UDC 004.93

D. Zhuravlov, O. Polshakova

DETECTION OF FACE SPOOFING ATTACKS ON BIOMETRIC IDENTIFICATION SYSTEMS

Abstract: The article considers methods for the detection and recognition of spoofing attacks on biometric protection systems using the human face, analyses their qualitative indicators and analyses the approach using convolutional neural networks that would allow to obtain the best HTER for the future protection system. The obtained result allowed to highlight the advantages and disadvantages in the design of an attack detection system in the considered area of application. The proposed algorithm of the spoofing attack detection system based on convolutional neural networks using image depth maps.

Keywords: anti-spoofing, face identification, convolutional neural networks, key points and features, biometric authentication and identification.

Introduction

Biometric identification of a person is a pretty old idea in the field of information technology for recognizing people, which was tried to be technically implemented. Passwords can be stolen, copied, forgotten, keys can be forged, but the unique characteristics of a person are much harder to forge and lose. Such characteristics can be fingerprints, voice, retinal vascular pattern, gait, etc. Modern human face recognition systems demonstrate a fairly high accuracy, with the advent of large data sets and complex system architectures, it has become possible to achieve face recognition accuracy of up to 0.000001 (one mistake per million), and they are also suitable for transfer to mobile platforms. But the main vulnerability of biometric identification systems is their security.

In order to fake data and impersonate another person in our technical reality, masks are most often used. Masks can be made of materials of completely different quality, from a printed photo of another person held in front of the face to very complex three-dimensional heated masks. They also try to deceive the computer system by presenting someone else's face in front of the camera instead of their own, in a photo or video that is shown on the screen of a tablet or smartphone.

A lot of attention to the topic of spoofing attacks was attracted by a successful attempt to hack the "Face ID" system on the iPhone “X” smartphone, using a complex

mask made of stone powder with special inserts around the eyes that simulate the warmth of a living face using infrared radiation [1]. The presence of such vulnerabilities is very dangerous for banking or government systems of user authentication by biometric data, where the penetration of an attacker causes significant losses. Accordingly, a set of measures to counteract such deception, we call anti-spoofing, it can be implemented in the form of various technologies and algorithms embedded in the module of the identification and authentication system. The most effective anti-spoofing system of measures for the HTER indicator will be determined in this paper.

Analysis of the problem of biometric identification by face

To determine the quality of the system, the HTER metric is often used, which is calculated as the sum of the coefficients of false positive and false negative identifications divided in half:

$$\mathrm{HTER} = (\mathrm{FAR} + \mathrm{FRR})/2 \quad (1)$$

It is worth noting that biometrics systems usually pay the most attention to the FAR indicator in order to do everything possible to prevent an intruder from entering the system and achieve good success in this. The flip side is the inevitable increase in FRR – the number of ordinary users mistakenly classified as intruders, and if for government, defense and other similar systems with sensitive data, this can be sacrificed, then mobile technologies, working with their huge scale, variety of subscriber devices are very sensitive to any factors that may force users to abandon services, so this should be paid special attention when developing a system.

The most popular means of spoofing attacks on facial biometric identification systems are masks or as this method is also called Mask attack. There is nothing easier than to put on a mask that fakes the image of another person's face and present the face to the identification system (fig. 1).

Figure 1. Mask attack example

You can also print a photo of yourself or someone else on a piece of paper and bring it to the camera, this type of attack is called a Printed attack (fig. 2). A more complex type of attack is Replay attack, when the system is presented with the screen of another device (tablet, smartphone), which plays a pre-recorded video with another person. The complexity of this method is compensated by the high efficiency of such an attack, since protection systems often use features based on the analysis of time sequences, for example, tracking blinking, head movements, facial expressions, breathing, etc. All this is easy to reproduce on video.

The last two types of attacks (Replay attack and Printed attack) have a number of characteristic features that allow them to be detected, and thus distinguish a tablet screen or a piece of paper from a real person.

Let us summarize the characteristic features that allow to identify these two types of attacks in Table 1.

Table 1

Printed attack	**Replay attack**
Reduction of image texture quality during printing	Moire
Artifacts of halftone image transmission when printing on a printer	Reflections (glare)
Mechanical artifacts of printing (horizontal lines)	Flat pattern (no depth)
Lack of local movements (for example, blinking)	Borders of an image may be visible
Borders of an image may be visible	

Analysis of existing spoofing attack detection systems

The first approach which will be appropriate to consider is based on the use of the features of image quality deterioration during printing or playback on the screen [2]. On printed images or images on the screens of digital devices, local patterns will be detected, albeit elusive to the eye, but they can be identified, for example, by counting local binary patterns (LBP, local binary pattern) for different zones of the human face image after selection from the frame. This system can be considered the primary source of the whole direction of face anti-spoofing detection algorithms based on image

analysis. If we consider this approach in more detail, then when calculating LBP, each pixel of the image and its eight neighbors are sequentially taken and their brightness is compared, if the brightness is greater than the central pixel, then one is assigned to the matrix corresponding to the size of the image pixel matrix, if less – zero. Thus, an 8-bit sequence is obtained for each pixel. According to the obtained sequences, a pixel-by-pixel histogram is built, which is fed to the input of the SVM classifier. The HTER efficiency for this approach is 15%, which means that a significant part of the attackers overcome the protection without much effort, although it should be noted that most of them are eliminated. The system with a HTER of 15% was tested on IDIAP's Replay-Attack dataset, which consists of 1200 short videos from 50 respondents and three types of attacks: printed attack, mobile attack, and high-definition attack.

An alternative approach to detect spoofing attacks was developed in 2015 by scientist Bukinafit Z. from the University of Oulu. Bukinafit developed an algorithm for alternative image splitting into channels other than traditional RGB, for the results of which local binary patterns were again calculated [3], which, as in the previous method, were fed to the input of the SVN classifier. HTER accuracy calculated on CASIA and Replay-Attack datasets was 3%.

Another existing method for detecting spoofing attacks is image moiré detection. K. Patel from the University of Michigan, USA published an article [4], where he proposed to search for image artifacts in the form of a periodic pattern caused by the superposition of two scans. This approach proved to be effective, showing an HTER of about 6% on the IDIAP, CASIA and RAFS datasets. This was also the first attempt to compare the performance of the algorithm on different datasets.

To detect attempts of Printed Attack, the logical solution was to try to analyze not one image, but their sequence taken from the video stream. For example, Anjos A. and his colleagues from the Idiap Research Institute in Switzerland proposed to extract features from the optical stream on adjacent pairs of frames [5], feed them to the input of a binary classifier and average the results. The approach proved to be quite effective, demonstrating HTER of 1.52% on their own dataset.

Proposed solution

In order to achieve the HTER, which will be smaller, i.e. better than in the considered systems, in this work it is proposed to consider the approach using convolutional neural networks or CNN-networks (CNN – convolutional neural

network). This choice is due to the fact that the neurons inside the CNN layer are connected to only a small part of the neurons of the previous layer, called the receptive field, this approach reduces the amount of RAM that the program uses to store data. Combining a large number of such layers together creates nonlinear filters that become increasingly larger (i.e., a larger area of pixel space is perceived), so that the neural network at the first stage creates a representation of small elements of the input, and then from these elements constructs a representation of larger areas. In CNN network all filters are repeated over the entire visual field of the image. Such repeated nodes apply a common parameterization, weight and bias vector, and form a feature map. That is, all neurons in the specified convolutional layer respond to the same feature within their receptive field. By repeating the nodes in this way, the image features can be detected regardless of their position in the visual field, and thus the property of invariance with respect to the shift is provided.

CNN neural networks achieved a significant reduction in error rate when used for face recognition, 97.6% recognition rate on 5600 still images of more than 10 subjects [6]. CNNs were used to assess video quality in an objective way, after manual training; the resulting system had a very low root mean square error [7].

In this paper, a system using convolutional neural network is considered as an experiment, which will extract human face images using depth maps. Depth map is quite a good feature to determine the plane where the image is located. The main advantage is that the image is on a sheet of paper, there is no "depth" by definition. In the work of Ataum (2017 year), many separate small areas were extracted from the image, depth maps were calculated for them, which were then merged with the depth map of the main image [8].

In a potential protection system, it is proposed to merge the results of two convolutional neural networks (Fig. 2), the first of which will calculate depth maps for frames (parts of the image), and the second – for the image as a whole. When training on datasets, a depth map equal to zero will be associated with the Printed Attack class, and a series of randomly selected areas will be associated with the three-dimensional face model. The depth map does not solve the problem as a whole, only some indicator function characterizing the "depth of the area" was used from it. It is planned that the system will show the HTER value < 1%. Three public datasets will be used to train the neural network – CASIA-MFSD, MSU-USSA and Replay-Attack.

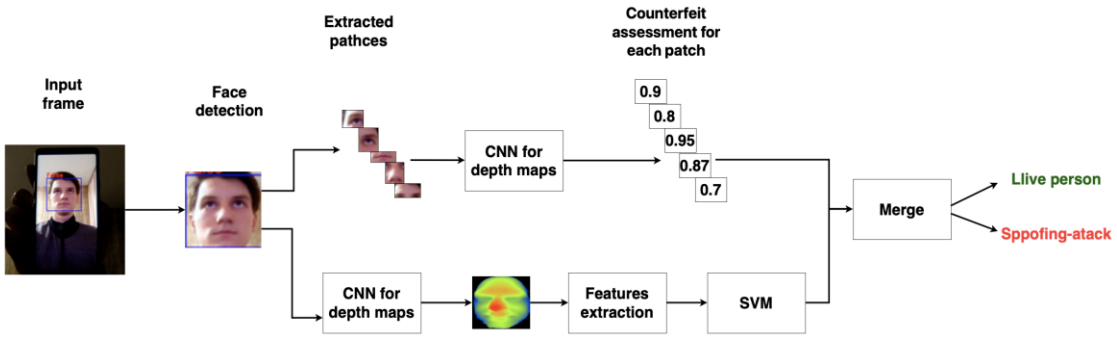

Figure 2. Architecture of the proposed anti-spoofing approach.

Conclusions

This paper analyzes the main types of spoofing attacks on the human face identification system and existing systems for detecting these attacks, their disadvantages and advantages, and presents the architecture of the anti-spoofing approach. As a potential solution, a system based on convolutional neural network using a depth map for classification of image features and an approach with the analysis of the image as a whole object, as well as the analysis of individual parts of this image as frames were proposed. At this stage of the system development, it is clear that to improve the performance, it will be necessary to merge several classification methods. Analysis of lesions, depth map consideration should be used together. A promising option for improving the protection system in biometrics may be the combination of other types of identification and authentication, which will help the system to provide an additional data stream, for example, recording of a person's voice and certain integrated approaches that allow to accommodate several technologies in a single system to detect spoofing attacks on the face identification system.

REFERENCES

1. Face id in businesstransactions. 2017. URL: http://www.bkav.com/d/top-news/-/view_content/content/103968/bkav%92s-new-mask-beats-face-id-in-twin-way-severity-level-raised-do-not-use-face-id-in-business-transactions.

2. Chingovska I. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing / Chingovska Ivana – Martigny, Suisse. 2017

3. Boulkenafet Z. Face anti-spoofing based on color texture analysis / Boulkenafet Zinelabidine – University of Oulu, Finland. 2015.

4. Keyurkumar P. Live Face Video vs. Spoof Face Video / Keyurkumar Patel – Michigan State University, USA. 2020.

5. Anjos A. Motion-based counter-measures to photo attacks in face recognition / Anjos André – Paris, France. 2014.

6. Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Networks. C. 555–559.

7. Callet, Patrick; Christian Viard-Gaudin; Dominique Barba (2006). A Convolutional Neural Network Approach for Objective Video Quality Assessment. IEEE Transactions on Neural Networks. C.1316–1327.

8. Atoum Y. Face Anti-Spoofing Using Patch and Depth-Based CNNs. 2020. URL: http://cvlab.cse.msu.edu/pdfs/FaceAntiSpoofingUsingPatchandDepthBasedCNNs.pdf.