GARP-Face: Balancing Privacy Protection and Utility Preservation in Face De-identification

Liang Du\textsuperscript{1}, Meng Yi\textsuperscript{1}, Erik Blasch\textsuperscript{2}, and Haibin Ling\textsuperscript{1}

\textsuperscript{1}Department of Computer and Information Science
Temple University
Philadelphia, PA, 19122
\{liang.du,Mengyi,hbling\}@temple.edu

\textsuperscript{2}Air Force Research Lab
Rome, NY, 13441
erik.blasch@rl.af.mil

Abstract

Face de-identification, the process of preventing a person’s identity from being connected with personal information, is an important privacy protection tool in multimedia data processing. With the advance of face detection algorithms, a natural solution is to blur or block facial regions in visual data so as to obscure identity information. Such solutions however often destroy privacy-insensitive information and hence limit the data utility, e.g., gender and age information. In this paper we address the de-identification problem by proposing a simple yet effective framework, named GARP-Face, that balances utility preservation in face de-identification. In particular, we use modern facial analysis technologies to determine the Gender, Age, and Race attributes of facial images, and preserving these attributes by seeking corresponding representatives constructed through a gallery dataset. We evaluate the proposed approach using the MORPH dataset in comparison with several state-of-the-art face de-identification solutions. The results show that our method outperforms previous solutions in preserving data utility while achieving similar degree of privacy protection.

1. Introduction

With advances in digital imaging technologies, it has never been easier to capture and share visual data as it is today. We may take photos or videos conveniently using cell phones or other digital devices, and immediately share them on online platforms. On the other hand, we ourselves are also often under the lenses of surveillance cameras, or filmed by other people, sometimes unknowingly. Accompanied with this digital convenience; however, is the potential privacy leakage for images to be used for identity theft.

An increasing amount of effort has been devoted towards addressing identity theft with imagery, from both academia and industry (See Section 1.1). Large, street-level image collections like Google Street View require automatic systems to detect and blur faces [3]. In television news, we see people whose faces are blocked or pixelated to protect their identities. We encounter a similar problem when distributing research datasets. Some medical face databases [33, 34] are not accessible to other groups, or require intensive manual post-processing for patient privacy.

A focus of previous studies is on reliably detecting facial regions[3]. Once a face is located, it will be either blurred (typically via Gaussian kernels) or blocked so as to obscure the identity. However, facial images contain rich information, such as gender, race and age, etc. This kind of identity-insensitive information is often the data utility that is desired to preserve in many applications involving visual understanding and data mining. Therefore, a successful face de-identification algorithm should balance protecting privacy and preserving utility.

Ad-hoc methods that simply blur the facial part in an image typically perform very well in terms of privacy protection, but cause serious loss of data utility as a side effect. This is obviously not desirable. Researchers have proposed some sophisticated methods for face de-identification [25, 27, 28, 7]. They can be roughly categorized into two categories, namely kSAME based methods [25, 27, 28] and face replacement [7].

The \textit{k}-SAME methods have made the pioneering attempt to borrow the \textit{k}-anonymity concept from privacy research in data mining to de-identify facial images [25, 27, 28]. These methods investigate the \textit{k} nearest neighbors of the query face in the face image set. In this way, the query
face is anonymized among at least $k$ candidates, namely $k$-anonymity. And it guarantees that after de-identification, face recognition accuracy is below $1/k$ [25]. Despite the guaranteed privacy gain and moderate consideration of data utility, the effectiveness of $k$-same methods in preserving data utility is questionable. Intuitively, the similarity of face images are correlated to consensus of utility. However, the face space is highly nonlinear, and affinity in a certain model space does not necessarily mean the fidelity in the semantic utility space. It is likely that faces similar in a general model space may have very different attributes, e.g. two faces may be close neighbors in a particular space, while one is male, and the other is female. Another potential weakness of $k$-same methods is that it is self de-identification done completely inside the original image set. It can be much easier to attack once the image set is leaked. We will address this security weakness by introducing a separate reference gallery.

**Face replacement** methods, such as the Face Swapping [7], replace a probe facial image with a “similar” face in a library $L$. These methods benefit from the use of a library in a way that they can get a similar face with a different identity. However, there are two limitations: First, from a security point of view, this method can be attacked by duplicating the process and inferring the identity of the de-identified face with high confidence as the one returns the same top nearest neighbor. Second, the replacing method, as well as the $k$-same methods, may alter the attributes like gender of a face or create large artifacts.

In this paper we propose a new de-identification framework named **GARP-Face** that bridges the semantic gap between facial appearance and data utility. Unlike previous methods which implicitly use appearance similarity as a measurement for data utility, we select three essential utilities: Gender, Age, and Race information and preserve them explicitly in de-identification. More specifically, we design a structured utility hierarchy based on observation on real image sets, and build a utility specific Active Appearance Model (AAM) for each category. These models are pre-trained on an external image gallery. We also formulate superfaces by aggregation of similar faces in each category. Given an input face image, GARP-Face first determines its gender, age and race attributes using modern facial analysis techniques, then the de-identified face is generated by blending with the GAR representative superface, which is most similar to the original face and has consistent attributes. The pipeline of GARP-Face is summarized in Figure 1.

Compared with previous research, our contributions are mainly two folds:
- Unlike previous approaches which either ignore related utilities or implicitly work on them, our approach explicitly thus more effectively preserves facial image utilities involving gender, age and race information.
- The identity of query faces are diluted using an external gallery. This extension to the $k$-anonymity model further enhances visual privacy protection.

To validate the effectiveness of the proposed GARP-Face algorithm, a de-identification experiment is conducted using the MORPH database [14] involving state-of-the-art face de-identification algorithms. The results clearly demonstrate the superiority of our approach in utility preservation while achieving similar degree of privacy protection.

The rest of this paper is organized as follows. Related work are surveyed in Sec. 1.1. In Sec. 2 we first

---

**Figure 1. Pipeline of Proposed method.** Given an input face image, first its utilities are determined by trained attribute classifiers. Then, it is modeled using corresponding utility specific AAM. Finally, de-identified face is formulated by blending the input and its closest neighboring superface in the selected sub-category.
discusses the formal definition and evaluation of face de-
identification, then introduce our utility preserving GARP
Face De-identifier. In Sec. 3 GARP is evaluated on the large
MORPH public face dataset. Finally, we conclude this pa-
per in Sec. 4.

1.1. Related Work

The problem of preserving privacy in data mining has
been intensively studied [15, 31, 4, 30, 6]. [30] studied
the problem of tradeoff between privacy and utility in data
privacy protection problem. [4] studied the possible draw-
backs of $k$-anonymity, showing that lack of diversity might
lead to the fail of privacy protection. In [15], an attribute
generalization supports as semantic hierarchy. For a sur-
vey of privacy protection, one can refer to [6]. Recently,
there are increasing interests in visual privacy protection
and gaining attraction through several initiatives, such as
COST action IC-1206 1. However, due to the lack of direct
semantic interpretation of visual information like images
and videos, there’s relatively few works on privacy protec-
tion in visual privacy protection. Privacy protection in vi-

1http://costic1206.uvigo.es/

sual analysis recently has increasing amount of research atten-
tion. A survey of privacy preserving video surveillance
is given in [5]. Chen et al. [8] studied the privacy preserv-
ing problem in the context of health care related surveil-

lance. Li et al. [10] uses coprime for privacy protected
video communication. Du and Ling [17] studied the prob-
lem of preservative license plate de-identification. Chan et
al. [1] proposed a method for counting people without ex-

plicit human detection. Upmanyu et al. [19] designed a
secure video sharing system inspired by the Chinese Re-

mainder Theorem which split each frame into a set of ran-
dom images. Schiff et al. [11] uses markers in surveillance
videos to detect persons whose identities are sensitive and
thereafter hide such information by masking. In [12] an
algorithm is presented to classify covert photos that often
convey a privacy leakage.

Recently, privacy-preserving biometric identification at-
tracts interests from many researches [36, 37, 38, 39, 40,
41, 42]. Ross and Othman [23] explore the use of visual
cryptography for imparting privacy to biometric data such
as face images. In [35], it introduced a face identification
system which reduces the privacy impact of camera based
surveillance. Boult [24] explores privacy protection through
invertible cryptographic obscuration.

Face de-identification is an important tool for visual pri-

vacy protection. Many face de-identification methods fo-
cus on face detection and simply blur or cover the detected
faces, which often brings unpleasant artifacts to the data and
damages data utility. Some researchers try to remedy this
problem by applying a more “careful” blurring or masking.
For example, a person de-identification method is proposed
in [22] to de-identify a person but retain his/her action in-
formation. It implicitly uses the human action as the data
utility. However, this utility is very limited and many im-
portant attributes (e.g., gender) are lost. In addition, how to
conduct a “sufficient blur” itself is non-trivial [3].

The $k$-same de-identification, which is based on the $k$-
anonymity framework introduced by Sweeney [18], guaran-
tees that each de-identified facial image represents at least
$k$ faces in the gallery, therefore limiting face recognition
performance to $1/k$. Its variants $k$-Same-Eigen, $k$-Same-M
(Model) [25, 27] and multi-factor models [28] adopt differ-
ent face modeling to produce results representative of the
entire $k$ gallery.

Replacing the facial image with a “similar” face in a li-

brary $L$ is another way for face de-identification. Bitouk et
al. [7] proposed an automatically face replacement method,
Face Swapping, which replaces a target face by a similar
face in a large pre-constructed library. It first detects all
faces in the input image, and selects similar candidate faces
from a face library. Then, it adjusts the input image to seam-
lessly blend in the top-ranked candidate faces.

As mentioned in the introduction, our work is closely
related to both $k$-same and face replacement methods. The
main difference between GARP and the $k$-same methods
lies in the separation of probe images and gallery images;
while GARP differs with face replacement by incorporating
the $k$-anonymity concept. More detailed discussion is given
in Sec. 2.

2. Utility Preserving Face De-identification

In this section, we first present the formal definition of
face de-identification and its contradictory requirement to
protect privacy while retaining utility. As well as setting up
the problem, this discussion also inspires our utility preserv-

ing GARP Face De-identifier. Then, we present the GARP
framework and which details on two key aspects: utility de-

termination and utility specific AAMs.

2.1. Face De-identification

Here we provide the definition of face de-identification
problem, and then introduce the evaluation criterion balanc-
ing privacy and utility concerns.

Given a set of probe faces $P$ and a set of reference faces
$R$, face recognition is a function $f : \{P \rightarrow R\}$ which
associates a probe face to a unique reference face. Face de-
identification can be viewed as a transformation function $\delta$
from the original face image set $\mathcal{I} = \{I_1, I_2, ..., I_l\}$ con-
taining $l$ face images to a set of de-identified face images
$\hat{\mathcal{I}} = \{\hat{I}_1, \hat{I}_2, ...\hat{I}_l\}$, so that

$$\delta(I_i) = \hat{I}_i, i = 1, ..., l$$  \hspace{1cm} (1)

The de-identification function $\delta$ intends to decrease recog-
nition accuracy and protect privacy.
The performance of a de-identification algorithm can be measured in two aspects: privacy gain (PG) and utility loss (UL) [31]. Privacy gain, also viewed as loss of the identify information from an attacker’s point of view, can be expressed as:

\[ PG(\delta, I) = \sum_{i=1}^{l} (P(i|\hat{I}_i) - P(i|I_i)), \]  

(2)

where, with slight overload of notation, we use \( i \) to denote the identity of face image \( I_i \) for the sake of conciseness; and \( P(i|I_i) \) denotes the probability of finding the true identity \( i \) given face image \( I_i \).

Utility of a dataset, whether de-identified or not, is inherently tied to the computation that one may perform on it [13]. Here, we evaluate the utility of the de-identified probe set \( I \) in terms of count querying, which has been widely used as a measurement of data utility [31, 13]. Let \( Q \) denote the count querying operation, then the utility loss can be defined as:

\[ UL(\delta, I) = \sum_{i=1}^{l} (Q(I_i) - Q(\hat{I}_i)). \]  

(3)

2.2. GARP De-Identification

The proposed GARP de-identification mainly consists of the following components: 1) Utility determination, 2) Utility-specific AAM models, and 3) a diverse face gallery. For utility, classifiers for selected attributes are trained to determine which utility specific model should be applied to the query face. Since the faces and utilities cannot be well captured by a single general model, we propose to build an attribute specific face model using AAM (Active Appearance Model) [29]. A large and diverse face gallery \( G \) is used for both training AAM models and attribute classifiers. Furthermore, the superfices are generated from \( G \) according to the utility class of the face to-be de-identified.

Figure 1 illustrates the pipeline of our de-identification procedure. Input is a face \( I \) containing explicit identity information: e.g. “Bob”, which is private; and underlying descriptive attributes: e.g. white, male, senior, which are useful and privacy insensitive. First, our pre-trained utility classifiers will extract the attributes of the face image. Next, based on the extracted attributes, we refer to the utility specific AAM model that associates with the particular attributes, and parameterizes the input image in that model space. Last, we refer to the superfices that associate with the particular attributes, and blend the input face with the closest superface to form the de-identified face. We will elaborate on utility determination in Sec. 2.3; and the utility specific AAMs in Sec. 2.4.

Our strategy of forming a superface by aggregation can be interpreted as a \( k \)-anonymity approach in the utility space. If an attribute-specific sub-gallery is divided into \( m \) clusters, it means that the whole attribute class of the whole population is represented with \( m \) super-faces. Then every de-identified face will be undistinguished with \( k' \) faces,

\[ k' = \frac{|C|}{m}, \]

Where \(|C|\) is the cardinality of utility of sub-gallery \( C \). Typ-
Figure 3. Mean face of the general AAM model and utility specific AAM models. On the first row is mean face of the general model. From row 2 to row 5 are mean faces for black male, black female, white male and white female respectively. In each row, from left to right are faces of youth, middle-aged and senior respectively.

2.3. Utility Determination

The utility of a facial image is the informative yet privacy insensitive attributes. Here we select three attributes of common interest: gender, race and age. One can also retain other attributes, or use more sophisticated classifiers to get better accuracy, under the similar framework.

The three attributes are organized in a structured manner. It is noticed that race could affect facial appearance to a very large extent, thus we first determine the race of a face, and then train race-specific gender classifiers and race-specific age classifiers [32]. The implementation of age classifiers is also done in a hierarchical way: two binary classifiers are trained to classifier the three classes: youth, middle-aged and senior. We first apply one classifier to determine if a photo is youth or not. If not youth, we apply the second classifier to distinguish between middle-aged and senior. The attribute hierarchy is shown in Figure 2.

The attribute-level classification problem has been recently studied for face verification and many other computer vision problems [2, 20]. In this paper, the gender and race classifiers are trained using adaboost classifiers and the Haar features. Age classifiers are trained using adaboost classifiers and Gabor features. These attribute classifiers produce satisfactory accuracies. Results are summarized in Table 1.

2.4. Utility Specific AAM Model

AAM model is a generative parametric model which have been successfully utilized in many face modeling and face tracking applications. It not only seeks to matches the shape of the model but also match the representation of texture over the object [29]. Here, we propose to use utility specific AAMs, in order to explicitly preserve data utility. For each of the 12 attribute classes, we manually label facial shape points for images in the reference gallery $G$, and train 12 separate AAM models. Then the parametrization of a face image is to minimize the mean square error (MSE) of the difference between a utility specific AAM model and the input image. Figure 3 demonstrates mean face of each identified image. The larger $m$ is, the more realistic and visually meaningful the de-identified images are. On the other hand, a larger $m$ leads to a longer computation time. Thus $m$ should be chosen according to the requirement of specific privacy application. Generally, $|C|$ is very large (e.g., in our experiments > 1000), GARP de-identification can achieve near-optimal privacy gain.

The gallery $G$ plays the role of modeling the facial image space. From $G$ we gain knowledge about the attribute space. Thus, $G$ should be diverse enough so that it can represent various types of data utility. Meanwhile, $G$ should be readily to be downgraded to a privacy preserved level in the underlying facial information hierarchy.
Table 1. Accuracy of structured attribute classifiers. Description in the parenthesis is the condition assumed to be true. E.g. Middle-aged (White, not Youth) denotes the accuracy of Middle-aged classifier, on the subset of white and not youth faces.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>0.0572</td>
</tr>
<tr>
<td>Gender (White)</td>
<td>0.0545</td>
</tr>
<tr>
<td>Gender (Black)</td>
<td>0.0552</td>
</tr>
<tr>
<td>Youth (White)</td>
<td>0.1275</td>
</tr>
<tr>
<td>Youth (Black)</td>
<td>0.2270</td>
</tr>
<tr>
<td>Middle-aged (White, not Youth)</td>
<td>0.0836</td>
</tr>
<tr>
<td>Middle-aged (Black, not Youth)</td>
<td>0.0846</td>
</tr>
</tbody>
</table>

AAM space corresponding to the 12 attribute classes.

3. Experiments

Our GARP de-identifier is evaluated against state-of-the-art methods using the publicly available MORPH database [14], which is a large dataset containing 55,000 unique images of more than 13,000 individuals, with diverse demographic information e.g., age, gender and race. Each image is associated with its attribute information. We randomly divide the dataset into a test set and a gallery set. The attributes we test in this experiment are age, race and gender. Age is divided into three groups: youth, middle-aged and senior. Race contains two categories: black and white. Gender contains male and female.

The gallery $G$ consists of 13620 individual facial images covering various facial attributes. The set of images to be de-identified is a subset $I$ of 1200 facial images randomly sampled from the MORPH database which is not overlapping with $G$.

We compare the proposed GARP de-identifier to two face de-identification algorithms. One is the $k$-same algorithm as described in [27]. This method works solely on the input image set, and de-identify each image using the affine combination of its $k$-nearest neighbors. We also implemented an model-based de-identifier which uses one general AAM model (See Procedure 1).

As mentioned in Sec. 2.2, GARP, as well as the other two approaches, fall into the $k$-anonymity framework. Thus, when choosing the same $k$, the recognition accuracy of de-identified faces after apply any of these algorithms share the same upper bound $1/k$. We can achieve desirable privacy protection level by setting proper $k$. In other words, the privacy gain of these methods are controllable and comparable. Therefore, here we set $k$ to 30, and focus on the utility side. The utility loss is measured following equation 3, with a small twist. We normalize it using the size of input image set $I$, so that $UL$ is within $[0, 1]$ thus more interpretable.

Table 2 shows the de-identification results. We evaluate utility loss on all aspects combined, and on each utility (age, gender, race) separately. From Table 2 and Figure 5, we can see that GARP ensures significantly lower data utility loss in all categories after de-identification, even with near optimal privacy gain achieved. Visual results are demonstrated in Figure 4.

### Procedure 1 General model-based De-identification

**Input:** Probe face set $I = \{I_1, I_2, \ldots I_l\}$; Reference gallery $G$;

**Output:** De-identified face set $I_d$;

1. Initialize an empty set $I_d$;
2. Train a general AAM model $M$ on $G$;
3. Generate superfaces $S = \{S_1, S_2, \ldots S_n\}$ on $G$;
4. for $i = 1$ to $l$ do
5. Represent $I_i$ using AAM model: $p_i = M(I_i)$;
6. Find the superface $S_k$ which is the closest neighbor of $p_i$, $k \in \{1, 2, \ldots, n\}$;
7. Blend $S_k$ into $p_i$ and add the de-identified face $\hat{I}_i$ to $I_d$;
8. end for

### Procedure 2 GARP-Face De-identification

**Input:** Probe face set $I = \{I_1, I_2, \ldots I_l\}$; Reference gallery $G$;

**Output:** De-identified face set $I_d$;

1. Initialize an empty set $I_d$;
2. Train attribute classifiers $C_s$ on $G$;
3. Divide $G$ into sub-galleries $G_1, G_2, \ldots G_m$ according to attributes;
4. Train a utility specific AAM $M_j$ on each $G_j$, $j \in \{1, 2, \ldots, m\}$;
5. Generate superfaces $S_j = \{S_{j1}, S_{j2}, \ldots S_{jn}\}$ on each $G_j$;
6. for $i = 1$ to $l$ do
7. Determine the attributes of $I_i$ using $C_s$, find the corresponding sub-category $j$;
8. Represent $I_i$ using AAM model $M_j$: $p_i = M_j(I_i)$;
9. Find the superface $S_{jk}$ which is the closest neighbor of $p_i$, $j \in \{1, 2, \ldots, n\}$;
10. Blend $S_{jk}$ into $p_i$ and add the de-identified face $\hat{I}_i$ to $I_d$;
11. end for
Table 2. Utility losses. Utility losses of $k$-same, general AAM model-based de-identification and proposed GARP-Face are measured using normalized count querying. The last row is the utility loss of three attributes combined together.

<table>
<thead>
<tr>
<th></th>
<th>$k$-same</th>
<th>Gen. AAM</th>
<th>GARP-Face</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>0.4818</td>
<td>0.3727</td>
<td>0.0897</td>
</tr>
<tr>
<td>Gender</td>
<td>0.1469</td>
<td>0.3139</td>
<td>0.1372</td>
</tr>
<tr>
<td>Age</td>
<td>0.3606</td>
<td>0.4056</td>
<td>0.0878</td>
</tr>
<tr>
<td>Combined</td>
<td>0.4897</td>
<td>0.5106</td>
<td>0.1173</td>
</tr>
</tbody>
</table>

4. Conclusion

Face de-identification is an important component in visual privacy protection. In this paper, we studied the objective of face de-identification, and developed a novel face de-identifier which explicitly addresses utility concern. We demonstrated that our GARP de-identifier outperforms other state-of-the-art methods, following the evaluation criterion combining privacy protection and utility preservation. In future research, we plan to apply the methodology to other kinds of visual information, including soft biometrics traits which could be auxiliary information for identification. Additionally, a more accurate attribute classifier can be developed to improve the quality of this de-identifier.

Acknowledgement

The authors would like to thank anonymous reviewers for valuable suggestions to improve the paper. The work is supported in part by the NSF Grants IIS-1218156 and CAREER Award IIS-1350521.

References


