Attributes Preserving Face De-Identification

Bin Yan, Mingtao Pei, Zhengang Nie
Beijing Institute of Technology
Beijing, 100081, P.R. China.
2120171086@bit.edu.cn,peimt@bit.edu.cn,zhengang.nie@bit.edu.cn

Abstract

In this paper, we propose a Face de-identification method to remove the identification information of a person while maintaining all the face attributes such as expression, age and gender. Motivated by the k-Same algorithm, our method consists of three steps: first, k face images are selected randomly. These k face images may contain same or different face attributes with the test face image. Secondly, ELEGANT model is employed to transfer attributes from the test face to the k selected faces. After attributes transferring, the k selected faces have the same attributes as the test face. Then we average the k selected faces as the de-identified image of the test face. Experimental results show that our method can de-identify a face image while preserving all of its attributes effectively.

1. Introduction

Nowadays, surveillance cameras are almost everywhere, such as deployed in buildings, airports, train stations and campus. Person are easily captured without their permission in public areas. The captured face images and videos arouses potential privacy concern from the general public[16]. Therefore, face de-identification[13], which aims to eliminate the identity information by modifying a face image, receives more and more attention, and many face de-identification methods are proposed[3, 12, 13, 1, 15].

The early researches on face de-identification focused on naive methods. Black box approach[3] uses a black box to cover the face region (especially eyes, mouth and nose region). Pixelation approach[12] conceals identity information by subsampling the face image. Blurring approach[14] de-identified face by smoothing face region with Gaussian filter. These naive methods thwart recognition systems in a simply way, but because all facial details are obscured, the result is of limited use.

The k-Same algorithm [13] conceals personal identity by creating new faces. Given a test face image and a face image set, k-Same algorithm selects k face images with closest Euclidean distances to the test face image from the face image set, then a new face image is created by averaging the k images. The k-Same algorithm can guarantee the k-anonymity. However, the face attributes are not preserved. K-same-select [6] improves k-Same algorithm by integrating utility into face de-identification. Face images are divided into different subsets, faces in each subset have the same utility(attributes), then, the k-Same algorithm is applied independently to the different subsets to preserve the face attributes. However, one face image may contain many attributes, the number of all the possible combinations of the attributes are very huge, which makes dividing face images into subsets containing every possible combinations of face attributes unrealistic. APFD [8] can de-identify a face image while preserving a large set of facial attributes. Given a test image, attribute classifiers are utilized to estimate N attributes, and k images with the highest attribute similarities are selected to generate the de-identified image. However, for a test image, it can not guarantee that k images with the same face attributes can be found. Therefore, APFD can not guarantee that all of the face attributes are preserved.

In this paper, we propose a face de-identification method to remove the identification information of a person while maintaining all the face attributes such as expression, age and gender. The framework of our method is shown in Fig1. Instead of dividing face images into different subsets with different face attributes, we select k face images randomly, and transfer the face attributes from the test face to the k selected faces by employing the ELEGANT model[18]. After attributes transferring, the k selected faces have the same attributes as the test face. Then we average the k selected faces as the de-identified image of the test face. Our method can guarantee the k-anonymity of the de-identified face image, and can preserve all of the test face's attributes.

The main contributions of this paper are as follows: First, we proposed a effective framework to de-identify face images with attributes preserving. Second, there is no need to divided face images set to different subsets, since the ELEGANT model[18] is employed.

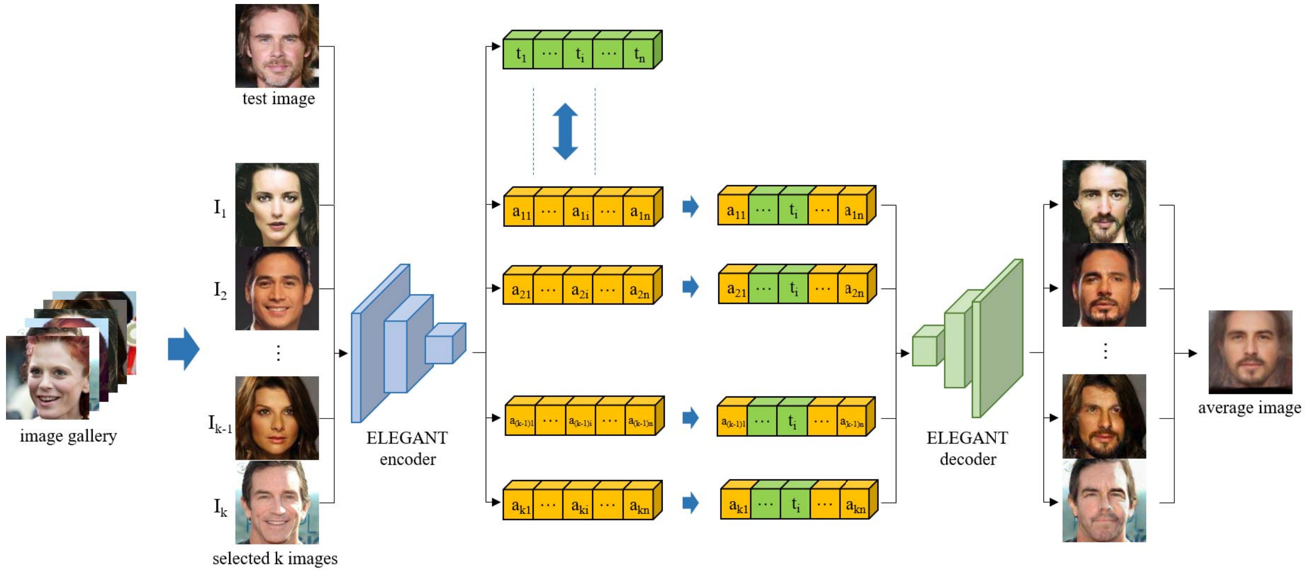

Figure 1. The framework of our method.

The rest of this paper is organized as follows: Section 2 gives a brief review of the related work. Section 3 demonstrates the details of our method. Section 4 shows the experimental results. And section 5 concludes the paper.

2. Related Work

2.1. Face De-Identification

Face de-identification is the process of altering raw face images obtained from surveillance data by replacing the faces with their modified representations [15]. Many face de-identification methods are proposed [17, 3, 12, 13, 1, 15]. These approaches can be divided into two categories: face image de-identification and face video de-identification.

Face image de-identification[8, 2] concerns about face de-identification in images. These methods try to wipe off the personal identity of face by concealing or using a substitute image, such as face replacement and k-same. The early face de-identification methods focus on the effect of de-identification only, leaving a low data utility [8]. The recent research pay more and more attention on the utility of data in the de-identifying process.

Face video de-identification[1, 5] concerns about face de-identification in videos, in which not only the concealing of personal identity, but also the spatial and temporal alignment and algorithm efficiency need to be concerned. In this paper, we focus on face de-identification in images.

2.2. K-same Series Methods

The k-Same[13] algorithm is a simple yet effective de-identification algorithm. The drawback of k-same is that too much information is lost during the process of de-identification. Especially, face attributes, such as age, gender and expression, could not be maintained in the de-identified images. The k-Same-M[7] and k-Same-select[6] try to preserve face attributes in the de-identification process. APFD[8] is also based on k-Same method. It can de-identify a face image while preserving a large set of facial attributes. As we mentioned before, k-Same-select and APFD requires that the face image set contains enough face images that have the same face attributes as the test face image, which is not realistic.

2.3. Face Attributes Transfer

Advances in generative adversarial networks and image-to-image translation in recent years have made it possible to manipulate face attributes. Zhu et al. [21] focus on face attributes translation task, which is considered as image-to-image translation. Different attributes are regarded as different domains, each domain contains one face attribute, face attributes translation can be viewed as domain translation. These approaches exploited the map between source image domain and target image domain. Li et al. [9] and Liu et al. [10] learns binary code for face attributes representation by hash function. However, these methods are not suitable for face attributes transfer task, cause of the lost of large amounts of information when binary code are used to represent face attributes. Zhong et al. [19] utilizes the mid-level CNN features to represent face attributes for both face recognition and facial attribute prediction. Cao et al. [4] proposes a Partially Shared Multi-task Convolutional Neural Network to learn face attributes representation, Xiao et al. [18, 20] learns face attributes in a conciser way that they encode face attributes in a gene mode and different parts in

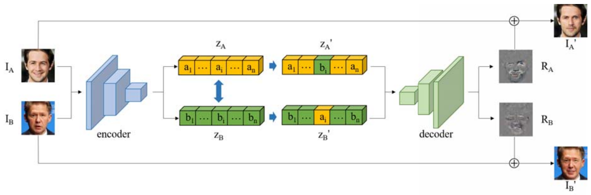

Figure 2. The ELEGANT framework.

the latent space represent different attributes. In this paper, we use the ELEGANT model [18] to represent and transfer face attributes from the test face to the k selected faces for face de-identification.

3. Attributes Preserving Face De-Identification

Our method is based on the k-Same method. That is, given a test face image and a face image set, we select k faces from the face image set, and use the k selected faces to de-identify the test face.

3.1. Face Attributes transfer

Face attributes transfer has been studied widely in recent years. Xiao et al. [18] regards the task of face attributes transfer as a type of conditional image generation. They propose a novel model to achieve face attributes transfer by exemplars, which is called ELEGANT(Exchanging Latent Encodings with GAN for Transferring attributes). The ELEGANT model encode face image into latent space by using an encoder. The latent encoding of face image is divided into separated parts, each part encodes a single face attribute. And a decoder is used to decode the latent encoding to corresponding face image. In the latent space, exchanging the corresponding parts of the two latent encodings means exchanging the relevant attributes of the two original face images. Fig 2 shows the framework of ELEGANT model.

Given two face images I_A and I_B, after encoding by using encoder, we can obtain latent encoding z_A, and z_B, corresponding to face image I_A and I_B, respectively

$$\begin{aligned} z_A &= [a_1, a_2, ..., a_i, ..., a_n] \\ z_B &= [b_1, b_2, ..., b_i, ..., b_n] \end{aligned} \tag{1}$$

where z'_A is the encoding of the non-mustache face image I_A, and z'_B is the encoding of face image I_B with mustache. After obtaining the new encodings, the decoder is employed to do the decode work.

$$\begin{aligned} Dec(z'_A) &= R_A, I_{A'} = I_A + R_A \\ Dec(z'_B) &= R_B, I_{B'} = I_B + R_B \end{aligned} \tag{2}$$

where $Dec()$ is the decoder, R_A and R_B are residual image, $I_{A'}$ and $I_{B'}$ are the new version face image that contain different face attributes with their original images. We call the pair of encoder and decoder the face attributes transferring generator. For more details of the ELEGANT model, please refer to [18].

3.2. Face De-Identification with attributes preserving

As shown in Fig1, given a test face image which contains a male face with mustache and neutral expression, we select k face images randomly from the face image gallery. The selected k face images contain faces with different attributes as the test face image, such as female faces, male faces without mustache and male face with smile expression. We use the ELEGANT model to transfer the face attributes from the test face to the k selected faces. After attributes transferring, the k selected faces have the same attributes as the test face. Then we average the k selected faces as the de-identified image of the test face. Our method can guarantee the k-anonymity of the de-identified face image, and can preserve all of the test face's attributes.

4. Experiments

We train the face attribute transferring generator on celebA [11]. The CelebA dataset contains 202,599 images of 10,177 identities and each image is labeled with 40 face attributes. All the images are aligned according to the 5-point landmarks and cropped to 256×256.

The ELEGANT model can transfer multiple face attributes at the same time. However, in our experiments, we find that given multiple face attributes, transferring them one by one can obtain better results than transferring them together. The reason maybe that learning face attribute transferring generator for multiple attributes is much harder than learning the generator for just one attribute. Therefore, in our experiments, we transfer only one attribute at a time. In other words, we train the corresponding generator for each attribute seperately.

For each attribute $a_i, i = 1, 2, ..., 40$, we divide the CelebA data into two sets according to whether the face images contain attribute a_i. If a face image contain face with attribute a_i, we put it into the positive sample set, otherwise, we put it into the negative sample set. Then we train the face attributes transferring generator for attribute a_i on the positive and negative data set corresponding to attribute a_i.

We build a face image set containing 200 face images which are randomly selected from CelebA dataset. Given a test image, we randomly select $k = 10$ images from the face image set. The k selected images may contain same or different face attributes with the test image. We employ the pre-trained face attributes generators to the k images to manipulate their face attributes, so that the k images contain the same face attributes with the test image. The face images after transferring the attributes are shown in Fig3.

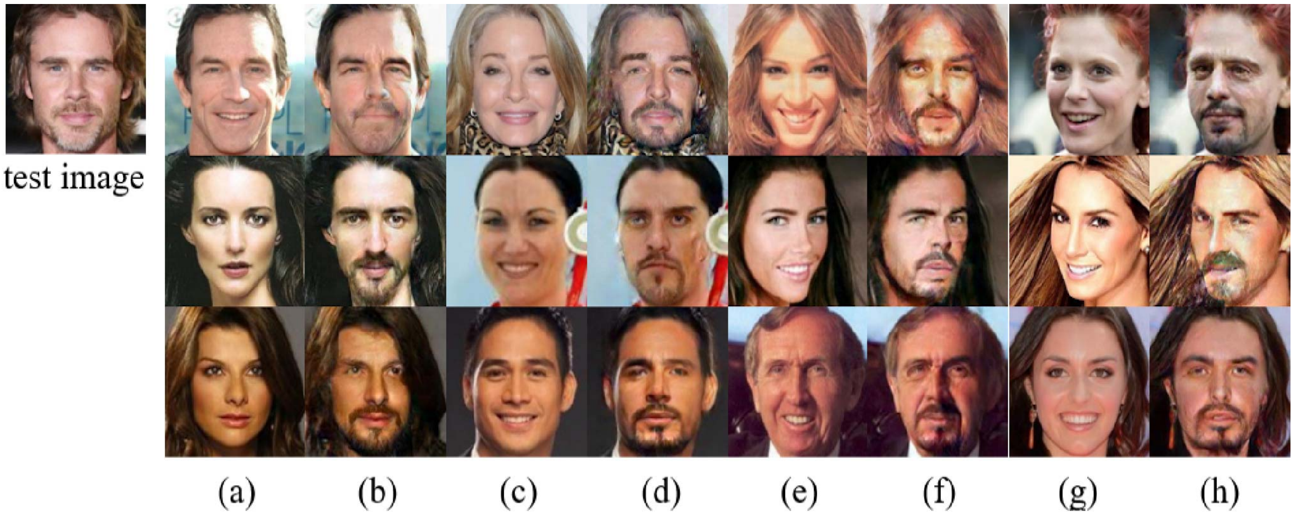

Figure 3. Results of face attributes tranferring. Column $(a), (c), (e)$ and (g) represent the original face images, column $(b), (d), (f)$ and (h) represent the face images after attributes transferring.

In Fig3, column $(a), (c), (e)$ and (g) represent the original face images, column $(b), (d), (f)$ and (h) represent the face images after attributes transferring. For the face image in the first row of column (a), we first add mustache on it, then we change its expression from smile to natural expression, and we obtain the result face image in the first row of column (b), which has the same attributes as the test image. From Fig3, we can see that face attributes can be effectively manipulated by the ELEGANT model.

Finally, the faces in the k selected face images and the test face image have the same face attributes. We simply employ the averaging method to obtain the average face image of the k selected images, and the average face image is regarded as the de-identified image. Fig4 shows the de-identification results of our method. In Fig4, column $(a), (d)$ and (g) are the test face image, column $(b), (e)$ and (h) are the de-identification results without face attributes transferring, and column $(c), (f)$ and (i) are the de-identification results with face attributes transferring. We can see that with face attributes transferring, the de-identified face contains the same face attributes with the test image, while without face attributes transferring, the de-identified face contains different face attributes with the test image.

Meanwhile, compared with APFD[8] and K-same-select [6], our method can obtain satisfactory de-identification results by just using a small face image set, while APFD and K-same-select need a large face image set for just few face attributes. For example, suppose we have $n = 10$ attributes, each attribute has 2 values, then for K-same-select method, it needs to divided the face image set into 1024 subsets, and each subset need to contain at least $k = 10$ images. That means for K-same-select method, it needs more than ten thousand face images, and requests that the attributes are uniformly distributed on these images, to de-identify faces with attributes preserving. While our method just need 200 face images and has no requirement on the attributes distribution.

5. CONCLUSIONS

In this paper we introduced a simply yet effective framework to de-identify face images with attributes preserving. Different with k-Same series attribute preserving methods, our method does not need to divide face images into different subsets with different face attributes. We only need to transfer face attributes from the test face to the k selected faces by employing the ELEGANT model, then the averaged face image of the k selected faces is regarded as the de-identified image of the test face. Our method can guarantee the k-anonymity of the de-identified face image, and can preserve all of the test face's attributes.

References

[1] P. Agrawal and P. J. Narayanan. Person de-identification in videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(3):299–310, 2011.

[2] V. Blanz, K. Scherbaum, T. Vetter, and H. Seidel. Exchanging faces in images. *Computer Graphics Forum*, 23(3):669–676, 2010.

[3] M. Boyle, C. Edwards, and S. Greenberg. The effects of filtered video on awareness and privacy. pages 1–10, 2000.

[4] J. Cao, Y. Li, and Z. Zhang. Partially shared multi-task convolutional neural network with local constraint for face attribute learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4290–4299, June 2018.

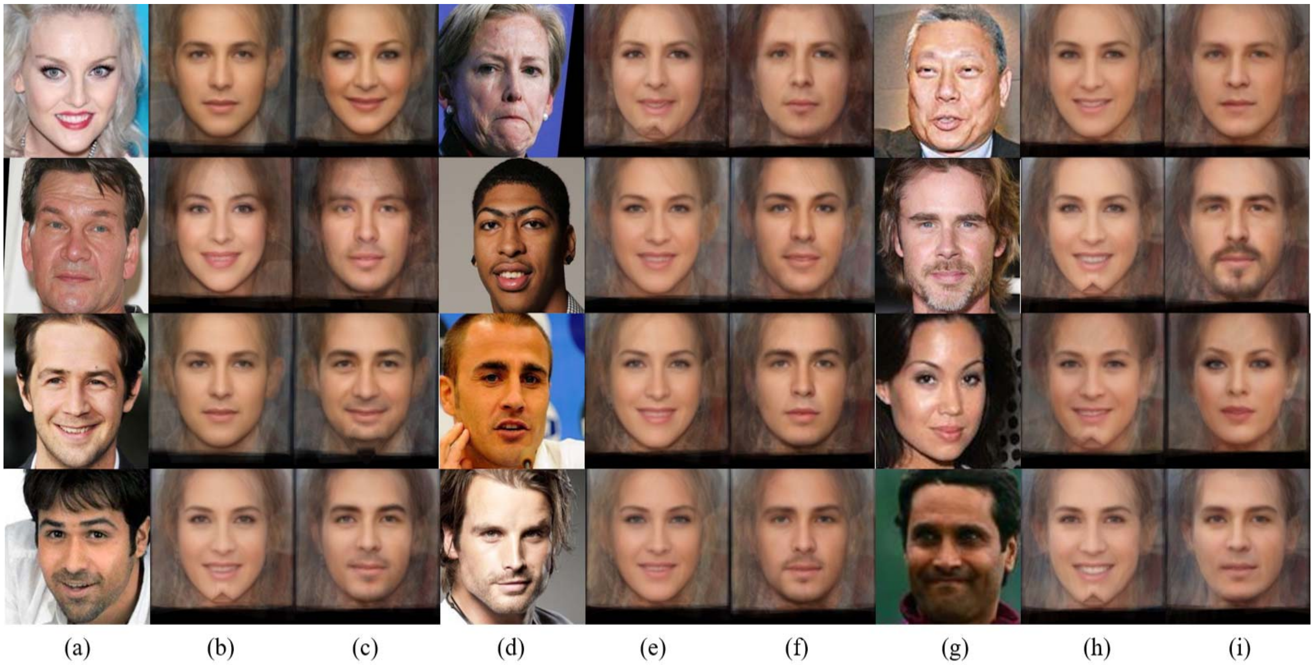

Figure 4. Experiment results of face de-identification with attributes preserving. Column $(a), (d)$ and (g) are the test face image, column $(b), (e)$ and (h) are the de-identification results without face attributes transferring, and column $(c), (f)$ and (i) are the de-identification results with face attributes transferring.

[5] K. Dale, K. Sunkavalli, M. K. Johnson, D. Vlasic, W. Matusik, and H. Pfister. Video face replacement. *Acm Transactions on Graphics*, 30(6):1–10, 2011.

[6] R. Gross, E. Airoldi, B. Malin, and L. Sweeney. *Integrating Utility into Face De-identification.* Springer Berlin Heidelberg, 2006.

[7] R. Gross, L. Sweeney, D. L. T. Fernando, and Baker. Model-based face de-identification. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, pages 161–161, 2006.

[8] A. Jourabloo, X. Yin, and X. Liu. Attribute preserved face de-identification. In *International Conference on Biometrics*, pages 278–285, 2015.

[9] Y. Li, R. Wang, H. Liu, H. Jiang, S. Shan, and X. Chen. Two birds, one stone: Jointly learning binary code for large-scale face image retrieval and attributes prediction. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3819–3827, Dec 2015.

[10] H. Liu, R. Wang, S. Shan, and X. Chen. Learning multifunctional binary codes for both category and attribute oriented retrieval tasks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6259–6268, July 2017.

[11] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. pages 3730–3738, 2014.

[12] C. Neustaedter, S. Greenberg, and M. Boyle. *Blur filtration fails to preserve privacy for home-based video conferencing*. ACM, 2006.

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

[14] S. Ribaric and N. Pavesic. An overview of face de-identification in still images and videos. In *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pages 1–6, 2015.

[15] B. Samarzija and S. Ribaric. An approach to the de-identification of faces in different poses. In *International Convention on Information and Communication Technology, Electronics and Microelectronics*, pages 1246–1251, 2014.

[16] A. Senior. *Privacy Protection in a Video Surveillance System.* Springer London, 2009.

[17] L. SWEENEY. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[18] T. Xiao, J. Hong, and J. Ma. Elegant: Exchanging latent encodings with gan for transferring multiple face attributes. 2018.

[19] Y. Zhong, J. Sullivan, and H. Li. Face attribute prediction using off-the-shelf deep learning networks. *CoRR*, abs/1602.03935, 2016.

[20] S. Zhou, T. Xiao, Y. Yang, D. Feng, Q. He, and W. He. Genegan: Learning object transfiguration and attribute subspace from unpaired data. 2017.

[21] J. Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman. Toward multimodal image-to-image translation. 2017.