FACE DETECTION HINDERING

Panteleimon Chriskos[⋆] *Jonathan Munro*[†] *Vasileios Mygdalis*[†] *Ioannis Pitas*[†⋆]

[†] Department of Electrical and Electronic Engineering, University of Bristol, UK
[⋆] Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece

ABSTRACT

In this paper, we develop a face detection hindering method, as a means of preventing the threats to people's privacy, automatic video analysis may pose. Face detection in images or videos is the first step in human-centered video analysis to be followed, e.g. by automatic face recognition. Therefore, by hindering face detection, we also render automatic face recognition improbable. To this end, we examine the application of two methods. First, we consider a naive approach, i.e., we simply use additive or impulsive noise to the input image, until the point where the face cannot be automatically detected anymore. Second, we examine the application of the SVD-DID face de-identification method. Our experimental results denote that both methods attain high face detection failure rates.

Index Terms— face detection, privacy protection, surveillance, face de-identification

1. INTRODUCTION

As a vast amount of visual media is daily shared, viewed and stored on-line, serious threats to the depicted persons' privacy may be posed. For example, World Wide Web monitoring systems that use face detection, tracking and recognition [1] [2] in shared videos or images could also be used to violate privacy. Another major threat to privacy is due to the wide use of video surveillance in public places by surveillance/traffic cameras [3] or drones, since, any person can potentially be identified on such videos and images. In this context, let us suppose that a malicious user attempts to recognize and track a specific individual in visual media automatically. The first step is to detect all faces in an image or video frame then recognize these facial regions of interest (ROIs) and retain only the ones that depict the targeted individual. In order to protect the targeted individual privacy, dedicated to hinder face recognition, also known as face de-identification methods have been proposed.

This project has received funding European Union's European Union Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE). This publication reflects only the authors' views. The European Commission is not responsible for any use that may be made of the information it contains.

To begin with, ad-hoc face de-identification methods [4] apply masks on facial regions, e.g., by employing black bars to cover the eyes or T-shaped masks that cover both the eyes and nose and/or mouth. Other mask shapes can be used, e.g. rectangular ones that reveal only the mouth and elliptical or circular ones that over the entire facial image ROI. Other ad-hoc methods low-pass filter the facial image ROI [4], apply random noise [3], use the negative of the facial image, or swap face sub regions, such as eyes, nose or mouth, belonging to different individuals [5]. Other de-identification methods spatially subsample a facial image, resulting in pixelation, or apply a threshold to the facial image pixels [4]. Variational adaptive filtering in conjunction with face key point detection has also been proposed to achieve face de-identification, while retaining the facial expression of the depicted individual [6]. In [7] Active Appearance Models and the k-Same-furthest model are used to retain facial expressions on de-identified images. A large family of face de-dentification methods implement the k-anonymity model [3] [8], so that any of the de-identified images can be misclassified as belonging to at least to k original individuals. The de-identified image is calculated by averaging the k facial images that are most similar to the input image. Furthermore, an objective function can be formulated and the optimal weights for averaging the k most similar images are learned via gradient descent [9]. In [10], the least similar k images known as k-Same-furthest, are used instead. Building upon previous work, apart from using the least similar k images, unique de-identified faces are generated a for each of the k original faces [11]. The particularities of specific face identification methods can be used in order to defeat them [12]. In [13], facial images are replaced by 3D morphable facial models. In [14], the initial face is replaced with a face from another person. Another face de-identification method reduces the number of eigenfaces used in reconstructing the facial images [15].

Although there is much research in face de-identification, this is not true in the case of face detector hindering. In this paper, we aim at hindering face detection in the first place, thus rendering face recognition improbable. Our proposed approach is to perform facial image corruption as little as possible, so that automatic face detection fails, while the face is still recognizable by humans. To this end, we examine the

application of two methods. First, we consider a naive approach, i.e., we simply use additive or impulsive noise to the input image, until the point where the face cannot be automatically detected anymore. Second, we examine the application of the so-called SVD-DID face de-identification method [16]. Our experiments denote that both methods have the potential of hindering robust face detection in images.

The rest of this paper is organized as follows. In section 2, additive or impulsive noise or the SVD-DID method are used to hinder face detection. Section 3 describes the experiment and results. Finally in Section 4, the conclusions are drawn.

2. HINDERING FACE DETECTION BY CORRUPTING THE FACIAL IMAGE

The most straight forward approach to hindering face detection is to apply noise to the image. Uniform noise: $n_u(i,j) = \beta(\eta(i,j) - 0.5)$ (where $\eta(i,j) \sim U[0,1]$), or Gaussian Noise: $n_g(i,j) \sim N(0,\sigma)$ can by used for image corruption $I_n = I + n$. Alternatively, impulsive noise $n_i(i,j) = \begin{cases} 255 \text{ or } 0 & \text{for } p > \eta(i,j) \\ I(i,j) & \text{for } p <= \eta(i,j) \end{cases}$, where $\eta(i,j) \sim U[0,1]$ can be used for the same purpose.

The intensity of the noise (and hence its visibility) can be changed by varying β, σ or p for uniform, Gaussian or impulsive noise, respectively. In all cases increasing the parameter values leads to an increase in the intensity of the noise. These noise patterns were then applied to the face region, found using the Viola and Jones face detector [17] on the original image.

The SVD-DID method [16] utilizes the Singular Value Decomposition (SVD) method to introduce artifacts in the output image. It was originally proposed for face de-identification. Here we prove experimentally that it can be used for hindering face detection.

Briefly, the SVD-DID method uses the facial image SVD matrix $\mathbf{A} \in \Re^{N \times M}$ factorization as a product of three matrices: the singular values matrix $\mathbf{S} \in \Re^{N \times M}$ and the singularvector matrices $\mathbf{U} \in \Re^{N \times M}$ and $\mathbf{V} \in \Re^{M \times M}$ [18]:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T. \quad (1)$$

The eigenvectors of matrix $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ form the columns of matrices $\mathbf{U}$ and $\mathbf{V}$ respectively. The singular values in $\mathbf{S}$ are the square roots of the eigenvalues of matrix $\mathbf{A}\mathbf{A}^T$. The SVD-DID method modifies the output image by altering the entries of matrices $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$. This is done in three distinct steps:

1. SVD Coefficient Zeroing (SVD-CZ). In the first step, we note that the largest singular values correspond to the majority of facial image energy. In this step, we remove this information by zeroing the first N_Z singular values in $\mathbf{S}$ ($N_Z \leq N \leq M$), thus producing a new $\mathbf{S}$ matrix referred to as $\mathbf{S}_{CZ}$. As the final facial image tends to become darker than the input image due to energy loss, this is counterbalanced by increasing the facial image pixel luminance at the end of the de-identification process, e.g. by adding a fixed luminance value to the output facial image pixels.

2. SVD Coefficient Averaging (SVD-CA). In the next step, the entries of the eigenvectors in matrices $\mathbf{U}$, $\mathbf{V}$ are low-pass filtered using an $m \times m$ circular averaging filter with $m = 2R + 1$ [19], where R is the radius of the circular filter, thus producing the matrices $\mathbf{U}_{AV}$ and $\mathbf{V}_{AV}$. The facial image reconstruction solely from these averaged matrices leads to poor image quality. To counterbalance this effect, the new matrices $\mathbf{U}_{AV}$ and $\mathbf{V}_{AV}$ are blended with the original $\mathbf{U}$, $\mathbf{V}$ matrices as follows:

$$\mathbf{U}_{CA} = \frac{\alpha * \mathbf{U}_{AV} + \mathbf{U}}{1+\alpha}, \mathbf{V}_{CA} = \frac{\alpha * \mathbf{V}_{AV} + \mathbf{V}}{1+\alpha}, \quad (2)$$

where the parameter α adjusts the trade-off between visual quality and face detection hindering potential. Similarly to the previous step, facial image darkening is counterbalanced by adding luminance to the output image pixels.

3. SVD Modified Sobel Filtering (SVD-MSF). The final step utilizes a modified Sobel filter in order to high pass filter [19] matrices $\mathbf{U}_{CA}$ and $\mathbf{V}_{CA}$. The modified Sobel filter coefficients have a 3×3 matrix form:

$$\mathbf{G} = \begin{bmatrix} d & 2d & d \\ 0 & 0 & 0 \\ -d & -2d & -d \end{bmatrix}, \quad (3)$$

where parameter d specifies the intensity of the high pass filtering. Finally matrices $\mathbf{U}_F$ and $\mathbf{V}_F$ are blended with the original matrices $\mathbf{U}$ and $\mathbf{V}$ according to (2) resulting in the matrices $\mathbf{U}_F$ and $\mathbf{V}_F$, to be used in the calculation of the output facial image matrix $\mathbf{A_d}$ of the SVD-DID method:

$$\mathbf{A_d} = \mathbf{U}_F\mathbf{S}_{CZ}\mathbf{V}_F^T. \quad (4)$$

3. EXPERIMENTAL FACE DETECTION HINDERING RESULTS

3.1. Experimental Setup

Experiments to assess the effectiveness of the above methods were performed on 653 401×321 pixel facial images depicting 15 different individuals from the XM2VTS [20] database. The facial images are close-ups frontal ones and have a neutral background. From dataset [21], a subset was also used containing 3471 images depicting 150 different individuals. The images were used both either RGB or 8-bit grayscale ones. In both cases the results were similar and as such only the results for the grayscale images are presented.

To quantify face detector hindering the face detection failure percentage F_p was used which is the number of images in

which a face is detected after applying the methods above, divided by the total number of facial images. However, even if, after the face obfuscation process, a face is detected, this does not mean that the actual face region has been detected. In order to quantify the accuracy of the face detector after obfuscation metric r is defined as:

$$r = \frac{|Im \cap Im'|}{|Im \cup Im'|} \quad (5)$$

where Im, Im' are the facial image regions (pixel sets) found on the original and the obfuscated image, respectively. The [17] face detector has been used in all experiments. Using this face detector on the XM2VTS images there were no false detections prior to corrupting the face ROI. In the other dataset some false face detections were present and the true face ROI was selected by selecting the largest of the detected regions which were verified manually. In the case of perfect alignment of Im and Im', this metric is equal to 1, whereas lower values indicate higher success in face detection hindering. To quantify this face detection inaccuracy over an entire facial imageset, the mean $\bar{r}$ is calculated $\bar{r} = \frac{1}{N_f}\sum_{i=1}^{N_f} r_i$, where N_f is the number of images in which a face is detected after obfuscation.

Results using noise corruption. They are presented in Table 1. It must be noted that, before the application of the SVD-DID method, the face detector had 100 % face detection accuracy. It can be concluded that applying Uniform and Gaussian noise on an image leads to poor results regarding face detector hindering. In the subset of the XM2VTS dataset, the corruption by uniform noise fails to prevent face detection, since the percentage F_p is equal to zero and metric $\bar{r}$ is high. Similar results are obtained for the second dataset, but, in this case, the percentage F_p is higher. Similar results are found for the Gaussian noise. However, facial image corruption by impulsive noise result in much better face detector hindering than in the two previous cases. For high values of parameter p, high F_p percentages can be attained and at the same time metric $\bar{r}$ is low. However this is achieved for $p = 0.5$ and $p = 0.8$ meaning that over half of the original image pixels are polluted by noise, leading to poor subjective facial image quality, as shown in Figure 1. As most face recognition algorithms, e.g. the subspace based ones [22], are sensitive to face localization and size errors, even relatively large values of $\bar{r}$ (e.g. $\bar{r} = 0.9$ in Table 1 means that the detected images are unrecognizable. All types of noise corruption are good in this respect. Such obfuscated images are much more presentable, as can be seen in Figure 2.

SVD-DID Results. The SVD-DID method is applied only on the facial regions in order to minimize image quality degradation. Table 2 presents F_p and $\bar{r}$ for this method. It can be deduced that for various SVD-DID parameter values, the face detector failure percentages are high. The lowest face detection failure percentage is equal to 85.76%, while the highest is equal to 99.08% for the XM2VTS subset. The corresponding percentages are 79.27% and 97.49%, respectively, for the other dataset subset. Representative obfuscated images, where face detection fails are shown in Figure 3. This difference is mainly due to the number of singular values that are zeroed during the SVD-CZ step, since, in the second case, a larger percentage of data energy is removed, while constructing the output image.

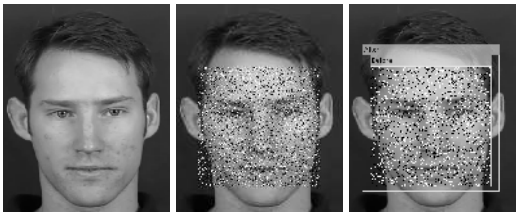
Table 1. Face detection failure percentages after adding noise

Uniform Noise				
Parameter	XM2VTS		Subset of [21]	
β	F_p	$\bar{r}$	F_p	$\bar{r}$
50	0.00%	0.953	11.08%	0.921
100	0.00%	0.894	16.43%	0.875
150	0.00%	0.848	24.59%	0.830
Gaussian Noise				
σ	F_p	$\bar{r}$	F_p	$\bar{r}$
25	0.00%	0.909	14.82%	0.891
50	0.15%	0.835	29.01%	0.820
75	8.73%	0.790	46.52%	0.764
Impulsive Noise				
p	F_p	$\bar{r}$	F_p	$\bar{r}$
0.2	7.20 %	0.800	45.03 %	0.765
0.5	95.71 %	0.584	93.39 %	0.448
0.8	98.93 %	0.115	98.13 %	0.264

a) b) c)

Fig. 1. Facial image a) obfuscation using impulsive noise with: b) $p = 0.2$, c) $p = 0.5$.

In all cases tabulated in Table 2, the added image pixel luminance is set to $+100$, to counter image darkening. The rest of the parameters also cause slight variations in face detection failure percentages and, depending on their combination, impact face detection failure percentages positively or negatively. As shown, the face detection failure percentages are high, meaning that the face detector fails to detect any face in an image. In fact, however, the face detector failures are even higher, since detecting a face in an image does not necessarily mean that a face has been correctly localized, thus leading in face recognition failures. Examples of such cases from the XM2VTS image subset are presented in Figure 5, where the SVD-DID method is applied for parameters values: $N = 2$, $lum = +100$, $\alpha = 0.2$, $d = 0.5$ and $R = 5$. Another example for the same parameters is displayed in Figure 4, where the SVD-DID method is selectively applied on faces in an



a) b) c)

Fig. 2. a) Original facial image, b) obfuscated image with impulsive noise, c) detected ROIs before and after image corruption.

a) b) c)

Fig. 3. Example of applying the SVD-DID method: a) original image b) face detection c) apply SVD-DID on face area ($N_Z = 1, \alpha = 0.5, d = 0.8, R = 5$). No face is detected in this image.

image captured with an aerial drone. In this example, the proposed method had been applied only for the person depicted on the left. It can be seen in Table 2 that, in the cases that a face is detected after obfuscation, the $\bar{r}$ values range from a maximum of 0.667 to a minimum of 0.032 for XM2VTS and 0.6686 and 0.0723 respectively for [21], thus as in the failure percentages, N_Z plays a major role in the values of $\bar{r}$.

These results confirm that the SVD-DID method it is very good at hindering both face detection and face recognition.

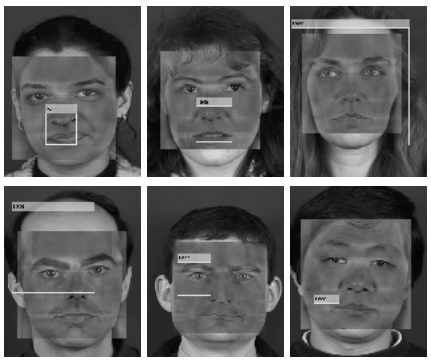


Fig. 5. Examples of false face detection after applying the SVD-DID method.

Table 2. Face detection failure percentages after applying SVD-DID

Parameter Values				XM2VTS		Subset of [21]	
N_Z	α	d	R	F_p	$\bar{r}$	F_p	$\bar{r}$
1	0.2	0.1	5	89.74%	0.591	79.27%	0.658
1	0.2	0.1	10	88.82%	0.594	79.77%	0.655
1	0.2	0.5	5	91.58%	0.582	79.97%	0.646
1	0.2	0.5	10	87.44%	0.644	79.85%	0.666
1	0.5	0.1	5	85.76%	0.655	81.20%	0.669
1	0.5	0.1	10	85.91%	0.658	81.29%	0.658
1	0.5	0.5	5	86.68%	0.645	81.96%	0.662
1	0.5	0.5	10	86.52%	0.666	81.52%	0.652
2	0.2	0.1	5	93.72%	0.408	93.45%	0.296
2	0.2	0.1	10	94.33%	0.456	93.22%	0.276
2	0.2	0.5	5	94.95%	0.437	93.45%	0.304
2	0.2	0.5	10	94.33%	0.424	93.60%	0.290
2	0.5	0.1	5	91.88%	0.482	94.18%	0.300
2	0.5	0.1	10	92.34%	0.543	94.01%	0.284
2	0.5	0.5	5	91.88%	0.470	94.39%	0.273
2	0.5	0.5	10	92.96%	0.528	94.33%	0.288
5	0.2	0.1	5	97.09%	0.040	96.84%	0.089
5	0.2	0.1	10	96.63%	0.041	96.78%	0.095
5	0.2	0.5	5	97.24%	0.039	97.08%	0.087
5	0.2	0.5	10	96.17%	0.037	96.58%	0.095
5	0.5	0.1	5	98.32%	0.033	97.40%	0.075
5	0.5	0.1	10	97.70%	0.037	97.31%	0.081
5	0.5	0.5	5	99.08%	0.032	97.19%	0.072
5	0.5	0.5	10	98.16%	0.040	97.49%	0.079

Fig. 4. Example of selectively applying the SVD-DID method on an image captured from an aerial drone.

4. CONCLUSIONS

The experimental results verify that the SVD-DID method is capable of hindering face detection. Its performance is much better than noise corruption. This can be proven since face detection failure percentages reach 99.08% and that the mean overlap of the detected face regions before and after applying SVD-DID is equal to 0.032. Future work in this area will focus on developing less visible and reversible face detector obfuscation methods so that, image quality does not suffer as much while the original image can be recovered. Furthermore the effectiveness SVD-DID will be assessed against more robust face detectors e.g. based on deep neural networks.

5. REFERENCES

[1] Alexandros Iosifidis, Anastasios Tefas, and Ioannis Pitas, "Person identification from actions based on dynemes and discriminant learning," in *Biometrics and Forensics (IWBF), 2013 International Workshop on*. IEEE, 2013, pp. 1–4.

[2] GN Stamou, M Krinidis, N Nikolaidis, and I Pitas, "A monocular system for automatic face detection and tracking," in *Visual Communications and Image Processing 2005*. International Society for Optics and Photonics, 2005, pp. 59602C–59602C.

[3] Ralph Gross, Latanya Sweeney, Jeffrey Cohn, Fernando De la Torre, and Simon Baker, "Face de-identification," in *Protecting Privacy in Video Surveillance*, pp. 129–146. Springer, 2009.

[4] Elaine Newton, Latanya Sweeney, and Bradley Malin, "B.: Preserving privacy by de-identifying facial images," in *IEEE Transactions on Knowledge and Data Engineering, IEEE TKDE*. Citeseer, 2005.

[5] Saleh Mosaddegh, Loic Simon, and Frédéric Jurie, "Photorealistic face de-identification by aggregating donors face components," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 159–174.

[6] Geoffrey Letournel, Aurélie Bugeau, V-T Ta, and J-P Domenger, "Face de-identification with expressions preservation," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4366–4370.

[7] Lily Meng, Zongji Sun, Aladdin Ariyaeeinia, and Ken L Bennett, "Retaining expressions on de-identified faces," in *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*. IEEE, 2014, pp. 1252–1257.

[8] Latanya Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[9] Amin Jourabloo, Xi Yin, and Xiaoming Liu, "Attribute preserved face de-identification," in *2015 International Conference on Biometrics (ICB)*. IEEE, 2015, pp. 278–285.

[10] Lily Meng and Zongji Sun, "Face de-identification with perfect privacy protection," in *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*. IEEE, 2014, pp. 1234–1239.

[11] Zongji Sun, Li Meng, and Aladdin Ariyaeeinia, "Distinguishable de-identified faces," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*. IEEE, 2015, vol. 4, pp. 1–6.

[12] Benedikt Driessen and Markus Dürmuth, "Achieving anonymity against major face recognition algorithms," in *IFIP International Conference on Communications and Multimedia Security*. Springer, 2013, pp. 18–33.

[13] Volker Blanz, Sami Romdhani, and Thomas Vetter, "Face identification across different poses and illuminations with a 3d morphable model," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 192–197.

[14] Dmitri Bitouk, Neeraj Kumar, Samreen Dhillon, Peter Belhumeur, and Shree K Nayar, "Face swapping: automatically replacing faces in photographs," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 39, 2008.

[15] P Jonathon Phillips, "Privacy operating characteristic for privacy protection in surveillance applications," in *International Conference on Audio-and Video-Based Biometric Person Authentication*. Springer, 2005, pp. 869–878.

[16] P Chriskos, O Zoidi, A Tefas, and I Pitas, "De-identifying facial images using singular value decomposition and projections," *Multimedia Tools and Applications*, pp. 1–34, 2016.

[17] Paul Viola and Michael J Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[18] Gene H Golub and Charles F Van Loan, *Matrix computations*, vol. 3, JHU Press, 2012.

[19] Ioannis Pitas, *Digital image processing algorithms and applications*, John Wiley & Sons, 2000.

[20] Kieron Messer, Jiri Matas, Josef Kittler, Juergen Luettin, and Gilbert Maitre, "Xm2vtsdb: The extended m2vts database," in *Second international conference on audio and video-based biometric person authentication*. Citeseer, 1999, vol. 964, pp. 965–966.

[21] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[22] Anastasios Maronidis, Dimitris Bolis, Anastasios Tefas, and Ioannis Pitas, "Improving subspace learning for facial expression recognition using person dependent and geometrically enriched training sets," *Neural Networks*, vol. 24, no. 8, pp. 814–823, 2011.