A method for circumventing the biometric identification system based on facial recognition

V O Antonov¹, N Y Svistunov¹, A V Garanzha¹ and I P Bychihin¹

¹North-Caucasus Federal University, Stavropol, Pushkin str., 1, 355017, Russia

E-mail svart.sky@gmail.com

Abstract. The article deals with the task of identifying wanted people using two-dimensional face recognition. A potential method for circumventing the recognition system is described. "Potential" in this case means the absence of precedents for the use of the described method in governmental systems. An illustration of the algorithm for implementing a potential method for disrupting the functioning of the reference samples database of the facial recognition system is given. The results of testing the effectiveness of the presented method on the Microsoft Azure platform are presented. The results of the analysis of the modified image for obvious signs of montage are presented. As a result, the conclusion is presented that it is impossible to implement effective and operational protection of the reference samples database from the presented potential method of circumventing the facial recognition system.

Keywords: Face recognition; Circumventing; Neural network; Image cloaking

1. Introduction

Facial recognition systems are rapidly becoming more widespread and relevant. They can be used in such tasks as: verification and access control, law enforcement, forensic examination. There are many facial recognition technologies available. All of them are based on the principle of uniqueness of geometric parameters: facial features and the shape of the skull of each person are individual [1]. This area is divided into two directions: 2D (two-dimensional) recognition and 3D (three-dimensional) recognition. For widespread use, for example in a citywide system for recognizing people on the wanted list, 2D recognition methods are used. The advantages of this method are: no need to use expensive scanning equipment; ease of obtaining information for analysis (image from conventional city surveillance cameras); using the most easily accessible biometric unit of a person as a reference for comparison – a face image.

There are also disadvantages: low statistical reliability, high requirements for the quality of lighting, necessarily a frontal image of the face, etc. [1]. However, facial recognition systems based on two-dimensional images are now more popular and widely used, especially in the task of identifying criminals on the streets of the city. Of course, there is also a demand for deception of such recognition systems.

Among the recent works devoted to face recognition, it is worth noting the work [2], in which a probability-theoretic model of a halftone image is constructed and a method of identifying a person from a photo of a face based on the optimal Bayesian rule is applied. In another paper [3], the problem of automatic recognition was solved on the basis of the principle of minimum information mismatch. The
method of finding the principal components from the normalized images of the training sample is described in [4]. Work [5] is devoted to the processing of three-dimensional images, but the task of 3D facial recognition has not yet been sufficiently investigated.

2. Facial recognition systems in Russian Federation

In early December 2018, it became known about the intentions of the Moscow city administration to deploy a citywide facial recognition system that will be able to find wanted criminals. The plan includes the creation of a "uniform video space" in which a powerful computer system will receive, store and analyze images from cameras installed in entrances, on streets, at train stations, on transport, at traffic lights and on city highways. According to information given at the beginning of 2019, the Moscow video surveillance network consists of 174 thousand cameras.

The algorithm implemented in the city video surveillance system is based on the use of neural networks and analyzes video recordings from city cameras in real time. The faces on the recordings are scanned so that, if necessary, they can be compared with information in various databases – for example, in the photo databases of law enforcement agencies when it comes to finding an offender.

The video surveillance system operates on a centralized basis [6]. There is a Uniform Data storage and Processing Center (UDSPC), where video streams from all 174 thousand city surveillance cameras flow. The face recognition function is implemented as a separate module of the video surveillance system in the UDSPC. If law enforcement agencies are faced with the task of finding criminals, they use their database of photos, with which, if necessary, they compare the faces that fall into the lenses of cameras. If the images match, the police get a signal about it.

At the same time, the accuracy of recognition depends on many factors: weather conditions, the location of the camera, etc. If the task is to identify a person from a photo from the wanted database and these photos are of good quality, then the system correctly recognizes the person with a probability of 95-97%.

Thus, it can be assumed that there is a possibility of sabotage of this database of reference images by employees who have the right to make changes to the database. Deception of biometric algorithms has two main goals: intruders either try to deceive the system so that it does not recognize them in principle, or to take them for another person. For example, an authorized employee may be asked to upload modified photos of a criminal to the reference database. As a result, the recognition system will no longer be able to react to the criminal on the streets of the city because it does not recognize him because of the distorted reference (standard) in the database, and other employees may not notice the substitution due to the external similarity of the photos.

In support of the assumption of sabotage, some statistics can be cited. According to research by the analytical center "Infowatch" (Russia), such types of incident as "access abuse" and "disloyal behavior" are more common in public authorities and law enforcement agencies than in financial organizations and industry [7]. A comparative infographic is shown in figure 1.
Moreover, there are already known cases of abuse of access to the UDSPC. In Moscow, a criminal case was opened against police officers who, according to investigators, illegally distributed information about citizens obtained from the facial recognition system.

3. Method of circumventing the reference database

2D (two-dimensional) facial recognition technology is based on flat two-dimensional images. Face recognition algorithms use anthropometric parameters of the face, graph models of faces or elastic 2D models of faces, as well as images with faces represented by a certain set of physical or mathematical features (Figure 2).

The considered method consists of distorting the geometry of a person's face in a photo in such a way as to confuse facial recognition systems, while allowing other people to recognize the person. In other literature, this phenomenon is called "cloaking" (Image cloaking, Photo cloaking). Testing has shown that the "cloaking" effect is difficult to recognize when training a neural network and it does not cause errors in training [8]. In other words, the human operators of the facial recognition system will not suspect anything amiss.

As a tool for manually performing the "cloaking" operation, it is proposed to use a graphic editor (for example, Adobe Photoshop [9]), but the final efficiency will depend on the talent of the performer. Automatic execution of "cloaking" can be performed by the De-identification algorithm authored by Facebook AI Research [10], or the Fawkes algorithm authored by the University of Chicago [8]. However, at the moment, the effectiveness of the latter is questionable [11]. It should be noted that the original purpose of these algorithms was not to circumvent governmental information systems by forgery of the reference photo in the database. The primary reason for creating such tools was the need to anonymize people in photos and protect their identity from unauthorized facial recognition systems.
The result is achieved by adding changes to the bitmap image. At its core, "cloaking" does not occur at the pixel level, but at the level of the feature space, that is, pixel changes actually have a deep nature. At the same time, the photo must be modified in a non-obvious way, that is, to preserve the main recognizable at first glance facial features in the image. This is necessary to reduce the probability of detecting a fake photo by a person (human operator). These features include: position, emotional expression, skin tone, hair color, distinguishing features (moles, scars), occlusion (bite), lighting and shadows.

A visual algorithm for a potential method of circumventing the biometric identification system based on facial recognition is presented in figure 3.

![Figure 3. Algorithm for implementing a potential method of circumventing the biometric identification system based on facial recognition](image)

A modification method (represented as a function for extracting anthropometric parameters $\Phi$ and donor bitmaps $D$) is applied to the reference photo of the intruder $Z$ to generate a new image $Z'$.

The database of reference photos of criminals receives a modified image $Z'$ by uploading this image by an accomplice of forgery (access abuse). Thus, there is a deliberately incorrect training of the intruder $Z$ recognition system. As a result, when the system receives new veritable images of the intruder $Z$ from city surveillance cameras, the recognition system either does not recognize the person at all, or "confuses" him with someone similar in anthropometric parameters from the image $Z'$. Please note that the presence of image $D$ in the reference database is not a mandatory factor. At the same time, $D$ must have the anthropometric parameters of the face as opposite as possible from $Z$.

An explanation of why the model trained on distorted photos does not recognize faces on the originals in 2D space, of the four features $A, B, U, T$, is presented in figure 4 [8]. On the left are the boundaries of decision-making when learning on the originals, on the right are the boundaries of decision-making when learning after "cloaking".

![Figure 4. Visualization of decision boundaries by the face recognition system](image)

According to the developers of such algorithms, the efficiency is 100%, and they easily deceive modern face recognition services from Microsoft (Azure), Amazon (Rekognition) and Megvii (Face ++) [8].

At the same time, it is worth noting that the fewer training (reference) images the system uses for each person, the lower its accuracy. The situation when the system uses only one reference image leads to a significant decrease in recognition accuracy. In the scientific community this problem appears under the abbreviations SSPP (Single Sample per Person) [12] or OSS (One Sample Size) [13]. Among the tasks of automatic face recognition, the problem of SSPP occupies a special place.
4. **Validating the effectiveness of the presented method**

To test the effectiveness of the presented method, the Microsoft Azure [14] platform was used. The source data is the original photo of the co-author of the article Ivan Bychihin, and the same photo modified with the help of Adobe Photoshop software [9]. The tools "Plastic" and "Restoring brush" were used. It is important to note that the distinctive features that are important for human perception are preserved. These features include: position, emotional expression, skin tone, hair color, distinguishing features (moles, scars), occlusion (bite), lighting and shadows.

The result of comparing images (Figure 5) is «The two faces belong to different people». At the same time, the Confidence parameter remained quite large (for identical photos it is equal to 1, for photos of dissimilar people it is close to 0).

![Face verification](image)

**Figure 5. The result of the images comparison on the Microsoft Azure platform**

Thus, it is safe to say that the algorithm of a potential method of circumventing the biometric identification system based on facial recognition will work.

5. **Forensic analysis of the modified reference (standard)**

Based on the result obtained, a question arises about methods for detecting a circumvent attempt. It can be assumed that a large number of changes are received in the reference database, so it is not possible to view each change manually. Then the only way is an automatic forensic analysis of the photo uploaded to the database. But at the moment, such tools do not exist, and to create them, you need a very clear result of using forensic filters on the image.

Based on this assumption, the modified image was analyzed for obvious signs of editing. The web tool Forensically was used for the analysis [15].

As a result of using the Noise Level Analysis, Error Level Analysis and Luminance Gradient tools, a clear sign of editing was detected only when considering the result of processing with the Error Level Analysis tool (Figure 6).
Thus, we can conclude that it is possible to automate the process of checking photos by the Error Level parameter and there is a prospect of creating such software.

6. Conclusion

During the review of the principle of functioning of Russian facial recognition systems, it was concluded that there is a potential possibility of unauthorized access or abuse of authority by an employee in order to bypass the algorithm for recognizing the faces of people who are wanted. "Potential" in this case means the absence of precedents for using the described method.

An experiment on the Microsoft Azure platform showed the effectiveness of the described method of sabotaging the database of standards of the facial recognition system.

The experiment on the rapid detection of a modified photo using the tools of forensic image analysis gave a result in one test out of three, which means that there is a prospect for automating the analysis process.

Thus, at the moment there is no effective and operational method of protecting the reference database from the presented potential method of circumventing the facial recognition system.

Developers and administrators of such databases are recommended to pay attention to the method presented in this article and tighten control over changes and additions made to the database.

References

[1] Antonchik A and Deryushev A 2009 Review of methods of face recognition in the image Minsk: Reports of Belarusian State University of Informatics and Radioelectronics

[2] Viola P 2004 Robust realtime face detection International Journal of Computer Vision pp 137-154

[3] Viola P 2001 Rapid object detection using a boosted cascade of simple features IEEE Conf. on Computer Vision and Pattern Recognition pp 511-518

[4] Turk M and Pentland A 1991 Face Recognizing Using Eigenfaces Proc. IEEE pp 586-591

[5] Baev A, Morozovskiy K and Rozhentsov A 2013 Modified generalized Hough transform for processing threedimensional images with unknown parameters of rotation and scaling (Avtometriya vol 2) pp 30-41 (in Russian)

[6] Official site of the Moscow Mayor Conversation with the head of the city video surveillance department of the Moscow Information Technology Department URL: https://www.mos.ru/dit/documents/view/220179220 (accessed: 10.09.2020) (in Russian)

[7] Information security in corporate information systems InfoWatch Analytical report URL: https://www.infowatch.ru/analytics/reports/4609 (accessed: 14.09.2020) (in Russian)

[8] Shawn S, Emily W, Jiayun Z, Huiying Li, Haitao Z and Ben Zhao 2020 Fawkes: Protecting
Privacy against Unauthorized Deep Learning Models 29th USENIX Security Symposium

[9] The official resource of the software provider Adobe Photoshop URL: https://www.adobe.com/ru/products/photoshop.html (accessed: 30.09.2020)

[10] Oran G, Lior W and Yaniv T 2019 Live Face De-Identification in Video Facebook Research

[11] Does the photo protection from face recognition systems works? URL: https://habr.com/ru/post/520194/ (in Russian)

[12] Tan X 2006 Recognition from a Single Sample per Person with Multiple SOM Fusion Lecture Notes in Computer Science vol 3972/2006 pp 128-133

[13] Štruc V 2009 Using Regression Techniques for Coping with the One-Sample-Size Problem of Face Recognition Electrotechnical review vol 76 pp 7–12

[14] Microsoft Azure AI service for analyzing faces in images URL: https://azure.microsoft.com/ru-ru/services/cognitive-services/face/

[15] Forensically Web-based tool for analyzing raster images URL: https://29a.ch/photo-forensics