Using Face Morphing to Protect Privacy

Pavel Korshunov
Multimedia Signal Processing Group
EPFL, Lausanne, Switzerland
pavel.korshunov@epfl.ch

Touradj Ebrahimi
Multimedia Signal Processing Group
EPFL, Lausanne, Switzerland
touradj.ebrahimi@epfl.ch

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. To overcome these shortcomings, in this paper, we propose a morphing-based privacy protection method and focus on its robustness, reversibility, and security properties. We morph faces from a standard FERET dataset and run face detection and recognition algorithms on the resulted images to demonstrate that morphed faces retain the likeness of a face, while making them unrecognizable, which ensures the protection of privacy. Our experiments also demonstrate the influence of morphing strength on robustness and security. We also show how to determine the right parameters of the method.

1. Introduction

Protection of visual privacy in video surveillance is an important and challenging task. Although many different privacy protection methods are available, none has all the following desired properties: (i) reversibility (possibility to undo protection on request from authorities), (ii) flexibility of application (independent of compression and video or image data format), (iii) robustness (high level of distortion to render images unrecognizable), (iv) security (recovery of the original data using a secret key) and resistance to malicious attacks, and (v) variable strength granularity (flexibility to protect data with different degrees of strength). Simple methods like blurring, pixelization, and masking are not reversible and insecure; encryption-based methods, such as proposed in [3, 14, 5] are secure but remove original pixel data, are opaque and fragile (change in one pixel destroys encryption); scrambling [11, 7] has advantages of encryption, while being robust to compression, but it is dependent on video or image compression; and anonymization methods like in [13] are often complex and require original data to be stored separately.

To overcome these shortcomings, we propose to use morphing-based method for protection of visual privacy, which, we believe, combines all the desired properties of a visual privacy protection method with no notable shortcomings. When conventional morphing is applied, pixels are interpolated into corresponding locations of the target image and their intensities are replaced with the corresponding pixels in the target image. In the proposed method, we perform only partial randomized morphing in both dimensions: interpolation and intensity. That means the resulted morphed image will be interpolated somewhere in-between original source and target images and intensities of the resulted pixels will be weighted accordingly, thus destroying original specific visual details in the image. Hence, the balance between privacy protection and surveillance task is transformed into the strength value by how much the pixels in an image are morphed. In the proposed morphing algorithm, a set of key points are determined in both original source and target images and a correspondence between these key points is established. Morphing is performed by dividing both images in triangles using Delaunay triangulation [2].

Since morphing is applied to pixel data, it is independent of compression methods. A security can be insured by using a secret key for seeding a pseudo-random algorithm, which is used for randomizing interpolation and intensity weights, as well as, for encryption of the key points used for triangulation. Recovering original image from the morphed one can be done by applying the inverse morphing transformation, given that the

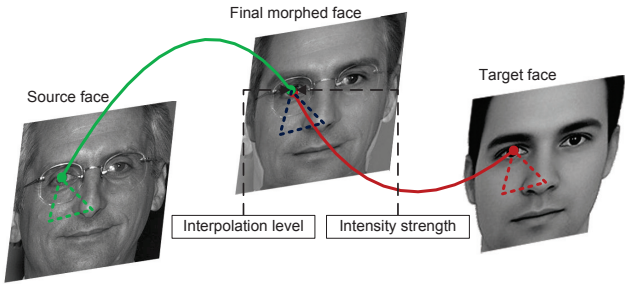


Figure 1: Overview of the proposed method.

target image, key points, and interpolation and intensity values are known.

We demonstrate feasibility of the proposed morphing method by applying it to faces of a standard FERET face dataset [10], since faces are among the most privacy sensitive regions. Location of each face is first detected with Viola-Jones [12] face detection algorithm. A set of key points, used as triangle vertices in the morphing transformation, is constructed from automatically detected eyes, nose, and mouth. To determine reversibility, robustness, and security of the method, we use Eigenfaces, Fisherfaces [1], and local binary patterns histograms (LBPH) [6] based face recognition algorithms. The recognition algorithms were run on the morphed and recovered faces to determine the efficiency of the proposed visual privacy protection tool. In an ideal scenario, a protected face would be visible as a face but would not be correctly identified by the recognition algorithm.

2. Morphing based privacy protection

In this section, we describe visual privacy protection method based on morphing. To demonstrate how such privacy protection works, we assume face to be a sensitive region to which the protection is applied. The following is the summary, illustrated in Figure 1, of morphing based privacy protection method:

- Automatically select key points in both original source and target (e.g., a standard human face) images by using face, eyes, nose, and mouth detections;
- For each pair of the corresponding points in two images determine some point in between, by using a given *level of interpolation* value;
- Divide images using Delaunay triangulation [2] with determined points as vertices of the triangles;
- Find coordinates of pixels in the final image by interpolating both source and target images according to each corresponding triangle;
- For each pixel in the final image, compute its intensity as a weighted sum of intensities between corresponding pixels in original source image and target image. Weights are determined by a given *intensity strength* value such as in

$$I_f = (1 - w_i)I_s + w_i I_t, \tag{1}$$

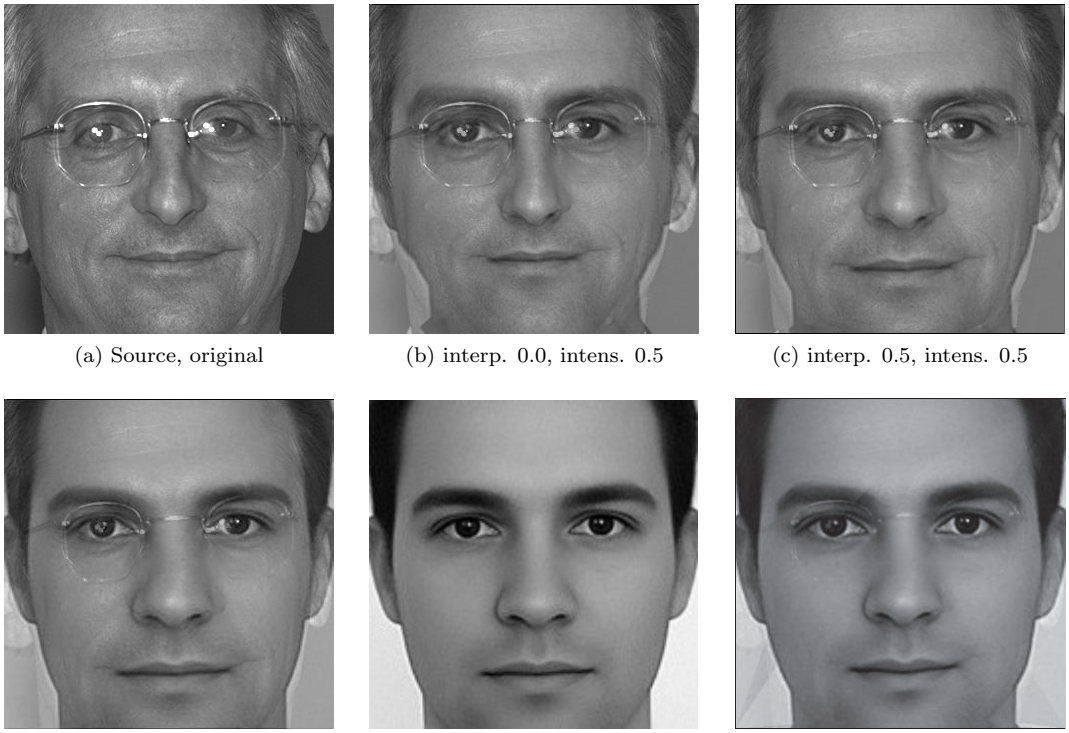
where I_f, I_s, I_t are final morphed, original source, and target faces respectively, and w_i is the intensity strength value.

As can be noticed form the algorithm's summary and Figure 1, two main values determine the final morphed face: interpolation level and intensity strength. Figure 2 demonstrate the effect of different such values on the morphed image. When interpolation level and intensity strength are zero, the resulted face is the same as the original, but the closer these values are to one, the more the resulted face looks like the target face. In our demonstrations and experiment, an average male face [4] was chosen as the target face, but, in practical applications, it can be any other face or facial avatar.

In a surveillance scenario, when a protected face needs to be recovered, an inverse of morphing operation, termed *unmorphing*, is applied. For the recovery of the original source face, the key points and the target face need to be known. The recovery algorithm is essentially the same as morphing, instead we simply estimate a starting face (source) by using known the 'middle' (morphed) and the end (target) faces (see 1).

This morphing-based visual privacy protection method is designed to overcome common shortcomings of other privacy protection techniques. Since morphing is simply a geometrical transformation of pixels, with pixels interpolated into weighted sum of known intensities, it is compression independent, as opposed to scrambling, while retaining the main features of the morphed region (such as face), as opposed to encryption privacy protection methods. Security of the proposed method can be ensured by encrypting the key points (the vertices of Dalaunay triangles) of the morphing algorithm and randomizing interpolation level and intensity strength values for each morphed triangle (see Figure 2f for illustration), as we discuss in more details in Section 2.2.

We use standard FERET dataset [10] (a subset of 100 faces) with provided ground truth for testing the proposed morphing-based privacy protection. Morphing was applied to faces in the dataset, which were detected with Viola-Jones face detection [12] algorithm. For vertices of Delaunay triangles, 18 key points were automatically selected based on the detected eyes (5



(a) Source, original (b) interp. 0.0, intens. 0.5 (c) interp. 0.5, intens. 0.5

(d) interp. 1.0, intens. 0.7 (e) Target, original (f) interp. 0.6-1.0, intens. 0.7-0.9

Figure 2: Examples of morphing the same face with different interpolation levels and intensity strengths.

points for each eye), nose (3 points), and mouth (5 points) in a face. To evaluate the robustness and security of the proposed algorithm, we run OpenCV[1] implementations of Viola-Jones face detection and three face recognition algorithms, which are based on Eigenfaces, Fisherfaces [1], and LBPH [6], on the morphed faces and recovered faces. We use gallery ('fa' set of faces) and probe ('fb' set of faces) images provided by FERET dataset for the recognition task.

2.1. Robustness and reversibility

The aim of the experiments is to determine whether the morphing preserves the generic features of a face, in which case the accuracy of a face detection algorithm would not be affected by morphing, i.e., faces would be detected as for original images, while the specific personal features are distorted, which would lead to the significant decrease in recognition accuracy.

To understand the effect of interpolation level and intensity strength values on the recognition accuracy, we vary these values from 0 to 1 with step 0.1, and for each pair of values, we run recognition algorithms on the resulted morphed faces from FERET dataset. Figure 3 demonstrates the resulted tradeoff for Fisherfaces-based approach (for other recognition algorithms the results are similar but omitted here due to space limitations). In the figure, vertical axis reflects the accuracy of recognition computed as rank one of cumulative match characteristic (CMC) [8], which is a standard measure of accuracy for identification task of recognition. Horizontal axis of the figure shows changes in intensity strength, while each plotted curve corresponds to different interpolation levels. Recognition accuracy corresponding to interpolation level and intensity strength values 0 is essentially the same as for original non-morphed images.

Figure 3 demonstrates a significant drop (until almost 0) in accuracy for higher values of intensity strength and interpolation level. It can be also noted that intensity strength affects recognition accuracy more significantly. This figure can help in determining values of these parameters in practical implementations.

To demonstrate reversibility of the morphing-based approach, we conducted similar experiment for recovered faces and plot corresponding recognition results

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

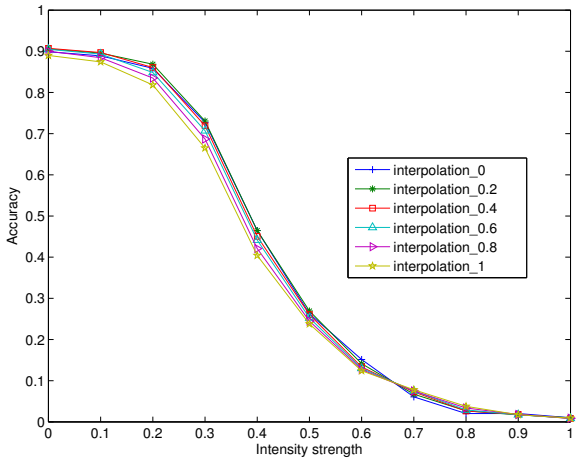


Figure 3: Accuracy of Fisherface recognition for morphed faces with different interpolation level and intensity strength values.

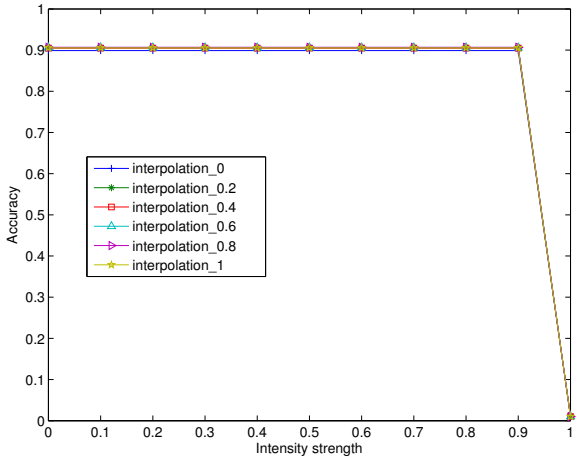


Figure 4: Accuracy of Fisherface recognition on faces recovered from morphed with different interpolation level and intensity strength values.

in Figure 4. In this figure, recovered faces were obtained by unmorphing the faces morphed for previous experiments. The figure demonstrates that recognition accuracy does not change compared to the accuracy for original images until the intensity strength value becomes 1.0. With intensity strength 1.0, the end-resulted morphed face does not contain even a fraction of intensities from the original source face (see Equation 1), and, therefore, the intensity values of the original face cannot be recovered from the morphed one at all. Hence, in practical applications, intensity strength of morphing-based privacy protection should always be strictly less than 1.0.

Figure 3 and Figure 4 show that morphing-based distortion of faces significantly decreases accuracy of recognition, while being reversible in most cases (except for intensity strength 1.0). These results are very promising, especially since it was previously shown that recognition algorithms perform well even in cases of significant distortions [9].

To understand the effect of morphing method on face detection, we also run Viola-Jones face detection algorithm on the images from FERET dataset with faces replaced by the morphed versions obtained by changing all combinations of interpolation level and intensity strength. Detection accuracy was the same for all morphed and recovered images, with exception when intensity strength is 1.0 for recovered images, as for the original images, which is 100% accuracy (FERET dataset is 'easy' for detection algorithm). There was only slighter increase in false positive (to 0.8%) for morphed faces. We did not plot these results to avoid cluttering the space with trivial plots.

2.2. Security

There are two ways to make the recovering of original faces from morphed with the proposed approach difficult: encrypt key points (vertices of Delaunay triangles) and randomize the choice of interpolation level and intensity strength values for each triangle. Since the number of key points is relatively small (we use 18 points in our experiments), their encrypted values brings little overhead to storage and can be embedded as an additional information inside an image, e.g., in JPEG image format, APPn markers can be used. Interpolation level and intensity strength can be random values generated in a certain interval by pseudo-random algorithm with a secret key as its seed. Figure 2f shows an example of such randomized morphing. To recover original faces, key points, target image, morphed image, and a series of interpolation level and intensity strength values are necessary to perform the same unmorphing operation as for non-randomized method.

Experiments in the previous section show that interpolation level and intensity strength values need to be properly selected for morphing to render faces unrecognizable. By looking on Figure 3, we can estimate intervals $0.5 - 1.0$ for interpolation level and $0.5 - 0.9$ for intensity strength as appropriate. These intervals should ensure that morphing-based protection will be robust, i.e., lead to low recognition for morphed and high recognition for recovered faces.

To validate our reasoning, we run recognition algorithms on randomized faces when different intervals for interpolation level and intensity strength values are used. The results for morphed and corresponding re-

Table 1: Recognition accuracy for randomized morphed/recovered faces, FERET dataset.

Parameters	**Eigen (morphed)**	**Eigen (recovered)**	**Fisher (morphed)**	**Fisher (recovered)**	**LPBH (morphed)**	**LPBH (recovered)**
Interp: 0.0-1.0, Intens: 0.0-0.9	0.38	0.91	0.39	0.91	0.36	0.90
Interp: 0.5-1.0, Intens: 0.5-0.9	0.07	0.90	0.07	0.90	0.08	0.90
Interp: 0.5-1.0, Intens: 0.6-0.9	0.08	0.90	0.07	0.90	0.04	0.88
Interp: 0.5-1.0, Intens: 0.7-0.9	0.03	0.90	0.03	0.91	0.04	0.88
Interp: 0.6-1.0, Intens: 0.5-0.9	0.06	0.91	0.06	0.91	0.08	0.91
Interp: 0.6-1.0, Intens: 0.6-0.9	0.06	0.91	0.06	0.91	0.05	0.88
Interp: 0.6-1.0, Intens: 0.7-0.9	**0.02**	**0.91**	**0.03**	**0.91**	**0.02**	**0.91**
Interp: 0.7-1.0, Intens: 0.5-0.9	0.08	0.91	0.08	0.91	0.05	0.91
Interp: 0.7-1.0, Intens: 0.6-0.9	0.06	0.90	0.05	0.90	0.06	0.87
Interp: 0.7-1.0, Intens: 0.7-0.9	0.04	0.91	0.04	0.91	0.03	0.89

Table 2: Recognition accuracy for faces unmorphed (with assumed interpolation 0.5 and intensity 0.5) from randomized morphed faces, FERET dataset. Illustration of a possible attack on the faces morphed with our method.

Parameters	**Eigen**	**Fisher**	**LPBH**
Interp: 0.5-1.0, Intens: 0.5-0.9	0.31	0.31	0.17
Interp: 0.5-1.0, Intens: 0.6-0.9	0.16	0.15	0.11
Interp: 0.5-1.0, Intens: 0.7-0.9	0.10	0.10	0.10
Interp: 0.6-1.0, Intens: 0.5-0.9	0.23	0.22	0.20
Interp: 0.6-1.0, Intens: 0.6-0.9	0.17	0.16	0.14
Interp: 0.6-1.0, Intens: 0.7-0.9	**0.07**	**0.07**	**0.09**
Interp: 0.7-1.0, Intens: 0.5-0.9	0.29	0.29	0.13
Interp: 0.7-1.0, Intens: 0.6-0.9	0.13	0.13	0.12
Interp: 0.7-1.0, Intens: 0.7-0.9	0.06	0.06	0.09

covered faces are shown in Table 1 for three recognition algorithms. First line of the table shows recognition when the most liberal intervals are used with interpolation level in $0.0-1.0$ and intensity strength in $0.0-0.9$. Recognition for morphed images is not low enough in this case. However, when we restrict intervals to $0.5-1.0$ ranges, recognition accuracy for morphed faces drops to acceptable low levels, while recognition of recovered faces remain high. The best pair of intervals, however, would be the one that leads to the highest gap between recognition accuracies for recovered faces and for morphed faces. This pair of intervals is highlighted in the table with bold.

As well known in cryptography, the security of the algorithm should depend only on secrecy of the private key. That means, when evaluating the security of our approach, we should assume that a potential malicious attacker knows both the morphing algorithm and the target face. In such case, with what accuracy can attacker recover the original faces from the randomized morphed versions? Key points can be estimated from morphed face by using the same automated algorithm that we use in our morphing method (based on automatically detected eyes, nose, and mouth locations). Instead of using random values for interpolation level and intensity strength, one can assume a fixed reason-

able value, for example 0.5, which is a mid point between 0 and 1. Using these assumptions, we run unmorphing algorithm on morphed images obtained using different randomization intervals. The recognition accuracies for these unmorphed faces are presented in Table 2. These results show that intensity strength is significantly more important parameter than interpolation level, as was first noted in our robustness findings in Section 2.1, and it should be chosen with care in practice. The table also demonstrates that the method is the best in withstanding the attack when using the pair of intervals $0.6 - 1.0$ for interpolation level and $0.7 - 0.9$ for intensity strength, as also chosen in Table 1.

3. Conclusion and future work

In this paper, we proposed a new morphing-based visual privacy protection algorithm for video surveillance. By using recognition and detection algorithms, we demonstrated that the proposed method keeps the general facial features intact, thus allowing undistracted conduct of surveillance task, while distorting the specific facial information rendering faces equally unrecognizable by three different recognition algorithm, this ensuring protection of privacy. The experimental results shown that the method has high reversibility, robustness, and security.

We also demonstrated that interpolation level has little influence on the robustness of the morphing based privacy protection. This finding may imply that purely geometrical-based methods that do not significantly change intensities of pixels in the image may not be very suitable for privacy protection. On the other hand, the choice of intensity strength value is crucial, and should be carefully chosen in practical implementations, for the proposed method to be robust and resistant to security attacks.

These results motivate us to perform subjective evaluations of the proposed morphing-based privacy protection using human subjects to see whether the subjective recognition aligns with obtained objective results. We would also like to test how resilient this method is to non-malicious attacks, such as compression.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.

[2] P. J. Benson. Morph transformation of the facial image. *Image and Vision Computing*, 12(10):691–696, 1994.

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

[4] C. Braun, M. Gruendl, C. Marberger, and C. Scherber. Beautycheck – ursachen und folgen von attraktivitaet. Technical report, Universitt Regensburg, Universitt Regensburg, 93040 Regensburg, Germany, 2001.

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

[6] C.-H. Chan, J. Kittler, and K. Messer. Multi-scale local binary pattern histograms for face recognition. In S.-W. Lee and S. Li, editors, *Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 809–818. Springer Berlin Heidelberg, 2007.

[7] F. Dufaux and T. Ebrahimi. Scrambling for privacy protection in video surveillance systems. *IEEE Trans. on Circuits and Systems for Video Technology*, 18(8):1168–1174, Aug 2008.

[8] 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*, pages 937–945, Guildford, UK, June 2003.

[9] P. Korshunov and W. T. Ooi. Video quality for face detection, recognition, and tracking. *ACM Trans. Multimedia Comput. Commun. Appl.*, 7(3):14:1–14:21, Sept. 2011.

[10] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104, Oct. 2000.

[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] 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*, volume 2, page 747, Vancouver, Canada, July 2001.

[13] 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*, pages 48–55, New York, NY, USA, 2004. ACM.

[14] 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)*, pages 593 –600, Aug. 29-Sep. 1 2010.