De-genderization by body contours reshaping

Natacha Ruchaud and Jean-Luc Dugelay
Eurecom
450 Route des Chappes, 06904 Sophia Antipolis
ruchaud@eurecom.fr

Abstract

This paper deals with privacy protection in video surveillance. More specifically, the main goal of this work is to make the gender of people no more recognizable while preserving enough information concerning body shape and motion of people for action classification. We denote this processing as de-genderization. *Regarding the current state-of-art methods, most of them have privacy filters only dedicated to de-identify people. These methods do not automatically imply the suppression of visual semantic traits such as gender. Therefore, we propose two approaches that modify the visual appearance of the body shape in order to* de-genderize *people while keeping the possibility to interpret the video. In both methods we start by extracting the contour points attached to the body shape of people in videos. Then we either mix the coordinates of the body shape and a predefined model, or we smooth the body shape by successive polygonal approximations based on convexity. Our results demonstrate that both proposed approaches protect the gender information while preserving the global body movement. The second approach based on convexity better preserves the visibility of human activities.*

1. INTRODUCTION

One main goal of privacy protection is to prevent uncontrolled dissemination of personal data. In video surveillance, mostly identity but also some semantic traits are sensitive information about people. For several years, de-identification [17, 18, 20, 22] was the major topic in privacy protection of biometrics. The goal of such privacy filters is to make the identity of people unrecognizable. However, images of people offer additional information, and soft biometric attributes such as age [5], gender [10] and ethnicity [9] can be extracted from face/body images. In some cases, it is possible to deduce identity from a bag of soft biometric traits. Thus, privacy protection is not only a problem of identity recognition. For a better protection, it would be preferable to remove all visual personal information about people from videos except the ones required by the application. For instance, if gender information is unnecessary, we should remove it. One main difficulty is to preserve at the same time other information, such as people actions. Therefore, in this paper, we introduce a new concept, *degenderization*, which means that the gender of a person is made unrecognizable while preserving the intelligibility of the scene.

More precisely, our concern is to ensure the protection of the identity and the gender while preserving the silhouette of the body in order to visualize the minimum of information that are required for the surveillance (i.e. the motion of arms and legs for activity recognition).

The rest of the paper is organized as follows: in the next section, the current state-of-the-art related to privacy filters is summarized. We prove in Section 3 that the traditional privacy techniques do not protect gender information even if the identity is no more extractable. In Section 4 we describe the proposed approach. We present and discuss of the results in Section 5. Finally, we draw some conclusion and perspective in Section 6.

2. Background and related work

Naive methods, like blurring or pixelization, are already used to anonymize people (e.g., Google Street View, FacePixelizer[1], ObscuraCam[2] on Android). In [15], authors prove that the gender remains visible by humans and by machines even with the application of these anonymization methods.

Authors in [17, 18] combine scrambling and steganography techniques. First, they encrypt the RoI (Region of Interest) and they shift the Most Significant Bits (MSBs) of the encrypted pixels to the Least Significant Bits (LSBs). Then, the bits from the edge of RoI (shape of the body) replace the MSBs of the resulting image in order to keep the scene understandable. In [1], the authors, first, find the foreground (i.e. body) using GrabCut and then apply a Gaussian blur on this foreground. The body shape is easily deduced from the two previous methods. Nevertheless, in this paper, we prove that the idea of keeping the body shape (as illustrated in 1) is not sufficient to protect the gender information.

In [13], authors hide privacy information inside the image itself and use an inpainting method to refill the privacy-sensitive regions. Despite the aesthetic of images, the body of people becomes invisible. Finally, in [11], instead of removing the privacy-sensitive regions, they replace the original body by another one from a database. The movements of the new body and the original one differ, and thus, compromises the monitoring.

People employ scrambling for its reversibility. With several levels of security, authors in [22], pseudo-randomly invert the sign of selected transform coefficients within JPEG framework. How-

[1] http://www.facepixelizer.com/
[2] https://guardianproject.info/apps/obscuracam/

ever, the tool can either protects the privacy by making the human activities not recognizable, or the contrary.

In [12], authors design a face morphing method which suppresses the gender, but keeps the face identity. However, the goal of privacy protection, in this paper, is to do both, protect the identity and some soft biometric traits as well.

All of the aforementioned methods, either fail to protect the identity or the gender information, or the body shape cannot be used any more to interpret the body movements.

When we record people from far, the recognition of gender performs much better by using the shape of the whole body rather than the face only [6]. That is why the proposed *de-genderization* focus on the whole body.

We demonstrate in the next section that gender information remains visible even after the application of traditional privacy protection methods.

3. Recognition in the presence of privacy filters

3.1. Algorithm baselines

For face recognition evaluation, we have selected the Eigen [21] with the Euclidean distance and Histogram of Oriented Gradient [4] (HoG) with a SVM as baselines. We trained these methods on a subset of Feret [14] and ScfaceData [7] database, named training set, and tested on another subset, named test set, where privacy protection is applied. As privacy protection we select the pixelization and Gaussian blur with three different strengths, and the extraction of the shape (explained in Section 4.1). For original test set (without the privacy protection), the accuracy of identity recognition is 96% for the Eigen method, and 94% for HoG.

For the gender detection evaluation, we use a pre-trained gender detection method by bodies using Convolutional Neural Networks (CNN), published in [2]. We evaluate this tool over 1081 videos from the HID database [19], including 864 male videos and 220 female videos. Each video contains one person only. For each frame of a video, we get the confidence-rate of the gender predicted. 1 corresponds to a male prediction and 0 to a female one. To obtain the predicted gender in each video, we average all confidence-rates, denoted by mc. Then, the accuracy of female and male detection are computed when we get the gender predictions of the 1081 videos (i.e. if $mc > 0.5$, the gender is predicted as a male otherwise as a female). We apply the gender classifier [2] over the original videos and on the videos where the pixelization, Gaussian blur and body shape filters are applied. The accuracy of the classifier on the original body images is 81 %.

3.2. Protect the identity but not the gender

The results in Table 1, show that the number of recognition errors introduced by anonymization (i.e. the differences between the accuracy of algorithms got with the original images and the accuracy of algorithms got with the protected images) increases by a factor of more than two compared to the gender recognition errors. For instance, the difference in face recognition performance with the Eigen baseline before and after pixelization for the strongest strength is 64%, meaning the recognition accuracy was 96% before and after 32% in this case. For the gender recognition, the number before and after are 81% and 72%, thus the difference is 9%. For these filters, the identity is protected while gender information still remains.

Table 1: The percentage of decrease in performances between the accuracy of the original images and the ones from the protected versions with three different strength for Pixelization and Blur

Algo	Pixelization	Gaussian blur	Shape
`Identity with Eigen %`	[4; 33; 64]	[7; 30; 64]	[95.9]
`Identity with HoG %`	[9; 40; 81]	[33; 60; 93]	[51]
`Gender %`	[1; 4; 9]	[3; 11; 20]	[0]

4. Our proposed De-genderization methods

We have designed two different *de-genderization* methods to transform a body shape. The first one distorts the body shape towards a reference model, and the second one approximates the silhouette. The strength of each method is tunable depending on the exact application.

4.1. Find the body shape of a person

Figure 1: Respectively, original and body shape image for a male and a female

For videos, we apply a mixture of Gaussian (MOG) [23] (i.e. an adaptive background subtraction method). For images, we apply GrabCut [16], an approach based on optimization by an iterative graph-cut [3] to find the foreground (i.e. body of people). OpenCV[3] includes these methods.

We remove noise and holes using basic morphological operations, and we apply a Canny edge detection.

These methods provide a black and white image where white pixels are associated with the silhouette of the body, as it is shown in Fig. 1.

4.2. Merging between the body shape and a reference model

We merge the coordinates of the body shape and the ones from a model by following the steps below.

Step 1: The first step consists in aligning a shape body image with the model. For this purpose, we resize the image of the body shape with the ratios computed in the equation 1. A bounding box denotes the smallest box containing the set of points from a specific object (i.e. in our case, the body shape). In the equation 1,

[3] http://docs.opencv.org/

we denote: i) the height and the width from the model shape (Mh, Mw) and those from the body shape (Bh, Bw), and ii) the upper-left bounding box coordinates for the model shape (Mx, My) and those for the body shape (Bx, By).

$$ratioHeight = \frac{Mh - Mx}{Bh - Bx},\ ratioWidth = \frac{Mw - My}{Bw - By} \quad (1)$$

Step 2: Let B denotes the set of points of a body shape and M the set of points from the model shape. $\forall p_i \in B$, we find the point p_j included in M such that the distance is minimum. The formula is given in (2) where (x, y) represent the coordinates of a point.

$$\min_{\forall p_j \in M} \sqrt{(p_i.x - p_j.x)^2 + (p_i.y - p_j.y)^2} \quad (2)$$

Step 3: For each couple (p_i, p_j) found in the step 2, we merge the coordinates of p_i with the ones of p_j following the formula 3, where α is a parameter which controls the strength of the merging and p the new point with (x', y') as coordinates. For α, the closer to 1, the closer to the original shape is the new shape, and the closer to 0, the closer to the model shape is the new shape. If α is equal to 1, the new shape is equal to the original shape and if α is equal to 0, the new points are included in a subset of M because $\forall p_i \in B$ (not $\forall p_j \in M$) we found a couple (p_i, p_j). The model shape used to obtain the results in Figs. 2 and 3, is the male one shown in Fig. 4a.

$$\begin{aligned} & \forall (p_i, p_j) \\ p(x', y') = (& p_i.x * \alpha + p_j.x * (1 - \alpha), \\ & p_i.y * \alpha + p_j.y * (1 - \alpha)) \end{aligned} \quad (3)$$

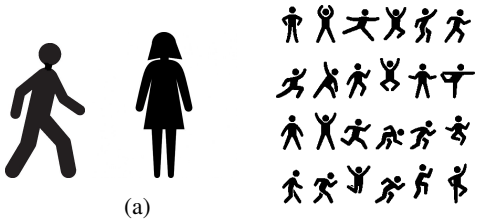
Step 4: In order to get the final shape, we link all the border segments according to the new coordinates found in step 3. See Fig. 2 for an example.

Figure 2: The purple shape (model), the yellow shape (original) and the blue shape (final) with $\alpha = 1, 0.8, 0.6, 0.4, 0.2, 0$

α	1	0.8	0.6	0.4	0.2	0
80.5	80.6	81.6	79.6	75.7	54	50

Figure 3: Original image and shape approximation using merging with the associated average accuracy of gender recognition (in the second row) according to the value of the parameter α

The average accuracy of gender recognition showing in Fig. 3 and the experiments explained in Section 5, show that the gender of the final shape is not correctly detected anymore because it converges systematically towards the gender of the reference model and the gender can be guessed with a probability of 50 %.



(a) (b)

Figure 4: (a) Selected models in our experiments, (b) Example of codebook of postures

In order to better preserve the movement of arms or legs, we could choose an appropriate model, frame by frame, from a codebook such as the one represented in Fig. 4b. Indeed, the visualization of arm and leg movements depend on the selected model. Thus, we have designed a second method based on the convexity of the body shape itself. In other words, it converges towards a polygonal simplification of the silhouette that helps to preserve human activities.

4.3. Polygonal approximation of a body shape

A convex set defines a region where for every pair of points in this region, each point on the segment that joins the pair of points is also inside that region. A convex hull defines the minimal convex set containing a set of points.

We replace the coordinates of the body shape (i.e. the coordinates of the white pixels) by the ones from the convex hull of several sets. Each set of points belongs to the points of the body shape and contains n neighbours.

Step 1: Let S denotes the set of points of a body shape. $\forall x \in S$, we find and draw the convex hull of $V \subset S$, where V is a subset of S containing x and its $n - 1$ nearest neighbours in terms of distance along the body contour.

Step 2: The step 1 produces several polygons of n points. We keep only the lines on the border (not the ones inside the new shape) and we obtain the new shape.

The green lines and points in the Fig. 5 define the original body shape and the blue lines the shape approximated by convexity with $n = 5$.

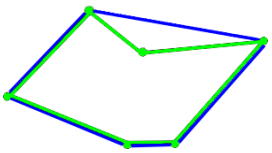


Figure 5: Approximation of 5 points using convexity

The higher is n the more convex is the shape (Fig. 6).

The experiments explained in Section 5 show the suppression

n	3	5	10	20	30
80.5	76	74.4	58.8	50	50

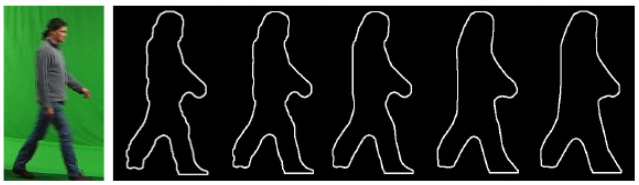


Figure 6: Original image and shape approximation using convexity with the associated average accuracy of gender recognition (in the second row) according to the value of the parameter n

of the gender of the final shape. Indeed, the machine systematically interprets the final shape as female.

5. Experimental Results

In this part, we prove that our methods hamper the gender detection like the inpainting [13], JPEG scrambling with high level [22] of protection and black box methods. Nevertheless, in the following, we conclude that the approximation using convexity preserves the human activities contrary to the others.

5.1. Evaluation of De-genderization

To evaluate the *de-genderization* of our two methods, we apply [2] (in 3.1, the description of the protocol) over the original videos, those generated by the first method (4.2) with $\alpha = 0, 0.2, 0.4, 0.6, 0.8, 1$ and those generated by the second method (4.3) with $n = 3, 5, 10, 20, 30$.

The Figs. 7 and 8 show the accuracy of male and female detection for original (the black point), shape (the pink point) body images compared to the approximation model (the blue points). The random case (the red points) can be considered as the results of the inpainting, scrambling and black box methods. For example, when $\alpha = 0.2$ using a male model (the point towards the top left corner in Fig. 7), the rate of correct detection for female is about 10 % whereas the one for male is about 95 %. We can note that the performances decrease for the two approximation models.

The Fig. 7 represents the results of the first method (4.2). We can observe that the closer to the model shape is the final shape, the more the tool classifies the gender of the final shape as the gender of the model shape. The Fig. 4a shows the selected reference models.

The Fig. 8 represents the results of the second method (4.3). We can observe that the higher is the number of n the more the classifier tags the body images as females. This is mainly due to the convex approximation producing more female forms, especially when the head has some typical women hairstyle and when in the lower body we can guess the presence of a skirt.

The approximation using the convexity is a better approach than the other one. As with an intermediate value of n (i.e. 10) the gender detection becomes weak whereas the motion of the arms and legs are much more visible than with the approximation using the merging. This can be observed in the video available at [4]. Moreover, for the first method, given the knowledge of reference

[4] https://youtu.be/rpuIDLrHx3g

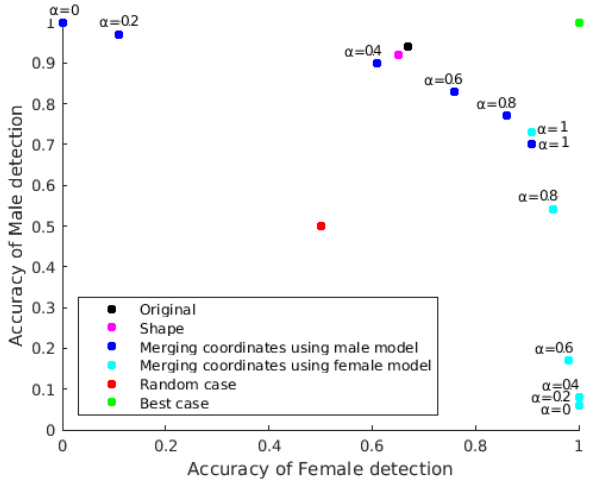


Figure 7: Results for the merging approach

model and α, an attacker might easily reverse the transformation and then retrieve the original shape.

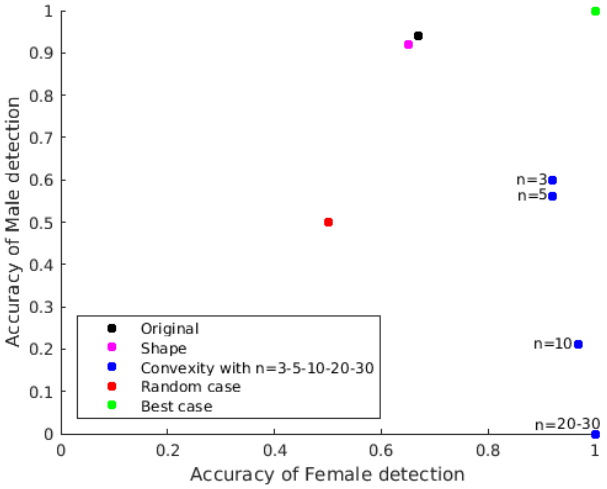


Figure 8: Results for the approximation by convexity approach

5.2. Evaluation of sports events classification

To evaluate the impact on the event classification, we chose to test our algorithm on sports events. We utilize Deepdetect[5] to classify the sports and [8] as dataset. Among the 143 sports available in [8], we select: acrobatic gym, walking, riding a horse, diving, golf, football, skateboarding and running. Next, for each selected images, we apply inpainting, JPEG high-Level scrambling, black box, shape detection (4.1) and approximations using convexity (4.3) with $n = 3, 5, 10, 20, 30$ (Figure 9), and feed again the classification tool with them.

The classification tool gives as results an ordered list of classes from the best to the worst one. Therefore, we compute the @k

[5] http://www.deepdetect.com/

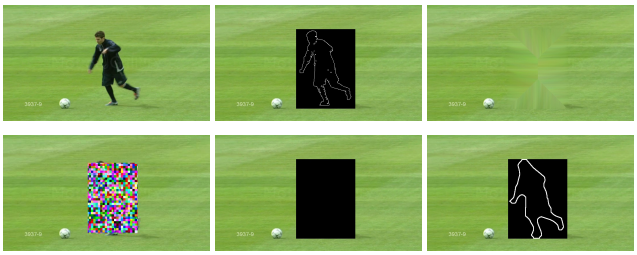
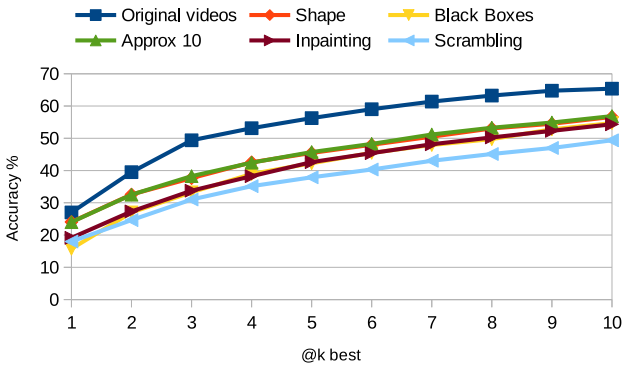
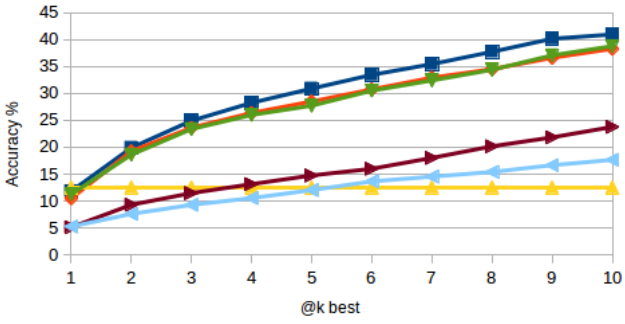


Figure 9: Respectively: Original image, shape, inpainting, scrambling, black box and approximated body with $n = 10$

accuracy curve with k=10. The curves, in Fig. 10, depict if the proper class is among the first ten results in the ordered list.



(a)



(b)

Figure 10: Accuracy@10 of sports events classification (a) on the whole images, (b) only on the RoI

The results denote similar performances for sports classification in terms of shape and approximated body ($\sim 10^{-2}$ %). Moreover, the body approximation method achieves the best score compared to the black box, the inpainting and the scrambling methods. However, the performances of sports classification on the whole images 10a for the black box, inpainting or scrambling filters, are closer than expected from the original ones. This is due to the background which helps a lot in the recognition of some sports (e.g. acrobatic gym, horse riding, diving, golf or football). Indeed, in 10b, where the tool recognizes sport only from the RoI, the performances drop more than 10 % compared to the original ones.

Thus, our approximation using convexity preserves the global movement of a person as well as people activities even with the absence of the background. In contrary to the methods that replace people with black or color boxes and preserve only the global motion.

5.3. Optimal parameter

We average the accuracy of male and female detection and the accuracy of the @k (from 1 to 10) best sports classification for each n. We illustrate the results in the table of the Fig. 11. This table reveals that both, the performances of gender detection and sports classification are decreasing. Nevertheless, the performances of sports classification slightly decrease (less than 1 %). Thus, the value of the parameter n does not much impact the sports classification. Indeed, the motion of the body is still preserved in any value of n as illustrated the images in Fig. 11. Consequently, the optimal value of n is the smaller value such as the accuracy of gender detection is equal to 50 %. According to our tests this value is $n = 18$.

n	3	5	10	20	30
Gender	76	74.4	58.8	50	50
Sports	45	44.93	44.74	44.27	44

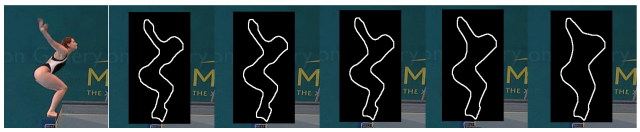


Figure 11: Average accuracy in % of gender detection and of sports classification according to n. Respectively: Original image and approximated body with respectively $n = 3, 5, 10, 20, 30$

6. Conclusions and Future works

In this paper, we present two different methods whose goal is not to only de-identify people but also to hide gender, denoted as *de-genderization*. The first method transforms a body shape by merging its coordinates with the ones of a predefined model whereas the second one approximates a body shape by applying polygonal approximations on it. We prove, in the experiments, that both methods, hamper the recognition of the gender. The gender, resulting from the first method, converges towards the gender associated with the model, whereas the second method feminizes the body shape. Unlike existing privacy filters, our method based on the polygonal approximation yields to hide the gender while keeping the action recognizable.

The drawback of these methods is their irreversibility. To have the possibility to recover the original image, we could combine our method with the one in [17]. Indeed, instead of inserting the body shape inside the MSBs of encrypted pixels we could use our images of body shape approximation.

A key point concerning the first method is about the design of the most appropriate codebook of models. It should include different postures and some visual traits (e.g. height, weight, clothes,

accessories) as well depending on the requirement of the exact application. The aim is to restitute the maximum of the authorized information.

It is known that identity or semantic traits can be extracted from gait. It would be a significant interest to also evaluate the impact of our new privacy filters on the performances attached to such tools.

We could design an adaptive approximation depending on the body parts. For example, we should apply a strong approximation on the hairstyle, shoulders and on the hips, but on the contrary apply a weak approximation on the legs and arms because these parts provide mostly information about human actions and few about gender.

We will investigate separately and jointly the design of some privacy filters dedicated to some other semantic traits (e.g. age, ethinicity, weight).

References

[1] P. Agrawal and P. Narayanan. Person de-identification in videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(3):299–310, 2011. 1

[2] G. Antipov, S.-A. Berrani, N. Ruchaud, and J.-L. Dugelay. Learned vs. hand-crafted features for pedestrian gender recognition. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, pages 1263–1266. ACM, 2015. 2, 4

[3] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 105–112. IEEE, 2001. 2

[4] O. Déniz, G. Bueno, J. Salido, and F. De la Torre. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 32(12):1598–1603, 2011. 2

[5] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32. 1

[6] E. Gonzalez, A. Dantcheva, R. Vera-Rodriguez, and al. Image-based gender estimation from body and face across distances. 2016. 2

[7] M. Grgic, K. Delac, and S. Grgic. Scface–surveillance cameras face database. *Multimedia tools and applications*, 51(3):863–879, 2011. 2

[8] L.-J. Li and L. Fei-Fei. What, where and who? classifying events by scene and object recognition. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007. 4

[9] X. Lu and A. K. Jain. Ethnicity identification from face images. In *Defense and Security*, pages 114–123. International Society for Optics and Photonics, 2004. 1

[10] E. Mäkinen and R. Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3):541–547, 2008. 1

[11] A. Nodari, M. Vanetti, and I. Gallo. Digital privacy: Replacing pedestrians from google street view images. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2889–2893. IEEE, 2012. 1

[12] A. Othman and A. Ross. Privacy of facial soft biometrics: Suppressing gender but retaining identity. In *Computer Vision-ECCV 2014 Workshops*, pages 682–696. Springer, 2014. 2

[13] J. K. Paruchuri, S.-C. S. Cheung, and M. W. Hail. Video data hiding for managing privacy information in surveillance systems. *EURASIP Journal on Information Security*, 2009:7, 2009. 1, 4

[14] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090–1104, 2000. 2

[15] N. Rachaud, G. Antipov, P. Korshunov, J.-L. Dugelay, T. Ebrahimi, and S.-A. Berrani. The impact of privacy protection filters on gender recognition. In *SPIE Optical Engineering+ Applications*, pages 959906–959906. International Society for Optics and Photonics, 2015. 1

[16] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, volume 23, pages 309–314. ACM, 2004. 2

[17] N. Ruchaud and J.-L. Dugelay. Efficient privacy protection in video surveillance by stegoscrambling. In WIFS, Rome, Italy, November, 2015. 1, 5

[18] N. Ruchaud and J. L. Dugelay. Privacy protection filter using stegoscrambling in video surveillance. In *MediaEval 2015 Workshop, Wurzen, Germany*, 2015. 1

[19] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. In *Applications and Science in Soft Computing*, pages 339–346. Springer, 2004. 2

[20] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002. 1

[21] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991. 2

[22] L. Yuan, P. Korshunov, and T. Ebrahimi. Secure jpeg scrambling enabling privacy in photo sharing. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 4, pages 1–6. IEEE, 2015. 1, 4

[23] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31. IEEE, 2004. 2