De-Identifying Facial Images Using Projections on Hyperspheres

Panteleimon Chriskos

Olga Zoidi

Anastasios Tefas

Ioannis Pitas

AIIA Lab
Aristotle University of Thessaloniki
Greece

Workshop on De-identification for privacy protection in multimedia
May 4 2015, in conjunction with IEEE FG 2015

Outline

- Introduction
- Projections on hyperspheres
- Support Vector Data Description
- Experiments
- Future directions

Introduction

- Sharing of personal image data
  - Social media
  - Medical data
  - Etc
- Face de-identification is an emerging need for privacy protection
- Different applications have different needs
  - Prevent automatic recognition, allow recognition from humans.
  - Prevent identity recognition allow data utility on other concept recognition (e.g., expressions, age, gender, etc)

Introduction

- A novel method was developed that successfully hinders automatic face recognition.

- The method:
  - maintains enough visual data so that viewers can identify the person or persons in a scene.
  - maintains the image quality so that the end product can be considered acceptable for everyday use.
  - de-identifies faces using projections on hyperspheres.
  - Can be easily extended for specific utility preservation and/or k-anonymity.

De-identification based on Projections

- The proposed face de-identification method utilizes projections on hyperspheres.
- The goal is to manipulate facial images in order to reduce facial identification by software agents.
- Two methods were developed:
  - Projection De-Identification on Origin (PDID-O)
  - Projection De-Identification on Mean Image (PDID-M)

Projections on Hyperspheres

- A hypersphere S^{n-1} centered at the origin is defined as:

$$S^{n-1} = \left\{ x \in \Re^n : \|x\| = R \right\}$$

where x is a point in the n-dimensional space and R is the radius of the hypershere.

- The projection of a point $x \in \Re^n$ onto S^{n-1} is given by:

$$P_{S^{n-1}}(x) = \frac{R}{\|x\|} x$$

Projection De-Identification on Origin

- The de-identified image I_{DID} is defined as:

$$I_{DID} = \frac{1}{2}\left(\frac{R}{\|I\|} I + \bar{I}\right)$$

where I is the original facial image, R is the hypershpere radius and $\bar{I}$ is the average facial image of the dataset calculated as:

$$\bar{I} = \frac{1}{N_{im}} \sum_{i=1}^{N_{im}} I_i$$

where N_{im} is the total number of facial images

Projection De-Identification on Mean

- The de-identified image I_{DID} is defined as:

$$I_{DID} = \left(\frac{R * \left(I - \bar{I} \right)}{\left\| I - \bar{I} \right\|} + \bar{I} \right)$$

where $\frac{R * \left(I - \bar{I} \right)}{\left\| I - \bar{I} \right\|}$ is the projection of the original facial image onto the hypersphere of radius R centered at $\bar{I}$

Selection of Radius R

- By choosing a small value for radius R the facial images are projected near the hypershpere center and further from their original positions.

- By choosing a large value for R the images are projected close to their original positions.

- Therefore, it is expected that for large values of R the identification error rates will be lower compared to the error rates for smaller values of R.

- But which values of R are considered large/small?

Automatic Selection of Radius R

- Automatic selection of radius R is achieved through the Support Vector Data Description (SVDD) method.

- SVDD is a method for defining the minimum bounding sphere that encompasses most of or all of the training vectors and as such it gives an estimate of the radius R that should be used.

Support Vector Data Description

- SVDD solves the following optimization problem:

$$\min_{R,\xi,\mathbf{u}} \quad R^2 + c\sum_{i}^{N} \xi_i$$

$$s.t. \quad ||\mathbf{x}_i - \mathbf{u}||_2^2 \leq R^2 + \xi_i$$

$$\xi_i \geq 0, \quad i = 1, 2, \ldots, N$$

where R is the radius, $\mathbf{u}$ is the center of the sphere, $\mathbf{x}_i$ are the facial image representations, ξ_i are slack variables, and c is a parameter that describes the importance of the error.

Support Vector Data Description

- The optimization problem can be solved by finding the saddle point of the following Lagrangian:

$$\mathcal{L}(R, \xi_i, \mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = R^2 + c\sum_{i}^{N} \xi_i - \sum_{i=1}^{N} \beta_i \xi_i$$

$$- \sum_{i=1}^{N} a_i \left(R^2 + \xi_i - ||\mathbf{x}_i - \mathbf{u}||_2^2\right).$$

and its optimality conditions:

$$\frac{\vartheta \mathcal{L}}{\vartheta \mathbf{u}} = 0 \Rightarrow \sum_{i=1}^{N} a_i \mathbf{u} = \sum_{i=1}^{N} a_i \mathbf{x}_i, \qquad \frac{\vartheta \mathcal{L}}{\vartheta R} = 0 \Rightarrow \sum_{i=1}^{N} a_i = 1$$

$$\frac{\vartheta \mathcal{L}}{\vartheta \xi_i} = 0 \Rightarrow a_i = c - \beta_i$$

Support Vector Data Description

- By using the Karush-Kuhn-Tucker theorem, the problem is formulated in its dual form as:

$$\max_{\boldsymbol{\alpha}} \sum_{j=1}^{N} a_i \mathbf{x}_i^T \mathbf{x}_i - \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \mathbf{x}_i^T \mathbf{x}_i,$$

under the condition $0 \leq \alpha_i \leq c$ and $\sum_i a_i = 1$.

- Finaly, the radius R is calculated as:

$$R^2 = \{\min ||\mathbf{x}_i - \mathbf{u}||_2^2, \mathbf{x}_i \text{ is a support vector or } a_i > 0\}$$

Experimental Setup

- The effectiveness of the Projection-DID method was tested on two facial image datasets:
  - XM2VTS database having 388 train samples, 256 test samples, and
  - Extended Yale B database having 1209 train samples, 1205 test samples
- Facial image representation was performed through:
  - pixel value vectorization
  - Linear Discriminant analysis (LDA)
- Three classifiers where used in the process:
  - the K-Nearest Neighbour Classifier (KNN)
  - the Nearest Centroid Classifier and
  - the Naive Bayes Classifier.

Experimental Procedure

- The de-identified image representation error and quality preservation were measured with the mean Mean Square Error (mMSE) metric:

$$mMSE = \frac{1}{N_{im}} \sum_{i=1}^{N_{im}} \left[\frac{1}{np} \sum_{j=1}^{np} \left(I_i - \hat{I}_j \right)^2 \right]$$

where np is the image pixel number, I is the de-identified facial image and $\hat{I}$ is the original facial image.

Experimental Results for PDID-O

TABLE I

ERROR RATES FOR PDID-O (XM2VTS)

	Classifiers			mMSE
Radius	KNN	NC	NBC	
10	93.21 %	93.21 %	97.36 %	0.06046
30	93.21 %	93.21 %	93.58 %	0.04818
50	90.57 %	90.57 %	93.58 %	0.03746
60	90.57 %	90.57 %	93.58 %	0.03268
67.4034	90.57 %	90.57 %	93.58 %	0.02939
70	90.57 %	90.57 %	93.58 %	0.02829
80	90.57 %	90.57 %	93.58 %	0.02428
100	49.06 %	48.30 %	61.89 %	0.01745
120	26.04 %	26.04 %	54.72 %	0.01216

TABLE II

ERROR RATES FOR PDID-O (YALE B)

	Classifiers			mMSE
Radius	KNN	NC	NBC	
5	94.94 %	94.19 %	92.94 %	0.04760
10	89.96 %	79.92 %	72.61 %	0.02878
15	60.83 %	88.13 %	82.57 %	0.02038
17.4241	48.30 %	90.37 %	86.14 %	0.02005
20	38.67 %	91.95 %	89.38 %	0.02239

Experimental Results for PDID-O

$R = 10$

$R = 30$

$R = 50$

$R = 70$

$R = 100$

$R = 120$

Experimental Results for PDID-M

TABLE III

ERROR RATES FOR PDID-M (XM2VTS)

	Classifiers			mMSE
Radius	KNN	NC	NBC	
4	96.23 %	96.23 %	96.23 %	0.01954
6	90.19 %	94.72 %	96.23 %	0.01804
8	90.19 %	90.19 %	90.19 %	0.01660
10	90.19 %	90.19 %	90.19 %	0.01522
12	66.04 %	71.70 %	90.19 %	0.01390
14	53.21 %	53.58 %	73.58 %	0.01265

TABLE IV

ERROR RATES FOR PDID-M (YALE B)

	Classifiers			mMSE
Radius	KNN	NC	NBC	
1	96.76 %	92.61 %	88.13 %	0.04384
2	95.02 %	89.21 %	83.32 %	0.03307
3	88.71 %	89.71 %	83.15 %	0.02396
4	76.51 %	89.96 %	81.74 %	0.01652
5	66.14 %	90.54 %	81.41 %	0.01075

Experimental Results for PDID-M

$R = 4$

$R = 6$

$R = 8$

$R = 10$

$R = 12$

$R = 14$

Conclusions

- The developed Projection-DID methods aim to limit the effectiveness of face identification methods while retaining adequate visual quality.
- By applying these methods a high level of privacy can be attained.
- The highest identification error rates achieved were:
  - [PDID-O]97.36% (XM2VTS) and 94.94 % (Yale B) and
  - [PDID-M] 96.23 % and 96.76% (Yale B) .
- Despite the high error rates, the end product of these methods can be characterized as acceptable for everyday use, rendering the Projection-DID method successful in protecting privacy and providing a visually acceptable output.

Future research directions

- Defining multiple hyperspheres centered at different image samples to retain k-anonymity.
- Defining multiple hyperspheres in order to keep specific required utility
  - One or more hyperspheres for each facial expression, gender, age interval, etc
- The hypersphere can be defined based on a privacy key to prevent “parrot recognition”

Acknowledgment

The research leading to these results has been partially supported from COST, Action IC1206 and the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 287674 (3DTVS).

Thanks for the attention

Q & A