incorporer soit :

- Plus tard, une fois que les données acquises sont stockées, une analyse basée sur le mouvement ou bien sur les changements pertinents peuvent être appliquées en vue de faire de l'extraction d'informations pertinentes (exemple : suivi de cibles) ou de la recherche basée sur le contenu. Ce type de traitements fait partie de ce qu'on appelle communément la vidéo surveillance intelligente.

- Directement après l'acquisition des images (soit dans la caméra elle même ou bien dans tout système qui traite immédiatement les données issues de la caméra). Les exploitations immédiates d'un tel traitement peuvent être :

- la compression des séquences vidéo : en effet, l'estimation du mouvement a toujours été une étape importante dans la réduction de la redondance temporelle dans les standards de codage les plus populaires, MPEG, H264 et MJPEG.

- la sélection de séquences vidéo pour stockage: l'analyse du mouvement ou des changements de scène permet de distinguer les scènes statiques des scènes dynamiques et de distinguer aussi les scènes à changements pertinents (mouvements intéressants) de celles dont les changements sont non pertinents (exemple : mouvements distrayants, objets non intéressants). Seules les séquences contenant des informations jugées intéressantes (suite à cette analyse) sont alors retenues pour le stockage. Ce qui réduit l'espace de stockage requis par une surveillance continue (24/24) et, au même temps, expose l'opérateur à la partie probablement essentielle et utile des séquences vidéo. Notons qu'il importe que cette notion de pertinence soit définie selon le contexte de chaque application. Par exemple, un objet en mouvement pertinent peut être une personne qui bouge ou bien un véhicule qui bouge selon qu'on est dans le cas de surveillance de l'intérieur d'un édifice à bureaux ou bien d'une autoroute.

V. La protection de l'information privée

Le déploiement à grande échelle de la vidéo surveillance ainsi que la surveillance continue d'individus dans leur milieu de travail et dans les lieux publics soulèvent de grandes inquiétudes relativement aux intrusions potentielles dans la vie privée d'individus innocents. Même si l'apport sécuritaire de la vidéo surveillance est largement apprécié et reconnu, ceci n'écarte pas sa perception comme une véritable menace sur la sphère privée. En effet, des craintes sont reportées sur les abus possibles que peut engendrer la disponibilité de quantités d'informations visuelles privées entre de mauvaises mains et leur utilisation à des fins autres que celles prévues initialement (espionnage, oppression politique, chantage, voyeurisme). Garantir la protection de la vie privée se pose alors comme une exigence cruciale à laquelle doivent se conformer les systèmes de vidéo surveillance modernes. L'atteinte d'une telle exigence peut se faire par le biais d'une coopération conjointe de plusieurs actions opérant à des niveaux législatifs, organisationnels et technologiques [1][2]. Aussi, l'annexe I de ce document constitue une bonne introduction à ce sujet. Bien que cette problématique ne soit qu'à ses débuts, plusieurs efforts de recherche ont d'ores et déjà été déployés afin de proposer des solutions technologiques aidant à l'amélioration de la protection des informations privées dans les séquences de vidéo surveillance capturées [1][7][13][14][19][20]. En général, la plupart des approches proposées ont en commun une démarche algorithmique qui peut se résumer en trois étapes :

- 1- Analyser la séquence vidéo et localiser des régions d'intérêt dont les détails sont à caractères privés.
- 2- Appliquer une procédure de dissimulation de l'information privée. Celle-ci peut être une obscuration (flou, brouillage), un camouflage ou bien carrément un effacement (extraction).
- 3- Mettre à disposition une stratégie de recouvrement de l'information privée sur demande d'autorités légales et compétentes. Par exemple, la dissimulation appliquée peut être réversible moyennant l'utilisation d'une clé secrète.

Dans leur publication [19], Newton et al. considèrent que les logiciels de reconnaissance automatique de formes faciales sont des menaces pour la protection de l'information privée. De ce fait, ils analysent les possibilités de débusquer ces algorithmes et réduire leur taux de réussite. Ceci a abouti à la proposition de l'algorithme k-same. Il s'agit d'un processus de dé-identification, qui prend un objet facial détecté dans la scène, effectue une analyse par similarité suivie d'un moyennage de composantes d'images et obtient une nouvelle face 'fictive' qui remplacera la face originale. La nouvelle face, d'apparence tout à fait normale, est alors méconnaissable et anonyme alors que tout le comportement des individus de la scène reste compréhensible et observable. La réversibilité du processus de dé-identification est annoncée comme possible, mais aucune implémentation n'est décrite. De plus, cette méthode présente une complexité assez élevée, ce qui risque de compromettre tout projet d'implémentation en temps réel. De son côté, Boulton [20] exploite tout simplement la cryptographie pour dissimuler l'information

privée d'une séquence d'images. Les régions pertinentes de l'image sont encryptées pour avoir des apparences aléatoires et bruitées. Une clé secrète peut être utilisée pour décrypter ces régions. Bien que la dissimulation soit tout à fait réalisable et que la réversibilité soit garantie, cette méthode ne spécifie pas la partie concernant la détection des régions pertinentes contenant les informations privées. En s'appuyant sur des concepts cryptographiques et des listes de contrôle d'accès, Senior et al. [1] conçoivent une console vidéo qui analyse la séquence vidéo capturée et délivre plusieurs séquences vidéo permettant un accès hiérarchique et contrôlé à l'information privée. Par exemple, un usager ordinaire peut accéder à une information statistique ; un usager privilégié peut accéder à une information privée limitée ; alors qu'une institution légale peut accéder à toutes les données originales en plus d'information additionnelles telles que les identités d'individus. La Privacy-Cam est une version réduite, autonome et portable sur un senseur hardware. Il est à noter que l'inconvénient majeur de cette méthode est la multiplicité des données vidéo requise au stockage.

Afin d'aider à renforcer la compréhension en ce qui a trait à la protection de l'information privée à l'aide de la technologie de traitement de l'information, nous nous proposons d'analyser plus en détail deux propositions de systèmes publiées en 2005 et 2006.

Dans leur publication [7], les chercheurs Zhang et al. proposent un système de vidéo surveillance qui permet de protéger l'information privée et d'authentifier les données visuelles de la séquence vidéo. Le travail, tel que décrit, se concentre beaucoup plus sur la dissimulation de l'information privée plutôt que la manière de l'extraire. Le mécanisme de protection se base sur l'extraction et le retrait de l'information privée, son codage et insertion cachée par marquage invisible ainsi que la possibilité de restitution par une autorité en possession de la clé secrète. Le système proposé est supposé opérer dans un environnement fermé et restreint où circulent des personnes connues et identifiables. Cet environnement pourrait être par exemple un hôpital ou une institution administrative. Parmi les exigences du système figurent sa capacité à distinguer ces personnes parmi tant d'autres et veiller à protéger leur vie privée de toutes sortes d'abus pouvant résulter d'une surveillance continue et intensive. De manière plus spécifique, le fonctionnement d'un tel système procède comme suit :

- Les individus apparaissant dans une scène capturée sont classifiés en deux groupes : personnes autorisées et personnes non autorisées
- Les régions correspondant aux personnes autorisées sont retirées des images et remplacées par un fond d'image adéquat
- Les régions ainsi extraites sont compressées et cryptées à l'aide d'une clé pour constituer un watermark d'informations privées.
- Le reste de la séquence est aussi compressé en utilisant la DCT puis le watermark y est inséré par une méthode qui respecte des critères de perceptibilité minimale.
- Une signature d'authentification est générée pour toute la séquence vidéo marquée. Cette signature est insérée au niveau de l'entête.

Lors du décodage, seul un usager en possession de la clé secrète sera capable de décrypter l'information privée à partir de la marque extraite. Ainsi, cet usager pourra

reconstruire la séquence vidéo en remplaçant une “approximation acceptable” de l’information privée dans la séquence. Pour un usager quelconque, la séquence vidéo est tout simplement décompressée et visualisée sans l’information privée.

Il est à noter que toute manipulation malicieuse de la séquence vidéo risque de compromettre le processus de reconstruction. D’où la nécessité de s’assurer au préalable que la séquence vidéo est authentique par le biais de sa signature.

Parmi les faiblesses du design de ce système on peut citer :

- la fragilité de l’authentification
- l’incapacité de localiser les régions manipulées dans l’image
- l’incapacité à réparer ou reconstruire les régions manipulées
- la dépendance forte de la reconstruction par rapport à l’authentification
- l’absence d’une définition claire d’une méthode de classification des objets personnes en vue de l’extraction de l’information privée.
- Absence d’une analyse claire des attaques possibles sur le schéma de marquage proposé.

En première analyse, ce système serait vulnérable aux attaques suivantes :

- Toute modification, même mineure, apportée à la vidéo compressée fait échouer l’opération d’authentification si celle-ci est fragile (voir la référence [4] de [7])
- Dans le cas où une modification apportée à la vidéo compressés altère le watermark caché, ceci va faire échouer l’opération de décryptage de l’information privée. Une approche de cryptage localisée serait dans ce cas à moindre risque.

Suite à l’analyse du système proposé, il serait utile de trouver une réponse adéquate à la question suivante : quel est l’avantage d’insérer la marque dans les coefficients DCT de la vidéo plutôt que les bits de poids faible du domaine spatial ? quelle réponse parmi les suivantes serait la plus juste ?

1. Pour assurer une bonne qualité de reconstruction
2. Pour assurer une meilleure robustesse de la marque contre les manipulations
3. Pour pouvoir compresser la vidéo sans altérer la marque (plusieurs algorithmes de compression JPEG utilisent la DCT quantifiée)

En complément de la réponse 3, il est à noter que les auteurs ont mis le doigt sur le défi que constitue l’insertion d’une marque volumineuse en nombre de bits. A cela, nous ajoutons le défi d’extraire la marque avec une grande précision. Ils ont ainsi procédé à des expérimentations qui consistent à trouver comment insérer la marque dans les coefficients DCT de façon à garantir un bon compromis entre la qualité de l’image marquée et le taux de compression vidéo obtenu.

Dufaux et al. [13,14] ont proposé un système de protection de l'information privée qui se base sur l'identification dans chaque image de régions d'intérêt (contenant une information jugée à caractère privée) et l'application d'une technique de brouillage contrôlée par une clé cryptographique. Le contenu de la séquence vidéo reste quand même compréhensible même si certains objets de la scène (par exemple des faces de personnes, ou des immatriculations de voiture) ne sont pas identifiables. Pour un usager en possession de la clé cryptographique, le processus de brouillage est réversible et il est donc tout à fait possible de restituer les régions brouillées à leur aspect original et de les visualiser de manière claire.

L'identification de régions d'intérêt est sensée se faire par analyse de scène (détection de faces, détection de changements...), mais aucune méthode particulière n'est spécifiée. Pour le reste des traitements, il est supposé que ces régions sont connues et déterminées par un masque binaire de segmentation (ground truth segmentation mask).

La méthode de brouillage qui est utilisée pour camoufler l'information jugée privée dans les séquences d'images de vidéo surveillance est déjà apparue dans [16,17] et repose sur l'inversion pseudo-aléatoire du signe de coefficients transformés appartenant aux régions que l'on veut brouiller. Cette méthode est très attrayante pour plusieurs raisons :

- La complexité de traitement est relativement bas pour un niveau de sécurité très acceptable.
- Le brouillage est une transformation complètement réversible
- Le brouillage est flexible puisque la variation du nombre d'inversion de signes permet le contrôle du taux de brouillage (faible à très fort).
- Plusieurs transformées peuvent être utilisées. La DCT et la transformée en ondelettes ont été particulièrement testées. Ce qui rend cette méthode de brouillage exploitable conjointement avec la plupart des standards de codage d'images et vidéo (JPEG, JPEG2000, motion JPEG, motion JPEG200, MPEG ou AVC).
- L'inversion de signe ne diminue pas l'efficacité de la compression lors du codage vidéo. Le bitstream augmente de moins de 10% dans les cas MPEG4 et JPEG2000.
- Le brouillage peut s'appliquer à n'importe quelle forme de région d'intérêt. Cependant, toute forme arbitraire doit être transmise comme une méta donnée du bitstream vidéo ou bien dans un canal séparé. Dans le cas de JPEG2000, la syntaxe réservée au codage ROI peut être exploitée.

Pour élever le niveau de sécurité du processus de brouillage, les auteurs proposent d'utiliser plusieurs matrices aléatoires (indiquant les positions d'inversion de signe) générées par le biais de plusieurs germes (seeds), ces derniers sont encryptés par l'algorithme RSA et encodé dans le bitstream de la vidéo compressée. Donc seule une clé cryptologique a besoin d'être gardée secrète.

VI. Conclusion

À l'issue de cette investigation, il s'avère que les séquences de vidéo surveillance peuvent être sécurisées par marquage tout en tolérant une coopération avec certains traitements qui peuvent soit précéder cette sécurisation (analyse des changements de scène, sélection vidéo, protection de l'information privée), ou bien s'opérer conjointement (compression) à celle-ci. Ce qui rend la conception d'une caméra intelligente avec sécurisation tout à fait plausible, pourvu que l'ensemble des traitements requis obéisse aux contraintes temps réel essentiellement dictées par les paramètres d'acquisition. Il est donc payant de s'investir sur une conception suivie d'une implémentation temps réel d'un dispositif prototype de traitement temps réel incorporable au sein d'une caméra visant essentiellement une sécurisation de type authentification conjointement à une protection de l'information privée.

RÉFÉRENCES

- [1] Senior, A.; Pankanti, S.; Hampapur, A.; Brown, L.; Ying-Li Tian; Ekin, A.; Connell, J.; Chiao Fe Shu; Lu, M.; "*Enabling video privacy through computer vision.*" IEEE Security and Privacy Magazine, Vol. 3, Issue 3, Page(s):50-57, May-June 2005.
- [2] A. Hampapur, L. Brown, J. Connell, S. Pankanti, A.W. Senior, and Y-L Tian, "*Smart surveillance: applications, technologies and implications.*" IEEE Pacific-Rim Conference On Multimedia, Singapore, Dec., 2003.
- [3] Ying-li Tian and Arun Hampapur, "*Robust salient motion detection with complex background for real-time video surveillance.*" IEEE Computer Society Workshop on Motion and Video Computing, Breckenridge, Colorado, January 5 and 6, 2005.
- [4] Arun Hampapur, Lisa M. Brown, Jonathan Connell, Max Lu, Hans Merkl, S. Pankanti, Andrew W. Senior, Chiao-fe Shu, and Ying-li Tian, "*Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking*" IEEE Signal Processing Magazine, Vol. 22, No. 2, March 2005.
- [5] Ying-li Tian and Arun Hampapur, "*Multiscale tracking for smart video surveillance.*" IEEE CVPR, San Diego, 2005.
- [6] Foresti, G.L.; Micheloni, C.; Snidaro, L.; Remagnino, P.; Ellis, T. "*Active video-based surveillance system: the low-level image and video processing techniques for implementation.*" IEEE Signal Processing Magazine, Vol. 22, No. 2, March 2005.

- [7] Zhang, W., S.-C. Cheung, and M. Chen, "*Hiding privacy information in video surveillance system.*" IEEE International Conference on Image Processing, ICIP 2005, pp. 868-871, Genoa, Italy, September 2005.
- [8] Cheung, S.-C. and C. Kamath. "*Robust techniques for background subtraction in urban traffic video.*" Proceedings of Electronic Imaging: Visual Communications and Image Processing(Part One), San Jose, California. Bellingham, WA:SPIE. (5308):881-892, January 2004.
- [9] Fonda, F. Pastore, S., "*Innovative image watermarking technique for image authentication in surveillance applications.*" IEEE International Workshop on Imaging systems and techniques, Niagara Falls, May 2005.
- [10] Yuk Ying Chung Fang Fei Xu, "*A Secure Digital Watermarking Scheme for MPEG-2 Video Copyright Protection.*" IEEE International Conference on Advanced Video and Signal based Surveillance AVSS, Sydney, Australia, November 2006.
- [11] Bartolini, F. Tefas, A. Barni, M. Pitas, I., "*Image authentication techniques for surveillance applications.*" Proceedings of the IEEE, Vol. 89, Issue 10, October 2004.
- [12] J. Apostolopoulos, S. Wee, F. Dufaux, T. Ebrahimi, Q. Sun and Z. Zhang, The Emerging JPEG 2000 Security (JPSEC) Standard, *IEEE Int. Symp. on Circuits and Systems (ISCAS)*, 2006.
- [13] F. Dufaux and T. Ebrahimi, "*Scrambling for Video Surveillance with Privacy.*" *IEEE Workshop on Privacy Research in Vision*, 2006.
- [14] F. Dufaux, M. Ouaret, Y. Abdeljaoued, A. Navarro, F. Vergnenegre and T. Ebrahimi, "*Privacy Enabling Technology for Video Surveillance.*" SPIE Mobile Multimedia/Image Processing for Military and Security Applications, 2006.
- [15] Cheung, S.-C. and C. Kamath, "*Robust techniques for background subtraction in urban traffic video.*" Proceedings of Electronic Imaging: Visual Communications and Image Processing (Part One), , San Jose, January 2004.
- [16] W. Zeng and S. Lei, "*Efficient Frequency Domain Video Scrambling for Content Access Control*", in Proc. ACM. Multimedia, Orlando, FL, Oct. 1999.
- [17] W. Zeng and S. Lei, "*Efficient frequency domain selective scrambling of digital video.*" in IEEE Trans. Multimedia, March 2003.
- [18] Ying-li Tian, Max Lu and Arun Hampapur, "*Robust and Efficient Foreground Analysis for Real-time Video Surveillance*" IEEE CVPR, San Diego, 2005.
- [19] Newton, E. Sweeny, L. and Malin B. "*Preserving Privacy by de-identifying facial images* ", IEEE Transactions on Knowledge and Data Engineering, 17(2) , February 2005.

- [20] Boulton, T.E. "*PICO: privacy through invertible cryptographic obscuration*", Proceedings of the Computer Vision for Interactive and Intelligent Environment CVIIE'05, November 2005.
- [21] Fonda, F. and Pastore, S. "*Innovative image watermarking technique for image authentication in surveillance applications*", IEEE International Workshop on Imaging Systems and Techniques, May 2005.
- [22] Atrey, P.K. Kumar, V. Kumar, A. Kankanhalli, M.S. "*Experiential Sampling based Foreground/Background Segmentation for Video Surveillance*", IEEE International Conference on Multimedia and Expo, ICME06, Toronto, July 2006.
- [23] Fridrich, J. and Goljan, M. "*Protection of digital images using self embedding*", Symposium on Content Security and Data Hiding in Digital Media, New Jersey Institute of Technology, May 14, 1999.
- [24] Lin, C. and Chang, S. "*A robust image authentication method distinguishing JPEG compression from malicious manipulation*", IEEE Trans. On Circuits and Systems of Video Technology, vol. 11, no. 2, February 2001, pp. 153–168.
- [25] Rey, C. and Dugelay, J.L. "*A survey of watermarking algorithms for image authentication*", EURASIP Journal on Applied Signal Processing, June 2002, pp. 613–621.
- [26] Sun, Q.B., Chang, S.F., Kurato, M. and Suto, M. "*A new semi-fragile image authentication framework combining ECC and PKI infrastructure*", ISCAS02, Phoenix, USA, May 2002.
- [27] Sun, Q.B. and Chang, S.F. "*Semi-fragile image authentication using generic wavelet domain features and ECC*", Proc. ICIP, Rochester, USA, September 2002.

ANNEXE

Protection de la sphère privée et vidéo surveillance

(par Virginie CARNIEL , extrait en juin 2007 de :
<http://ditwww.epfl.ch/SIC/SA/SPIP/Publications/spip.php?article1132>)

Au cours de ces dernières années, les pays industrialisés ont subi un nombre croissant d' actes terroristes et ont vu le taux de criminalité de leurs zones urbaines augmenter de façon conséquente. En réponse à ces menaces diverses et pour assurer la sécurité de leurs concitoyens, les autorités publiques ont massivement déployé des systèmes de vidéo surveillance dans les points stratégiques que cela soit dans les aéroports, les banques, les transports publics ou dans les zones urbaines. Or, ce type de mesures soulève immédiatement la question de la protection de la sphère privée et le risque de déviance vers une société à la Big Brother. Ceci d' autant plus que les données récoltées par les systèmes de vidéo surveillance pourraient être et ont été utilisées de manière abusive par des opérateurs à des fins de voyeurisme, de chantage ou de discrimination. Pour que les individus puissent jouir d' une sécurité accrue tout en ayant la garantie de la protection de leur sphère privée, EMITALL Surveillance SA (www.emitall.com), une jeune start-up de Montreux, a développé une technologie dans le domaine de la vidéo surveillance intelligente qui permet de répondre à ces deux préoccupations.

Contexte politique

Les gouvernements du monde entier sont concernés par le risque accru d' insécurité. Afin d' anticiper les actions terroristes et dans le but de permettre

l'identification de suspects, les autorités investissent des montants toujours plus élevés dans la sécurité avec un fort accent sur les moyens de vidéo surveillance. Or, certains pays, à l'instar de la Suisse, de la France, du Danemark, de l'Allemagne et du Canada, sont sensibles par tradition aux questions de la protection des données et recherchent des solutions pour garantir la préservation de l'anonymat de leurs citoyens.

En Suisse, le préposé fédéral à la protection des données a émis des recommandations sur l'utilisation de la vidéo surveillance (www.edsb.ch).

- La vidéosurveillance ne peut être effectuée que si cette atteinte à la personnalité est justifiée par le consentement des personnes concernées, par un intérêt prépondérant public ou privé ou par la loi (principe de licéité).
- La vidéosurveillance doit être un moyen adéquat et nécessaire à la réalisation de l'objectif poursuivi, à savoir la sécurité, notamment la protection contre les atteintes aux biens et/ou aux personnes. Elle ne peut être retenue que si d'autres mesures moins attentatoires à la vie privée telles que verrouillages complémentaires, renforcement des portes d'entrée, systèmes d'alarme, s'avèrent insuffisantes ou impraticables (principe de proportionnalité).

Toujours selon ce principe de proportionnalité, les données personnelles enregistrées par une caméra doivent être effacées dans un délai particulièrement bref. En effet, la constatation d'une infraction aux biens ou aux personnes aura lieu dans la plupart des cas dans les heures qui suivent sa perpétration. Un délai de 24 heures apparaît donc suffisant au regard de la finalité poursuivie dans la mesure où aucune atteinte aux biens ou aux personnes n'est constatée dans ce délai. Ce délai peut être plus long dans certains cas de vidéo surveillance. En Suisse comme dans bon nombre d'autres pays, le débat sur la vidéo surveillance est au coeur de l'actualité politique et les autorités cantonales et communales sont de plus en plus souvent amenées à légiférer sur le sujet.

Une solution pour la protection de la sphère privée

Pour répondre à la préoccupation des autorités politiques de garantir une sécurité accrue tout en préservant la sphère privée, EMITALL Surveillance SA, une jeune société montreuusienne spécialisée dans les technologies de la vidéo surveillance intelligente, a développé un logiciel spécifiquement conçu pour être intégré dans les plates-formes de vidéo surveillance. Cette technologie logicielle permet la détection automatique d' événements (personnes ou objets en mouvement par exemple) tout en brouillant automatiquement et sélectivement les régions correspondantes (lesdits personnes ou objets ne peuvent donc plus être identifiés), garantissant ainsi l' anonymat des personnes filmées par les caméras de vidéo surveillance.

La force novatrice de cette technologie réside dans le module d' analyse vidéo qui identifie les régions d' intérêt sensibles, telles que personnes ou plaques d' immatriculation par exemple, qui les brouille ensuite et surtout qui permet la réversibilité de l' opération, à savoir une ouverture de l' image grâce à une clé d' encryptage.

La technologie de brouillage est compatible avec la plupart des techniques de compression vidéo telles que Motion JPEG, Motion JPEG 2000, MPEG-4 ou AVC/H.264. D' autre part, le niveau de distorsion introduit peut aller d' un flou léger à un bruit complet, ceci ayant pour conséquence que seules les informations sensibles sont brouillées alors que le reste de la scène demeure compréhensible.

Le brouillage est contrôlé par une clé secrète d' encryptage qui permet aux personnes autorisées d' inverser le processus et d' ouvrir les images brouillées. La possession de la clé étant dépendante du système juridique en place, elle sera en général détenue par une autorité compétente à l' instar d' un juge d' instruction ou de toute autre force légale. En revanche, toute personne non autorisée et ne possédant pas la clé, ne pourra pas accéder aux données en clair et ne pourra reconnaître les individus filmés ou autres informations sensibles puisque non identifiables. La technologie d' EMITALL

Surveillance a été spécifiquement développée pour pouvoir être intégrée dans les plates-formes de vidéo surveillance publique existantes ou futures.



- La caméra capte l'image.
- Dans la scène ci-dessus, les personnes sont automatiquement détectées et brouillées.
- Les images sont enregistrées et stockées en mode brouillé.
- Une clé secrète d'encryptage protèges les objets brouillés et est détenue par l'autorité compétente.
- En cas de suspicion d'acte délictueux, l'autorité compétente peut prendre la décision d'ouvrir les données cryptées.
- Chaque objet est protégé par une clé spécifique
- En utilisant une ou plusieurs clés, l'objet désigné est ouvert et permet l'identification du suspect sans perte de qualité des données stockées.

Différentes intensités de brouillage peuvent être appliquées sur la scène

Dans la scène 1, l' illustration en page suivante montre une scène urbaine avec un brouillage de forte intensité qui permet de reconnaître les silhouettes des individus et les contours des voitures sans permettre l' identification. Dans ce cas, le système a détecté tous les objets en mouvement et les a brouillés.

Dans la scène 2, on voit cette fois un brouillage de moyenne intensité qui permet de voir plus de détails sans toutefois pouvoir identifier les personnes filmées.



Scène 1



Scène 2

Technologie

L'approche utilisée consiste à détecter des régions d'intérêt, les brouiller et les protéger par une clé de cryptage en travaillant dans le domaine transformé (*transform domain*). Cette approche est générique et peut être appliquée à toute technique de codage par transformée (transform-coding) telles que celles basées sur la transformée en cosinus (*Discrete Cosine Transform* -DCT) ou transformée en ondelettes (*Discrete Wavelet Transform* -DWT). La décision d'effectuer le brouillage dans le domaine transformé est justifiée par l'efficacité optimale des taux de compression obtenus. En effet si l'on appliquait le brouillage avant la compression, on risquerait de perdre de l'efficacité dans le processus de compression, de même que si l'on appliquait le brouillage après la compression, on rencontrerait des difficultés à garder la syntaxe du *codestream* conforme.

Le brouillage est obtenu en inversant les signes des coefficients durant la compression. Cette technique est flexible et permet d'ajuster le niveau de distorsion introduit, en passant d'un flou léger à un bruit complet.

Les régions d'intérêt peuvent correspondre soit à des zones prédéfinies dans la scène ou être automatiquement estimées en utilisant l'analyse vidéo. La

segmentation automatique d' objets dans la vidéo peut poser problème, or en utilisant des techniques telles que la détection de visage, la détection de changements, la détection de peau ou le tracking, ou encore une combinaison de ces diverses méthodes, le résultat sera probant.

Ces techniques de brouillage offrent bon nombre d' avantages. En effet, la sortie du système consiste en un flux de données uniques protégé. Ce même flux de données est transmis à tous les clients indifféremment de leur contrôle d' accès et d' identification. D' une part, les clients autorisés, en possession de la clé d' encryptage, peuvent *dé-brouiller* le flux de données et retrouver l' image d' origine sans distorsion et d' autre part, les personnes non autorisées ne verront que l' image brouillée. La technique développée est ainsi très flexible. Elle a peu d' impact sur les performances d' encodage et requiert peu de puissance de calcul alors qu' elle s' adapte à la plupart des standards de compression vidéo existants.



Illustration de Motion JPEG 2000 - Discrete Wavelet Transform (DWT) - Intra-frame coding

Conclusion

Protéger l' anonymat des individus tout en offrant la possibilité de les identifier dans le cadre d' une enquête officielle donne un moyen performant et un outil efficace aux autorités pour éviter les abus de la vidéo surveillance et surtout leur permet d' augmenter l' acceptation de telles installations auprès des citoyens. Le fait que les images soient enregistrées et stockées brouillées permet aussi de garder les données plus longtemps avec l' assurance d' un anonymat garanti. Ceci démontre que les technologies peuvent apporter des solutions pour la protection de la sphère privée et même influencer les législateurs. En effet, la Commune du Grand-Saconnex a voté récemment un texte qui stipule que des

installations de vidéo surveillance pourront être installées à la condition expresse que celles-ci comporte une technologie de brouillage réversible.

Références

- F. Dufaux and T. Ebrahimi, **Scrambling for Video Surveillance with Privacy**, Proc. of IEEE Workshop on Privacy Research In Vision, New York, NY, June 2006. <http://www.emital.com/template/fs/documents/scambling.pdf>.
- F. Dufaux, M. Ouaret Y. Abdeljaoued, A. Navarro, F. Vergnenegre and T. Ebrahimi, **Privacy Enabling Technology for Video Surveillance**, in *SPIE Proc. Mobile Multimedia / Image Processing for Military and Security Applications*, Orlando, FL, April 2006. <http://www.emital.com/template/fs/documents/privacy.pdf>.