Photorealistic Face de-Identification by Aggregating Donors' Face Components

Saleh Mosaddegh, Loïc Simon, Frédéric Jurie

Photorealistic Face de-Identification by Aggregating Donors' Face Components

Saleh Mosaddegh, Loic Simon and Frédéric Jurie

CNRS UMR 6072 - University of Caen - ENSICAEN
firstname.lastname@unicaen.fr

Abstract. With the adoption of pervasive surveillance systems and the development of efficient automatic face matchers, the question of preserving privacy becomes paramount. In this context, automated face de-identification is revived. Typical solutions based on eyes masking or pixelization, while commonly used in news broadcasts, produce very unnatural images. More sophisticated solutions were sparingly introduced in the literature, but they fail to account for fundamental constraints such as the visual likeliness of de-identified images. In contrast, we identify essential principles and build upon efficient techniques to derive an automated face de-identification solution meeting our predefined criteria. More specifically, our approach relies on a set of face donors from which it can borrow various face components (eyes, chin, *etc.*). Faces are then de-identified by substituting their own face components with the donors' ones, in such a way that an automatic face matcher is fooled while the appearance of the generated faces are as close as possible to original faces. Experiments on several datasets validate the approach and show its ability both in terms of privacy preservation and visual quality.

1 Introduction

A large number of cameras oversee public and semi-public spaces today. It raises concerns on the unintentional and unwarranted invasion of the privacy of individuals caught in the videos. To address these concerns, automated methods to de-identify individuals in these videos are necessary [1]. De-identification does not aim at removing all the information involving the individuals. Its ideal goals are to obscure the identity of the subject without obscuring the action or the rest of the scene.

Finding the right trade-off between privacy and awareness has a long history in the computer vision literature. As noted by Hudson and Smith [16], systems which attempt to support awareness in distributed media immediately face several important challenges. First among these is the widely recognized issue of privacy. They believe there is a fundamental trade-off between providing awareness information and preserving privacy. In general, the more information transmitted about one's actions, the more potential for awareness exists among those receiving the information. At the same time, however, the more information transmitted, the more potential for violation of one's privacy exists.

Fig. 1. Comparing blurring with our approach (from left to right): original image, image de-identified with naive blur, image automatically de-identified with our approach. While looking natural, the photo does not reveal explicitly the identity of the person.

Ideally, face, silhouette, gait and other characteristics need to be obscured. Our approach focuses however on face de-identification on a per-frame basis. Our main concern is to ensure privacy while preserving the likelihood of the produced de-identified image. One typical illustration of this process is illustrated in Fig.1. The image looks natural but the face shown on the photo can not be identified because it has been replaced by an artificial computer-generated face. Obviously, the new face should not reveal the identity of someone else, which would be the case if using face swapping algorithms (*e.g.* using [19]). The new face must be a fully artificial face, looking natural and similar to the face to be de-identified.

We formulate this goal in a principled way by expressing our response as the solution of an optimization problem. Our approach relies on the use of a bank of face donors from which the algorithm is allowed to pick components (such as eyes, nose, *etc.*). The optimal face, for a given face matcher algorithm, is the image which fools the face matcher (*i.e.* the distance between the original face and the de-identified one is greater than a threshold) while the difference between the two images is visually as small as possible (in this paper the visual difference is computed as the PSNR). In addition, another constraint guarantees that the forged face is not recognized by the face matcher for any of the donors used to de-identify the image. The comparison of Fig. 1, obtained using our approach, with existing de-identification approaches shown Fig. 2 motivates this work. Besides, in order to make the solving tractable, we reduce the search space to faces that are likely to be relevant to our purposes using different heuristics.

Let us list several key aspects of our approach. First, the de-identification can be applied to faces outside the donor database (*e.g.* Fig. 1). Besides, the proposed approach does not requires having the best face matcher available at the moment. Indeed, the face matcher is meant to give the direction along which faces should be modified to make their recognition harder. Obviously, better face matchers can lead to de-identified images with smaller visual differences (in

(a) De-indentification images from [10]: original image, pixelization with a block size of 16, Gaussian blur with a standard deviation $\sigma = 8$, scrambling by random sign inversion, scrambling by random permutation.

(b) Original/de-identified image, live-shadow algorithm of [29].

(c) Original/de-identified image, k-same algorithm of [23].

(d) Original/de-identified image, multi-factor $(\epsilon; k)$-map algorithm algorithm of [12].

(e) Original/de-identified image, Driessen and Durmuth's algorithm [9].

Fig. 2. Visual comparison of recently published de-identification approaches.

terms of PSNR). Another strong aspect of the method is the lack of required pre-processing. In other words, the queried image does not need to be normalized beforehand.

The rest of the paper is organized as follows: after presenting the related work (section 2) the details of our formulation are presented in section 3, followed by an experimental validation (section 4) in which the proposed approach is validated using images from three different datasets.

2 Related work

Enforcing privacy by de-identifying faces in images has a long history in the computer vision literature. As noted by [5], one common technique, often seen on news broadcasts, is to *pixelize* people's faces, replacing them by large pixels (squares). More advanced and general techniques have also been developed *e.g.* Hudson and Smith [16] who described a shadow-view filter giving the visual impression of ghostly shadow moving in a static scene, or Crowley *et al.* [8] who used eigen-filters to analyze a scene and reconstruct its images in a *socially-correct* way.

While identity masking techniques have been widely used in the media, there has been little empirical investigation of their effectiveness in protecting the identity of innocent passers-by children or crime witnesses from people familiar with them. Zhao and Stasko [29] examined four filters by asking volunteers to identify which of five actors were featured. Before the experiment started, volunteers were shown portraits of the actors. In [5], Boyle *et al.* analyzed how a blur and a

pixelize image filter might impact both awareness and privacy in images. They examined how well observers of several filtered video scenes extract particular awareness cues. Their results suggest that the blur filter, and to a lesser extent the pixelize filter, have a level suitable for providing awareness information while safeguarding privacy. More recently, Lander *et al.* [18] evaluated the effectiveness of pixelization and blurring on masking the identity of familiar faces. They concluded that privacy may not be fully preserved depending on whose identity is being concealed. Another important issue is to ensure anonymity while preserving the rest of the information [1]. Finally, it is worth noting that simple image manipulations can be retro engineered, allowing to reconstruct faces [6].

Besides these simple blurring and pixelization techniques, more advanced techniques have been introduced, most of them based on the well-known eigenfaces representation [9, 23], or some variants [12]. Newton *et al.* [23] proposed k-Same, a privacy enhancing algorithm based on the concept of k-anonymity to face image databases. The algorithm determines similarity between faces of the database, clusters similar faces, and creates a new face by aggregating the faces of a cluster. Gross *et al.* [12] proposed a factorization approach to separate identity and non-identity related factors, allowing to only replace the factors expressing the identity by the cluster's aggregation, while keeping the non-identity factors untouched to better preserve facial expressions. Dufaux and Ebrahimi [10] present an effective scrambling techniques to foil face recognition. Very recently, Driessen and Dürmuth [9] focussed on the preservation of the human recognition as a top requirement, by finding the modification of the image which has the lowest distortion (in the image space) while changing the signature to a desired value. This is done by projecting the face on the manifold spanned by some eigenfaces. Then, modifying the signature amounts to changing this projection. And since this is a linear process, mapping back this modification into the image space is simply achieved by modulating the eigenfaces components. The image part orthogonal to the space spanned by the eigenfaces is kept untouched.

The approaches mentioned in the previous paragraphs present two drawbacks. First, they do not produce photorealistic face images; they all look unnatural and attract the attention (see Fig. 2). Second, the k-Same principle is addressing a question which is very different from ours: the goal is to sanitize a database before publishing it, in such a way that searching for the most similar face to a query image will output k identities. Our approach is different as we assume the social network has already published a collection of pictures, and it is up to the user to process any novel picture before posting it.

Face swapping can be seen as an interesting solution for addressing the aforementioned limitations. In [3], given an input image, Bitouk *et al.* detect all the faces occurring in the image, align them, select candidate face images from a face library, adjust the pose, lighting, and color of the candidate face images to match the appearance of those in the input image, and seamlessly blend in the results. A user study validates the high quality of the replacement results. Zhu *et al.* [30] extended the approach by proposing a better alignment approach. Lin *et al.* [19] addressed the case of face replacement with large-pose differences. However, face

swapping raises other issues related to privacy, as the face template used for the replacement can easily be recognized.

Face synthesis provides yet another way to produce de-identified images. Taking inspiration from exemplar-based texture synthesis [11], Mohammed *et al.* [22] generated realistic images of faces using a model trained from real examples, describing textures with local non-parametric models. This approach paves the way to very realistic faces with interesting inpainting applications. The approach is however limited to frontal views and the use of such a technique for face de-identification has not been investigated so far. A somehow similar approach was introduced in [28] in the context of face hallucination.

In addition, several loosely related techniques may prove useful in improving the status of face de-identification. For instance, approaches based on landmark detection [31] such as active appearance models [7, 20] provide effective ways to align faces with different poses, hence alleviating the frontal view restriction. Parsing a face into meaningful components has been addressed in different ways, *e.g.* in [26]. Based on parsing, exemplar-based synthesis can be adapted into replacing parts of a face with corresponding components taken from a database. Eventually, leveraging seamless blending [25] allows to avoid artefacts.

In conclusion, despite the existence of a large body of related techniques, face de-identification remains insufficiently explored. We argue that at least two fundamental principles should be endorsed, namely (i) face de-identification should be guided by face matching, as face matching algorithms are getting closer and closer to human performance [27] and (ii) faces should look natural and artefact-free. To the best of our knowledge, the question of ensuring both criteria has not been satisfactorily addressed in the literature. Therefore we build upon powerful techniques such as automatic landmark detection, exemplar-based synthesis and seamless blending to provide an adapted solution to face de-identification.

3 A new method for synthesizing artificial faces using face component donations

3.1 Problem statement

Let I be loosely defined as the set of images containing one single face per image. We also assume that we have at our disposal a face matcher, *i.e.* a function $F(x_s, x_m)$ from $I \times I$ to $\mathbb{R}$, associating two images x_s and x_m with a scalar value such that $F(x_s, x_m) > \mu$ if and only if the two faces are believed to represent different persons. We also assume that a set of N face images of different people acting as *face donors* is available from which we can harvest face components (eyes, noses, *etc.*). This set is denoted by $\mathcal{FD}$.

Starting with an image x_o to be de-identified, our objective is to produce a modified image x_m by minimizing the following problem:

$$\begin{aligned} x_m^* = \quad & \underset{x_m \in I}{\operatorname{argmin}} \quad \|x_m - x_o\| \\ & \text{subject to} \quad F(x_m, x_o) > \mu \\ & \qquad\qquad \forall x_i \in \mathcal{FD}, F(x_m, x_i) > \mu \end{aligned} \tag{1}$$

In other words, we want to produce a novel face that meets several criteria. First it must not be confused by the face matcher with any of the donors nor with the original face. Second, it must remain as similar to the original face as allowed by the previous constraints.

In some applications (*e.g.* social networks), another exemplar x'_o of the same person as x_o may be available and published already. In that case, we introduce a slight modification where we verify that the face matcher does not recognize x'_o instead of x_o which is not meant to be published. The new variant is as follows:

$$\begin{aligned} x_m^* = \quad & \underset{x_m \in I}{\operatorname{argmin}} \quad \|x_m - x_o\| \\ & \text{subject to} \quad F(x_m, x'_o) > \mu \\ & \qquad\qquad \forall x_i \in \mathcal{FD}, F(x_m, x_i) > \mu \end{aligned} \tag{2}$$

To distinguish between the two formulations we will refer to them as self-deidentification for the former ($x'_o = x_o$) and pairwise de-identification for the later. For the sake of generality, our approach will be described for a generic x'_o.

3.2 Overview of the approach

We address the previous problem by using a face component cloning algorithm. We assume that faces are made of a set of C spatially delimited face components that can be cloned individually. Let $c = (c_1, \cdots, c_i, \cdots, c_c) \in [\![0, N]\!]^C$ an index vector expressing the origin within the donor bank of the different components of the artificially generated face. More precisely, $c_i = 0$ means the face component i is unchanged, $c_i = k$ with $0 < k \leq N$ means that the i-th component of the face has to be cloned from the k-th image of the face donor bank.

Let's denote by $FG(x_o, c, \mathcal{FD})$ a face generator algorithm that can generate an image from a source image x_o by cloning the face components indexed in c from the donor bank. The previous problem is then made simpler by restricting the search space from I to the range of faces created by FG when starting from the original face. It yields the following approximation:

$$\begin{aligned} x_m^* = \quad & \underset{x_m \in FG(x_o)}{\operatorname{argmin}} \quad \|x_o - x_m\| \\ & \text{subject to} \quad F(x_m, x'_o) > \mu \\ & \qquad\qquad \forall x_i \in \mathcal{FD}(x_m, x_i) > \mu \end{aligned} \tag{3}$$

where with a slight abuse of notation $FG(x_o) = \{FG(x_o, c, \mathcal{FD}) / c \in [\![0, N]\!]^C\}$.

input: $x_o, x'_o, \mathcal{FD}$
output: x^*_m

function OPTIMIZE($x_o, x'_o, \mathcal{FD}$)
  $x^*_m \leftarrow x_o$
  $\forall 1 \leq i \leq C$, $K_i \leftarrow [\![1, N]\!]$ *keep track of candidate donors for each component*
  while $\exists i$ s.t. $K_i \neq \emptyset$ and not CHECKCONSTRAINT(x^*_m, x'_o, $\mathcal{FD}$) **do** *Main Loop*
    for $i := 1$ to C **do** *Loop on components*
      $k_i \leftarrow \underset{k \in K_i}{\text{argmin}} \|x_o - FG(x^*_m, e_{i,k}, \mathcal{FD})\|$ *First heuristic*
      $y_i \leftarrow FG(x^*_m, e_{i,k_i}, \mathcal{FD})$
    end for
    $i^* \leftarrow \underset{1 \leq i \leq C}{\text{argmax}} F(x_0, y_i)$ *Second heuristic*
    $K_{i^*} \leftarrow K_{i^*} \setminus \{k_{i^*}\}$
  end while
  return x^*_m
end function

Algorithm 1: Optimization routine.

3.3 Optimization strategy

The optimization problem in Eq. 3 cannot easily be solved by using standard optimization toolboxes let alone brute force. It has the form of a constrained problem, where the energy functional is admittedly convex, but where the constraints can display any intricate behaviour. As a matter of fact, these constraints involve $F(x_m, x'_o)$ and despite this latter being often interpreted as a distance function, it is actually a black-box on which we do not exert any control.

Instead of falling back on brute force, that would be overly inefficient, we propose a greedy alternative presented in Algorithm 1. We iteratively replace one new component at a time, using one particular donor. We denote by $e_{i,k} = (0, \cdots, 0, k, 0, \cdots, 0)$ the index vector corresponding to the substitution of component i from donor k. Two greedy heuristics are implemented for the selection of the component and the donor. More precisely, the donor choice is leveraged in order to keep the energy functional small, and the component choice is meant to maximize our chances to de-identify the original face. The iterations continue until the de-identification constraints are met or if there is no donor left for any of the components. In the latter case, the optimization has failed.

3.4 Face generator

The overall working of the face generator is depicted in Fig. 3. For a given target image, we use up to 4 donors in order to replace the content of different regions of interest (ROI) associated with the components. There is a single connected ROI for three components, namely the mouth+chin component, the nose+cheeks one and the eyebrows. For the eyes on the other hand two connected ROIs are extracted. To perform the replacement of a given ROI we first cleanly align the donor image with the target inside the ROI and then we apply Poisson blending [25].

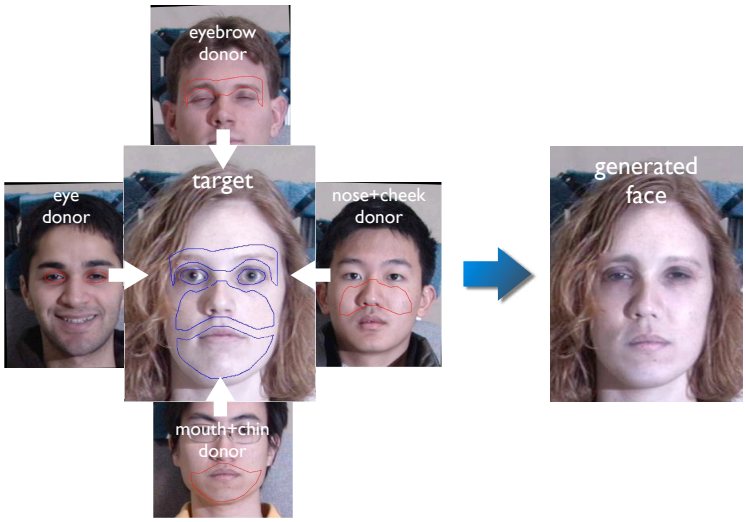


Fig. 3. The face generator implemented in this article. All the components are being replaced using a different donor. Each donor image is registered so that the region of interest (ROI) of the concerned component (red contour) is cleanly aligned with the same ROI in the target image. Artefacts are mitigated thanks to Poisson blending.

Landmark detection. We start our process by automatically detecting landmarks. They will be used in the alignment procedure and in generation of the ROIs. In our current implementation, the landmarks are detected following the approach of [31]. In total, the set of landmarks is composed of 68 points. As shown in Fig. 4 (a) each one of the 4 components is attached a fixed subset of the landmarks. This association is represented through a color code in the figure.

Face alignment. In order to improve accuracy, the alignment phase is performed per ROI. It is guided by the subset of relevant landmarks. More precisely, we estimate an optimal similarity transformation to register the target landmarks contained inside the ROI with their corresponding landmarks in the donor. The optimal transformation is obtained through classical Procrustes analysis. The obtained registration process is similar to the approach of [4].

Composing faces. So far we did not explain how the ROIs are actually built. The leading purpose here is to create a set of non overlapping ROIs covering most of the features of the face while leaving narrow bands of the original face uncovered. Such bands are useful to ensure the overall consistency among the blended donor components. The process to generate the ROIs is described in Fig. 4. This process is entirely automated and is composed of 4 steps. Step (a) corresponds to the landmark detection. In (b), we generate additional landmarks bound to define the contours of a first version of the ROIs depicted in (c). In this first version, the ROIs are tightly joined at some locations. We therefore shrink

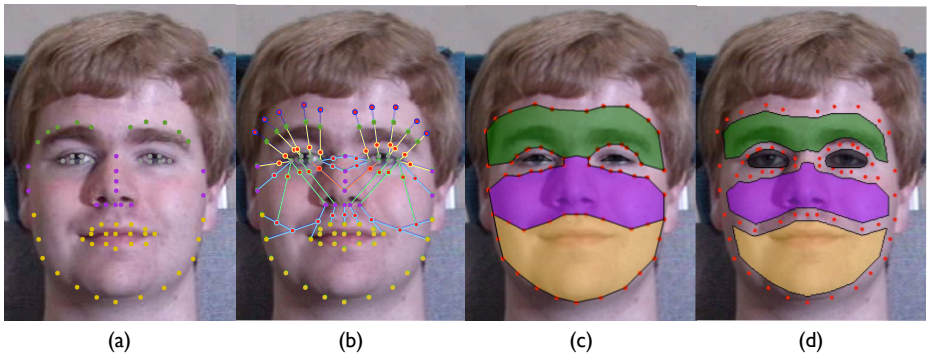


Fig. 4. Mask creation is conducted in 4 steps: (a) Landmarks detection and association with the face components (color coded) - (b) Landmark completion by barycentric averages - (c) mask initialization based on a predefined subset of landmarks - (d) Final mask creation by erosion or dilation. Two color codes are used in this picture (see the text for details).

them by applying a morphological erosion with a 6 pixel disk. This is not applied to the eyes ROI, because it is actually well separated from the surrounding ROIs. On the contrary, we apply a dilation in that case. The final ROIs are depicted in (d) using the same color code as in (a) and (c): green for the forehead, gray for the eyes, purple for the nose and cheeks, and yellow for mouth and chin.

The process to generate the additional landmarks (step (b)) is slightly technical. They are produced by using barycentric averages between two existing landmarks (original ones or already generated new landmarks). The weights used in the barycentric operations and the pairs of landmarks have been hand-tuned once for all in order to maximize the chance to capture meaningful textures in each ROI. The chosen segments and the produced landmarks are depicted in Fig. 4 (b). The new landmarks are filled in red. And the color of the link between the support pairs of landmarks represents the value of the weights. Each weight is chosen among a list of 5 possible values $\frac{1}{5}$ (green), $\frac{1}{4}$ (orange), $\frac{1}{2}$ (light-blue), $\frac{2}{3}$ (yellow) and $-\frac{2}{3}$ (dark blue). In the last case, the generated landmark is actually extrapolated rather than interpolated.

Additional remarks. At this point, it may appear odd that the eyes are handled in a different way compared to the other components. This is true in several respects. First, the component is broken in two separate ROIs. This choice is justified by the goal of generating a natural looking face. Such a purpose requires that both eyes are accurately located and scaled. Performing the alignment separately on each eye greatly simplifies that task. Furthermore, the initial ROI of each eye is fixed in a conservative way (that is tightly around the original landmarks) and then scaled up while the exact opposite strategy is applied to the other components. The rationale here is that the original eyes landmarks are much more reliably located around the actual boundary of the eye than the

Fig. 5. A few examples of face component's ROIs generated on faces from MUCT database (left-hand side) and MULTIPIE database (right-hand side).

generated ones. It is therefore easier to capture the whole shape of the eye by dilating the region delimited by the original landmarks.

Ensuring that the generated ROIs do not break distinctive features in several parts is an important achievement in our context. It brings the guarantee that in most situations, the collage will not produce structural artefacts. To illustrate the visual performance of the ROIs extraction process, we show in Fig. 5 several examples selected from two different databases.

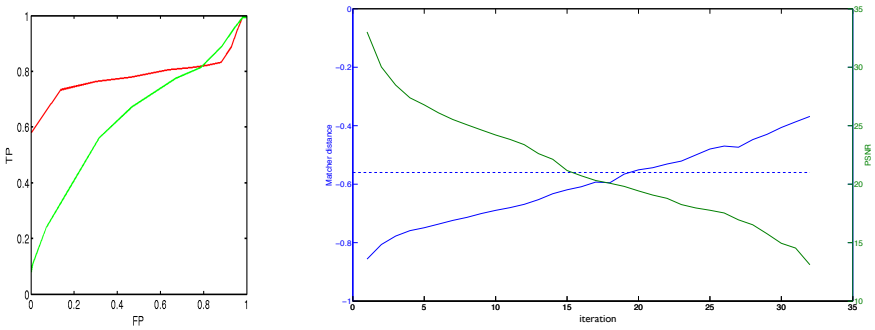
4 Experimental results

The objective of the experimental validation is twofold: one one hand, we want to show that the optimization algorithm is working as expected *i.e.* that at each iteration the distance from the original face is monotonically increased (from the perspective of the face matcher) while the modifications of the image are as small as possible (from the perspective of the PSNR). On the other hand, we want to show that resulting images are visually plausible and artifact free.

4.1 Experimental settings

The experimental validation is mostly done on a subset of the MultiePie dataset [13] containing 2184 frontal images of 346 people (men and women), with 11 images for each individual, under different facial expressions and illuminations. The choice of this database is led by the higher image quality compared to more common datasets such as LFW [15] which is an inevitable requirement for donor database. The images are annotated based on whether the person is wearing sun glasses (*i.e.* the eyes are visible or not). We also present some experiments on images drawn from the MUCT [21] (3755 images of 276 different subjects) and PUT [17] (9971 images of 100 subjects) datasets (using similar settings to MultiePie). During the component replacement we use MultiPie's annotations to reject eye and eyebrows components covered by sunglasses hence only using mouths and noses. In addition, we do not consider the donors with glasses while replacing eyes. Finally, we reject any donor images belonging to the target image. The set of donors contains around 100 faces in the presented experiments.

The face matcher used throughout our approach is inspired by [24] and computes face signatures as histograms of LBP descriptors [2]. Once the descriptors



(a) ROC profile for the face matcher before (red) and after (green) applying our approach.

(b) Evolutions of the PSNR (green) and the face matcher distance (blue) with the iterations. The dashed line indicates the value of the de-identification threshold μ.

Fig. 6. Face matcher's de-identification performance.

of two images are computed, the distance of the two faces is evaluated as the cosine distance. In other words, if we denote by ϕ and ϕ' the descriptors of the images x and x', then $F(x, x') := \frac{\langle \phi | \phi' \rangle}{\|\phi\| \|\Phi'\|}$. As mentioned in the introduction, while more recent methods (*e.g.* [27]) give better performances, this simple verification pipeline has the great advantage of being much faster to compute (which is an important advantage when working with a large donor data-set), while being good enough for giving the direction in which an image should be modified to alter the identity as much as possible without altering the image too much. The threshold μ (see section 3) is set using a validation set.

4.2 Quantitative results and algorithm's convergence

To demonstrate the de-identification power of our approach we first present ROC profiles in Fig. 6(a). In order to produce these curves, 262 positive pairs (2 images of the same person) and the same number of negative pairs (two images of different persons) were randomly selected from the MULTIPIE dataset. In each positive pair, one of the two images is de-identified with our approach while negative pairs are untouched. The red curve shows the true positive detection rate against the false positive one, as obtained by the face matcher, before de-identification. The alternate curve (in green) shows the same statistics when the face matcher is applied after de-identification of the database. As expected, de-identification effectively renders the face matcher inefficient since the ROC profile becomes closer to the top-right diagonal curve. Such a profile is typical of a system that provides purely random outputs.

Our optimization is guided by the aim of applying as mild visual degradations to the original image as is allowed by the de-identification itself. Although PSNR is a debatable choice for measuring the amount of visual perturbation, it is the

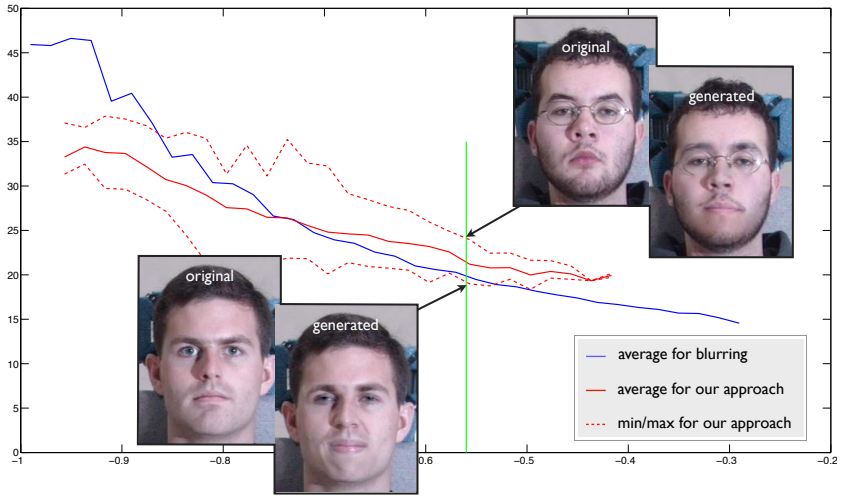


Fig. 7. PSNR against face matcher distance. The solid curves correspond to average performance on the MULTIPIE database, and the dashed ones to the best and worst cases (maximum and minimum PSNR). The blue curve corresponds to de-identification by blurring, and the red one to our approach. The threshold μ is indicated via the green line. For this value of the face Matcher, we show the original and de-identified faces corresponding to the worst and best PSNR.

one that we have favored in our formulation. Therefore Fig. 6(b) presents the evolution of the PSNR and of the face matcher distance against the number of iterations of the optimization process. The curves are actually obtained by averaging 132 different runs. As expected, the face matcher distance increases with the iterations while the PSNR decreases. Interestingly, in average, de-identification requires around 20 iterations. This means that for each component there are several donors which look very similar to the subject under consideration.

In Fig. 7, we refine our analysis by comparing the needed amount of texture deformation to reach a certain face matcher distance with our method and with naive blurring. The relevant part of the curves is the rightmost one, since it corresponds to the highest level of de-identification. In particular, one should focus on the curves behavior near and beyond the de-identification threshold μ. From that perspective our method is uniformly superior to blurring. To emphasize the fact that our approach performs evenly in all experiments, we have also added worst and best case curves in dash. In both cases, the original and de-identified face corresponding to the de-identification threshold are also presented.

4.3 Visual inspection of output images

Although the good behavior of a de-identification method with respect of PSNR is a desirable property, it is actually much more important to assess if the produced images are visually realistic. Making quantitative assessment about such

realism is not an easy task. Instead, we provide in Fig. 8 a selection of examples (the first 2 rows are from MULTIPIE, the next 3 ones from MUCT and the last 2 ones from PUT). For each example, a sequence of 6 images is presented. Among such a sequence, the original image is marked with a blue frame and the first to reach the prescribed de-identification level is marked with a yellow frame. The remaining four images were picked randomly and were placed accordingly to their iteration. Amazingly, even long after de-identification is reached the generated faces continue to look very convincing.

Some examples demonstrate a striking versatility of the face generator. For instance in the third row, the subject was gradually transformed into an individual of seemingly different ethnic origin without noticeable artefacts. As a side note, the presented sequences suggest the creation of criminal identikits as an unexpected application of our method. In such a scenario, a witness would start from an exemplar face looking globally similar to the target criminal (similar face silhouette and global features). Then, the profiled face would be refined by iteratively replacing the components of the current image with those of a selection of donors. The set of donors would be determined under the guidance of the witness and based on appropriate criteria.

As a more conventional application of our work, we consider now the typical situation where a user of a social network is about to share a photo and wants to prevent the faces from being tagged. Figure 1 presented on the first page serves as a good example. The same principle could be also applied to web-services such as Google Streetview. Indeed, for legal reasons, such services are bound to occult the identity of the passers-by. So far, this is done based on aggressive blurring so that the images do not remain realistic.

5 Conclusions

Relying on a bank of face components from which face elements can be borrowed, and exploiting the power of Poisson editing techniques, this paper presents a simple yet very efficient and fully automatic pipeline for the de-identification of face images. The proposed algorithm produces optimally de-identified faces with respect to a given face matcher. The so-obtained faces have the property of being as close as possible to the original ones while having the guarantee they fool the face matcher. As a proof of concept a standard face matcher was used, but recent state-of-the art matchers such as [27], considered to be close to human performance, would allow to produce even more convincing images. The approach has been validated on three different datasets for which we obtained impressive results. Without requiring any prior normalization, our approach can handle arbitrary faces provided that it is within 15 degrees from the frontal pose. In order to tackle this limitation, an extension based on the extraction of a 3D template similar to [14] is currently studied.

Acknowledgments

This work was partly supported by the ANR-SECULAR project.

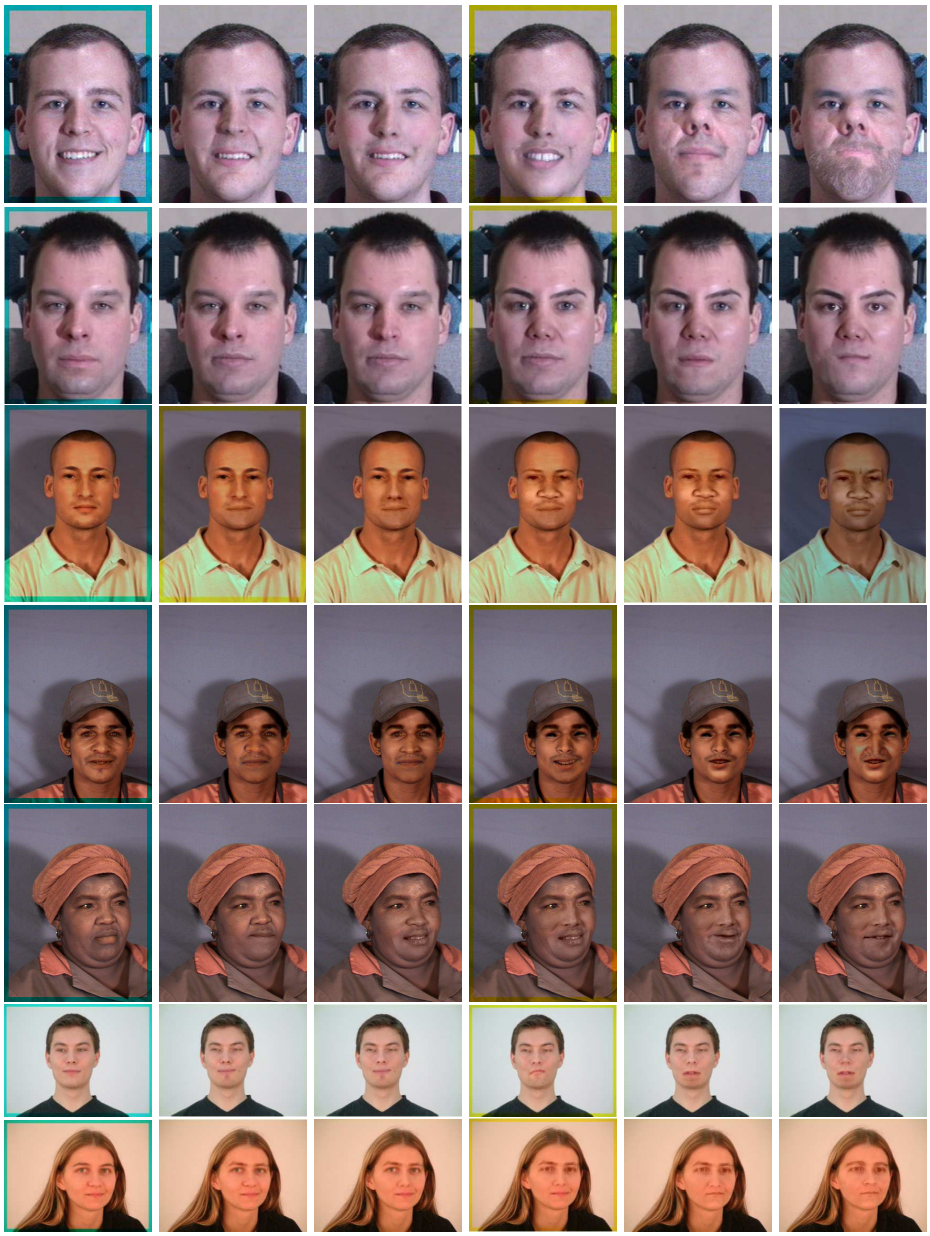


Fig. 8. A few randomly selected images forged during optimization procedure. The faces are presented in the same order as they were produced during the iterations. The original image is marked with a blue frame. Similarly the forged face that passes the de-identification test is marked with a yellow frame. The right-most faces give a glimpse of the degree of de-identification possibly achieved by letting the algorithm run further (or equivalently by imposing a more challenging de-identification threshold μ).

References

1. Agrawal, P., Narayanan, P.J.: Person De-Identification in Videos. IEEE Transactions on Circuits and Systems for Video Technology 21(3), 299–310 (2011)
2. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 2037–2041 (2006)
3. Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P., Nayar, S.K.: Face swapping: automatically replacing faces in photographs. In: ACM Transactions on Graphics (TOG). vol. 27, p. 39 (2008)
4. Bonnen, K., Klare, B.F., Jain, A.K.: Component-based representation in automated face recognition. IEEE Transactions on Information Forensics and Security 8(1), 239–253 (2013)
5. Boyle, M., Edwards, C., Greenberg, S.: The effects of filtered video on awareness and privacy. In: ACM conference on Computer supported cooperative work. pp. 1–10. New York, USA (Dec 2000)
6. Cavedon, L., Foschini, L., Vigna, G.: Getting the face behind the squares: Reconstructing pixelized video streams. In: Proceedings of the 5th USENIX Conference on Offensive Technologies. pp. 5–5. WOOT'11, USENIX Association, Berkeley, CA, USA (2011)
7. Cootes, T.F., Edwards, G.J., Taylor, C.J., et al.: Active appearance models. IEEE Transactions on pattern analysis and machine intelligence 23(6), 681–685 (2001)
8. Crowley, J.L., Coutaz, J., Bérard, F.: Perceptual user interfaces: things that see. Communications of the ACM 43(3), 54–60. (Mar 2000)
9. Driessen, B., Dürmuth, M.: Achieving Anonymity against Major Face Recognition Algorithms. In: Communications and Multimedia Security. pp. 18–33. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
10. Dufaux, F., Ebrahimi, T.: A framework for the validation of privacy protection solutions in video surveillance. In: IEEE International Conference on Multimedia and Expo (ICME). pp. 66–71 (2010)
11. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. pp. 341–346 (2001)
12. Gross, R., Sweeney, L., De la Torre, F., Baker, S.: Semi-supervised learning of multi-factor models for face de-identification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (2008)
13. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and Vision Computing 28(5), 807–813 (2010)
14. Hassner, T.: Viewing real-world faces in 3d. In: IEEE International Conference on Computer Vision. pp. 3607–3614 (2013)
15. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007)
16. Hudson, S.E., Smith, I.: Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In: ACM conference on Computer supported cooperative work. pp. 248–257. New York, New York, USA (1996)
17. Kasinski, A., Florek, A., Schmidt, A.: The put face database. Image Processing and Communications 13(3–4), 59–64 (2008)
18. Lander, K., Bruce, V., Hill, H.: Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. Applied Cognitive Psychology 15(1), 101–116 (2001)

19. Lin, Y., Lin, Q., Tang, F., Wang, S.: Face replacement with large-pose differences. In: ACM international conference on Multimedia. pp. 1249–1250. ACM, New York, New York, USA (Oct 2012)
20. Matthews, I., Xiao, J., Baker, S.: 2d vs. 3d deformable face models: Representational power, construction, and real-time fitting. International journal of computer vision 75(1), 93–113 (2007)
21. Milborrow, S., Morkel, J., Nicolls, F.: The muct landmarked face database. Pattern Recognition Association of South Africa (2010)
22. Mohammed, U., Prince, S.J., Kautz, J.: Visio-lization: generating novel facial images. In: ACM Transactions on Graphics (TOG). vol. 28, p. 57 (2009)
23. Newton, E.M., Sweeney, L., Malin, B.: Preserving privacy by de-identifying face images. IEEE Transactions on Knowledge and Data Engineering 17(2), 232–243 (2005)
24. Nguyen, H., Bai, L.: Cosine similarity metric learning for face verification. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) Computer Vision ACCV 2010, Lecture Notes in Computer Science, vol. 6493, pp. 709–720. Springer Berlin Heidelberg (2011)
25. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. In: ACM Transactions on Graphics (TOG). vol. 22, pp. 313–318 (2003)
26. Smith, B.M., Zhang, L., Brandt, J., Lin, Z., Yang, J.: Exemplar-based face parsing. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3484–3491 (2013)
27. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)
28. Yang, C.Y., Liu, S., Yang, M.H.: Structured face hallucination. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1099–1106 (2013)
29. Zhao, Q.A., Stasko, J.T.: Evaluating image filtering based techniques in media space applications. In: ACM conference on Computer supported cooperative work. pp. 11–18. New York, New York, USA (Nov 1998)
30. Zhu, J., Van Gool, L., Hoi, S.C.: Unsupervised face alignment by robust nonrigid mapping. In: IEEE 12th International Conference on Computer Vision. pp. 1265–1272. IEEE (2009)
31. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2879–2886 (2012)