Encryption of Digitized Humans: Chasing Privacy in the Internet of Things

KLIMIS NTALIANIS[1] and NIKOLAOS MASTORAKIS[2]

[1]Department of Marketing
Athens University of Applied Sciences (TEI of Athens)
Agiou Spyridonos, Egaleo, Athens
GREECE
kntal@teiath.gr

[2]Industrial Engineering Department
Technical University of Sofia,
Sofia,
BULGARIA
mastor@tu-sofia.bg

Abstract: - Video surveillance will greatly increase in the era of the Internet of Things, raising severe concerns about people's privacy. Since video surveillance systems are invasive, it is very challenging to balance between privacy and system's functionalities. Traditional algorithms for protection of visual privacy usually encrypt and decrypt the whole video, without considering video objects. However this approach has several drawbacks including lack of scalability of access and security. In this paper a digitized human encryption scheme is proposed which overcomes the aforementioned shortcomings. In particular initially a fast face detection algorithm is adopted which provides initial information to the proposed body detection module. Afterwards a chaotic encryption scheme is designed to provide security and fast scrambling of the extracted digitized humans. Experimental results on real world videos exhibit the promising performance of the proposed scheme.

Key-Words: - chaos, encryption, digitized human, face and body detection, surveillance, Internet of Things

1 Introduction

In 1999 Kevin Ashton from MIT's AutoID Lab said: "*If we had computers that knew everything there was to know about things—using data they gathered without any help from us—we would be able to track and count everything, and greatly reduce waste, loss and cost. We would know when things needed replacing, repairing or recalling, and whether they were fresh or past their best. We need to empower computers with their own means of gathering information, so they can see, hear and smell the world for themselves, in all its random glory*". This was the opening remark for the Internet of Things, which seems to be rapidly implemented.

The fundamental design principle of the Internet of Things (IoT) is to connect every entity (living or not) to a super-network. In this sense, all entities will be able to exchange information with each other. From a technical point of view, the architecture is based on data communication tools such as RFID-tagged items, wireless sensors and cameras, all of which posses a unique Electronic Product Code (EPC) [1] or similar handles.

On the other hand, some of the major issues of the IoT include security, privacy and safety of people. In particular let us imagine that wireless surveillance cameras will dramatically grow in number, in the era of the IoT. According to [2], networked security cameras are the most likely to have vulnerabilities when it comes to securing IoT devices and they are considered as particularly dangerous. This is due to the fact that if malicious users can receive their contents, then they will be able to know e.g. which homes are currently empty, personal outdoor moments of people (e.g. kissing, crying, fighting, shouting etc), where exactly are the guards of a building, localization of people etc. However, this very sensitive information should be received only by official authorities.

Towards this direction, in this paper we propose a digitized human encryption scheme, where surveillance or other content is processed by a smart camera before been transmitted. In particular,

initially humans appearing in a surveillance recording or other video are detected and segmented by applying a fast and generic face and body detection method. Next a chaotic encryption scheme is proposed, which produces a unique encryption key for each digitized human object. Then each object is encrypted with its personal key. Finally all different objects are combined to provide the overall encrypted content. Here it should be mentioned that – if necessary – the background content can be encrypted using weaker encryption, since sensitive content usually appears in the foreground. By this way: (a) the complexity of the encryption process can be reduced, (b) in case of network congestion, only the foreground information can be transmitted and (c) scalability of access is achieved, since different digitized humans can be decrypted by different levels of authorization. Furthermore the proposed scheme provides stronger encryptions compared to traditional frame-based algorithms, since objects' topology makes decryption harder for malicious users.

The rest of this paper is organized as follows: in Section 2 state-of-art approaches are presented. Section 3 provides an overview of the proposed face and body detection algorithm, while Section 4 focuses on the encryption scheme. Experimental results on real data are provided in Section 5 and Section 6 concludes this paper.

2 Previous Work

Many approaches aiming to protect personal privacy in surveillance video distort, remove, or hide visual information, which can be used for personal identification. These techniques are different in terms of complexity (ease-to-use), effectiveness of the privacy protection (how hard to identify a protected person), reversibility (possibility to undo protection), usage flexibility (can be used with compressed or uncompressed video), etc. In this section some of these approaches are mentioned.

Since visible identifiable face is a major threat to privacy in video surveillance, many researchers have focused on face de-identification techniques.

In [3] the identities of digitized humans are protected by obscuring their face with a colored ellipse. The authors claim that such protection allows observation of people's actions in full details, while hiding their identity. Arguing that de-identification of faces is not enough for adequate protection of a human's privacy, in [4] a technique for obscuring the whole body silhouette is proposed, which is based on the edge and motion model. Going a step further, in [5] the authors propose to completely remove the silhouette of the moving person from the scene, so that his/her identity is concealed. In [6] objective evaluation of several primitive privacy filters is performed. In this paper the authors demonstrated that such filters cannot adequately protect from successful face recognition. A similar work regarding the robustness of face recognition algorithms to distortions can also be found in [7]. In [8] a framework is defined to evaluate the performance of face recognition algorithms under various obfuscation methods. In [9] advanced scrambling-based privacy filters are proposed. This technique is based on randomized modifications of the compressed video stream encoded as a series of JPEG and JPEG 2000 images. In [10], ROI code-blocks are trimmed down to the lowest quality layer of the codestream. Subsequently, the quality of the ROI can be decreased by limiting the video bit rate.

In [11] the Privacy through Invertible Cryptographic Obscuration scheme is proposed, where face areas are encrypted. The process is reversible for authorized users in possession of a secret encryption key. Another similar work is presented in [12], where a permutation-based encryption technique in the pixel domain is introduced. In [13] the authors focus on the compression blocks based encryption mechanism in order to face the real-time constraints. In [14] ROIs are scrambled using chaotic encryption. The chaos-based scheme generates pseudo-random numbers, using one initial secret seed. Other important schemes include [15]-[19]. However, although very interesting, the aforementioned techniques either do not provide strong encryption, or limit their focus only to the face area. Furthermore they do not consider chaotic encryption.

3 Face and Body Detection

In order to achieve privacy, in this paper a face and body detector is initially incorporated, where face detection is performed by using the Normalized Pixel Difference feature while body detection is straightforwardly accomplished by a probabilistic model.

3.1 Face Detection

In this paper the method proposed in [20] is adopted for face detection. This algorithm is very fast and efficient, while it can work in several

challenging environments. In particular let $f(x,y)$ be the Normalized Pixel Difference feature between two pixels in an image:

$$f(x,y) = \frac{x-y}{x+y} \quad (1)$$

where x, $y \geq 0$ are intensity values of the two pixels x and y, and $f(0,0)$ is defined as 0 when $x = y = 0$.

The NPD feature measures the relative difference between two pixel values. The sign of $f(x,y)$ indicates the ordinal relationship between the two pixels x and y, and the magnitude of $f(x,y)$ measures the relative difference (as a percentage of the joint intensity $x + y$) between x and y.

The NPD feature is antisymmetric, so either $f(x,y)$ or $f(y,x)$ is adequate for feature representation, resulting in a reduced feature space. Additionally the sign of $f(x,y)$ is an indicator of the ordinal relationship between x and y. Ordinal relationship has been shown to be an effective encoding for object detection and recognition [21] because ordinal relationship encodes the intrinsic structure of an object image and it is invariant under various illumination changes [21]. Furthermore the NPD feature is scale invariant, which leads to robustness against illumination changes. Finally the NPD feature $f(x,y)$ is bounded in [-1,1].

On the other hand the classic Viola-Jones face detector [21] learns representative features by boosted stumps. However interactions between different feature dimensions cannot be captured, while the simple thresholding ignores higher-order information contained in a feature. Therefore, a quadratic splitting strategy and a deeper tree structure are considered. Specifically, for a feature x, the tree node splitting is:

$$(ax^2 + bx + c) < t \quad (2)$$

where a, b, c are constants and t is the splitting threshold. This corresponds to checking whether x is in a range $[\theta_1, \theta_2]$ or not, where θ_1 and θ_2 are two learned thresholds. Compared to the original linear splitting $x < t$, Eq. (2) considers both the first-order and second-order information of x, enabling a better interpretation of the splitting rule. In particular, three kinds of object structures can be learned:

$$-1 \leq \frac{x-y}{x+y} \leq \theta < 0 \quad (3)$$

$$0 < \theta \leq \frac{x-y}{x+y} \leq 1 \quad (4)$$

$$\theta_1 \leq \frac{x-y}{x+y} \leq \theta_2 \quad (5)$$

where $\theta_1 < 0$ and $\theta_2 > 0$. Eq. (3) applies if x is notably darker than y, while Eq. (4) covers the case when x is notably brighter than y.

In practice, instead of solving Eq. (2) for quadratic splitting, the feature range is quantized into L discrete bins and an exhaustive search is carried out to determine the two optimal thresholds, where the weighted mean square error is applied as the optimal splitting criterion.

Furthermore, the quadratic splitting is applied to learn a deep tree, instead of a stump or a shallow tree for face detection. This way, several NPD features are optimally combined together to represent the intrinsic face structure.

3.2 Body Detection

Detection of the body area can be achieved using topological attributes that relate the locations of face and body. Initially the centre, width and height of the estimated face region, denoted as $c_f = [c_x\ c_y]^T$, w_f and h_f respectively, are computed. Human body is then localized by means of a probabilistic model, the parameters of which are estimated according to c_f, w_f and h_f.

In particular, if $r(B_i)=[r_x(B_i)\ r_y(B_i)]^T$ is the distance between the i-th block, B_i, and the origin, with $r_x(B_i)$ and $r_y(B_i)$ the respective x and y coordinates, the product of two independent 1-dimensional Gaussian p.d.fs is used to model the location of human body. Thus, for each block B_i of an image, a probability $P(r(B_i)|\Omega_b)$ is assigned, expressing the degree of block B_i belonging to the human body class, say Ω_b

$$P(\mathbf{r}(B_i)\,|\,\Omega_b) = \frac{\exp(-\frac{1}{2\sigma_x^2}(r_x(B_i)-\mu_x)^2)\exp(-\frac{1}{2\sigma_y}(r_y(B_i)-\mu_y)^2)}{(2\pi)\sigma_x\sigma_y} \quad (6)$$

where μ_x, μ_y, σ_x and σ_y are the parameters of the human body localization model; these parameters are calculated based on the information derived from the face detection task, taking into account the relationship between human face and body. In our experiments, the parameters of the human body localization model are estimated with respect to the face region as follows:

$$\mu_x = c_x,\ \mu_y = c_y + h_f,\ \sigma_x = w_f,\ \sigma_y = h_{f/2} \quad (7)$$

Similarly to human face detection, a block B_i belongs to the body class Ω_b, if the respective probability, $P(r(B_i)|\Omega_b)$, is high, using a similar threshold as in the face detection case. The computed face and body masks can be properly used to extract human objects.

4 Chaotic Encryption

After the extraction of digitized humans, the

proposed chaotic digitized human encryption scheme is activated, which consists of a chaotic pseudo-random bit generator and a chaos-based cipher module. In the proposed scheme, for each video object a different key of size 256 bits is used, leading to a symmetric cipher. Each key is generated by a chaotic pseudo-random bit generator (C-PRBG). C-PRBGs based on a single chaotic system can be insecure for this reason we propose a PRBG based on a triplet of chaotic systems. The basic idea of the C-PRBG is to generate pseudo-random bits by mixing three different and asymptotically independent chaotic orbits. Towards this direction let $F_1(x_1, p_1)$, $F_2(x_2, p_2)$ and $F_3(x_3, p_3)$ be three different one-dimensional chaotic maps: $x_1(i+1) = F_1(x_1(i), p_1)$, $x_2(i+1) = F_2(x_2(i), p_2)$, $x_3(i+1) = F_3(x_3(i), p_3)$, where p_1, p_2, p_3 are control parameters, $x_1(0)$, $x_2(0)$, $x_3(0)$ are initial conditions, and $\{x_1(i)\}$, $\{x_2(i)\}$, $\{x_3(i)\}$ denote the three chaotic orbits. Then a pseudo-random bit sequence can be defined as:

$$k(i) = \begin{cases} 1, & F_3(x_1(i), p_3) > F_3(x_2(i), p_3) \\ k(i-1), & F_3(x_1(i), p_3) = F_3(x_2(i), p_3) \\ 0, & F_3(x_1(i), p_3) < F_3(x_2(i), p_3) \end{cases} \quad (8)$$

After generating a pseudo-random key for each digitized human, the cipher module is activated. Initially the pixels of each digitized human are scanned from top-left to bottom-right providing plaintext pixels P_i. Next, a simple chaotic stream cipher and two simple chaotic block ciphers (with time variant S-boxes) are combined to implement a complex product cipher.

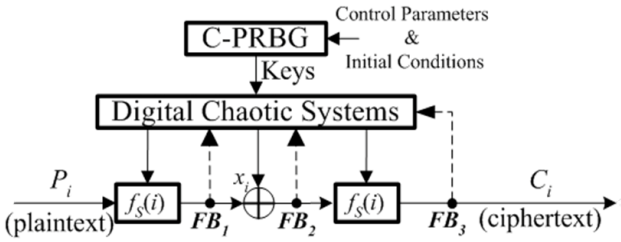


Figure 1: The encryption procedure

Considering Figure 1, the encryption procedure is defined by:

$$C_i = f_S(\{f_S(P_i, i) \oplus x_i\}, i) \quad (9)$$

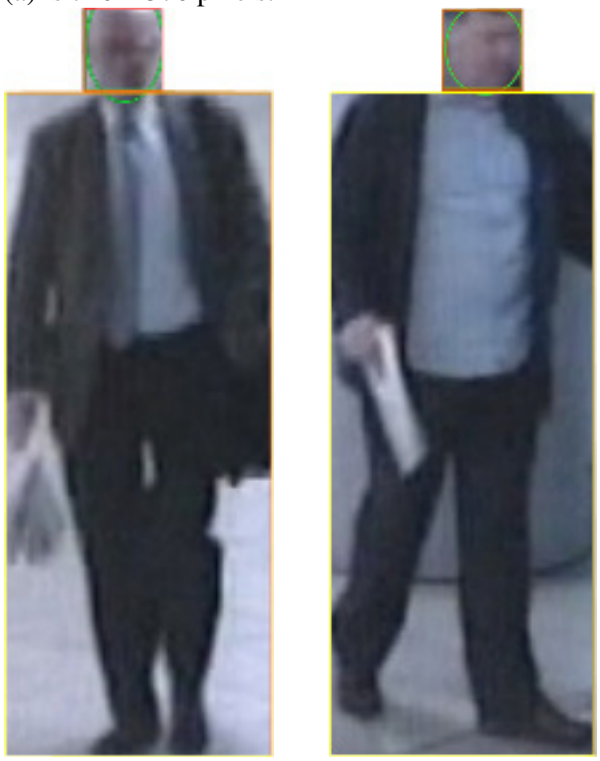
where symbol $\oplus$ represents the XOR function, $f_S(\cdot, i)$ are time-variant $n \times n$ S-boxes (bijections defined on $\{0, 1, \ldots, 2^n - 1\}$) and x_i is produced from the states of three chaotic systems. Here, f_S are also pseudo-randomly controlled by the chaotic systems. The secret key provides the initial conditions and control parameters of the employed chaotic systems. The increased complexity of the proposed cipher against possible attacks is due to the mixed feedback (internal and external): $f_S(P_i, i)$ at $\mathbf{FB_1}$, $f_S(P_i, i) \oplus x_i$ at $\mathbf{FB_2}$ and ciphertext feedback C_i at $\mathbf{FB_3}$, which lead the cipher to acyclic behavior.

The procedure is terminated after all human objects are encrypted. Finally by combining the encrypted human objects the final content is generated, which does not provide any clue about the contours of the video objects or the structure of the visual information.

Figure 2: (a) a frame from the AVSS07 typical surveillance set (b) detected face for the left digitized human (c) detected face for the right digitized human (d) body detection for the left and right digitized humans.

5 Experimental Results

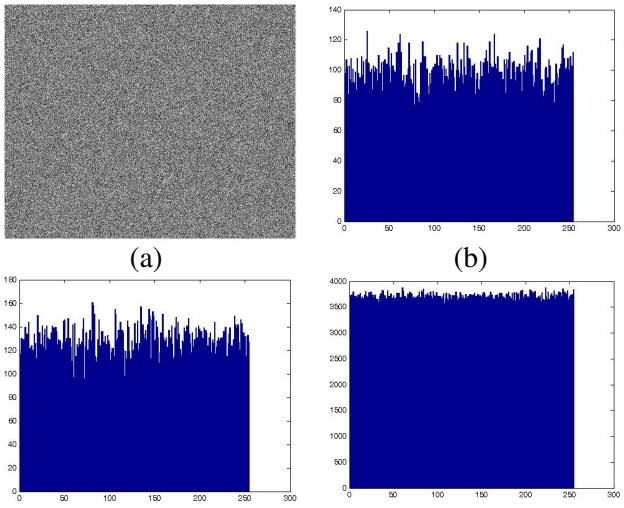
For evaluation purposes the proposed digitized humans' encryption scheme is examined in terms of security and efficiency. In particular the proposed approach is applied to the video surveillance sequences of the AVSS07 (Advanced Video and Signal based Surveillance 2007 datasets – Andrea Cavallaro) [23]. One frame of these sequences is depicted in Figure 2(a). The frame size of Figure 2(a) is 720 × 576 pixels.



(a) (b)

Figure 3: (a) extracted left digitized human (b) extracted right digitized human

In case of Figure 2(a) initially the face module is activated providing the detected faces of (Figures 2(b) and 2(c)). Next the body module is activated having as initial seeds the two human faces. Results of the body module are provided in Figure 2(d). As shown in Figures 2(b)-(d) the human faces are accurately detected, laying inside red rectangles, while human bodies are not very accurately detected (laying in yellow rectangles), since also background blocks are included. This is due to the fact that the incorporated probabilistic model takes into consideration only distances and not also visual information of the digitized humans. As a result, non-accurate but very fast face and body detection is accomplished. The extracted digitized humans are shown in Figures 3(a) and 3(b).



(a) (b) (c) (d)

Figure 4: (a) overall encrypted content, (b) histogram of the encrypted left foreground digitized human, (c) histogram of the encrypted right foreground digitized human, (d) histogram of the encrypted background.

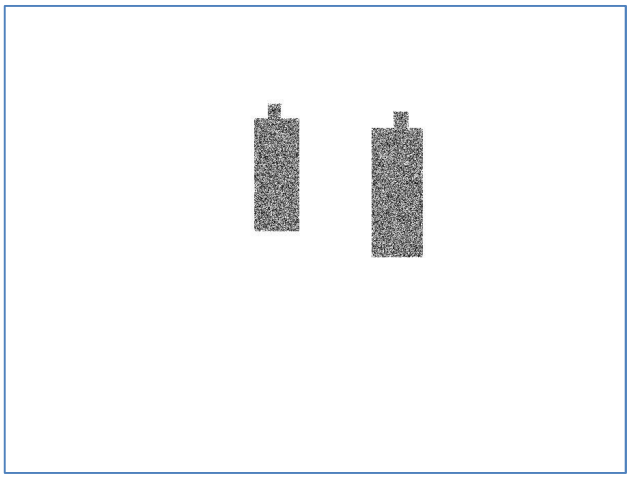


Figure 5: Decryption results in case of known contours (but the key differs by just 1 bit).

Afterwards the encryption module is activated, where in the performed experiments the three incorporated chaotic maps (both in the C-PRBG module and the cipher module) are piecewise linear chaotic maps (PWLCMs) of the form:

$$F(x,p)=\begin{cases} x/p, & x\in[0,p) \\ (x-p)/\left(\frac{1}{2}-p\right), & x\in[p,\frac{1}{2}] \\ F(1-x,p), & x\in(\frac{1}{2},1] \end{cases} \quad (10)$$

where 0 < p < ½, and initial control parameters p_1=0.15, p_2=0.27 and p_3=0.43. Next the pixels of each digitized human are properly arranged in sequential order and a different 256-bit key is

produced by the C-PRBG. After encrypting both digitized humans as well as the background, rearrangement is performed and all encrypted content is synthesized to provide the final frame. In Figure 4(a) the overall encrypted content is depicted. As it can be observed, the final content does not provide any clues relevant to the number or location of digitized humans. This fact is further clarified in Figures 4(b), 4(c) and 4(d) where the histograms of the left foreground digitized human, the right foreground digitized human and the background are presented. All histograms approximate the histogram of a table with random values. This is very important, since the encrypted content approximates the statistics of random pixels, independently of the plaintext.

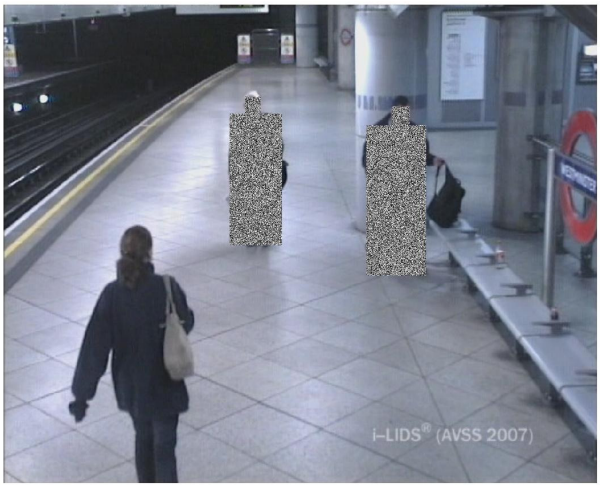


Figure 6: Decryption of the background (authorization for low level personnel)

Afterwards the encrypted content can be transmitted. In this case the smallest n-1 contours (for n digitized humans) are transmitted together with the encrypted content.

Now the security of the proposed scheme is further examined. Let us assume that an unauthorized user knows the contours of the VOs and applies brute force attack. If the exact keys are detected then the content is decrypted. However even if the keys differ by just one bit, the content will not be decrypted (Figure 5). Finally in case the user knows only the contour and the key of the background he/she can decrypt only the background (Figure 6). This scalability feature is a very important merit of the proposed scheme, since camera operators in the IoT may provide to their low-level personnel only background information. Authorization for decrypting digitized humans may be kept for high-level personnel.

6 Conclusion

In this paper a chaotic encryption scheme is designed to confront the problem of digitized humans' privacy in the era of the Internet of Things. The case of surveillance sequences is examined, where a fast face and body approach is proposed. Experimental results on real life data show: (a) that the final encrypted content does not provide any clue relative to the number and locations of digitized humans and enables scalable access to the visual material, (b) even in case of brute-force attack, if unauthorized users do not know the exact contours, it is difficult to decrypt the content.

References:

[1] https://en.wikipedia.org/wiki/Electronic_Product_Code, accessed: 18th Jan. 2017.

[2] https://www.zscaler.com/blogs/research/iot-devices-enterprise, (By: Deepen Desai director of security research at Zscaler), accessed: 15th Nov. 2016.

[3] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Goldberg, "Respectful cameras: detecting visual markers in real-time to address privacy concerns," in International Conference on Intelligent Robots and Systems (IROS 2007), p.p. 971–978, Oct. 2007.

[4] D. Chen, Y. Chang, R. Yan, and J. Yang, "Protecting Personal Identification in Video," Protecting privacy in video surveillance. Springer-Verlag, p.p. 115–128, 2009.

[5] S.-C. S. Cheung, M. V. Venkatesh, J. K. Paruchuri, J. Zhao, and T. Nguyen, "Protecting and Managing Privacy Information in Video Surveillance Systems," Protecting privacy in video surveillance, Springer-Verlag, p.p. 11–33, 2009.

[6] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," IEEE Trans. on Circuits and Systems for Video Technology, vol. 18, no. 8, pp. 1168–1174, Aug 2008.

[7] P. Korshunov and W. T. Ooi, "Video quality for face detection, recognition, and tracking," ACM Trans. Multimedia Comput. Commun. Appl., vol. 7, no. 3, pp. 1–21, Sept. 2011.

[8] F. Dufaux and T. Ebrahimi, "A framework for the validation of privacy protection solutions in video surveillance," in Proceedings of IEEE International Conference on Multimedia & Expo (ICME 2010), Singapore, July 2010.

[9] F. Dufaux and T. Ebrahimi, "Video surveillance using JPEG 2000," in proc. SPIE Applications of Digital Image Processing

XXVII, vol. 5588, Denver, CO, p.p. 268–275, Aug. 2004.

[10] I. M. Ponte, X. Desurmont, J. Meessen, and J.-F. Delaigle, "Robust human face hiding ensuring privacy," in in Proc. of International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Montreux, Switzerland, Apr. 2005.

[11] T. E. Boult, "PICO: Privacy through invertible cryptographic obscuration," in IEEE Workshop on Computer Vision for Interactive and Intelligent Environments, Lexington, KY, p.p. 27–38, Nov. 2005.

[12] P. Carrillo, H. Kalva, and S. Magliveras, "Compression independent reversible encryption for privacy in video surveillance," EURASIP J. Inf. Secur., vol. 2009, pp. 5:1–5:13, Jan. 2009.

[13] A. Pande, P. Mohapatra, and J. Zambreno, "Securing multimedia content using joint compression and encryption," IEEE Multimedia, vol. PP, no. 99, 2012.

[14] S. Rahman, M. Hossain, H. Mouftah, A. El Saddik, and E. Okamoto, "Chaos-cryptography based privacy preservation technique for video surveillance," Multimedia Systems, vol. 18, p.p. 145–155, 2012.

[15] P. Korshunov and T. Ebrahimi, "Using warping for privacy protection in video surveillance," in 18th International Conference on Digital Signal Processing (DSP), Santorini, Greece, June 2013.

[16] P. Korshunov and T. Ebrahimi, "Towards optimal distortion-based visual privacy filters," in IEEE International Conference on Image Processing, ICIP'2014, Paris, France, 2014.

[17] Bonetto, M., Korshunov, P., Ramponi, G., and Ebrahimi, T., Privacy in Mini-drone Based Video Surveillance, Workshop on De-identification for privacy protection in multimedia, May 2015.

[18] N. Ruchaud and J.-L. Dugelay, "Privacy protection filter using stegoscrambling in video surveillance." In MediaEval 2015 Workshop, Wurzen, Germany, Sept. 2015.

[19] S. ¸Cift¸ci, P. Korshunov, A. O. Akyuz, and T. Ebrahimi. ¨ Using false colors to protect visual privacy of sensitive content. In SPIE Human Vision and Electronic Imaging XX, pages 93941L–93941L–13, 2015.

[20] S. Liao, A. K. Jain, and S. Z. Li, "A Fast and Accurate Unconstrained Face Detector," IEEE Trans. PAMI, Vol. 38, No. 2, Feb. 2016.

[21] P. Sinha, "Qualitative representations for recognition," in Proc. 2nd Int. Workshop Biol. Motivated Comput. Vis. Workshop, pp. 249–262, 2002.

[22] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., p.p. I-511–I-518, 2001.

[23] http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html, accessed: 11th Jan. 2017.