DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-181

Automatic Face Anonymization in Visual Data: Are we really well protected?

Natacha Ruchaud and Jean-Luc Dugelay
Eurecom
450 Route des Chappes, 06904 Biot Sophia Antipolis, France

Abstract

With the proliferation of digital visual data in diverse domains (video surveillance, social networks, medias, etc.), privacy concerns increase. Obscuring faces in images and videos is one option to preserve privacy while keeping a certain level of quality and intelligibility of the video. Most popular filters are blackener (black masking), pixelization and blurring. Even if it appears efficient at first sight, in terms of human perception, we demonstrate in this article that as soon as the category and the strength of the filter used to obscure faces can be (automatically) identified, there exist in the literature ad-hoc powerful approaches enable to partially cancel the impact of such filters with regards to automatic face recognition. Hence, evaluation is expressed in terms of face recognition rate associated with clean, obscured and de-obscured face images.

Figure 1: Respectively, "20 minutes" a French magazine using pixelization filter, "crimes" a French program using blurring filter and Street view by google using blurring filter.

Introduction

The widespread use of cameras and social networks in everyday life enforce concerns about personal privacy violation. Because significant recent improvements have been made in the field of pedestrian detection [4], face detection [26], human recognition [1, 25] and image restoration, questions about respect of privacy become more and more important. Also the improvement in image sensors helps to increase performances attached to identification techniques (e.g. a person can be recognized even far away from a camera).

Traditional privacy filters have already been designed and seem to protect enough privacy of people e.g. blurring or pixelization in media as illustrated Figure 1. In addition, some applications have been created like ObscuraCam [1] on Android where faces can be hidden by pixelization or blackener, and FacePixelizer [2] on Google plus where faces can be hidden by pixelization, blurring or blackener with different levels of intensity. Google street view also protects privacy of people by blurring faces as illustrated in Figure 1. Of course, there exist more sophisticated filters like morphing [13], warping [13], scrambling [18] but they are rarely used in practice by media.

This is why, we consider, at first, the four following filters:

- Blackener filter is obtained by applying the following formula:

$$ImgBlackened = originalImg * (1 - \alpha) \tag{1}$$

with α representing the opacity, the bigger is α the stronger is the impact of the filter.

- Pixelization can be perceived as a downsampling of the image without modifying the size. Image is split by N*N non overlapping squares, with N a parameter manually tuned. All pixels inside those squares are replaced by the average value included in each square.

$$Ip(x,y) = \frac{1}{b^2}\sum_{i=0}^{b-1}\sum_{j=0}^{b-1} I(\lfloor\frac{x}{b}\rfloor + i, \lfloor\frac{y}{b}\rfloor + j), \tag{2}$$

where x and y are the pixel coordinates and b is the block size.

- Gaussian blur is produced by the convolution of the image and the gaussian function [19]:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

where x and y are the pixel coordinates, and σ is the standard deviation of the Gaussian distribution.

- Gaussian noise uses the probability density function p of a Gaussian random variable z given by:

$$p_G(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \tag{4}$$

where z represents the grey level, μ the mean value and σ the standard deviation. [2]

Secondly, for the testing part, the four following filters have been added:

-Averaging neighbours of circle area.

-Motion blur with different strengths defined by the length and the angle of the motion that we want.

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

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-181

-Speckle noise [8] with varying standard deviation.

-Salt and Pepper with different density noise.

The level of privacy protection is controlled by varying parameters. The experiments in [14] demonstrate that, in general, an increase in strength of privacy filters leads to an increase in privacy (i.e., reduction in terms of recognition rate).

An illustration of those filters is shown in Figure 2 and the name of their associated tuning parameters are summed up in Table 1.

Filter	Parameter
Blackener	opacity
Pixelization	size of squares
Gaussian Blur	standard deviation
Gaussian Noise	standard deviation
Average Blur	neighbouring area
Motion Blur	length and angle of the motion
Speckle noise	standard deviation
Salt and Pepper Noise	noise density

Table 1: Table 1: Privacy filters with the name of their associated strength.

In [14], authors objectively demonstrated that face recognition highly decreases for blackened, pixelated and blurred faces. In the rest of the paper, face images where privacy filters are applied, are referred as obscured face images. The domain of image restoration enables to partially recover original face images, referred as de-obscured face images later in the paper. Indeed, image restoration improves quality of images or reconstructs corrupted images. Several methods like de-blurring [24], de-noising [10, 16, 22], superResolution [12, 20] are potentially efficient but require an a priori knowledge about the exact corruption.

To the best of our knowledge, there is no method to detect the category of the filter used to obscure face images. Hence, the key step consists in identifying automatically the category and the associated strength of the filter that have been used to obscure faces. Concerning this preliminary but mandatory step, we propose the following approach. First, we select a method to classify obscured faces from not obscured faces. Non obscured face images are defined as clean face images in the rest of the paper. Then a second approach is designed to classify the type of filter. As soon as the filter is identified, a last step would consist in defining the strength. Supposing an a priori knowledge of the category and the strength of the filter, found in the previous steps, we demonstrate that a de-obscured face operation (i.e. image restoration methods) can be efficiently performed and therefore privacy filters become much less effective.

The rest of the paper is organized as follows: in the next section, we explain the proposed obscured face detection against clear face following by the categorization of the filter and the estimation of the filter strength. Then we describe image restoration methods used to de-obscured face images. In the last section, a detailed evaluation of our proposed workflow for filter classification is explained and knowing the category and the strength of the privacy filter we demonstrate that image restoration allows to recover significant face recognition rates. Finally, we briefly conclude and give an outlook on possible future works.

Figure 2: Obscured face images with filters. From left to right on the top: clean faces, blackener, pixelization, gaussian blur, gaussian noise; on the bottom: average blur, motion blur, speckle noise, salt and pepper noise.

System overview

Detection of obscured face images

Histogram of oriented gradient, referred to as HoG in the paper, is widely used in pedestrian recognition [4] as well as for faces [5] and object recognition [7]. First, HoG computes gradients of image, then calculates histogram of oriented gradients (between *0* and *180°* with *9* bins) on each sub part of an image defined by size of cell (*[8, 8]* in our experiments) and concatenates all histograms. All images should be of the same size. In our experiments we used *112* for the height and *92* for the width.

Clean face images and obscured images generate several differences for oriented gradients. Indeed, pixelization creates more block effects among direction to *0* or *90°* appear. For noise, all directions got almost the same frequency. Blurring and blackener create dominant orientations.

HoG features with a linear SVM classifier are computed to train a model which differentiates clean faces from obscured faces.

Categorization of the filter

This step consists to detect the category of the filter used to obscure face images. Principal Component Analysis (PCA) [25], referred to as Eigen in this paper, is more sensitive to light, scale and translation variations. So Eigen is less robust against blackener filter which modifies the darkness of images, compared to other filters. Local Binary Pattern Histogram [1], referred to as 'LBPH' in the paper, is also employed to classify texture that is appropriate in the present case as privacy filters create some specific texture patterns except blackener filter which removes textures by masking them.

So instead of making a classification between blackener, pixelization, blurring and noise we first compute a classification with Eigen features between blackener against all other filters and then a classification with LBPH to distinguish the three reminder filters.

Estimation of the filter strength

Restoration methods are efficient when not only the exact filter is a priori known but also the strength that has been used. Therefore we propose to automatically sort the strength of a filter by the following approaches:

-Blackener filters are sorted into three strengths according to the image's darkness (mean of image pixels).

-Pixelization filters are sorted into three strengths according to the percentage of straights lines in edge images.

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-181

-Blurring filters are sorted into four strengths according to the percentage of edges. Point spread function (PSF) estimation for Image Deblurring [3] is mainly used but unfortunately this method does not work for all type of blur, in particular motion blur. This is why, an approach has been designed in our work.

-Noising filters are sorted into four strengths according to their noise estimation using the detail coefficients of the Discrete wavelet transform (DWT) [10, 22].

To prove that our proposed filter categorization and strength estimation help image restoration methods, we implement basic and advanced image restoration methods (presented in the next section). Then, we test face recognition after an image restoration method which is selected according to the filter category and applied with and without strength estimation. Results clearly show that strength classification increases performances of face recognition.

Image restoration

In this section, we describe image restoration methods which have been selected in our experiments.

-De-blackened: Formula 5 allows to obscure faces with blackener.

$$ImgBlackened = cleanImg * (1 - \alpha) \tag{5}$$

with α representing the opacity between *0.99* and *0.7*. Inverse formula is computed to find an approximated version of the clean image:

$$cleanImg \sim ImgBlackened/(1 - \alpha) \tag{6}$$

Opacity, α, in equation 6, is computed depending on the estimation of the filter strength.

-De-pixelization method 1: Bicubic interpolation [12] is the most simple and popular method in superResolution domain. First pixelated faces are down sampled in a smaller size depending on the estimation of the filter strength to be re expressed as a super-Resolution problem. Then the down sampled faces are resized with bicubic interpolation to its original size.

-De-pixelization method 2: a superResolution by adaptive sparse domain selection and an adaptive regularization [6] are applied on pixelated face images. Depending on the estimation of the filter strength, the size of the point-spread function is found.

-De-blurring method 1: the principle of unsharp method is to compute an edge image *g(x, y)* from an input image *f(x, y)* and *fsmooth(x, y)* a smoothed version of *f(x, y)*:

$$g(x,y) = f(x,y) - fsmooth(x,y) \tag{7}$$

And the sharpen image is calculated as following:

$$fsharp(x,y) = f(x,y) + k * g(x,y) \tag{8}$$

where k is a scaling constant. We fix k depending on the estimation of the filter strength.

-De-blurring method 2: In order to de-blur, an estimation of the unknown blur is performed with a maximum a posteriori estimation [15] on blurred faces. The number of iteration inside this method depends on the estimation of the filter strength.

-De-noising method 1: Formula 9 and 10 represent the Wiener de-noising algorithm [16] which is reiterated on noised face images depending on the estimation of the filter strength.

$$b(x,y) = \mu + \frac{\sigma^2 - v^2}{\sigma^2}(f(x,y) - \mu), \tag{9}$$

where v is the local estimated variances, f a noised image, b the de-noising image, x and y the pixel coordinates.

$$\mu = \frac{1}{NM}\sum_{x,y\in\eta} f(x,y), \sigma^2 = \frac{1}{NM}\sum_{x,y\in\eta} f^2(x,y) - \mu^2 \tag{10}$$

where η represents the local neighbourhood of each pixel and N, M the image size. Finally, bicubic interpolation is computed to first reduce and then enlarge image in order to delete remaining noise.

-De-noising method 2: We select a denoising method based on wavelet decompositions [17]. The number of decompositions depends on the estimation of the filter strength.

Figure 3 shows the workflow of the proposed method.

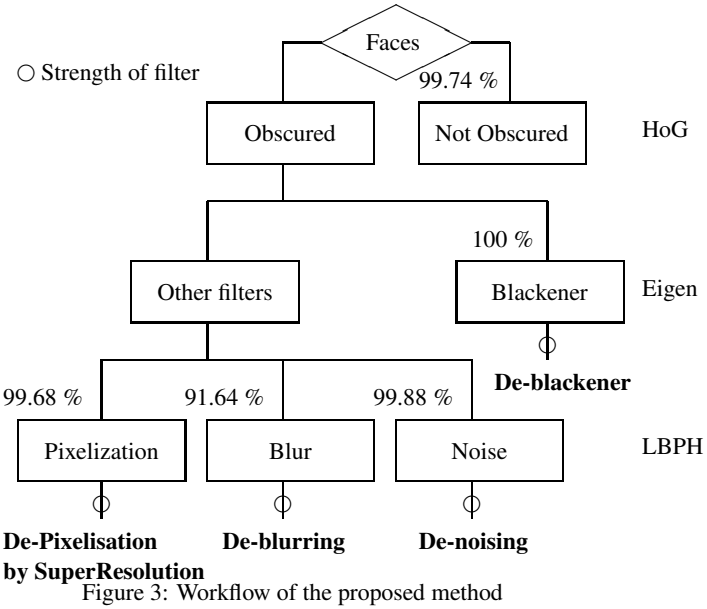


Figure 3: Workflow of the proposed method

Experimental results

Our framework has been evaluated in terms on percentage of correct classification. Three popular datasets are selected: Feret [21], ScFace [9] and ORL [23]. Faces have been already cropped using Viola and Jones [26] faces detector.

Evaluation of filters classification

All steps of filter classification, previously explained, are trained using Feret, ScFace and ORL face datasets whereas only faces from Feret are employed for the testing part. Blackener, pixelization, gaussian blur and gaussian noise are applied on clean faces with different level of strength. Table 2 sums up the number of faces which have been used in the training step for each method.

Our filter classification method have been tested on 13 810 faces which contains clean (1149), blackened (2302), pixelized

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-181

Method	Original	Obscured
HoG+SVM	3 644	11 606

Method	Blackener	Other filters
Eigen+Euclidean distance	442	5 967

Method	Pix	Blurring	Noising
LBPH+Euclidean distance	1 326	1 768	1 105

Table 2: Number of faces used in training set.

		Prediction				
		Orig	Black	Pix	Blur	Noise
Ground Truth	Orig	99.7%	0.3%			
	Black	0%	100%	0%	0%	0%
	Pix	0%	0%	99.7%	0.3%	0%
	Blur	8.4%	0%	0%	91.6%	0%
	Noise	0%	0.1%	0%	0%	99.9%

Table 3: Confusion matrix

(3138), gaussian blur (3768) and gaussian noise (3453) faces with different level of strength which had not necessary been used in the training set. The percentages of correct and wrong classification are represented by the confusion matrix in Table 3. According to these results, Pixelized faces which are wrongly classified are detected like blurred faces, Blurred faces which are wrongly classified as original faces and noisy faces which are wrongly classified as blackened faces. Indeed, Pixelization faces, for low strength, look like blurred faces. Blurred faces, for low strength, look like clean faces. Noisy faces, for strong strength, are closer to blackened faces.

In order to test the robustness of our framework, other types of noise (speckle, salt and pepper) and blur (average and motion blur) have been applied on clean faces. Then 11 510 other noised and blurred faces are classified by our method. Number of testing faces for other blur and noise are detailed in Table 4 as well as percentages of correct classification respectively.

	Blurring	Noising
Numb	5 755	5 755
%	97.85	99.1

Table 4: Percentages of correct classification for other type of blur and noise.

Image restoration

In this part only Feret dataset is used. In the training set, 265 people have been selected with 2-8 images per people, 879 images in total. In the testing set, 112 people have been selected with 1-3 images per people, 212 images in total. Blackener (opacity: 0.1, 0.2, 0.3), pixelization (size of averaging: 3, 4, 5, 6, 7, 8, 9, 10), gaussian blur (standard deviation: 2, 3, 4, 5, 8), Gaussian noise (variance: 0.001, 0.005, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.3) have been applied on the selected faces.

The robustness of face recognition against privacy filters differs from one algorithm to another one. This is why, to be representative of the state-of-the-art, three different face recognition algorithms, LBPH+Euclidean distance, HoG+SVM and Eigen+Euclidean distance, previously explained, have been selected as face recognition algorithms with respectively, 92.45 %, 94.34 % and 93.4 % of accuracy for clean face images. Moreover, LBPH and Eigen are often used as baselines by the biometric community to compare impact of obscuration on face recognition [14] but also to estimate face recognition before and after a superResolution method [11]. HoG is also employed as face recognition algorithm [5].

The difference between rate of good recognition on clean faces and obscured faces before image restoration in blue, after image restoration without strength classification in yellow or brown and after image restoration with strength prediction in red or green are shown, quantitatively, in Figures 4, 5, 7, 9, 11 and qualitatively, in Figures 6, 8, 10, 12. The lower is the curve the closer performances are to the clean face images and the better is the recognition. According Figures 5, 7, 9, 11, performances after image restoration with strength classification (in red and green) are the best. For instance, in Figure 5, a strong increase for Eigen recognition algorithm is shown between before (in blue) and after applying de-blackener (in red).

However, we notice that for de-blurring and de-noising, without strength classification, LBPH face recognitions obtain better results. This can be explained because LBPH is not robust against noise and sensitive when texture changes. Indeed, the de-noising methods as well as the de-blurring methods with strength classification add some details which do not exist in clean images therefore the textures change. If the strength is unknown, the results of the de-blurring and de-noising remain less sharp.

Moreover, as soon as the category of filter is known, the most appropriate recognition algorithm which is the most robust against a filter (the lowest curves in all Figures) can be chosen. In our case, we will select HoG after de-blackener, see Figure 5, and Eigen for the other filters after image restoration, see Figure 7, 9, 11. Hence, we obtain similar rates of face recognition to the clean face images. The proposed categorization of filters and estimation of the filter strength help to tune image restoration methods. Therefore, if privacy is protected by blackener, pixelization, blurring or noising filters, our proposed method can be applied and therefore privacy is no longer effective.

Conclusions and Future work

We have designed a framework which enables, in a first step, to detect the presence of a privacy filter and in the second step, to classify the type of filter (blackener, a pixelization, blurring and noising) and its strength . Using an appropriate tool (de-blackener, superResolution, de-blurring, de-noising), we can reverse the process and simulations show that the performances of face recognition are closer than the ones which are obtained for the clean faces. Hence privacy of people can be revealed and is no longer protected.

Videos are more and more popular and protection can be removed more efficiently using several frames from the same person. It could be interesting to apply our framework on videos which are more crucial for privacy concerns than still images.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[2] D. P. Cattin. Image restoration: Introduction to signal and image processing. *MIAC, University of Basel. Retrieved*, 11, 2013.

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-181

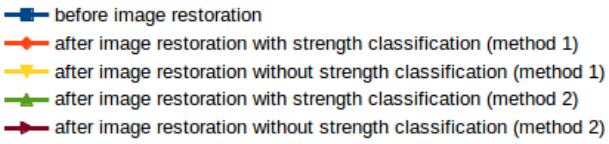


Figure 4: Legend for Figures 5, 7, 9, 11. The difference between rate of good recognition on clean faces and obscured faces.

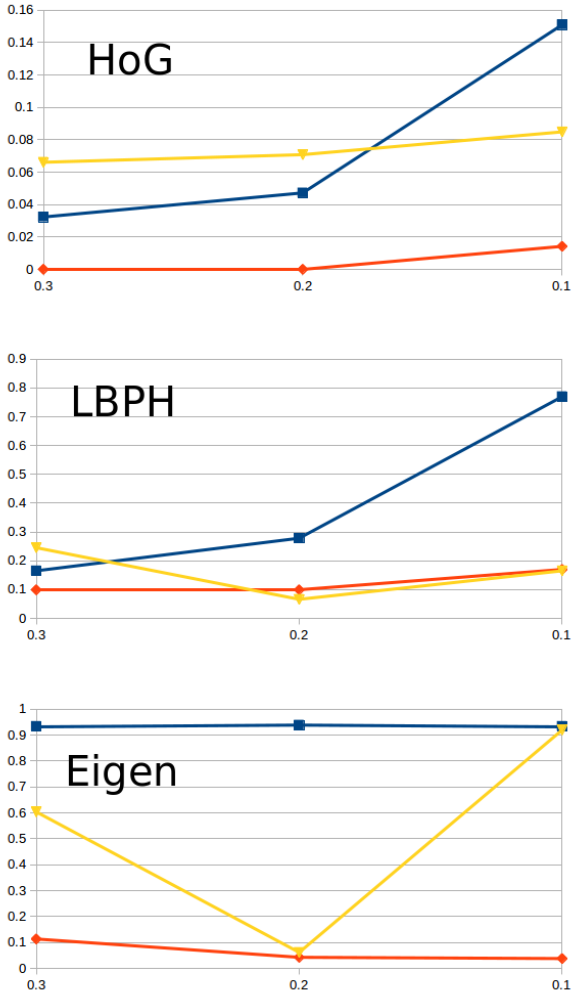


Figure 5: Impact of the de-blackener method depending on different opacity.

Figure 6: Respectively the Clean face image (first picture), Mask face image with $\alpha = 0.9$ (second picture), de-blackener without (third picture) and with (last picture) strength classification.

[3] B. Chalmond. Psf estimation for image deblurring. *CVGIP: Graphical Models and Image Processing*, 53(4):364–372, 1991.

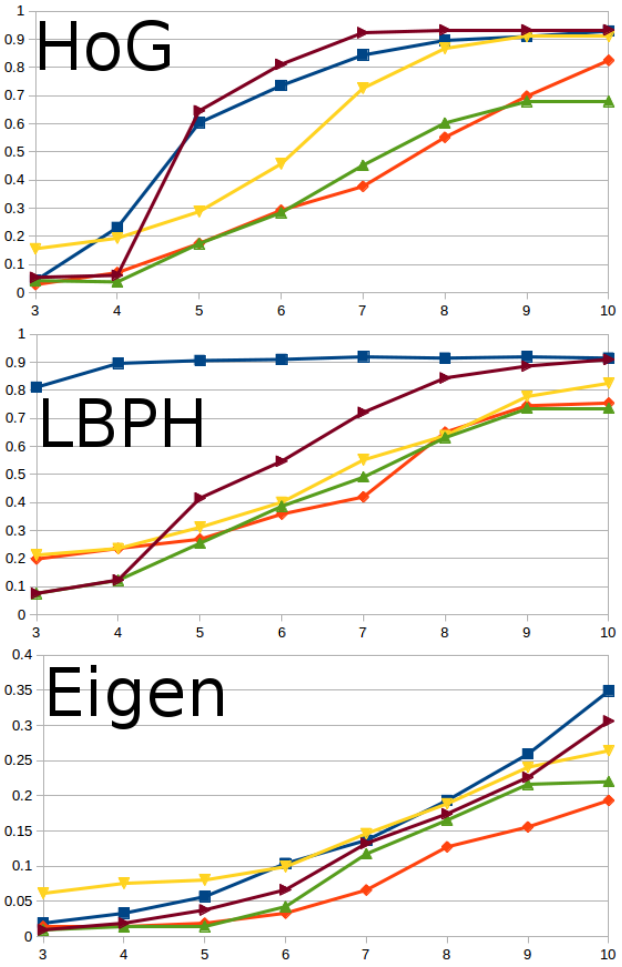
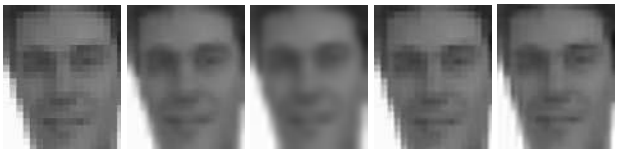


Figure 7: Impact of the superResolution methods depending on different size of squares.



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

Figure 8: Respectively Pix face image with size of squares = 5 (a), superResolution without (b) and with (c) strength classification for the first method, superResolution without (d) and with (e) strength classification for the second method.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[5] 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.

[6] W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *Image Processing, IEEE Transactions on*, 20(7):1838–1857, 2011.

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-181

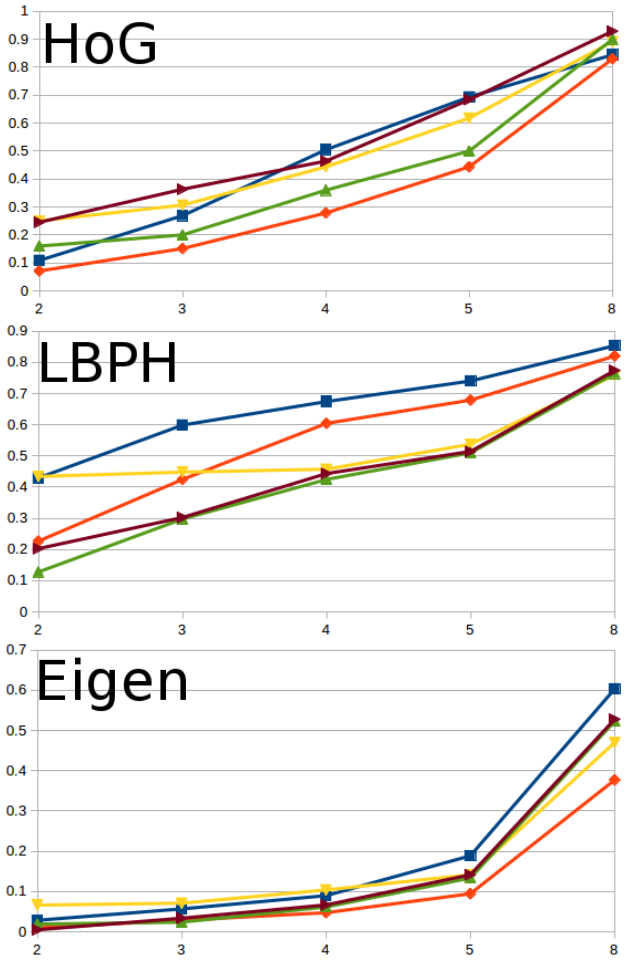


Figure 9: Impact of the de-blurring methods depending on different standard deviation.

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

Figure 10: Respectively Blurred face image with $\sigma = 2$ (a), de-blurring without (b) and with (c) strength classification for the first method, de-blurring without (d) and with (e) strength classification for the second method.

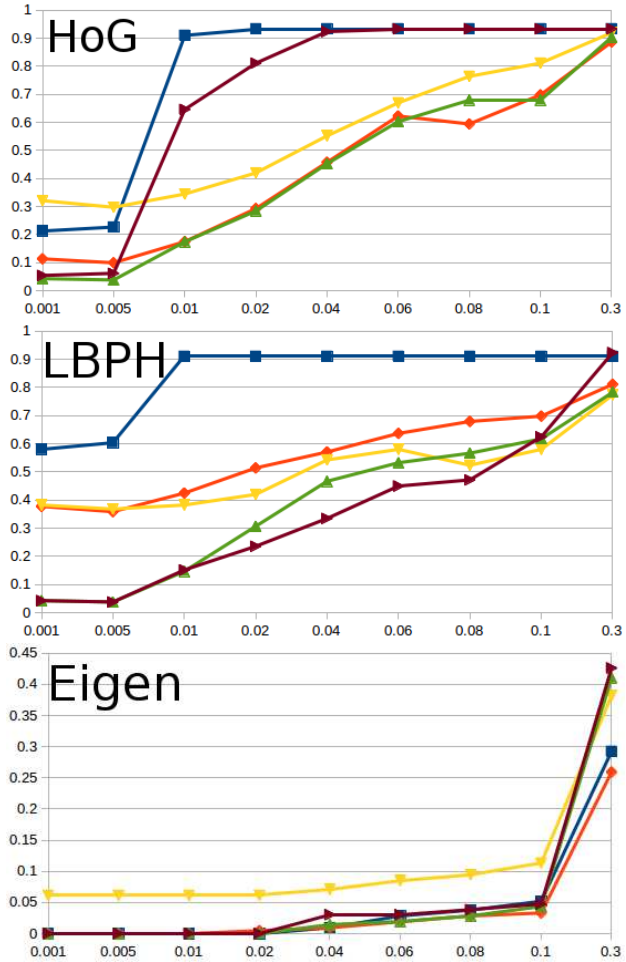


Figure 11: Impact of the de-noising methods depending on different standard deviation.

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

Figure 12: Respectively Noise face image with $\sigma = 0.01$ (a), de-noising without (b) and with (c) strength classification for the first method, de-noising without (d) and with (e) strength classification for the second method.

[7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.

[8] M. Forouzanfar. This work has been published by scitech publishing in" principles of waveform diversity and design" book available at http://www. scitechpub. com/wdd/. please cite this book chapter as follows: M. forouzanfar and h. abrishami-moghaddam, ultrasound speckle reduction in the complex wavelet domain, in principles of waveform diversity and design, m. wicks, e. mokole, s. blunt, r. schneible, and v.

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

[10] P. Hedaoo and S. S. Godbole. Wavelet thresholding approach for image denoising. *International Journal of Network Security & Its Applications*, 3(4):16–21, 2011.

[11] H. Huang and H. He. Super-resolution method for face recognition using nonlinear mappings on coherent features. *Neural Networks, IEEE Transactions on*, 22(1):121–130, 2011.

[12] R. Keys. Cubic convolution interpolation for digital image processing. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(6):1153–1160, 1981.

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-181

[13] P. Korshunov and T. Ebrahimi. Towards optimal distortion-based visual privacy filters. In *IEEE International Conference on Image Processing*, number EPFL-CONF-197087, 2014.

[14] P. Korshunov, A. Melle, J.-L. Dugelay, and T. Ebrahimi. Framework for objective evaluation of privacy filters. In *SPIE Optical Engineering+ Applications*, pages 88560T–88560T. International Society for Optics and Photonics, 2013.

[15] J. Kotera, F. Šroubek, and P. Milanfar. Blind deconvolution using alternating maximum a posteriori estimation with heavy-tailed priors. In *Computer Analysis of Images and Patterns*, pages 59–66. Springer, 2013.

[16] J. S. Lim. Two-dimensional signal and image processing. *Englewood Cliffs, NJ, Prentice Hall, 1990, 710 p.*, 1, 1990.

[17] M. A. Mayer, A. Borsdorf, M. Wagner, J. Hornegger, C. Y. Mardin, and R. P. Tornow. Wavelet denoising of multiframe optical coherence tomography data. *Biomedical optics express*, 3(3):572–589, 2012.

[18] A. Melle and J.-L. Dugelay. Scrambling faces for privacy protection using background self-similarities. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 6046–6050. IEEE, 2014.

[19] M. Nixon. *Feature extraction & image processing*. Academic Press, 2008.

[20] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.

[21] 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.

[22] S. D. Ruikar and D. D. Doye. Wavelet based image denoising technique. *IJACSA) International Journal of Advanced Computer Science and Applications*, 2(3), 2011.

[23] F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pages 138–142. IEEE, 1994.

[24] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *ACM Transactions on Graphics (TOG)*, volume 27, page 73. ACM, 2008.

[25] 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.

[26] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

Author Biography

Natacha Ruchaud *obtained her engineering degree in Vision, Image and Multimedia from Polytech Nice (2014). Since then she is a PhD student in the Multimedia Department at Eurecom, Sophia-Antipolis. Her work focuses on the privacy preserving biometrics and security in video Surveillance but she has many interests for identity, gender and action recognition.*

Professor Jean-Luc DUGELAY *obtained his PhD in Information Technology from the University of Rennes in 1992. His thesis work was undertaken at CCETT (France Télécom Research) at Rennes between 1989 and 1992. He then joined EURECOM in Sophia Antipolis where he is now a Professor in the Department of Multimedia Communications. His current work focuses in the domain of multimedia image processing, in particular activities in security (image forensics, biometrics and video surveillance, mini drones), and facial image processing.*