CAN I PROTECT MY FACE IMAGE FROM RECOGNITION?

G. A. Kukharev [1], K. Maulenov [2], N. L. Shchegoleva[3,a]

[1] *Saint Petersburg Electrotechnical University "LETI", ul. Professora Popova 5, 197376 St. Petersburg, Russia*

[2] *Non-profit limited company "A.Baitursynov Kostanay Regional University", Baytursynov Street 47, Kostanai, 110000, Republic of Kazakhstan,*

[3] *Saint Petersburg State University, 7– 9 Universitetskaya emb., Saint Petersburg, 199034, Russia*

E-mail:, [a] n.shchegoleva@spbu.ru

The "Fawkes" procedure is discussed as a method of protection against unauthorized use and recognition of facial images from social networks. As an example, the results of an experiment are given, confirming the fact of a low result of face image recognition within CNN, when the Fawkes procedure is applied with the parameter mode = "high". Based on a comparative analysis with the original images of faces, textural changes and graphical features of the structural destruction of images subjected to the Fawkes procedure are shown. In addition to this analysis, multilevel parametric estimates of these destructions are given and, on their basis, the reason for the impossibility of recognizing images of faces subjected to the Fawkes procedure, as well as their use in deep learning problems, is explained. The structural similarity index (ISSIM) and phase correlation of images are used as quantitative assessment tools. It is also noted that facial images subjected to the Fawkes procedure are well recognized outside of deep learning methods. For this purpose, models of two simple systems for recognizing face images subjected to the Fawkes procedure are proposed, and the results of the experiments performed are presented. It is argued that the use of simple face image recognition systems in a computer complex with CNN will make it possible to train such complexes and destroy the myth about the possibility of protecting face images. In conclusion, the question is posed as to whether it is possible to protect your face from recognition.

Keywords: Social networks, unauthorized access to photo, deep learning, face image protection, de-identification, Fawkes procedure, deterministic recognition methods

Georgy Kukharev, Kalybek Maulenov, Nadezhda Shchegoleva

1. Introduction

The end of the 20th century and the beginning of the 21st century were marked by the emergence of different platforms, Q&A services, sites for storing and sharing photos, which was the beginning of the creation of social networks around the world. Initially, social networks covered local groups of schoolchildren and students, music lovers and people with the same profession, and later began to unite various people looking for friends and interlocutors into common social groups. With the development of social networks, people began not only to exchange short messages with each other, but also practically began to live on social networks, presenting themselves and their friends in endless photo and video films. For example, only one social network "Facebook" [1] currently supports almost three billion accounts of active users. Very quickly, the total volume of photo images of faces and photo portraits of users across all social networks, photo sharing platforms and hosting exceeded tens of billions. And all these photos became available, which was used by companies developing face recognition technologies.

For example, the Russian company NTech.Lab with the FindFace technology implemented a search for people's accounts on the VKontakte social network, and the American company Clearview AI, has collected more than three billion photographs from social networks Facebook and Venmo, video hosting YouTube and other similar platforms, and sites [2].

Based on the collected photos, such as, for example, the MegaFace base [3], IT companies have taught neural networks to recognize people by faces, creating two negative precedents. The first is the unauthorized collection of photos of citizens, and the second is the creation of recognition systems trained on this data collection, and the sale of these systems to private companies. This is what led to the possibility of recognizing people from various social sites without the knowledge and consent of these people, and in fact – led to the invasion of the personal space of citizens around the world – that is, to a violation of their privacy.

Naturally, these precedents have caused a wide public outcry and the search for solutions to protect personal photos from recognition. The beginning of this was laid in the solutions for de-identification of images of faces "Face De-identification" – a procedure for distorting the shape or texture of the original image so that a person could understand this face, but not a computer [4, 5]. In this case, both complex methods of distortion were used (for example, modeling face changes based on triangulation procedures), and simple ones – smoothing or noise filters. However, this solution was suitable for private and corporate databases of face images, but not for social networks, since users left their best photos in them and did not want to distort the shape or texture of faces.

The Fawkes procedure has become a new revolutionary solution for de-identification of face images, which implements such a transformation of face images that makes them unsuitable for use in "deep learning" technology [6]. The developers of this procedure claim that during the Fawkes transformation, the face images do not undergo large distortions of the texture, but are destroyed so that they become useless in the CNN training task, and, therefore, will not be recognized by them. Details of these characteristics can be found in the article cited above. Further, the authors of the Fawkes procedure placed in article [7] the addresses of access to their programs and descriptions on the parameters of managing the Fawkes-transformation process, and also invited everyone to use these programs to protect their images before posting them in open social networks.

2. What changes in the image after the Fawkes procedure

Let's try to answer the question of what changes in the original image after the Fawkes procedure, how to see these changes and evaluate them numerically. First of all, we note that if we have only the result of the FAWKES transformation, but there is no original image, then indeed (as the authors write) we will not be able to see and / or evaluate anything. At the same time, it is not possible to use the corresponding pairs of images of faces from articles [6, 7], since hard copies of images contain various additional distortions both in size and in the color gamut of texture and shape. Our methodology will be based on comparing original images with the results of their FAWKES

transformation for electronic original images from the CUFS database [8] and electronic photos of LETI students.

Figure 1a shows an example of an original image and the same image that has passed the Fawkes conversion procedure. The coordinates of the key (anthropometric) points calculated from them are plotted on the faces. Parametric estimates of the differences in textures of these images are also given here [9, 10]: the structural similarity index (Index SSIM = 0.99) and the maximum phase correlation (max Phase Correlation = 0.96). And these estimates testify to the almost complete similarity of the two presented images of faces.

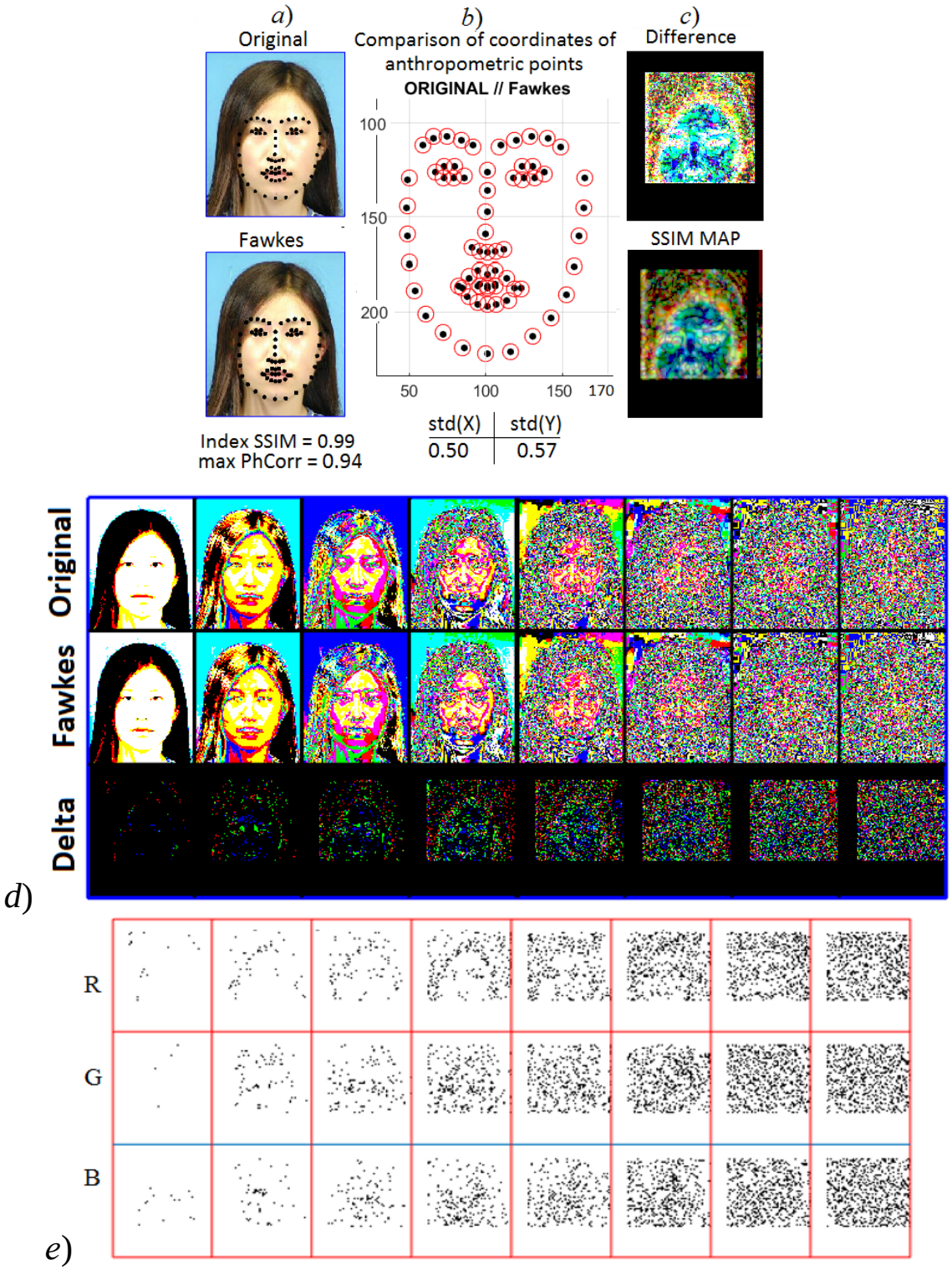

Figure 1. Forms of presentation of the difference between the Original and FAWKES image

In fig. 1b shows the coordinates of the key points of both persons, as well as estimates of the standard deviation of the coordinates of the key points for both persons. These deviations are about half a pixel, which can be attributed to both the differences in the position of the coordinates and the errors in their measurements. In the latter case, to summarize, we note that we do not observe visible changes between the texture and shape of the original image and the image that underwent the Fawkes conversion procedure. This important fact was reported by the authors of the Fawkes procedure.

In fig. 1c shows the absolute difference of textures (Difference) and the matrix of structural similarity (SSIM MAP [9]), between these images. And here we clearly see that there are differences between them (!), Which cover the upper part of the faces. Note by the way that the absolute difference of textures gives not a worse information image than SSIM MAP, although the latter is calculated much more difficult than obtaining the difference (Difference) of two images.

Obviously, the noted differences are "somewhere inside" the images that have passed the Fawkes transformation procedure. Indeed, all changes take place in the bit layers of the image. So in fig. 1d, eight master image "color bit layers" (CBLs), eight CBLs of the Fawkes image, and all eight CBL differences between them are represented. Let's consider them. In each CLS, two types of areas can be seen. Black areas (have zero values), and mark the fact that these areas completely match in the original and converted image. And the colored areas were formed as a result of changes in the original image using the Fawkes procedure and the difference we made between both images.

These areas can be seen more clearly in Fig. 1c, which shows 24 bit layers for the DSC of the difference result. For clarity, these layers are shown in inverse form: the black dots are the differences between the two images, and the white fields are the areas not changed by the Fawkes procedure. As you can see, the greatest changes (or destruction) occurred in LSB layers – in layers with a minimum weight equal to 1. Changes in it from 1 to 0 and vice versa will correspond to a change in image brightness by an amount equal to 1/255. In the next layers to the left of the LSB layer, these changes affect the brightness by a factor of 2/255, etc. In the leftmost column of bit layers, this ratio is 128/255 – or nearly half the luminance range. And, with this in mind, in the Fawkes procedure, the number of changed values in the bit layers decreases from layer to layer from left to right. But even in this case, changes are noticeable in the area of the eyebrows, bridge of the nose, nose and upper lip, as well as the ovals of the faces at eye level, which can be seen in Fig. 2. Here the left images in each pair are originals, and the right ones are the result of their transformation. If you enlarge these images, you can see all the flaws of the Fawkes transformation on these beautiful faces.

Figure 2. Faces after FAWKES procedure

Naturally, users of social networks, posting their best photos, would not want such transformations. And not every ordinary person will be able to perform such a transformation on his own ... and do it every day in the future !!! We conducted additional research and learned, within the framework of the simplest and most popular CNN, to recognize FIs that have undergone the FAWKES procedure with the highest destruction parameter. In addition, we also found that Fawkes face images are well recognized outside of deep learning techniques. Some ideas for these approaches will be presented in the presentation.

We carried out a number of experiments and found that if we use “non-convolutional” algorithms for processing the result of the Fawkes transformation even with the highest destruction parameter, then using the methods presented in [10], we can obtain a set of features that is not sensitive to changes performed by the Fawkes procedure. The system recognition results are shown in

Figure 3. This allows us to conclude that facial images subjected to the Fawkes procedure are well recognized outside of deep learning methods.

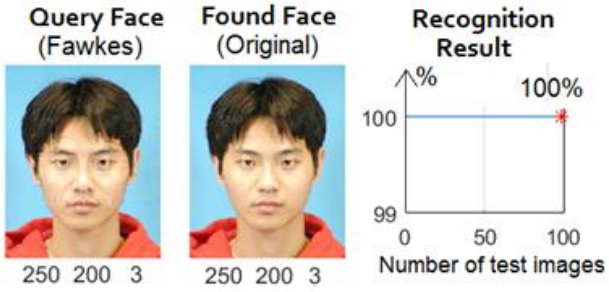

Figure 3. System recognition results

3. Conclusion

The article discussed the Fawkes procedure – as a method of protection against unauthorized use and recognition of facial images from social networks. The textural changes and graphical features of the structural destruction of images subjected to the Fawkes procedure are shown. The reason is investigated that makes it difficult to use images of faces subjected to the Fawkes procedure in deep learning and recognition problems. It is shown that images of faces subjected to the Fawkes procedure are well recognized outside of deep learning methods, which determines the further development of existing protection methods and the emergence of new ones.

References

[1] Top 15 Most Popular Social Networking Sites and Apps [2021]. https://www.dreamgrow.com/feinternational/

[2] HILL, K. The secretive company that might end privacy as we know it. //The New York Times (January 18 2020).

[3] MegaFace and MF2: Million-Scale Face Recognition. http://megaface.cs.washington.edu/

[4] N. L. Shchegoleva Face image models for criminalistics // Izvestia ETU "LETI" V. 7/2016, P. 37– 47.

[5] R. Gross, L. Sweeney, J. Cohn, F. de la Torre, and S. Baker. Preserving Privacy by De-identifying Facial Images In: Protecting Privacy in Video Surveillance, A. Senior, editor. Springer, 2009.

[6] S. Shan, E. Wenger, J. Zhang, H. Li, H. Zheng, B. Y. Zhao. Fawkes : Protecting Privacy against Unauthorized Deep Learning Models// arXiv:2002.08327v2 [cs.CR] 23 Jun 2020. https://arxiv.org/pdf/2002.08327.pdf

[7] Image "Cloaking" for Personal Privacy// https://sandlab.cs.uchicago.edu/Fawkes /

[8] CUHK Face Sketch Database (CUFS)/ http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image %quality assessment: From error measurement to structural similarity" %IEEE Transactios on Image Processing, vol. 13, no. 1, Jan. 2004.

[10] G. Kukharev, E Kamenskaya, Y. Matveev, N. Shchegoleva. Methods of facial images processing and recognition in biometrics, edited by M. Khitrov. St. Petersburg, Politechnika, 2013 – 388 p.