Using Warping for Privacy Protection in Video Surveillance

Pavel Korshunov and Touradj Ebrahimi
Multimedia Signal Processing Group, EPFL, Lausanne, Switzerland

***Abstract*—The widespread use of digital video surveillance systems has also increased the concerns for violation of privacy rights. Since video surveillance systems are invasive, it is a challenge to find an acceptable balance between privacy of the public under surveillance and the functionalities of the systems. Tools for protection of visual privacy available today lack either all or some of the important properties such as security of protected visual data, reversibility (ability to undo privacy protection), simplicity, and independence from the video encoding used. In this paper, we propose an algorithm based on well-known warping techniques (common for animation and artistic purposes) to obfuscate faces in video surveillance, aiming to overcome these shortcomings. To demonstrate the feasibility of such an approach, we apply warping algorithm to faces in a standard Yale dataset and run face detection and recognition algorithms on the resulted images. Experiments demonstrate the tradeoff between warping strength and accuracy for both detection and recognition.**

***Index Terms*—Privacy protection, video surveillance, warping**

I. Introduction

Recent rapid adoption of digital video surveillance systems, especially in public spaces and communities, has significantly increased the concern for protection of individual privacy. Typical surveillance systems are non-discriminative, surveying everyone and everything, which poses a threat to the human rights, to privacy, and fundamental individual freedoms [1][2]. Growing number of privacy abuses and ignorance of privacy laws increase the public tension preventing full acceptance of surveillance systems. Many privacy advocates worry that the abuses of video surveillance may outweigh its benefits. Moreover, with the latest progress in video analytics (detection, recognition, and tracking), in combination with personal information from web and social networks (which is more and more easily available), the emerging multi-modal surveillance systems pose a serious threat to fundamental rights to privacy. Therefore, there is a strong demand in user-centric solutions [3], which specifically focus on protection of privacy in video surveillance systems.

However, preserving privacy while performing surveillance is challenging. Often, these goals contradict each other. Since video surveillance systems are invasive by their design, it is hard to find an acceptable balance between privacy of public under surveillance and the surveillance tasks at hand, such as detection of suspicious individuals, objects, and events. A desired method for visual privacy protection would, therefore, retain basic characteristics or integrity of visual data, i.e., a protected face would remain visible, but remove personal information, i.e, specific facial features allowing to identify an individual would be hidden. In addition, a practical privacy protection method would have the following properties: (i) low complexity (easy to use), (ii) reversibility (possibility to undo protection), (iii) flexibility of application (independent of compression and video or image data format), (iv) security (recovery of the original data using secret key), and (v) variable strength granularity (flexibility to protect data with different degrees of strength).

Although many different privacy protection methods have been proposed for preserving privacy while retaining the surveillance objective, none of them fulfill all above mentioned properties. Such simple methods like blurring, pixelization, and masking are not reversible and insecure; encryption-based methods, such as proposed in [4] are secure but destroy integrity of the original pixel data; scrambling [5] is dependent on video or image compression; and anonymization methods like in [6] are often complex and require original data to be stored separately.

To overcome such shortcomings, we propose using a geometrical transformation (or warping) for protection of visual privacy. Pixels in the protected region can be shifted into slightly different locations, thus destroying visual details and relationships between neighboring pixels of an image. Hence, the balance between privacy protection and surveillance task is transformed into how much the pixels in an image are warped. To keep complexity low, we propose estimating warping transformation matrix based on the grid of key pixels that are shifted to a random distance. The rest of the pixels are warped using the estimated matrix, while the resulted gaps are interpolated with bicubic algorithm. Such approach allows us to distort the specific information while retaining the general shapes of the region, and the degree of distortion can be adjusted by changing the shifting distance of the key pixels. The number of key pixels characterizes the complexity of the algorithm. Since warping is applied to pixel data, it is independent to compression. A security can be insured by using a secret key for seeding a pseudo-random algorithm, as well as, for encryption of the key pixels used for transformation matrix estimation. Unwarping can be done by applying the inverse transformation matrix.

We demonstrate feasibility of the proposed warping method by applying it to faces of a standard Yale face dataset (see Figure 1), since faces are among the most privacy sensitive regions. Location of each face is first detected with Viola-Jones [7] face detection algorithm. A grid of key pixels, used

as a base for the warping transformation, is constructed from detected eyes, nose, and mouth. We then run face detection and face recognition (Fisher linear discriminant analysis) algorithms on the warped faces to determine whether the warping preserves the integrity of a face, such that the accuracy of a face detection algorithm is not affected by it, and whether the specific personal features are distorted, such that recognition accuracy is decreased. In an ideal scenario, a protected face would be detected by the detection algorithm but would not be recognized by the recognition algorithm.

In the next section, we discuss the related work on existing methods for visual privacy protection, emphasizing their advantages and disadvantages. In Section III, we give an overview of the proposed warping method and experiments conducted. In Section IV, we present the experimental results and discuss our findings. We conclude the paper with Section V.

II. Related 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 (easy-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. We review some of them in this section. Since visible identifiable face is a major threat to privacy in video surveillance, many researchers have focused on face de-identification techniques.

One way to preserve privacy of individuals under video surveillance is to make their facial information unintelligible by distorting the corresponding pixels, such that it would prevent face recognition techniques or human observers to identify persons. Such approaches include replacement of faces in a video frame with some simple shapes or pixelization. For instance, in [8] people's identities are protected by obscuring their face with a colored ellipse. The authors argue that such protection allows observation of the people actions in full details while hiding their identity. Other naïve approaches also include blurring and face masking for hiding the faces of the people in the video. These types of techniques are simple to implement in video or images, making them popular to use for hiding personal identities and other private information on TV, in social networks, or other potentially privacy infringing internet services. However, these type of filters irreversibly distorts video data at the pixel level, making it impossible to use video in situations, when, for example, due to a court order, identity of a person in the video needs to be retrieved.

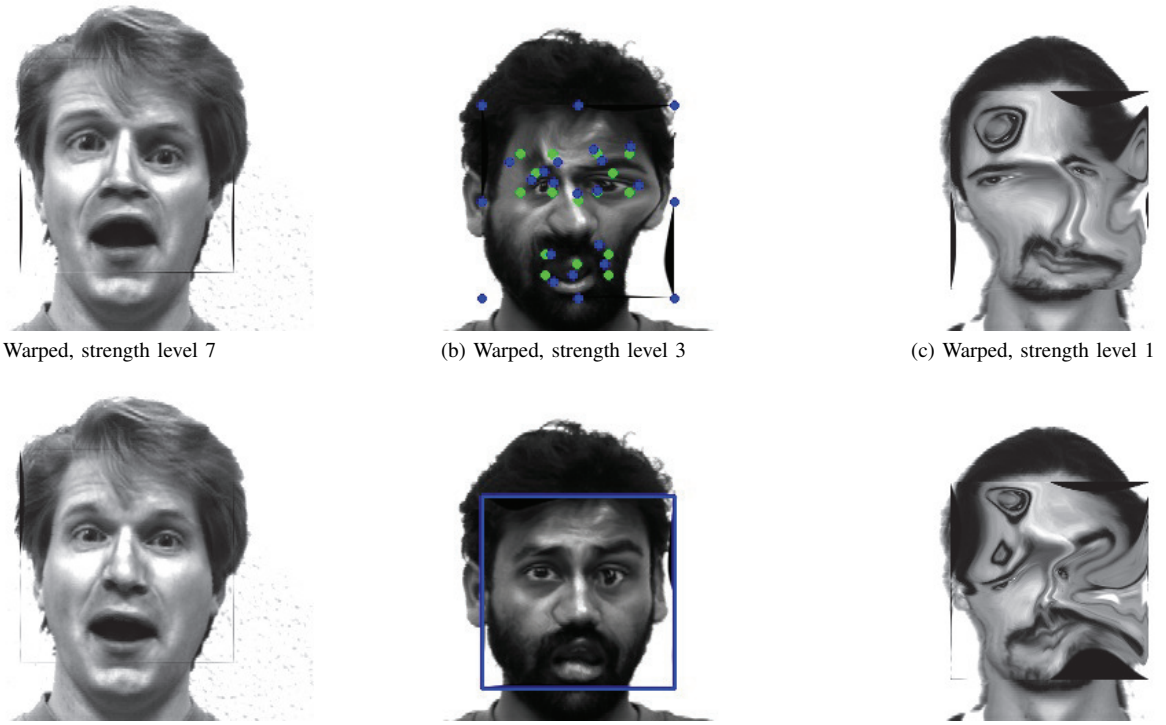
Arguing that de-identification of faces is not enough for an adequate privacy protection, the technique for obscuring of the whole body silhouette is proposed in [9], which is based on the edge and motion model. Going further, in [6] and [10] it is proposed to completely remove the silhouette of the moving person from the scene to hide its identity. Both approaches rely on RFID tags for pinpointing of the people locations, with [10] focusing on an efficient inpainting algorithm and encrypting the removed silhouette inside the original video bitstream. Similar to the previously described face de-identification methods, obscuring or removal of silhouettes distorts an original data content, which is not suitable in many surveillance applications. One way to solve this problem is to extract and transmit these silhouettes via separate secure channels to an observer authorized to view privacy related information.

Aiming to avoid constraints of the distortion-based methods, more advanced scrambling-based privacy filters are proposed in [5] and [11]. These techniques are based on randomized (seeded with a secret key) modifications of the compressed video stream encoded as a series of JPEG and JPEG 2000 images. Conditional access control techniques are proposed in [5] to scramble ROIs, e.g., corresponding to people or faces. The scrambling is applied either in wavelet-domain or codestream-domain. In [11], code-blocks corresponding to ROI 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.

Two efficient region-based transform-domain and codestream-domain scrambling techniques are proposed in [12] to hide privacy-sensitive information in MPEG-4 compressed video. In the first approach, the sign of selected transform coefficients is pseudo-randomly inverted during encoding. In the second approach, bits of the codestream are pseudo-randomly flipped after encoding. In [13], the region-based transform-domain scrambling is extended to H.264/AVC. In particular, to discriminate between scrambled and unscrambled regions, the technique exploits the Flexible Macroblock Ordering (FMO) mechanism of H.264/AVC to define two slice groups composed of Macroblocks (MB) corresponding to the foreground and background respectively.

The main advantage of scrambling-based privacy protection techniques is that they are reversible. By knowing a secret key, which could be stored and transmitted securely, one can decode the video back to undistorted state. Another advantage is that the appearance of the scrambled region is not completely distorted, making the viewing experience less distractive (compared to a black box instead of a region for instance). However, scrambling-based filters require high processing power. Also, since scrambling modifies the specific internal coefficients of a given compression algorithm, a separate scrambling tool needs to be designed and implemented for every video encoder and decoder used.

Another way to protect a sensitive region securely is to encrypt it. With Privacy through Invertible Cryptographic Obscuration (PICO) proposed in [4], data corresponding to faces is encrypted in order to conceal identity. The process is reversible for authorized users in possession of a secret encryption key. In other words, it does not undermine the objective of surveillance, as a subject can still be identified by decrypting the face, provided an appropriate warrant is issued. Similarly, a permutation-based encryption technique in the pixel domain is introduced in [14]. TrustCam is



(a) Warped, strength level 7 (b) Warped, strength level 3 (c) Warped, strength level 1

(d) Unwarped, strength level 7 (e) Unwarped, strength level 3 (f) Unwarped, strength level 1

Fig. 1: Examples of warping and unwarping for different faces and different levels of warping strength. Green dots in Figure (b) are original locations of key pixel grid, and blue dots are these pixels randomly shifted.

presented in [15], which is a video camera with onboard hardware security solution Trusted Platform Module (TMP), which implements trusted computing. This built-in chip allows establishing secure connection between cameras and observing stations, as well as applying SHA-1 based encryption to the sensitive regions, such as faces and license plates.

The idea of encrypting or scrambling face regions was developed further by [16], where the authors focus on the compression blocks based encryption mechanism. The authors argue that the conventional encryption methods are not suitable due to the real-time constraints, limited computational and network resources. Instead, they suggest adding a special parameter inside an encoder compression block, which would enable encryption and secret key generation.

Rahman *et al.* [17] also argue that conventional encryption mechanisms, such as RSA, are ineffective (too slow, complex key-exchange mechanism, etc.) for encryption of large amount of data such as ROIs in video. The authors propose to scramble ROIs using Chaos cryptography. Chaos-based encryption generates pseudo-random numbers with one initial secret seed (that both scrambling and de-scrambling algorithms must share) and masks (replaces) the original data with these numbers, which makes it relatively simpler on the large amounts of data.

As the last two methods ([16] and [17]) imply, encryption is a computationally heavy operation, and it is especially challenging when used in video streams in real-time video surveillance systems. Also, these privacy protection methods render encrypted regions as boxes of white noise instead of the meaningful features, which may hinder the ability of the observer to perform the required surveillance tasks.

III. Proposed method and experiments overview

Developing appropriate privacy protection for video surveillance systems is a challenging problem. By proposing warping based privacy protection approach, we aimed to overcome the shortcomings of such advanced visual privacy protection tools like scrambling and encryption. Both classes of methods are secure and flexible, but while encryption replaces a protected region with white noise, scrambling, in order to work, requires heavy modification of a compression encoder. Since warping is a mere geometrical transformation of pixels, which means pixels are simply shifted and their intensities are interpolated, it is compression independent, as opposed to scrambling. Also, unlike encryption, warping can preserve (depending on the strength level) general facial features and likeness of a face.

We use standard Yale dataset with provided ground truth for testing the feasibility of proposed warping-based privacy protection. Warping is applied automatically to faces that are detected with Viola-Jones face detection [7] algorithm implemented in OpenCV[1].

Warping algorithm is done as following (with unwarping being the same algorithm inversely applied to the warped

[1]http://opencv.willowgarage.com/wiki/

image):

- Select a set of key points in the image
- Randomly shift these points (i.e., change their coordinates) by adding or substracting random value with weight depending on the *warping strength*. The resulted coordinates constitute the desired destinations for the selected point in the target warped image.
- Based on the original and destination coordinates of the key point, compute transformation matrix.
- Apply the transformation to each pixel in the image, using 'cubic' interpolation.

For estimation of the transform matrix, we first construct a grid of key pixels from automatically detected (with the same algorithm that is used for face detection) eyes, nose, and mouth. Several points around each detected facial feature are picked, as well as, the sides of the face (see Figure 1b for illustration). These grid points are locally shifted to the distance generated by pseudo-random algorithm. The maximal possible shifting distance represents how strong the warping effect is, as illustrated by Figure 1a (mild effect) and Figure 1c (extremely strong effect).

Since in different images faces have different sizes in pixels, the value of distance should depend on the size of a given face. Hence the choice of appropriate distance is important and non-trivial. In our experiments, we varied the distance proportionally to the size of the eye in the face (which is one fifth of facial width, according to standard head proportions). Original locations of the grid pixels are mapped using bicubic interpolation to the shifted locations, thus determining the transformation matrix. The transformation matrix is applied to the rest of the pixels in a face using 'remap' function implemented in OpenCV with bicubic interpolation.

In a surveillance scenario, when a protected face needs to be recovered, an unwarping operation is applied. If the transformation matrix is known, the original face can be estimated from warped data by applying the inverse transform. The challenge with warping transformation is that most of the pixels in the warped image are interpolated between the transformed key points on the grid. Interpolation results in a crude estimation of the pixels' locations and intensities. Applying the reverse operation (unwarping) to the warped pixels increases an approximation error further, resulting in an unwarped image that is an estimation of the original. The stronger the warping, the higher estimation error, as demonstrated by Figures 1c and Figures 1f, where unwarping could not recover the original face.

To evaluate the proposed warping algorithm, we run face detection (Viola-Jones algorithm) and face recognition (Fisher linear discriminant analysis) algorithms, both of which are implemented in OpenCV library, on the warped faces from the Yale dataset. The aim of the experiments is to determine whether the warping preserves the generic features of a face, such that the accuracy of a face detection algorithm is not affected by it, and whether the specific personal features are distorted, such that recognition accuracy is decreased. In an ideal scenario, a protected face would be detected by the detection algorithm but would not be recognized by the recognition algorithm.

Since the choice of the distance to which the pixels are warped determines the strength of warping, it is important to find an appropriate distance for a given application scenario. Therefore, in our experiments, we varied the strength of warping by changing the maximal distance to which the grid key pixels can be shifted. The strength level is inversely proportional to the size of an eye in the detected face. The larger the strength level, the smaller the distance to which the pixels are shifted and smaller the warping effect. The smaller the strength level, the larger such distance and stronger the warping effect.

We vary the strength level of warping from 1 (maximally strong warping) to 15 (minimal warping effect). To find detection accuracy, for each strength level, the number of detected faces is recorded. Using available ground truth, we obtain the number of correctly detected faces and divide it by the recorded total number of faces to get the detection index, which is the measure of accuracy for face detection.

For face recognition, for each warping strength level, we compute rank one of cumulative match characteristic (CMC) [18], which is a standard measure of accuracy for identification task of face recognition algorithm. In such identification task, images in gallery have faces that are assumed to be known at the moment of recognition and images in probe set contain faces that are being recognized by the algorithm. We assume original unchanged images to constitute a gallery set. Probe set would then consist of warped (in scenario when protection is 'on') and unwarped (when protection is removed) images. Since warping algorithm randomly shifts pixels in images, to avoid bias in our recognition results, we generated 5 sets of warped and unwarped images from Yale dataset and ran recognition algorithm 5 times. The overall result is the average of recognition accuracies for all the 5 runs.

IV. Results and discussion

Experimental results for Viola-Jones [7] face detection algorithm are presented in Figure 2. X axis shows the strength of the applied warping with smaller value leading to stronger warping and higher value resulting in less warping effect (see Figure 1 for illustration). Note that the detection accuracy for the unwarped images with strength level 15 is the same as for the original images, when no warping/unwarping transformation is applied. Figure 2 demonstrates that face detection algorithm detects equally accurately both warped and unwarped faces, with a little gap appearing at warping strength level 5, until the rapid decrease in accuracy starting with level 2. It means that until faces are distorted to the level similar to Figure 1b and Figure 1e, the face detection performs well.

Similarly for the face recognition algorithm [19], recognition results for different strength of warped/unwarped images are shown in Figure 3. Recognition accuracy for the unwarped images with strength level 15 is the same as recognition accuracy for the original images, when no warping/unwarping transformation is applied. Recognition accuracy of the warped

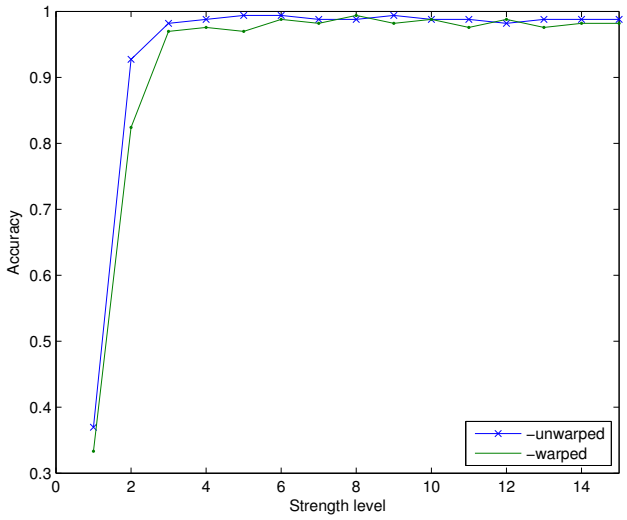


Fig. 2: Accuracy of face detection for different strengths of warped and unwarped images.

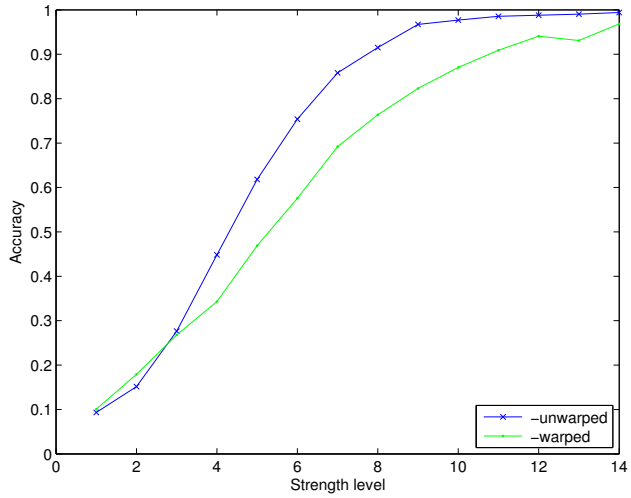


Fig. 3: Accuracy of face recognition for different strengths of warped and unwarped images.

images is lower than for original and decreases rapidly as warping strength increases. This rapid drop in recognition accuracy for warped images means that warping strength has a strong affect on the recognition algorithm distorting the fine details that allow more accurate recognition.

Unwarping on the other hand managed to recover those features maintaining the recognition accuracy unchanged until warping strength 9. Although, it would be preferable that the accuracy for unwarped images would continue being high for even stronger warping strength, but, as discussed in the previous section, the approximating nature of unwarping algorithm introduces large enough errors, preventing it to recover original images when strong level of warping is used.

However, the strong gap between 'warped' and 'unwarped' curves for the strength levels above 8 show that the warping/unwarping algorithm can be successfully used in practical, especially given a more realistic and challenging dataset for recognition compared to Yale dataset. The overall high recognition accuracy demonstrated by the algorithm (even for high warping strength levels) can be explained due to its robustness to the high distortion levels, as also supported by [20].

These results motivate us to (i) perform subjective evaluations of warping privacy protection using human subjects to see whether the subjective recognition rate is similarly high as accuracy of recognition algorithm, and (ii) refine our warping approach to improve the approximation accuracy of the unwarping recovery method, which will increase the gap between warped and unwarped recognition curves in Figure 3.

V. Conclusion and Future Work

In this paper, we proposed a new warping-based visual privacy protection algorithm for video surveillance systems. Warping is an effective approach for privacy protection, because of its simplicity, flexibility of application and degree of strength, and how it distorts local pixel information preserving the general likeness of a warped object. We evaluated the warping based privacy protection algorithm on the standard Yale faces dataset.

Experiments demonstrate the tradeoff between warping strength and accuracy for both detection and recognition algorithms. The results also show that despite high robustness of the face recognition algorithm, recognition accuracy decreases after a certain warping strength is achieved, while detection accuracy does not change. This pivotal point can be used in practical surveillance applications, since warping with this strength retains the likeness of a face, while distorting fine details and features that contain information allowing to identify a person.

These results motivate us to (i) perform subjective evaluations of warping privacy protection using human subjects to see whether the subjective recognition rate is similarly high as accuracy of recognition algorithm, and (ii) refine our warping approach to improve the approximation accuracy of the unwarping recovery method.

Acknowledgment

This work was conducted in the framework of the EC funded Network of Excellence VideoSense.

References

[1] S. Chesterman, "Privacy and surveillance in the age of terror," *Survival: Global Politics and Strategy*, vol. 52, no. 5, pp. 31–46, Sep 2010.

[2] B. Sheldon, "Camera surveillance within the uk: Enhancing public safety or a social threat?" *International Review of Law, Computers & Technology*, vol. 25, no. 3, pp. 193–203, Oct 2011, special Issue: COUNTER-TERROR STRATEGIES, HUMAN RIGHTS AND THE ROLES OF TECHNOLOGY.

[3] T. Winkler and B. Rinner, "User-centric privacy awareness in video surveillance," *Multimedia Syst.*, vol. 18, no. 2, pp. 99–121, 2012.

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

[5] F. Dufaux and T. Ebrahimi, "Video surveillance using JPEG 2000," in *proc. SPIE Applications of Digital Image Processing XXVII*, vol. 5588, Denver, CO, Aug 2004, pp. 268–275.

[6] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian, "Privacy protecting data collection in media spaces," in *Proceedings of the 12th annual ACM international conference on Multimedia*. New York, NY, USA: ACM, 2004, pp. 48–55. [Online]. Available: http://doi.acm.org/10.1145/1027527.1027537

[7] P. Viola and M. Jones, "Robust real-time face detection," in *Proceedings of the ICCV 2001 Workshop on Statistical and Computation Theories of Vision, ICCV'01*, vol. 2, Vancouver, Canada, Jul. 2001, p. 747.

[8] 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)*, Oct 2007, pp. 971–978.

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

[10] S.-C. S. Cheung, M. V. Venkatesh, J. K. Paruchuri, J. Zhao, and T. Nguyen, *Protecting privacy in video surveillance*. Springer-Verlag, 2009, ch. Protecting and Managing Privacy Information in Video Surveillance Systems, pp. 115–128.

[11] 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.

[12] 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.

[13] ——, "H.264/AVC video scrambling for privacy protection," in *in Proc. IEEE International Conference on Image Processing*, San Diego, CA, Oct 2008.

[14] 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. [Online]. Available: http://dx.doi.org/10.1155/2009/429581

[15] T. Winkler and B. Rinner, "Trustcam: Security and privacy-protection for an embedded smart camera based on trusted computing," in *Proceedings of Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS 2010)*, Aug. 29-Sep. 1 2010, pp. 593 –600.

[16] A. Pande, P. Mohapatra, and J. Zambreno, "Securing multimedia content using joint compression and encryption," *IEEE Multimedia*, vol. PP, no. 99, p. 1, 2012.

[17] 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, pp. 145–155, 2012, 10.1007/s00530-011-0246-9. [Online]. Available: http://dx.doi.org/10.1007/s00530-011-0246-9

[18] P. J. Grother, R. J. Micheals, and P. Phillips, "Face recognition vendor test 2002 performance metrics," in *Proceedings of the 4th International Conference on Audio Visual Based Person Authentication, AVBPA'03*, Guildford, UK, Jun. 2003, pp. 937–945.

[19] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognition Letters*, vol. 24, pp. 3079–3087, Dec. 2003.

[20] 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. 14:1–14:21, Sep. 2011. [Online]. Available: http://doi.acm.org/10.1145/2000486.2000488