



FACE DE-IDENTIFICATION WITH EXPRESSIONS PRESERVATION

Geoffrey Letournel, Aurélie Bugeau, Vinh-Thong Ta, Jean-Philippe Domenger

► To cite this version:

Geoffrey Letournel, Aurélie Bugeau, Vinh-Thong Ta, Jean-Philippe Domenger. FACE DE-IDENTIFICATION WITH EXPRESSIONS PRESERVATION. International Conference on Image Processing (ICIP) 2015, Sep 2015, Quebec, Canada. hal-01187654

HAL Id: hal-01187654

<https://hal.science/hal-01187654>

Submitted on 27 Aug 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FACE DE-IDENTIFICATION WITH EXPRESSIONS PRESERVATION

G. Letournel^{1,3}, A. Bugeau^{1,3}, V.-T. Ta^{2,3}, J.-P. Domenger^{1,3}

¹ Univ. Bordeaux, LaBRI, PICTURA, UMR 5800, F-33400 Talence, France.

² Bordeaux INP, LaBRI, PICTURA, UMR 5800, F-33400 Talence, France.

³ CNRS, LaBRI, PICTURA, UMR 5800, F-33400 Talence, France.

ABSTRACT

This paper addresses an application that has not been much explored, the de-identification of faces with expressions preservation in images. With the huge amount of images and videos shared on the Internet, protecting the identity of people in the data becomes crucial. Removing the identity information is often referred as de-identification. In this paper, we propose a novel de-identification process that preserves the important clues on the face for further behavior or emotions analysis. It is a difficult problem because obtaining the anonymity implies deteriorating the main face components. At the opposite, analyzing the expressions requires keeping enough information on the face such as, for instance, the gaze or the corners of the lips. Our approach relies on face and key points detection, followed by a variational adaptive filtering. Experimental results show the potential of the proposed method and open new insights for image dissemination or video broadcasting.

Index Terms—De-identification, Facial Expressions Recognition, Adaptive Filtering

1. INTRODUCTION

The dissemination of images is conditioned by the preservation of the privacy of each person. The anonymity is usually obtained by adding blur on the face or with a technique of pixelization. Examples include television reports, Google StreetView service and surveillance system in a medical context. This de-identification problem has been well described in the literature (we refer the interested reader to [1]).

A naive degradation of the face removes all the signs of expressivity and prevents any further content analysis. In order to maintain the possibility to analyze the data, our de-identification system must preserve not only the seven primary expressions but, for instance, the gaze and the corner of the lips.

This study has been carried out in the frame of “the Investments for the futur” Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02). The authors would like to thank Maria Catherina Manes Gallo (TELEM) and Veronique Lespinet-Najib (ENSC) for the useful discussions regarding quantitative validations.

De-identification with expressions preservation has a real interest in the study of communicative processes between people from audio-visual data and has many applications. For instance, in marketing, the de-identification would enable studying the behavior of a person in supermarkets. In the context of medical or law education it would provide a tool to disseminate real data and propose concrete case studies illustrating the theoretical training. In the context of medical care, it could help monitor patients at risk. In the context of surveillance, the use of cameras in public places create serious problem regarding privacy. Nevertheless, the semantic content must be freely distributed [2].

Main contributions. In this paper, we address the following ambiguous problem: how to de-identify a person while keeping the most informative data in the face? The main contributions are (i) We address the unexplored field of research of face de-identification with facial expressions preservation. (ii) We propose a new automatic method based on a variational adaptive filter including a total variation term. This method does not impose any constraints such as the poses, the lighting conditions or, the background of the considered images. (iii) We analyze qualitatively and quantitatively our results.

2. RELATED WORKS

2.1. Face Identification and Expressions Recognition

Several studies on face recognition by the human visual system in cognitive sciences (including psychology) have been conducted. They show that the identification is first performed by scanning horizontal lines (eyebrows, eyes and lips) and further, the shadows and lights parts (*e.g.*, nose, cheeks or forehead) [3, 4]. [5] shows that the eyebrows mainly play the role of markers for estimating the distances between the different parts of a face while the eyes are truly discriminating areas for recognition. To confirm the importance of the eyes, the mouth and the nose, studies show that other face regions, such as the chin, forehead and neck [6] are not important in the identification process.

One of the first method that enables the computer to identify faces was based on Eigenfaces [7] that uses a set of eigen-

vectors to decompose a face and compares it to others. Most recent face recognition systems can be decomposed into three steps [8]: face detection (*e.g.*, [9]), features extraction (*e.g.*, [10][11][12]), identification or verification (*e.g.*, [8][10]).

In the literature, facial expressions are classified into seven categories (neutral, happiness, sadness, surprise, anger, disgust and fear). The most used model to watch emotions is the FACS (Facial Action Coding System) [13].

Cognitive studies focussing on facial expressions show that the most commonly used face structures are placed on the eyebrows, the eyes and the mouth. Structures such as the cheeks, chin and the nose are not used [14].

2.2. Face De-identification

Naive de-identification methods are based on a deteriorating of the image by blur, pixelization methods or, with a deletion of part of the face.

Figure 1 shows an example of such methods. The facial clues contributing to emotions analysis are destroyed. More sophisticated methods try to enhance facial features (*e.g.* face edges), to change the color of the skin by color inversion or, to use cartoons methods [15]. These approaches can guaranty (partially) the anonymity but the expressions analysis become extremely difficult. For instance, when looking at a face with invert colors, the direction of the eyes seems to have changed. Moreover that de-identification is not efficient enough, the initial data can be easily recovered.



Fig. 1. De-identification without expressions preservation on an image from the GTAV database [16]

In [17], the authors propose to use the facial expression recognition to transfer a person’s face to the target one. Face transfer was also proposed in [18] based on speech-related mouth articulations, facial expressions and 3D pose from monocular video or film footage. An extension of this method was designed in [19]. It includes a coarse spatio-temporal synchronization between source and target videos. Another recent work in this trend concerns the face re-enactment [20].

The works more related to our objective are [21, 1]. The *k*-Same framework [21] finds the similarity between faces in a database and, creates a new face by averaging image components. In order to avoid alignment errors, its extension *k*-Same-M fits an active appearance model to the images and, applies *k*-Same on this model parameters. Both methods require each subject to be only present once in the database. The same authors proposed further [1] to use multi-factor models to solve this limitation. The results are very interesting but

averaging the clues of different people in the final face reconstruction leads to misinterpretations in the attitude analysis. The clues of one individual are indeed mixed with other ones (see Figure 5).

3. DE-IDENTIFICATION WITH EXPRESSIONS PRESERVATION

From the observations of the previous section, we propose a novel method, that de-identifies a person by adaptively blurring all the face structures (in particular the nose) while preserving the communicative clues of the face, *i.e.*, the eyes, the eyebrows and the mouth. The complete proposed pipeline is summarized in Figure 2.

3.1. Face Detection and Facial Landmarks Localization

The face can not be entirely deteriorated. Some areas must be preserved in order to keep the facial expressions. Hence, the first step of our pipeline consists in detecting the face and the regions of interest such as the eyes, the mouth or the nose. We use the recent method of [9] that detects the face and its landmarks simultaneously. The face landmarks are connected to each others, if they are in the same region of interest, *e.g.*, the eyebrows or the mouth. Figure 2 shows a result where the red “+” represents the facial landmarks detected.

The face is the most recognizable part of the body but hair also plays an important role in the identification process. In order to make sure that the entire head is processed, we automatically extract a circle around it by using the estimated pose and position of the nose and the chin.

3.2. Variational Adaptive Filtering

In the following, we present the proposed variational adaptive filtering, based on a total variation (*TV*) regularization and a distance map computation.

Total Variation based Functional. Let Ω be the set of all the pixels inside the circle previously extracted. We denote $u_0 : \Omega \mapsto U = [0, 255]^3$ the initial image and $u : \Omega \mapsto U$ the result of our adaptive filtering. Both images are defined in the RGB color space ($u = (R, G, B)$). In this work, we propose to regularize the pixels in Ω with a *TV* term by minimizing the following functional:

$$J(u) = TV(u) + \frac{\lambda}{2} \int_{\Omega} \|u - u_0\|^2 \quad (1)$$

The parameter λ controls the trade-off between the regularization term and the data fidelity term. The spatial *TV* is defined as: $TV(u) = \int_{\Omega} \sqrt{\|\partial_x u\|^2 + \|\partial_y u\|^2}$

The advantage of a *TV* regularization is to preserve image discontinuities, *i.e.*, the face contours. It avoids blurring the important contours that must be preserved but destroys small distinctive signs such as wrinkles or beauty spots. Bilateral

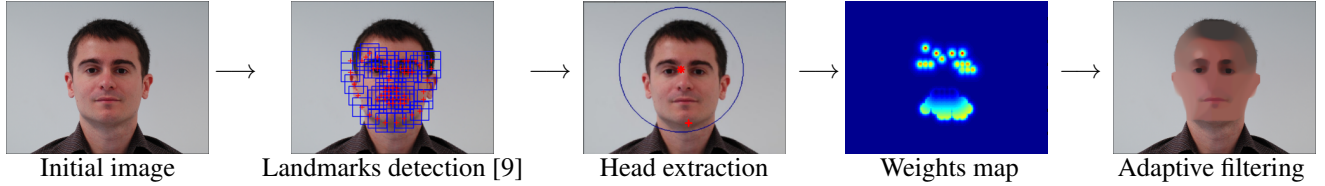


Fig. 2. De-identification with expressions preservation pipeline

filtering could have been used as well but recasting our problem in term of an energy minimization problem may enable adding more constraints and more terms in the future.

Adaptive Filtering. In order to preserve the information on the most important clues, we modify Equation (1) and, integrate an adaptive λ for each pixel $\mathbf{x} = (x, y) \in \Omega$. The functional (1) becomes:

$$J_s(u) = \int_{\Omega} |\nabla u(\mathbf{x})| d\mathbf{x} + \frac{1}{2} \int_{\Omega} \gamma(\mathbf{x}) \|u(\mathbf{x}) - u_0(\mathbf{x})\|^2 d\mathbf{x} \quad (2)$$

where the weight function $\gamma : \Omega \rightarrow \mathbb{R}^+$ adapts the filtering to the detected region of interest on the face. The eyes, the eyebrows and the mouth need to be preserved for emotion analysis while the nose can be suppressed for face de-identification as it is not an important emotional clue. The positions of these three clues are directly obtained from the landmarks of the face detector. Only the desired landmarks are retained.

Let $A = \{\text{mouth, eyes}\}$ and $B = \{\text{nose}\}$ be the regions of interest and, let γ_i with $i \in \{A, B\}$ be the associated weight. For the set A , we want the filtering to decrease when the pixels are close to the landmarks of this set. More specifically, we want $\gamma_i(\mathbf{x}) = \lambda_i$ to be maximal if \mathbf{x} is on a landmark of A (low regularization) and, to decrease exponentially when the pixels are far from this set (strong blurring). The Euclidean distance from a pixel $\mathbf{x} \in \Omega$ to the nearest landmark in each clue is denoted $D_i(\mathbf{x})$. For the set B , at the opposite, the regularization must be strong on the nose and, smaller for distant pixels. The adaptive weights are defined as:

$$\forall \mathbf{x} \in \Omega \quad \gamma_i(\mathbf{x}) = \begin{cases} \lambda_i \exp\left(-\frac{D_i(\mathbf{x})}{\sigma_i}\right) & \forall i \in A \\ 1 - \exp\left(-\frac{D_i(\mathbf{x})}{\sigma_i}\right) & \forall i \in B \end{cases}.$$

Figure 2 shows an example of weight map. The final weight function in (2) is defined as:

$$\forall \mathbf{x} \in \Omega \quad \gamma(\mathbf{x}) = \min \left[(\gamma_{\text{nose}}(\mathbf{x}), \max(\gamma_{\text{mouth}}(\mathbf{x}), \gamma_{\text{eyes}}(\mathbf{x}))) \right].$$

The functional (2) is minimized using the primal-dual algorithm defined in [22].

4. EXPERIMENTAL RESULTS

This section presents qualitative and, human and software-based quantitative results.

We have set experimentally $\sigma_i = 1.7$ for $i \in A$ and $\sigma_i = 0.2$ for $i \in B$, $\lambda_{\text{eyes}} = 1$ and $\lambda_{\text{mouth}} = 0.5$ for all our experiments.

4.1. Qualitative Analysis

To validate our method we use the Labeled Faces in the Wild (LFW) dataset [23]. LFW contains 13,233 face images of 5,749 different individuals, captured in unconstrained environments.

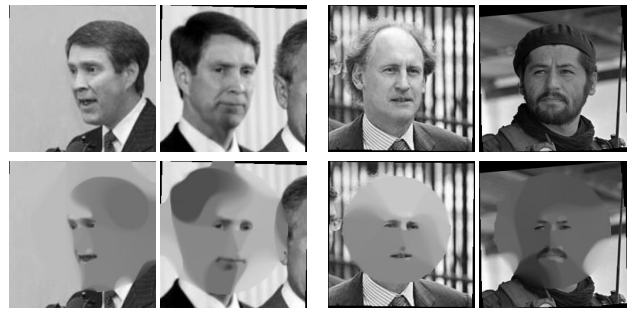


Fig. 3. Example images from the LFW dataset [23]. First row: initial images, second row: de-identified faces. At left: a match pair, at right: a mismatch pair.

On the first row of Figure 3 four initial images from the dataset are presented. The corresponding de-identified faces are shown on the second row. The expressions are well preserved. Even if the gaze direction and the corner of the lips are maintained, eyes and mouth are preserved, they were sufficiently degraded to de-identify the faces.



Fig. 5. A comparison with results taken from [1]

Figure 5 shows a qualitative comparison with k -Same [21] and $\text{MF}(\epsilon, k)$ [1] methods. The images were extracted from the associated paper [1]. While these two methods achieve an important decrease of the recognition rates, they are not able to correctly preserve all the signs of expression. For instance,



Fig. 4. Influence of the level of filtering on an image of the LFW dataset

in the first row images, the eyes are lost, and, in the second row images, the smile is lost, while, in the last column, they are well preserved with our method.

Figure 4 presents the resulting images when gradually decreasing the level of blur on the face by increasing λ_i and decreasing σ_i . This figure demonstrates that even with a strong filtering, our method is able to maintain the expressions on the face.

4.2. Quantitative Analysis

4.2.1. Software-based Analysis

To validate our de-identification method, a quantitative evaluation has been performed. The LFW contains pairs of faces that either match (the faces belong to the same person) or mismatch (they do not belong to the same person). Figure 3 shows an example of a *match* and a *mismatch*. Given a pair of images, the result of the face verification methods on this dataset aims at retrieving if this pair was a match or a mismatch. If they succeed for one pair, this pair is said to be detected as positive. We follow the restricted unsupervised framework [24], to evaluate our approach through a 10-fold cross validation process on the LFW aligned images supplied by Wolf et al. [25].

In [26], another algorithm dedicated to face de-identification but without any constraints on expressions preservation was also proposed. We follow their idea for testing our de-identification method based on the face verification method [11].

In the following, the initial pairs, the pairs with two de-identified images, and the images pairs with one initial and one de-identified will be denoted NA-NA (Non-Anonymous), A-A (Anonymous), and A-NA, respectively. When taking initial images (NA-NA), $85.4 \pm 1.7\%$ of the pairs in LFW are correctly detected as positive. Without retraining the classifier, we apply the testing procedure by replacing one initial image of each pair with a de-identified one leading to A-NA pairs. The recognition rate then falls to $56.4 \pm 1.6\%$.

We have also reproduced this experiment after replacing half of the training images with the corresponding de-identified images. The training set is now composed of A-A, A-NA and NA-NA pairs. The recognition rates after testing become: $80.8 \pm 1.5\%$ when testing with NA-NA pairs, $69.9 \pm 1.9\%$ with A-A pairs and, $67.9 \pm 2.0\%$ with A-NA

pairs. The result demonstrates that even when we enrich the learning dataset with de-identified images, our approach prevents sufficiently from recognizing a de-identified face.

4.2.2. Human Analysis

We also conducted experiments on subjects to evaluate if humans are able to recognize the face and its expression after de-identification with our method.

For the identity preservation, the experiment was conducted on 20 subjects. We first presented a de-identified face and next a set of 9 initial faces. The subject was asked if the de-identified face was present within these 9 images and where it was. This same question was asked for 6 different faces. Only $24 \pm 15.9\%$ of right answers were provided. This test illustrates the capacity of our method to de-identify the faces even for humans.

Similarly, another test was run to validate the expression preservation. 24 images were successively shown. For each, the subject was asked to choose an expression among 4 proposed. In order to compare the expression recognition for initial images and for de-identified ones, two groups of 19 persons were formed. The first one was only tested on the initial faces while the second one only on the de-identified ones. On $81 \pm 13.0\%$ of the faces, the same expression was recognized in the initial and de-identified faces.

5. CONCLUSION AND FUTURE WORKS

This paper introduced a new area of research, namely the de-identification of faces with expressions preservation. A variational model for adaptive filtering was introduced to process images. The filtering preserves the most important non-verbal communication signs such as the eyes, the gaze, the lips and their corners. Both qualitative and quantitative evaluations demonstrate the quality of our method. Furthermore, this work was realized in collaboration with researchers in human sciences. Future works include the accurate head segmentation and video de-identification. The researchers from human sciences are satisfied of our results in term of expressions preservation and, deeper experiments on machines will be conducted to validate the emotions preservation of our approach. The quantitative validation will also be performed on a larger scale.

6. REFERENCES

- [1] R. Gross, L. Sweeney, J. F. Cohn, F. De la Torre, and S. Baker, “Face de-identification,” in *Protecting Privacy in Video Surveillance*, pp. 129–146. Springer, 2009.
- [2] A. Prati and R. Cucchiara, “Video analysis for ambient intelligence in urban environments,” *Intelligent Env.: Methods, Algorithms and Applications*, p. 143, 2008.
- [3] S. C. Dakin and R. J. Watt, “Biological bar codes in human faces,” *Journ. of Vision*, vol. 9, no. 4, pp. 2, 2009.
- [4] M. Keil, ““I look in your eyes, honey”: Internal Face Features Induce Spatial Frequency Preference for Human Face Processing,” *PLoS Computational Biology*, vol. 5, no. 3, pp. e1000329, 2009.
- [5] C. Saavedra, P. Smith, and J. Peissig, “The relative role of eyes, eyebrows, and eye region in face recognition,” *Journ. of Vision*, vol. 13, no. 9, pp. 410–410, 2013.
- [6] S. Corrow, T. Donlon, J. Mathison, V. Adamson, and A. Yonas, “Predicting face recognition skills in children from holistic face processing and eye tracking,” *Journ. of Vision*, vol. 13, no. 9, pp. 104–104, 2013.
- [7] M. Turk and A. Pentland, “Face recognition using eigenfaces,” in *IEEE Conf. on Computer Vision and Pattern Recog.*, 1991, pp. 586–591.
- [8] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *IEEE Conf. on Computer Vision and Pattern Recog.*, 2014, pp. 1701–1708.
- [9] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *IEEE Conf. on Computer Vision and Pattern Recog.*, 2012, pp. 2879–2886.
- [10] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen, “Fusing robust face region descriptors via multiple metric learning for face recognition in the wild,” in *IEEE Conf. on Computer Vision and Pattern Recog.*, 2013, pp. 3554–3561.
- [11] S. Hussain, T. Napoléon, and F. Jurie, “Face recognition using local quantized patterns,” in *British Machine Vision Conf.*, 2012.
- [12] C. Lu and X. Tang, “Surpassing human-level face verification performance on lfw with gaussianface,” *arXiv preprint arXiv:1404.3840*, 2014.
- [13] Paul Ekman and Wallace V Friesen, “Facial action coding system,” 1977.
- [14] P. Ekman, W.V. Friesen, and J.C. Hager, *Facial Action Coding System (FACS): Manual*, A Human Face, Salt Lake City (USA), 2002.
- [15] A. L. Brooks, “Subject anonymisation in video reporting. is animation an option?,” in *Int. Conf. Disability, Virtual Reality & Associated Technologies*, 2012, pp. 431–433.
- [16] F. Tarrés and A. Rama, “GTAV Face Database,” <http://gps-tsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm>.
- [17] H. Mercier, J. Peyras, and P. Dalle, “Occluded facial expression tracking,” in *Image Analysis*, pp. 72–81, 2007.
- [18] D. Vlasic, M. Brand, H. Pfister, and J. Popović, “Face transfer with multilinear models,” *ACM Trans. on Graphics*, vol. 24, no. 3, pp. 426–433, 2005.
- [19] K. Dale, K. Sunkavalli, M. K. Johnson, D. Vlasic, W. Matusik, and H. Pfister, “Video face replacement,” *ACM Trans. on Graphics*, vol. 30, no. 6, pp. 130, 2011.
- [20] P. Garrido, L. Valgaerts, O. Rehmsen, T. Thormählen, P. Pérez, and C. Theobalt, “Automatic face reenactment,” in *IEEE Conf. on Computer Vision and Pattern Recog.*, 2014.
- [21] E. M. Newton, L. Sweeney, and B. Malin, “Preserving privacy by de-identifying face images,” *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.
- [22] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journ. of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [23] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.
- [24] G. B. Huang and E. Learned-Miller, “Labeled faces in the wild: Updates and new reporting procedures,” *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep.*, pp. 14–003, 2014.
- [25] L. Wolf, T. Hassner, and Y. Taigman, “Multiple one-shots for utilizing class label information,” in *British Machine Vision Conf.*, 2009, pp. 1–12.
- [26] B. Bhattarai, A. Mignon, F. Jurie, and T. Furon, “Puzzling face verification algorithms for privacy protection,” in *IEEE Workshop on Information Forensics and Security*, 2014.