Towards Real-World Face De-Identification

Ralph Gross and Latanya Sweeney

***Abstract*—A wide range of technological advances have helped to make extensive image and video acquisition close to effortless. As a consequence many applications which capture image data of people for either immediate inspection or storage and subsequent sharing have become possible. Along with these improved recording capabilities, however, come concerns about the privacy of people visible in the scene. While algorithms have been proposed to de-identify images, currently available methods are still lacking. In this paper we propose a general framework for the de-identification of images which subsumes a number of previously introduced approaches. Unlike the ad-hoc methods currently used in the field our algorithms aim at providing privacy guarantees. In experiments on illumination- and expression-variant face datasets we show that the proposed algorithms achieve the desired privacy protection while minimally distorting the data.**

I. Introduction

Due to recent advances in both camera technology as well as supporting computing hardware and software[1], image and video capture has become ubiquitous, ranging from simple web cams pointing out a window to video surveillance networks offering instant access to hundreds or thousands of cameras around inner cities [28]. Along with the increased usage of cameras however often come concerns about the privacy of people visible in the images [10], [25]. This question was raised most recently in connection with the introduction of Google Street View, a map service which offers high-resolution street-level images of a number of cities, in some cases depicting people in embarrasing situations [16].

For many applications the situation is often portrayed as a mutually exclusive choice between functionality on one side and privacy on the other. However, many uses of image and video capture do not require knowledge of the *identity* of people visible in the scene. In [31] a system is described which tracks the number of people appearing on a street corner in New York City by counting faces in the context of a bio-terrorism surveillance application. Similarly, Senior et al. [27] propose a video surveillance system which displays identity-obscured video to a security guard while storing the raw video cryptographically secured for subsequent retrieval by law enforcement, if necessary.

These examples make the case for the de-identification of image data. Privacy protection methods are well established for field-structured data [32], especially medical data [30]. While a number of algorithms have been proposed to achieve privacy protection for images as well, current methods are still lacking. In this paper we propose a novel *general* framework for the de-identification of images which subsumes a number of previously introduced approaches. Special emphasis will be placed on providing privacy guarantees for the resulting images as well as preserving as much of the original signal as possible. The work presented here concentrates on face images. While other modalities have been used for (automatic) human identification and verification from images or video (e.g. iris [6], ear [24], and gait recognition [4], [26]) face recognition is the most mature field, looking back at more than 30 years of research [18], [34].

The remainder of this paper is organized as follows. In Section II we give an overview of related work on image de-identification methods. We then introduce our framework and privacy protection models in Section III. In Section IV we discuss two de-identification algorithms which implement privacy protection models defined in Section III.

II. Related Work

Currently available image de-identification algorithms fall into one of two groups: ad-hoc distortion methods and the k-Same [22] family of algorithms implementing the k-anonymity protection model [30]. In this section we describe both approaches in detail (Sections II-A and II-B). In Section II-C we demonstrate shortcomings of k-Same as motivation to our new framework introduced in Section III.

A. Ad-hoc De-Identification Algorithms

Across a number of different communities including human computer interaction, computer vision, and computer supported cooperative work (CSCW), the problem of protecting privacy of people visible in images has been addressed. The majority of approaches employ simple obfuscation methods such as blurring (smoothing the image with e.g. a Gaussian filter with large variance) or pixelation (image subsampling) [2], [17], [21], [33]. While these algorithms are applicable to all images, they lack a formal privacy model. As a consequence no guarantees can be made that the privacy of people visible in the images is actually protected. Privacy protection is evaluated, if at all, only in human subject studies. It has been shown that these naïve algorithms are easy to defeat [22] and typically neither preserve privacy nor the utility of the data [12].

Alternatively, other approaches mask the areas of an image deemed sensitive [8], [9], [11], [19]. Similarly, the PrivacyCam architecture proposed by Senior et al. [27] suppresses automatically segmented foreground objects in the scene and cryptographically secures access to the video stream(s) produced by the system. As a consequence of the masking, most of the characteristics of the foreground objects are lost. Different

The authors are with the Data Privacy Laboratory, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[1]Relevant improvements include all system components such as processor speed, data storage capacity, speed of wireless communication, image compression algorithms, etc.

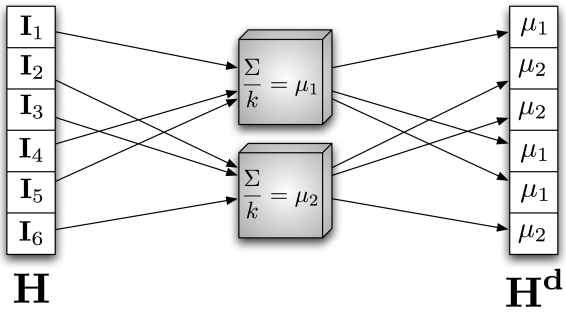


Fig. 1. Overview of the k-Same algorithm. Images are de-identified by computing averages over the closest neighbors of a given face in H and adding k copies of the resulting average to H^d.

versions of naïve de-identification algorithms are available commercially.[2]

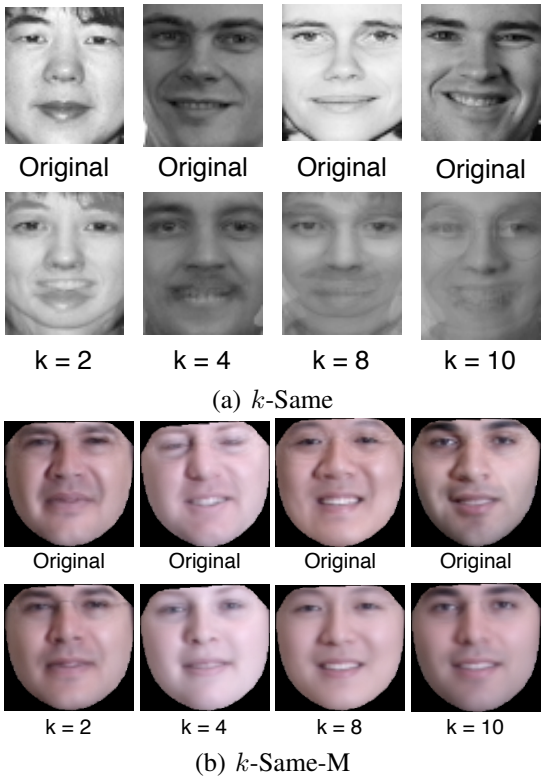
Phillips [23] proposed an algorithm for privacy protection of facial images through reduction of the number of eigenvectors used in reconstructing images from basis vectors. A direct trade-off between privacy protection and data utility is established through the introduction of the privacy operating characteristic (POC), a plot similiar to a receiver operating characteristic (ROC) often used in pattern classifier design [7].

A different approach to dealing with privacy issues in image processing was recently proposed by Avidan and Butman [1]. They use secure multi-party computation methods to perform image analysis tasks such as face detection without revealing the image content to the entity processing the data. While this approach avoids the need to de-identify images, the most secure version of the algorithm is comparatively slow.

B. The k-Same Framework

The k-Same family of algorithms [12], [15], [22] implement the k-anonymity protection model [30] for face images. Given a *person-specific*[3] set of images $H = \{\mathbf{I}_1, \ldots, \mathbf{I}_M\}$, k-Same computes a de-identified set of images $H^d = \{\mathbf{I}_1^d, \ldots, \mathbf{I}_M^d\}$ in which each $\mathbf{I}_i^d$ indiscriminately relates to at least k elements of H. It can then be shown that the best *possible* success rate for a face recognition algorithm linking an element of H^d to the correct face in H (independent of the algorithm used) is $\frac{1}{k}$. See [22] for details. k-Same achieves k-anonymity protection by averaging the k closest faces for each element of H and adding k copies of the resulting average to H^d. See Figure 1 for an illustration of the algorithm.

While k-Same provides provable privacy guarantees, the resulting de-identified images often contain undesireable artifacts. Since the algorithm directly averages pixel intensity values, even small alignment errors of the underlying faces cause "ghosting" effects. To overcome this problem a model-based extension to k-Same, referred to as k-Same-M was introduced in [15], which fits an Active Appearance Model (AAM) [5], [20] to input images and then applies k-Same on the AAM model parameters. The resulting de-identified images are of much higher quality than images produced by k-Same while the same privacy guarantees still hold. See Figure 2 for examples.



(a) k-Same

(b) k-Same-M

Fig. 2. Examples of de-identified face images. Faces shown in (a) were de-identified using the appearance-based version of k-Same. Due to misalignments in the face set, ghosting artifacts appear. Faces in (b) were de-identified using k-Same-M, the model-based extension of k-Same. In comparison, the images produced by k-Same-M are of much higher quality.

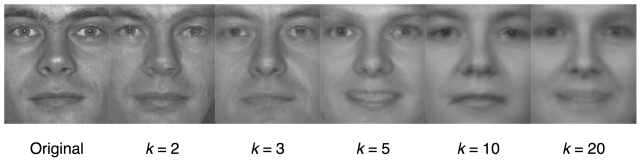
k-Same selects images for averaging based on raw Euclidean distances in image space or Principal Component Analysis coefficient space [22]. In order to use additional information during image selection such as gender or facial expression labels, k-Same-Select was introduced in [12]. The resulting algorithm provides k-anonymity privacy protection while preserving data utility as evidenced by both gender and facial expression recognition experiments. See Figure 3 for examples comparing the k-Same and k-Same-Select algorithms.

C. Shortcomings of the k-Same Framework

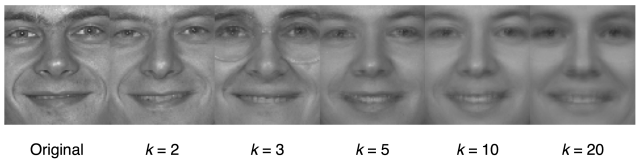
k-Same assumes that each subject is only represented once in the dataset H, a condition which is often not met in practice. Since k-Same uses the nearest neighbors of a given image during de-identification, the presence of multiple images of the same subject in the input set can lead to lower levels of privacy protection. To demonstrate this we report results of

[2]Eptascape (http://www.eptascape.com) overlays the output of a tracking system with a mask. Emitall scrambles compression coefficients in regions of interest http://www.emitall.com/.

[3]In a person-specific set of faces each subject is represented by no more than one image.



(a) k-Same



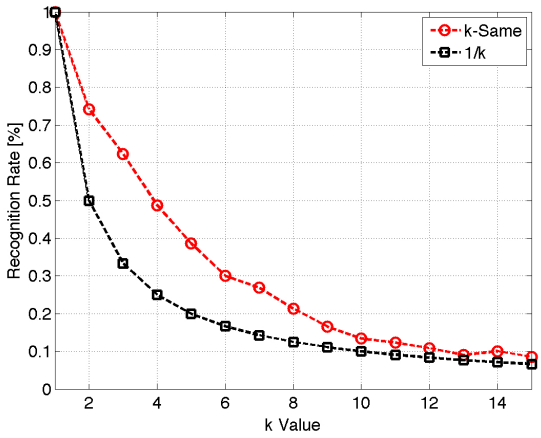
(b) k-Same-Select

Fig. 3. Examples of applying k-Same and k-Same-Select to expression variant faces. Since k-Same-Select factors facial expression labels into the image selection process, facial expressions are preserved better (notice the changing expression in the first row). Both algorithms provide k-anonymity privacy protection.

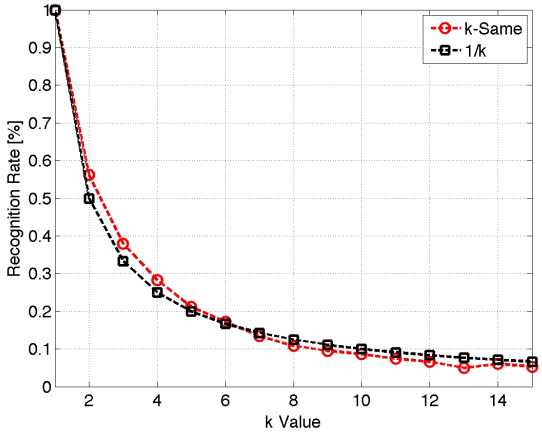
Fig. 4. Examples of images used in recognition experiments. We use frontal images of more than 200 subjects from the CMU Multi-PIE database under five illumination conditions, displaying a range of facial expressions (shown here are neutral, squint, smile, and disgust). In the experiments, faces are represented by the appearance coefficients of an Active Appearance Model [5], [20].

experiments on the CMU Multi-PIE database [14]. Each face in the dataset is represented using the appearance coefficients of an Active Appearance Model [5], [20]. See Figure 4 for examples. Recognition is performed by computing the nearest neighbors in the appearance coefficient space. We employ images of 203 subjects in frontal pose and frontal illumination, displaying neutral, surprise, and squint expressions. k-Same fails to provide adequate privacy protection. Figure 5 shows face recognition accuracies for varying levels of k. Accuracies stay well above the $\frac{1}{k}$ rate guaranteed by k-Same for datasets with single examples per class We obtain similar results even when class information is factored into the k-Same de-identification process. We can conclude that k-Same does not provide sufficient privacy protection if multiple images per subject are included in the dataset.

k-Same operates on a *closed* face set H and produces a corresponding de-identified set of faces H^d. Many potential application scenarios for de-identification techniques involve processing individual images or sequences of images. k-Same is not directly applicable in these situations. Due to the definition of k-Same, extensions for open-set de-identification are not obvious.



(a) k-Same on expression-variant face data



(b) k-Same on illumination-variant face data

Fig. 5. Rank-1 recognition accuracies of images de-identified using k-Same. The underlying image set contains multiple faces per subject. In (a) we show recognition accuracies after applying k-Same to a subset of the CMU Multi-PIE database containing multiple expressions (neutral, surprise, and squint) of each subject. Recognition accuracies exceed $\frac{1}{k}$ by far, indicating lower levels of privacy protection. In (b) we show recognition accuracies after applying k-Same on an illumination-variant subset of Multi-PIE. Again, accuracies exceed $\frac{1}{k}$ for lower levels of k.

III. Formal Models for Face Image De-Identification

In this section we describe our proposed framework for image de-identification. We start by providing background definitions in Section III-A. We then introduce the proposed de-identification framework in Section III-B. In Section III-C we describe relevant application scenarios.

A. Basic Definitions

While applications discussed here use face images or image sequences, all mathematical derivations are in terms of general vectors. In order to ensure comparability between vector dimensions, faces have to be registered and aligned. We use one of two procedures for face alignment: *appearance-based coding* where we align faces using manually established

feature points or *model-based coding* where a previously learned Active Appearance Model [5], [20] is fit to the image and the resulting model parameter vector is used to encode the image. In the following we assume all face images to be encoded as m-dimensional vectors.

At the core of our framework is the notion that the set of all images of a person's face can be described compactly using a model learned from data.

Definition III.1. *Face Model*
We refer to the parametric or non-parametric model representation computed from a set of faces of a subject s as F_s.

Definition III.2. *Subject-Specific Image Sets*
The subject visible in an image is indicated using subscriptes when necessary, e.g. $\mathbf{I}_s$ for an image $\mathbf{I}$ of subject s. If a set of images $\Gamma = \{\mathbf{I}_{s,1}, \ldots, \mathbf{I}_{s,t}\}$ of the same subject s is used, we assume it to be known that the images are coming from the same subject, although the subject identity is not generally known. We then refer to Γ as being subject-specific.

The stated goal of face de-identification is to thwart face recognition. We perform recognition by evaluating the probability of input faces given a set of known models.

Definition III.3. *Face Recognition*
Given a *gallery* set of face models $\mathcal{G} = \{F_1, F_2, \ldots, F_l\}$ known to the algorithm we evaluate $p(F_i|\mathbf{P}), i = 1, \ldots, l$ for the *probe* image $\mathbf{P}$ and assign $\mathbf{P}$ to the subject j for which $p(F_j|\mathbf{P}) > p(F_i|\mathbf{P}), \forall i \neq j$. This process extends naturally to probe sets of images.

We then define face de-identification as a transformation function.

Definition III.4. *Face De-Identification*
Face de-identification is defined as a function $\Psi_{\mathcal{R}} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ with respect to a reference set $\mathcal{R} = \{R_1, \ldots, R_m\}$ of face models. $\Psi_{\mathcal{R}}$ associates each input image $\mathbf{I}$ with a vector of equal dimensionality: $\Psi_{\mathcal{R}}(\mathbf{I}) = \mathbf{I}^d$. Ψ may be parameterized so that the functional mapping is $\Psi_{\mathcal{R}} : \mathbb{R}^m \times \mathbb{R} \rightarrow \mathbb{R}^m$.

The principal challenge in designing de-identification methods lies in striking a balance between privacy on one side and data usability on the other side. The following definition captures our notion of data usability:

Definition III.5. *Data Utility*
We define data utility as function $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}$ which assignes a utility score to an m-dimensional image vector.

Examples for data utility functions for face image data include gender and facial expression classification accuracies as well as raw image distances. Figure 6 shows an overview of the definitions.

B. Privacy Protection Models

Intuitively, the goal of this work is to find de-identification functions Ψ for which $p(F_i|\Psi(\mathbf{I_i})) < p(F_i|\mathbf{I_i})$ with $\Phi(\Psi(\mathbf{I_i})) \approx \Phi(\mathbf{I_i})$, i.e. protect privacy while preserving data utility. More formally we define the goal as follows:

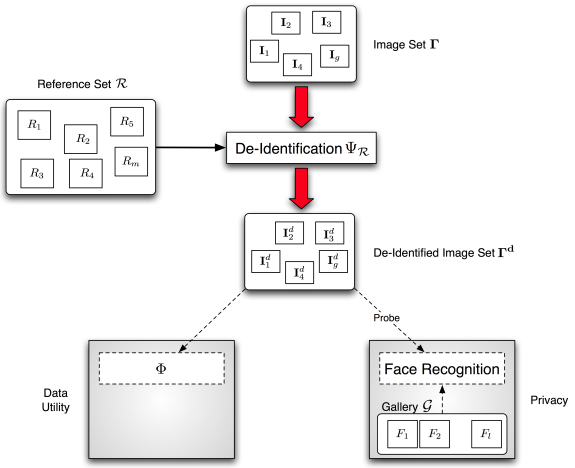


Fig. 6. Overview of the definitions. Images $\mathbf{I}_i$ in the input set Γ are de-identified to images $\mathbf{I}_i^d$. During de-identification the reference set $\mathcal{R}$ is used. Privacy protection is measured by using the de-identified image set Γ^d as probe in face recognition experiments with the gallery set of face models $\mathcal{G}$. The utility of de-identified faces is measured using the data utility function Φ.

Definition III.6. *Image De-Identification Problem*
Given a subject-specific set of images $\mathbf{\Gamma} = \{I_{s_1,1}, I_{s_1,2}, \ldots, I_{s_1,l_1}, I_{s_2,1}, \ldots, I_{s_n,l_n}\}$ find a de-identification function $\Psi_{\mathcal{R}} : \mathbf{\Gamma} \rightarrow \mathbf{\Gamma}^d$ with $\mathbf{\Gamma}^d = \{I^d_{s_1,1}, I^d_{s_1,2}, \ldots, I^d_{s_1,l_1}, I^d_{s_2,1}, \ldots, I^d_{s_n,l_n}\}$ so that $\mathbf{\Gamma}^d$ adheres to the privacy protection model Θ with respect to the reference set $\mathcal{R}$ of face models.

In the following we define three privacy protection models: ϵ-map, wrong-map and (ϵ, k)-map, which are closely related to the privacy protection models proposed in [29], [30].

Definition III.7. *ϵ-map*
We say that $\Psi_{\mathcal{R}} : \mathbf{\Gamma} \rightarrow \mathbf{\Gamma}^d$ provides ϵ-map protection if $p(R_i|\Psi(\mathbf{I})) < \epsilon, \forall R_i \in \mathcal{R}, \forall \mathbf{I}^d \in \Gamma^d$.

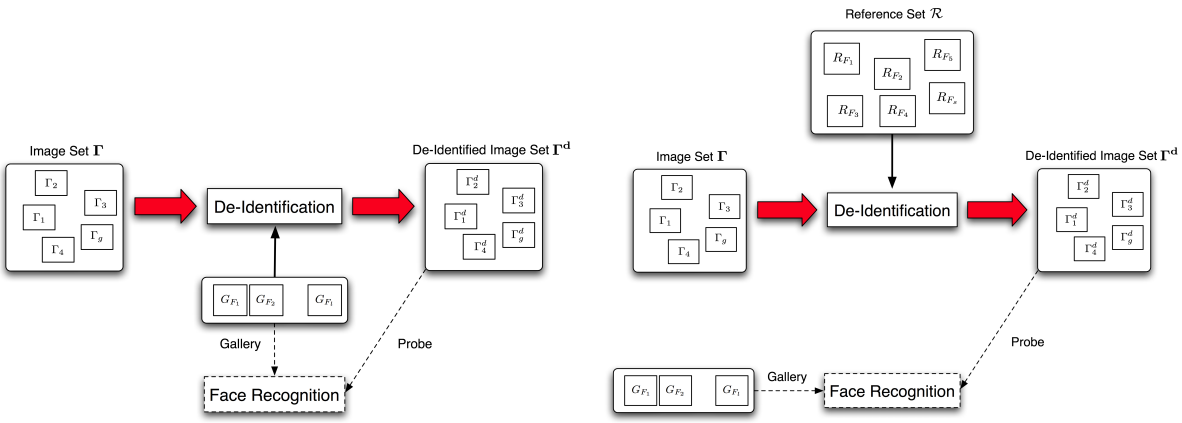
As we will argue later, all ad-hoc de-identification methods can be interpreted as naïve implementations of ϵ-map protection. As strategy, ϵ-map could be described as "*make it look like noone*".

Definition III.8. *wrong-map*
$\Psi_{\mathcal{R}} : \mathbf{\Gamma} \rightarrow \mathbf{\Gamma}^d$ provides wrong-map protection if $p(R_i|\Psi(\mathbf{I}_j)) > p(R_j|\Psi(\mathbf{I}_j)), \forall \mathbf{I}_j^d \in \Gamma^d$.

As strategy, wrong-map could be described as "*make it look like someone else*".

Definition III.9. *(ϵ, k)-map*
$\Psi_{\mathcal{R}} : \mathbf{\Gamma} \rightarrow \mathbf{\Gamma}^d$ provides (ϵ, k)-map protection if $\forall\ \mathbf{I}_i^d \in \Gamma^d\ \exists R_{i_1}, \ldots, R_{i_k} \in \mathcal{R}$ with $\|p(R_{i_j}|\mathbf{I}_i^d) - p(R_{i_l}|\mathbf{I}_i^d)\|^2 < \epsilon$.

As strategy, (ϵ, k)-map could be described as "*make it look like everyone*". (ϵ, k)-map is similar in spirit to the (c, t)-isolation concept proposed in [3]. Depending on the specific application needs one or multiple de-identification strategies might be appropriate.



(a) Office Building (b) Street Corner

Fig. 7. De-identification application scenarios. In the office building scenarion (a) faces to be de-identified come from a known set of subjects so the gallery can be used. In the street corner scenario (b) subjects to be de-identified are not known a priori so a generic reference face set has to be used.

C. Application Scenarios

As described in Section II-B the previously proposed k-Same algorithm is defined for static, closed sets of faces [22]. Typical real-world scenarios in which facial images might be captured, however, are often different. We are particularly interested in two scenarios:

- Office Building
  The reference face model set used during de-identification is the gallery. An example for this scenario is an office building in which faces to be de-identified come from a known set of people, the employees working at that location. See Figure 7(a) for an illustration.
- Street Corner
  Here the de-identification algorithm does not have access to the gallery. As a consequence a generic reference face set is used. An example for this scenario is a street corner at which it is not known a priori which people will appear. See Figure 7(b) for an illustration.

IV. De-Identification Algorithms

In this section we describe algorithms for ϵ-map (Section IV-A) and (ϵ, k)-map protection (Section IV-B). We will assume all face models *a priori* to be equally likely and replace the *a posteriori* probabilities $p(R_i|\Psi(\mathbf{I}))$ with the class-conditional density functions $p(\Psi(\mathbf{I})|R_i)$, following Bayes' rule [7].

A. Algorithm for ϵ-map Protection

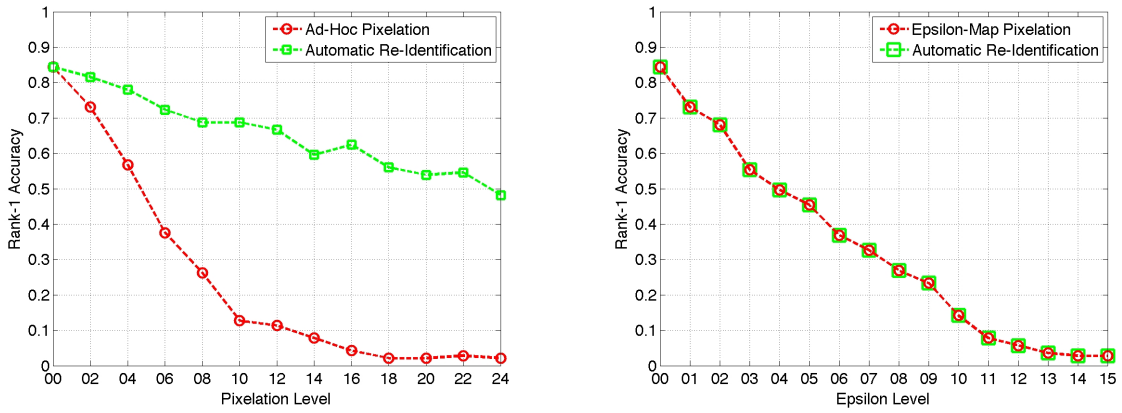
As pointed out in Section II naïve de-identification methods typically apply simple distortion methods such as blurring or pixelation [2], [17], [21], [33]. Intuitively, these approaches strive to implement ϵ-map protection (see Definition III.7), However, since ad-hoc methods do not factor in reference models, privacy is typically not appropriately protected. The algorithm we propose here simply exhaustively searches for the minimal parameter l for a given de-identification function Ψ to fulfill ϵ-map protection. See box IV.1 for a definition of the algorithm. For de-identification functions which remove increasing amounts of information from images for higher parameter settings (e.g. blurring with increasing kernel size), the algorithm is guaranteed to converge, in the worst case to a uniform image. Note that we included a randomization step which adds small, visually undetectable perturbations to the image. This renders common re-identification techniques ineffective (see below).

input : Face image set Γ, reference face model set $\mathcal{R}$, de-identification function Ψ parameterized by $l \in C_\Psi = \{l_1, \ldots, l_t\}$, privacy parameter ϵ
output: De-identified face set Γ^d

for $\mathbf{I} \in \Gamma$ **do**
  $i \leftarrow 0$
  while $i < t$ **do**
    if $p(\Psi(\mathbf{I}, l_i)|R_j) < \epsilon \; \forall R_j \in \mathcal{R}$ **then**
      break
    else
      $i \leftarrow i + 1$
    end
  end
  $\mathbf{I}^d \leftarrow Rand(\Psi(\mathbf{I}, l_i))$
end

Algorithm IV.1: Algorithm for ϵ-map protection.

We evaluate the algorithm using expression-variant images from 150 subjects from the CMU Multi-PIE database [14]. We use a simple normal distribution model over image space distances to compute $p(\Psi(\mathbf{I}, l_i)|R_j)$ from the distance between the de-identified image and the reference model. Here a single gallery image is used as reference model (application scenario (a) in Figure 7). Figure 8(a) shows rank-1 recognition accuracies of a nearest-neighbor classifier on images de-identified using ad-hoc pixelation. In the experiments we used



(a) Ad-Hoc Pixelation (b) ϵ-map Pixelation

Fig. 8. Rank-1 recognition accuracies of a nearest-neighbor classifier for different pixelation algorithms. (a) shows results from the ad-hoc algorithm which applies the same level of pixelation to each image. Comparatively high pixelation levels are necessary to ensure sufficient privacy protection. The algorithm is easily defeated using automatic re-identification. (b) shows results from the ϵ-map algorithm (with ϵ ranging from $1e-3$ to $1e-7$. Automatic re-identification is prevented using a random perturbation step in the algorithm. A more gradual reduction in recognition accuracies is achieved while requiring comparatively lower levels of pixelation than the ad-hoc algorithm.

neutral expressions as gallery and smile expressions as probe. Comparatively high levels of pixelation are necessary to ensure sufficient privacy protection for the ad-hoc algorithm. Pixelated images can be re-identified effectively by automatically detecting the pixelation level, applying the same pixelation to the gallery and running the classifier. This is termed parrot recognition in [22]. In contrast the random perturbation step in the ϵ-map algorithm prevents automatic parrot recognition, so re-identification is thwarted (see Figure 8(b)). The ϵ-map algorithm offers a more gradual reduction in recognition accuracies with decreasing ϵ levels, while requiring comparatively lower levels of pixelation then the ad-hoc algorithm.

B. Algorithm for (ϵ, k)-map Protection

The goal for (ϵ, k)-map protection is to let the de-identified image "blend in with the crowd" [3] by modifying it to be equally likely under multiple face models. Intuitively we could encode this goal directly by finding the de-identified vector $\mathbf{I}^d$ as solution to a minimization problem which equalizes the likelihood of $\mathbf{I}^d$ under all face models R_i in the reference set:

$$\mathbf{I}^d = \arg\min_{\mathbf{I}'} \sum_{m,n} \|p(\mathbf{I}'|R_m) - p(\mathbf{I}'|R_n)\|^2$$

This equation however is minimized by the trivial solution $\mathbf{I}' = 0$ and it does not factor in the magnitude of change from the original data vector. A better approach is therefore to find the de-identified vector $\mathbf{I}^d$ as update to the original data vector $\mathbf{I}$: $\mathbf{I}^d = \mathbf{I} + \Delta\mathbf{I}$ and compute the update $\Delta\mathbf{I}$ as solution to the minimization problem

$$\arg\min_{\Delta\mathbf{I}} \sum_{m,n} \|p(\mathbf{I}+\Delta\mathbf{I}|R_m) - p(\mathbf{I}+\Delta\mathbf{I}|R_n)\|^2 + \lambda\|\Delta\mathbf{I}\|^2 \quad (1)$$

with reference face models R_m, R_n. The resulting de-identified vector $\mathbf{I}^d$ is equally likely under all models and can therefore not be reliably identified. The regularization term $\|\Delta\mathbf{I}\|^2$ ensures that a solution is found which minimally alters the original data vector.

Many possibilities exist for the modelling of the class-conditional density function p. In its simplest form we can use a unit variance normal distribution over reference images which reduces Eqn. (1) to an expression minimizing the pairwise distances

$$\arg\min_{\Delta\mathbf{I}} \sum_{i,j} \left| \|\mathbf{I} + \Delta\mathbf{I} - \mathbf{I}_i\|^2 - \|\mathbf{I} + \Delta\mathbf{I} - \mathbf{I}_j\|^2 \right|^2 + \lambda\|\Delta\mathbf{I}\|^2 \quad (2)$$

for the reference images $\mathbf{I}_i, \mathbf{I}_j$. For a given set of images $\mathbf{I}_i, \mathbf{I}_j$ and an input image $\mathbf{I}$ we can compute $\Delta\mathbf{I}$ directly as least-squares solution to the equivalent expression

$$\sum_{i,j} \|2\Delta\mathbf{I}^T\mathbf{I}_j - 2\Delta\mathbf{I}^T\mathbf{I}_i + c_{i,j}\|^2 + \lambda\Delta\mathbf{I}^T\Delta\mathbf{I} \quad (3)$$

with $c_{i,j} = \|\mathbf{I}_i\|^2 - \|\mathbf{I}_j\|^2 + 2\mathbf{I}^T\mathbf{I}_j - 2\mathbf{I}^T\mathbf{I}_i$. All constraints for $\Delta\mathbf{I}$ can be combined into a single linear system

$$\begin{bmatrix} \mathbf{Q}_{2,1,\lambda} \\ \mathbf{Q}_{3,1,\lambda} \\ \vdots \\ \mathbf{Q}_{n,n-1,\lambda} \end{bmatrix} \Delta\mathbf{I} = \begin{pmatrix} -\frac{1}{2}c_{1,2}\Delta\mathbf{I}_{2,1} \\ -\frac{1}{2}c_{1,3}\Delta\mathbf{I}_{3,1} \\ \vdots \\ -\frac{1}{2}c_{n-1,n}\Delta\mathbf{I}_{n,n-1} \end{pmatrix} \quad (4)$$

with $\mathbf{Q}_{j,i,\lambda} = \Delta\mathbf{I}_{j,i} \times \Delta\mathbf{I}_{j,i} + \frac{\lambda}{4}\mathcal{I}$ and the identity matrix $\mathcal{I}$. The system in Eqn. (4) can be solved directly, achieving privacy protection at the $\epsilon = 0$ level. The constraints for a set of images can be combined into an even larger linear system and then solved concurrently.

We evaluate the algorithm using images of 249 subjects from the CMU Multi-PIE database [14] recorded in frontal pose and displaying neutral expressions. Images of five illumination conditions per subject are included in the dataset. Here, one illumination image is used as probe and four images per

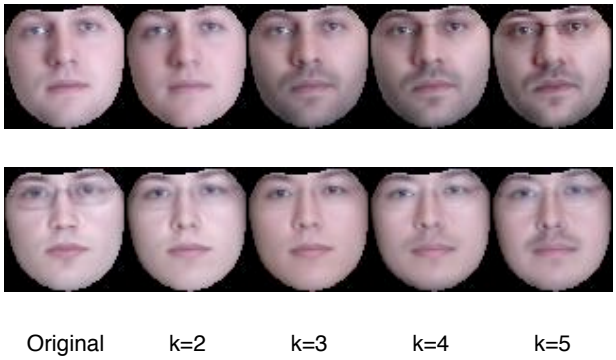


Fig. 9. Example images showing the results of applying the proposed algorithm to images of the Multi-PIE database.

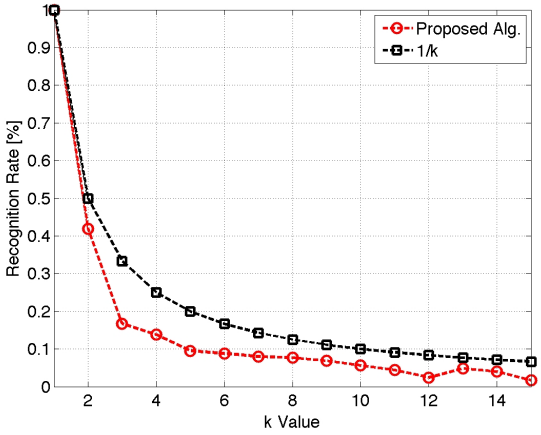


Fig. 10. Rank-1 recognition performance using a nearest-neighbor classifier on images de-identified with the (ϵ, k)-map algorithm with $\epsilon = 0$. The underlying dataset contains five images each of 249 subjects from the CMU Multi-PIE database. Performance stays below the $\frac{1}{k}$-level unlike in the case of images de-identified using k-Same (see Figure 5(b)).

subject are used in the gallery. Figure 9 shows example images of faces at different levels of de-identification. Figure 10 shows rank-1 recognition accuracies of a nearest-neighbor classifier on images de-identified with the (ϵ, k)-map algorithm (with $\epsilon = 0$). Even though the reference set contains multiple images per subject, the algorithm is able to produce appropriately de-identified images.

V. Discussion

In this paper we introduced a general framework for image de-identification and described three privacy protection models: ϵ-map, wrong-map, and (ϵ, k)-map. Our framework subsumes previously proposed ad-hoc methods such as pixelation and blurring and extends the k-Same family of algorithms. We described two algorithms implementing the ϵ-map and (ϵ, k)-map models. In experiments using images from the CMU Multi-PIE database we demonstrated successful de-identification as measured using a simple nearest-neighbor classifier. In previous work we experimentally showed that commercial face recognition systems did not perform better than predicted by the protection model [12]. We plan on repeating these experiments with the newly proposed de-identification algorithms.

The de-identification algorithms discussed here can operate directly on images as well as on parameter vectors extracted using Active Appearance Models [5], [20]. As a consequence, integration of de-identification into our real-time face tracking system [13], [20] is straightforward.

VI. Acknowledgements

We would like to thank the members of the Data Privacy Lab for discussions as well as the anonymous reviewers for their comments. This work was supported by the National Institute of Justice, Fast Capture Initiative, under award number 2005-IJ-CX-K046.

References

[1] S. Avidan and M. Butman. Blind vision. In *European Conference on Computer Vision*, 2006.

[2] M. Boyle, C. Edwards, and S. Greenberg. The effects of filtered video on awareness and privacy. In *ACM Conference on Computer Supported Cooperative Work*, pages 1–10, Philadelphia, PA, Dec 2000.

[3] S. Chawla, C. Dwork, F. McSherry, A. Smith, and H. Wee. Toward privacy in public databases. In *2nd Theory of Cryptography Conference (TCC)*, pages 363–385, 2005.

[4] R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *International Conference on Automatic Face and gesture recognition*, 2002.

[5] T. Cootes, G. Edwards, and C.J. Taylor. Active appearance models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(6), 2001.

[6] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits And Systems for Video Technology*, 14(1):21–30, 2004.

[7] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience, 2000.

[8] F. Dufaux and T. Ebrahimi. Scrambling for video surveillance with privacy. In *IEEE Workshop on Privacy Research in Vision*, 2006.

[9] F. Dufaux, M. Ouaret, Y. Abdeljaoued, A. Navarro, F. Vergnenegre, and T. Ebrahimi. Privacy enabling technology for video surveillance. In *Proc. SPIE 6250*, 2006.

[10] EPIC. Video surveillance. http://www.epic.org/privacy/surveillance/.

[11] D.A. Fidaleo, H.-A. Nguyen, and M. Trivedi. The networked sensor tapestry (nest): a privacy enhanced software architecture for interactive analysis of data in vide-sensor networks. In *Proceedings of the ACM 2nd international workshop on video surveillance and sensor networks*, 2004.

[12] R. Gross, E. Airoldi, B. Malin, and L. Sweeney. Integrating utility into face de-identification. In *Workshop on Privacy Enhancing Technologies (PET)*, June 2005.

[13] R. Gross, I. Matthews, and S. Baker. Active appearance models with occlusion. *Image and Vision Computing*, 24:593–604, 2006.

[14] R. Gross, I. Matthews, J. Cohn, S. Baker, and T. Kanade. The CMU multi-pose, illumination, and expression (Multi-PIE) face database. Technical Report TR-07-08, Carnegie Mellon University, Robotics Institute, 2007.

[15] R. Gross, L. Sweeney, F. de la Torre, and S. Baker. Model-based face de-identification. In *IEEE Workshop on Privacy Research in Vision*, 2006.

[16] M. Helft. Google zooms in too close for some. *New York Times*, June 1 2007.

[17] S. Hudson and I. Smith. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In *ACM Conference on Computer Supported Cooperative Work*, pages 1–10, Boston, MA, Nov 1996.

[18] T. Kanade. *Picture processing system by computer complex and recognition of human faces*. PhD thesis, Kyoto University, 1973.

[19] I. Martinez-Ponte, X. Desurmont, J. Meessen, and J.-F. Delaigle. Robust human face hiding ensuring privacy. In *Workshop on the integration of knowledge, semantics and digital media technology (WIAMIS)*, 2005.

[20] I. Matthews and S. Baker. Active appearance models revisited. *IJCV*, 60(2), 2004.

[21] C. Neustaedter, S. Greenberg, and M. Boyle. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Transactions on Computer Human Interactions (TOCHI)*, 2005. in press.

[22] E. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying facial images. *IEEE Transactions on Knowledge and Data Engineering*, 17(2):232–243, 2005.

[23] P. J. Phillips. Privacy operating charateristic for privacy protection in surveillance applications. In *AVBPA*, 2005.

[24] Y. Ping and K. Bowyer. Empirical evaluation of advanced ear biometrics. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[25] The Royal Academy of Engineering. *Dilemmas of privacy and surveillance*, 2007.

[26] S. Sarkar, P. J. Phillips, Z. Liu, I. Robledo, P. Grother, and K. Bowyer. The human ID gait challenge problem: data sets, performance, and analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 27(2):162–177, February 2005.

[27] A. Senior, S. Pankati, A. Hampapur, L. Brown, Y.-L. Tian, A. Ekin, J. Connell, C.-F. Shu, and M. Lu. Enabling video privacy through computer vision. *IEEE Security & Privacy*, 3(5), May/June 2005.

[28] B. Sharp. Crime-catching police cameras to scan city soon. In *Rochester Democrat and Chronicle*. August 15, 2007.

[29] L. Sweeney. Computational data privacy protection. Technical Report LIDAP-WP5, Laboratory for International Data Privacy, Carnegie Mellon University, 2000.

[30] L. Sweeney. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 10(5):557–570, 2002.

[31] L. Sweeney and R. Gross. Mining images in publicly-available cameras for homeland security. In *AAAI Spring Symposium*, 2005.

[32] L. Willenborg and T. de Waal. *Elements of statistical disclosure control*. Springer Verlag, 2000.

[33] Q. Zhao and J. Stasko. Evaluating image filtering based techniques in media space applications. In *ACM Conference on Computer Supported Cooperative Work*, pages 11–18, Seattle, WA, Nov 1998.

[34] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.