ASePPI : protéger la vie privée tout en préservant l'utilité de la vidéosurveillance

Natacha RUCHAUD[1], Jean-Luc DUGELAY[1],

[1]Eurecom
Campus SophiaTech, 450 Route des Chappes, 06410 Biot, France
ruchaud@eurecom.fr, dugelay@eurecom.fr

Résumé – L'utilisation des systèmes de vidéo surveillance augmente de plus en plus chaque année ce qui soulève de nombreuses questions liées à la protection de la vie privée. Dans cet article, nous présentons ASePPI, une nouvelle méthode de brouillage intégrée au codec H.264/AVC qui adapte le niveau de protection de la vie privée grâce au DC que nous modifions automatiquement en fonction de la résolution de la région à protéger. Les coefficients DCT originaux sont cryptés et cachés dans les coefficients AC. En comparaison avec les méthodes existantes, notre procédé offre un meilleur compromis entre la protection de la vie privée et la visibilité de la scène. De plus, l'impact sur le flux compressé est négligeable.

Abstract – The usage of video surveillance systems increases more and more every year and protecting people privacy becomes a serious concern. In this paper, we present ASePPI, an Adaptive Scrambling enabling Privacy Protection and Intelligibility. We integrate our process within the H.264 codec. The proposed approach automatically adapts the level of protection by modifying the DC coefficient according to the resolution of the region of interest. We encrypt and hide the original DCT coefficients in the AC coefficients. Compared to existing methods, our framework provides a better trade-off between the privacy protection and the visibility of the scene for monitoring. Moreover, the impact on the compressed stream is negligible.

1 Introduction

Les systèmes de détection et de reconnaissance, combinés aux réseaux omniprésents de caméras, accentuent les problèmes de confidentialité. Protéger la vie privée des personnes dans des vidéos de surveillance est complexe, car les actions doivent rester visibles. Ainsi, le défi à relever dans cet article est de gérer le compromis entre la protection de la vie privée (c.-à.-d. rendre l'identité des personnes méconnaissable) et l'utilité de la vidéosurveillance (c.-à.-d. garder une bonne compréhension et visibilité de la scène).

Certains systèmes utilisent des méthodes basiques et non réversibles pour anonymiser des personnes telles que le flou gaussien, la pixellisation ou un masque noircissant.

Les auteurs de [9, 8], déplacent dans le domaine spatial les bits les plus significatifs (MSB) des pixels cryptés vers les bits les moins significatifs (LSB). Ensuite, les bits de l'image contours sont insérés dans les MSB de l'image résultante afin de garder la scène compréhensible grâce aux contours. Si les images générées par cette méthode sont compressées elles ne seront plus réversibles. De nos jours, presque toutes les vidéos sont compressées. Il est donc préférable que les algorithmes de traitement d'images soient compatibles avec la compression.

Le profil de base du codec standard en compression vidéo, H.264/AVC, prend en charge les intra (I) et inter images prédictives (P) ainsi que le codage entropique avec des codes de longueur variable adaptatifs au contexte (CAVLC). Les images I contiennent uniquement des blocs intra qui sont uniquement prédits à partir de blocs de la même image. Les images P contiennent des blocs intra et inter, ces derniers sont prédits à partir de blocs d'une image de référence.

Pour protéger la vie privée, Dufaux et Ebrahimi [4] proposent de permuter aléatoirement les signes des coefficients non nuls de chaque bloc des zones à protéger pour le codec MPEG-4. Cependant, l'effet de brouillage est relativement faible, en particulier sur les images de haute résolution, ce qui ne protège pas toujours la vie privée.

Pour améliorer la protection de la vie privée, Wang et al. [12] proposent de crypter les modes intra prédictifs (IPM) en plus des signes des coefficients non nuls (SNC).

La plupart des méthodes de cryptage utilisées pour protéger la vie privée ([12], [5], [11]) produisent des images fortement brouillées empêchant la surveillance. Le cryptage est une méthode réversible et efficace pour protéger la vie privée, mais elle a du mal à gérer le compromis avec la surveillance.

Ruchaud et Dugelay [10] gèrent ce compromis en appliquant la fonction OU exclusif bit à bit pour chaque coefficient DCT (DC + AC) avec un nombre pseudo-aléatoire dans le cadre du codec JPEG (opérant sur des images fixes et non sur des vidéos). Ils décalent ensuite les coefficients chiffrés d'une position pour attribuer une valeur de leur choix pour le coefficient DC ce qui permet de mieux maîtriser le contenu de l'image finale.

Nous reprenons l'idée d'utiliser le DC pour gérer le compro-

mis entre la compréhension de la scène et la protection de la vie privée, tout en cryptant les coefficients d'origines. La principale utilité des vidéos de surveillance est de visualiser clairement les évènements en cas de litige. Notre méthode est réversible et rendue compatible avec le codec vidéo H.264/AVC. De plus, notre procédé adapte automatiquement le niveau de protection de la vie privée en fonction de la taille des régions d'intérêt.

Dans la section suivante, nous décrivons l'approche proposée. Nous présentons et discutons les résultats dans la section 3. Enfin, dans la section 4, nous concluons et donnons des perspectives pour la suite de nos travaux.

2 Description de la méthode ASePPI

L'intégration dans le codec H.264/AVC est réalisée uniquement pour le canal de luminance (Y) à l'intérieur de chaque bloc appartenant à la région d'intérêt, dénoté RoIs (e.g, les visages). Les blocs sont prédits à partir de blocs non brouillés lors de l'encodage, on crypte seulement l'erreur de prédiction après transformation et quantification de ses coefficients. Nous utilisons deux méthodes proposées par Tong, Dai et al. [3] pour éviter les erreurs de décodage produites dans la zone non cryptée par le brouillage : "Mode Restricted Intra Prediction" (MRIP) et "Search Window Restricted Motion Estimation" (SWRME).

2.1 Cryptage des blocs résiduels des images I

Pour chaque bloc résiduel des images I, nous cryptons les coefficients (DC et ACs) pour protéger l'information originale. Les coefficients chiffrés sont ensuite décalés d'une position vers les hautes fréquences. Ainsi, nous choisissons une nouvelle valeur du DC en fonction de nos critères (i.e. protection de la vie privée et visibilité de la scène). Le coefficient le moins significatif est perdu volontairement pour permettre le dédoublement du DC, un dédié à gérer l'apparence finale de l'image et l'autre est crypté et stocké dans le premier coefficient AC (pour la restitution).

2.1.1 Cryptage du DC

Afin de limiter le bruit dû à une valeur élevée du DC brouillé qui est caché dans le premier coefficient AC, nous le cryptons suivant l'algorithme 1 avec s(DC) égal à -1 si le signe du DC est négatif et +1 si positif. Ce cryptage permet de garder les valeurs cryptées dans un même ordre de grandeur/valeur que celles originales.

2.1.2 Permutation des AC

Nous extrayons, selon le parcours en zigzag, les coefficients AC (sauf le dernier). Les coefficients AC avant le dernier coefficient non nul sont permutés aléatoirement en utilisant l'algorithme Knuth shuffle [2]. Le dernier coefficient non nul est utilisé pour marquer la fin de la permutation.

Algorithme 1: Cryptage du DC

Générer un nombre aléatoire (RN);
if *(*$|DC| < 16$*)* **then**
  X = 16;
else
  $X = 2^n$;
if *(*$DC \neq 0$*)* & *((*$|DC| \neq (RN \mod X)$*))* **then**
  DC_e = ($|DC| \oplus (RN \mod X)$)*sign(DC);
else
  DC_e = DC;
with $n = \lfloor \log_2 |DC| \rfloor$ an integer

2.1.3 Cacher les coefficients brouillés dans les AC

Nous décalons les coefficients brouillés d'une position pour rendre disponible celle du DC. Par exemple, les coefficients extraits initialement [31 (DC), 0, -2, -1, -1, -1, 0, 0, -1, EOB] deviennent les coefficients brouillés suivants : [DC, 24 (DC crypté), -1, 0, -2, -1, 0, 0, -1, -1, EOB]. Nous transposons ces coefficients en un bloc en fonction du parcours en zigzag et choisissons la valeur DC avec la formule définie dans 2.1.4.

2.1.4 Choix de la nouvelle valeur du DC

Tandis que les coefficients chiffrés apparaissent sous forme de bruit dans la région protégée, la valeur du DC est dédiée à restituer une partie des informations d'origine. Si nous gardons uniquement le DC (i.e. la luminance moyenne) de chaque bloc résiduel, l'image devient pixelisée de la taille de ces blocs (4*4). Nous allons insérer le même DC pour plusieurs 4*4 blocs résiduels. Par exemple, si on veut une image pixélisée avec une taille de blocs de 16*16, les 4*4 sous-blocs résiduels auront le même coefficient DC.

L'équation (1) représente la relation entre la taille des blocs voulus, notée S, et le nombre de blocs voulus, noté Nb, en fonction du nombre de pixels (h x w) du RoI.

$$Nb = \frac{h * w}{S * S} \tag{1}$$

Nous re écrivons l'équation (1) en fonction de Nb et avec S un multiple de 4 car la taille des blocs résiduels est de 4*4.

$$S = round\left(\frac{\sqrt{\frac{h*w}{Nb}}}{4}\right) * 4 \tag{2}$$

Plus il y a de blocs plus la qualité de l'image sera élevée et la visibilité de la scène meilleure. L'objectif est de trouver le nombre de blocs, Nb, qui préserve la visibilité des évènements tout en minimisant les performances de reconnaissance faciale. Pour cela, nous évaluons les performances de reconnaissance de visage en faisant varier Nb et la taille du RoI (h et w).

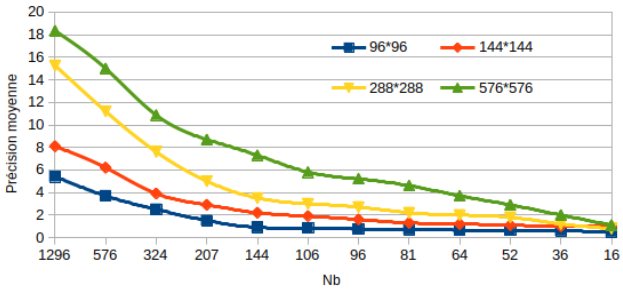

FIGURE 1: Précision de la reconnaissance d'identité (%).

Nous avons utilisé l'algorithme de reconnaissance de visage OpenFace [1] disponible en ligne et les visages des 158 personnes qui avaient le plus d'images dans la base de données LFW Face [6]. Nous avons séparé aléatoirement, dix fois, les images utilisées pour entraîner (75 % de la base) et celles pour tester (25 % de la base). Le taux d'identification moyen pour les images originales est de 84 %.

Nous avons sélectionné $Nb = 106$ parce que le taux d'identification est inférieur à 6% à cette valeur d'après la Figure 1. Cependant, nous pouvons changer la valeur de Nb selon l'application pour avoir plus ou moins de protection (e.g. pour cacher l'âge d'une personne, la valeur Nb pourrait être plus élevée).

2.2 Cryptage des blocs résiduels des images P

Certains blocs des images P à l'intérieur du RoI peuvent être prédits à partir de blocs non brouillés et donc ces blocs seraient correctement décodés. Par conséquent, nous permutons aléatoirement les coefficients AC de chaque blocs résiduels des images P comme dans 2.1.2 et laissons le DC tel quel contrairement aux blocs résiduels des images I.

3 Évaluation

Nous comparons la méthode proposée (ASePPI) avec celle du cryptage des signes des coefficients non nuls (SNC) et avec celle où en plus les modes intra prédictifs sont cryptés (SNC + IPM). Nous appliquons ces méthodes uniquement sur le canal de luminance pour être comparables avec notre approche et utilisons différentes valeurs de QP et IP. QP est le paramètre de quantification et IP définit le nombre d'images entre deux images I (intra).

3.1 Protection de la vie privée VS dégradation

Nous évaluons le taux d'identification avec OpenFace sur six séquences vidéo (i.e. 'foreman', 'suzie', 'akiyo', 'carphone', 'claire' et 'miss america' disponibles en ligne [1]) en taille CIF en utilisant le même procédé que pour la Section 2.1.4. Nous obtenons 100 % d'exactitude (la précision moyenne) pour les visages originaux, 18.9 % pour les visages protégés avec ASePPI (avec IP = 5, 10, 30 et QP = 24) et 38.7 % avec SNC. Les résultats montrent donc que notre méthode améliore la protection de la vie privée de 19,8 % de plus que SNC. En effet, nous pouvons visuellement distinguer la position des yeux, du nez et de la bouche avec l'application du SNC (voir 2(c)) alors que les deux autres méthodes (SNC + IPM et ASePPI) empêche la visualisation de tous ces détails (voir 2(e) et 2(g)).

Néanmoins, utiliser à la fois SNC et IPM empêche la compréhension globale de la scène ou des actions humaines. Par exemple, dans 2(f) il n'est pas évident de voir une personne

1. http://trace.eas.asu.edu/yuv/

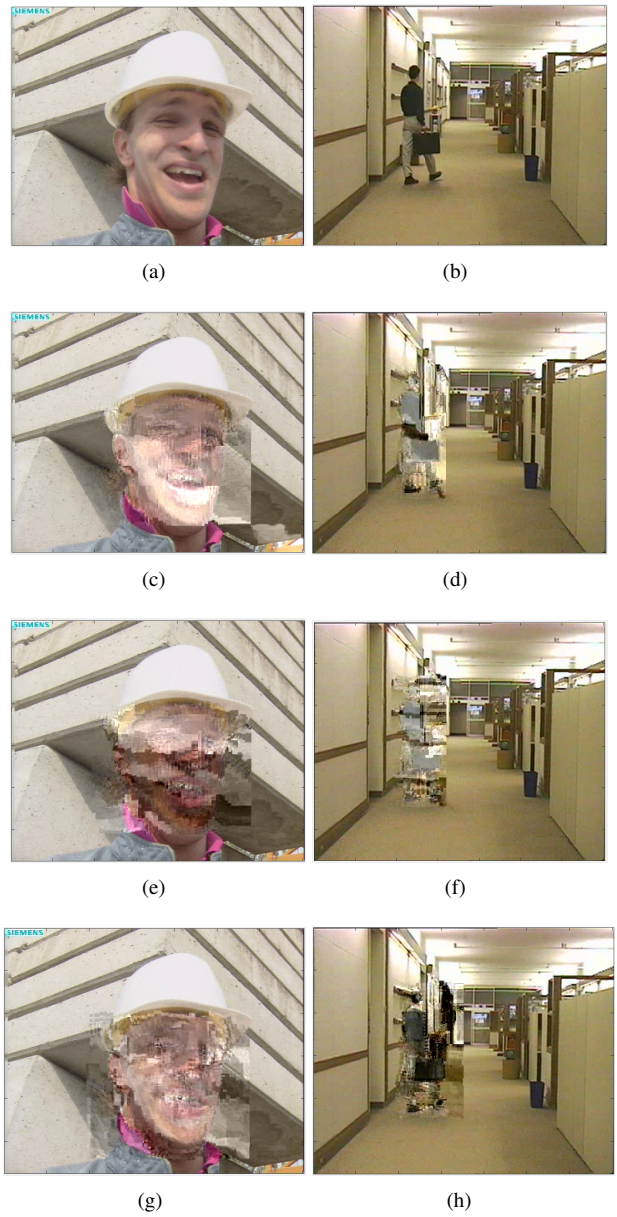

FIGURE 2: Avec la taille CIF, QP= 24 et IP= 5 : (a) la 15e image originale de la séquence 'foreman' (P), (b) la 40e image originale de la séquence 'hall' (P), (c) et (d) cryptée par SNC, (e) et (f) cryptée par SNC+IPM, (g) et (h) cryptée par ASePPI.

portant son sac tandis que dans 2(h) la forme de la tête et des pieds peuvent être visuellement distinguées ainsi que le sac. Nous évaluons également le taux de dégradation des vidéos protégées à l'aide de deux mesures : le "Peak Signal-to-Noise Ratio" (PSNR) pour mesurer la quantité de dégradation et l'"Edge Similarity Score" (ESS) [7] pour évaluer le degré de ressemblance des contours entre deux images. Nous appliquons ces métriques sur le RoI d'origine et le RoI brouillé des séquences 'foreman', 'suzie' et 'hall' de taille CIF, pour chaque QP = 18 et 24 et IP = 1,5,10,30. L'application de la méthode SNC + IPM dégrade, en moyenne, 8,4 % en plus les images par rapport à celle d'ASePPI et son degré de ressemblance des contours est 15,92 % moins important.

Ainsi, ASePPI protège suffisamment l'identité tout en conservant un minimum d'information sur la scène.

3.2 Impact sur le flux compressé

En moyenne, pour les séquences 'foreman', 'suzie' et 'hall' de taille QCIF, le pourcentage de bits ajoutés par notre processus par rapport au profil de base est de 6,62 %. Par exemple, pour la séquence 'foreman', avec QP = 18 et IP = 10, le nombre de bits à sauvegarder est 44862,7 pour le profil de base et 48114 avec l'intégration de notre processus ce qui produit $100 - 100 * 44862.7/48114$ % de bits en plus, soit 6,76 %.

La baisse des performances du PSNR en pourcentage pour les images reconstruites (à l'aide de la clé secrète) par rapport à celles originales est calculée de la même manière que pour le pourcentage de bits en plus et est de 0,25 %.

D'après nos résultats, l'impact sur le flux compressé reste négligeable.

4 Conclusion

Contrairement aux méthodes existantes, l'application de notre méthode ASePPI, améliore la protection de la vie privée dans le but de conserver le minimum d'informations requises par la surveillance. Notre approche adapte automatiquement l'intensité de la protection de la vie privée en fonction de la résolution des régions d'intérêt. La qualité des vidéos reconstruites est très proche de celle des originales et le processus génère un petit pourcentage de bits supplémentaires. Le code de notre approche est disponible sur Github [2].

Nous avons prouvé que notre méthode protège les visages face aux algorithmes de reconnaissances faciales. Cependant, nous prévoyons d'évaluer la résistance face à d'éventuelles attaques, par exemple, par force brute en supposant que la personne connaît l'algorithme de protection mais pas la clé.

Dans ce papier, nous avons surtout évalué objectivement les performances de notre méthode à l'aide de métriques. Nous voudrions également analyser subjectivement l'efficacité de la protection de la vie privée et de la dégradation qu'elle engendre.

2. https://github.com/NatachaRuchaud/ASePPI

Références

[1] B. Amos, B. Ludwiczuk, and M. Satyanarayanan. Openface : A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.

[2] D. Chafaï and F. Malrieu. Permutations, partitions, et graphes. In *Recueil de Modèles Aléatoires*, pages 57–68. Springer, 2016.

[3] F. Dai, L. Tong, Y. Zhang, and J. Li. Restricted h. 264/avc video coding for privacy protected video scrambling. *Journal of Visual Communication and Image Representation*, 22(6) :479–490, 2011.

[4] F. Dufaux and T. Ebrahimi. Scrambling for privacy protection in video surveillance systems. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8) :1168–1174, 2008.

[5] N. Khlif, T. Damak, F. Kammoun, and N. Masmoudi. Motion vectors signs encryption for h. 264/avc. In *Advanced Technologies for Signal and Image Processing (ATSIP), 1st International Conference on*, pages 1–6. IEEE, 2014.

[6] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua. Labeled faces in the wild : A survey. In *Advances in face detection and facial image analysis*, pages 189–248. Springer, 2016.

[7] Y. Mao and M. Wu. A joint signal processing and cryptographic approach to multimedia encryption. *IEEE Transactions on Image Processing*, 15(7) :2061–2075, 2006.

[8] N. Ruchaud and J. L. Dugelay. Efficient privacy protection in video surveillance by stegoscrambling. In *WIFS 7th IEEE International Workshop on Information Forensics and Security*, 2015.

[9] N. Ruchaud and J. L. Dugelay. Privacy protection filter using stegoscrambling in video surveillance. In *MediaEval*, 2015.

[10] N. Ruchaud and J.-L. Dugelay. Privacy protecting, intelligibility preserving video surveillance. In *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*, pages 1–6. IEEE, 2016.

[11] A. Unterweger, K. Van Ryckegem, D. Engel, and A. Uhl. Building a post-compression region-of-interest encryption framework for existing video surveillance systems. *Multimedia Systems*, 22(5) :617–639, 2016.

[12] Y. Wang, F. Kurugollu, et al. Privacy region protection for h. 264/avc with enhanced scrambling effect and a low bitrate overhead. *Signal Processing : Image Communication*, 35 :71–84, 2015.