Бекназарова Саида Сафибуллаевна
Тошкент ахборот технологиялари университети, т.ф.д., профессор
Жаумитбаева Мехрибан Караматдин қизи
Тошкент давлат юридик университетининг Ихтисослаштирилган филиали ходими

ANTI-SPOOFING МОДЕЛИ АСОСИДА ИНСОН ЮЗИНИ БИОМЕТРИК ИДЕНТИФИКАЦИЯЛАШНИ АНИҚЛАШ ТИЗИМИ

Бекназарова Саида Сафибуллаевна
Профессор, доктор технических наук, Ташкентский университет информационных технологий
Жаумитбаева Мехрибан Караматдин қизи
Сотрудник Специализированного филиала, Ташкентский государственный юридический университет

СИСТЕМА БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ ЛИЦА ЧЕЛОВЕКА НА ОСНОВЕ МОДЕЛИ ANTI-SPOOFING

Beknazarova Saida
Tashkent University of Information Technologies doctor of technical science, professor
Jaumitbaeva Mekhriban
Staff member Specialized branch of Tashkent State Law University

RECOGNITION SYSTEM OF BIOMETRIC IDENTIFICATION OF A PERSON'S FACE BASED ON THE ANTI-SPOOFING MODEL

Abstract: *Biometric identification is often referred to as pure or real authentication, since it uses not a virtual, but a biometric attribute that is actually relevant to a person. Passwords can be stolen, spied on, forgotten, keys can be forged. But the unique characteristics of the person himself are much more difficult to fake and lose. This can be fingerprints, voice, drawing of blood vessels of the retina, gait, etc. Face recognition looks like a very promising direction for use in the mobile sector. If everyone includes long been used to using fingerprints, and technologies for working with voice are gradually and rather predictably developing, then with face identification the situation is rather unusual and worthy of a small excursion into the history of the issue.*

Keywords: *recognition systems, technology of biometric identification, person's face, the anti-spoofing model, authentication*

Today's detection systems show huge accuracy. With the advent of large datasets and complex architectures, it has become possible to achieve face recognition accuracy of up to 0.000001 (one error in a million!), And they are already suitable for portability to mobile platforms. Their vulnerability became the bottleneck.

In order to personify another person in our technical reality, and not in the film, masks are most often used. They also attempt to fool the computer system by presenting someone else instead of their face. Masks come in very different qualities, from a printed photo of another person held in front of their face to highly complex 3D heated masks. Masks can either be presented separately in the form of a sheet or screen, or worn on the head.

The availability of such vulnerabilities is really hazardous for banking or government systems of user authentication by face, where an intruder's penetration entails significant losses.

We can to call an attempt to deceive the identification system by presenting it with a fake biometric parameter a spoofing attack.

And, we will call the complex of protective measures to resist such deception anti-spoofing. It can be implemented in the form of a variety of technologies and algorithms built into the pipeline of the identification system.

The ISO offers a somewhat extended set of terminology, with terms such as presentation attack – attempts to force the system to incorrectly identify a user or to allow him to avoid identification by displaying a picture, recorded video, and so on. Normal (Bona Fide) – corresponds to the normal algorithm of the system, that is, everything that is NOT an attack. Presentation attack instrument means a means of performing an attack, for example, an artificially manufactured part of the body. And finally, Presentation attack detection – automated means of detecting such attacks. However, the standards themselves are still in development, so it is impossible to talk about any well-established concepts. Terminology in Russian is almost completely absent.

To determine the quality of the system, the HTER metric (Half-Total Error Rate) is often used, which is calculated as the sum of the coefficients of erroneously allowed identifications (FAR-False Acceptance Rate) and erroneously prohibited identifications (FRR – False Rejection Rate) divided by in half. HTER = (FAR + FRR) / 2.

It is worth saying that in biometric systems, the most attention is usually paid to FAR, in order to do everything possible to prevent an attacker from entering the system. And they are making good progress in this (remember one millionth from the beginning of the article?) The flip side is the inevitable increase in FRR – the number of ordinary users mistakenly classified as intruders. If this can be sacrificed for government, defense and other similar systems, then mobile technologies, working with their huge scale, variety of subscriber devices and, in general, user-perspective oriented, are very sensitive to any factors that can force users to refuse services. If you want to reduce the number of phones smashed against the wall after the tenth consecutive denial of identification, the FRR is worth looking into!

The most popular means of deception are masks. Nothing is more obvious than putting on another person's mask and presenting the face to an identification system (often referred to as a Mask attack).

You can also print a photo of yourself or someone else on a sheet of paper and bring it to the camera (let's agree to call this type of attack Printed attack).

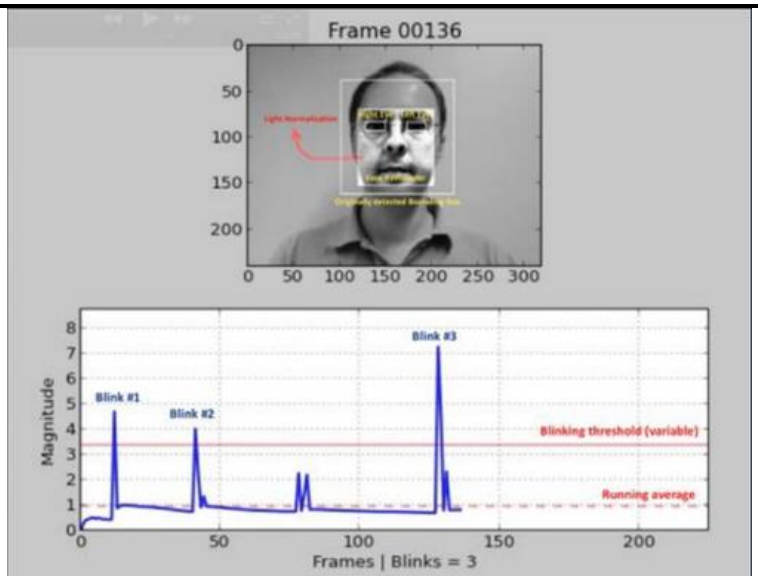
Slightly more complex is the Replay attack, when the system is presented with the screen of another device, on which a previously recorded video with another person is played. The complexity of execution is compensated by the high efficiency of such an attack, since control systems often use signs based on the analysis of time sequences, for example, tracking blinking, micro movements of the head, the presence of facial expressions, breathing, and so on. All this can be easily reproduced on video.

Both types of attacks have a number of characteristic features that allow them to be detected, and thus distinguish a tablet screen or sheet of paper from a real person.

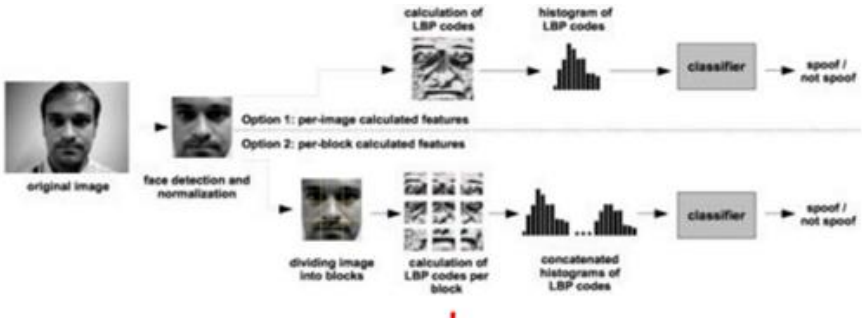
Let's summarize the characteristic features that make it possible to identify these two types of attacks in the table:

Printed attack	**Replay attack**
Decreased image texture quality when printed	Moire
Halftone transmission artifacts when printing on a printer	Reflections (glare)
Mechanical print artifacts (horizontal lines)	Flat picture (no depth)
Lack of local movement (eg, blinking)	Image borders may be visible
Image borders may be visible	

One of the oldest approaches (works of 2007, 2008) is based on the detection of human blinking by analyzing the image using a mask. The point is to build some kind of binary classifier that allows you to select images with open and closed eyes in a sequence of frames. This can be the analysis of the video stream using landmark detection, or using some simple neural network. And today this method is most often used; the user is prompted to perform some sequence of actions: shake his head, wink, smile, and so on. If the sequence is random, it is not easy for an attacker to prepare for it in advance. Unfortunately, for an honest user, this quest is also not always surmountable, and engagement drops sharply.



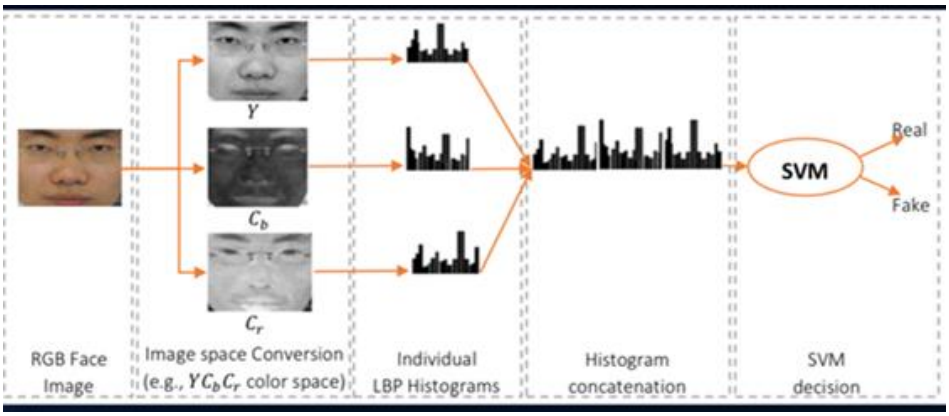
You can also use the features of picture quality degradation when printing or displaying on the screen. Most likely, even some local patterns, albeit elusive to the eye, will be found in the image. This can be done, for example, by counting local binary patterns (LBP, local binary pattern) for different areas of the face after extracting it from the frame. The described system can be considered the founder of the whole direction of face anti-spoofing algorithms based on image analysis. In a nutshell, when calculating the LBP, each pixel of the image, eight of its neighbors are sequentially taken and their intensities are compared. If the intensity is greater than the central pixel, one is assigned, if less – zero. Thus, an 8-bit sequence is obtained for each pixel. The obtained sequences are used to construct a pixel-by-pixel histogram, which is fed to the input of the SVM classifier.



Local binary patterns, histogram and SVM.

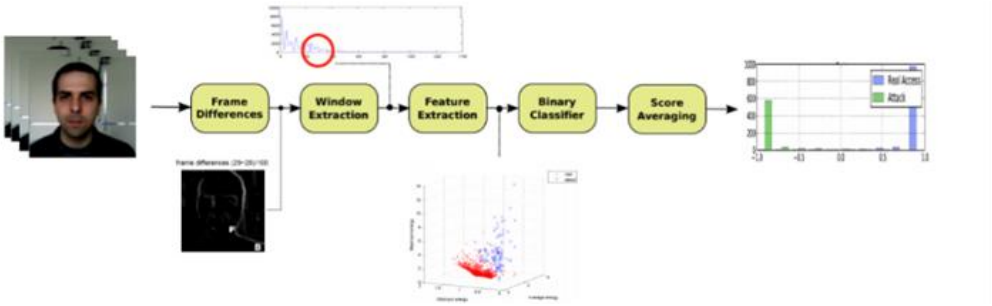
The HTER efficiency indicator is “as much” as 15%, which means that a significant part of attackers overcome the protection without much effort, although it should be admitted that many are eliminated. The algorithm was tested on the IDIAP Replay-Attack dataset, which is composed of 1200 short videos of 50 respondents and three types of attacks – printed attack, mobile attack, high-definition attack.

The ideas of image texture analysis were continued. In 2015, Bukinafit developed an algorithm for alternatively dividing the image into channels, in addition to the traditional RGB, for the results of which local binary patterns were again calculated, which, as in the previous method, were fed to the input of the SVN classifier. The accuracy of HTER, calculated on the CASIA and Replay-Attack datasets, was an impressive 3% at that time.



HTER 2.9%. (2015) Boulkenafet Z. et al. Face Anti-Spoofing Based on Color-Texture Analysis

To detect attempts to present a photo, the logical solution was to try to analyze not one image, but their sequence taken from the video stream. For example, Anzhos and his colleagues proposed to extract features from the optical stream on adjacent pairs of frames, to feed it to the input of a binary classifier and average the results. The approach proved to be quite effective, showing an HTER of 1.52% on their own dataset.



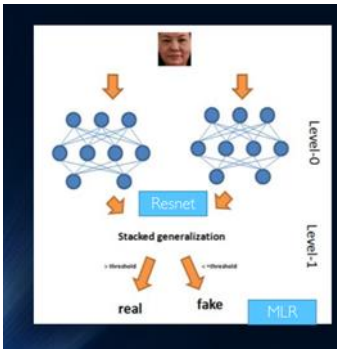
The method of tracking movements looks interesting, which is somewhat outside of the generally accepted approaches. Since in 2013 the principle “feed a raw image to the input of the convolutional network and adjust the mesh layers until the result” was not common for modern projects in the field of deep learning, Bharadwaj consistently applied more complex preliminary transformations. In particular, he applied the Eulerian video magnification algorithm known from the work of scientists from MIT, which was successfully used to analyze color changes in the skin depending on the pulse. Replaced LBP with HOOF (histograms

of optical flow directions), having correctly noted that as soon as we want to track movements, and features we need appropriate, and not just texture analysis. The same SVM, traditional at that time, was used as a classifier. The algorithm showed extremely impressive results on the Print Attack (0%) and Replay Attack (1.25%) datasets.

The “first sign” can be considered the method of analyzing depth maps in separate areas (“patches”) of the image. Obviously, the depth map is a very good indication of the plane in which the image is located. If only because the image on a sheet of paper has no “depth” by definition.

Unfortunately, the availability of a large number of excellent frameworks for deep learning has led to the emergence of a huge number of developers who are trying to solve the face anti-spoofing problem head-on in a familiar way of ensembling neural networks. Usually it looks like a stack of feature maps at the outputs of several networks, pre-trained on some widespread dataset, which is fed to a binary classifier.

In general, it is worth concluding that to date, quite a lot of works have been published, which, on the whole, demonstrate good results, and which are united by only one small “but”. All of these results are demonstrated within one specific dataset!



Approach	Replay-attack-dev(EER)%	Replay-attack-test(HTER)%	CASIA-FASD-test-EER (%)
IQA based (Galbally and Marcel, 2014)	---	---	32.40
Motion (Pereira et al., 2013)	11.60	11.70	26.60
LBP + SVM (Yang et al., 2013)	8.55	11.75	18.50
LBP-TOP + SVM (Pereira et al., 2012)	7.88	7.60	10.00
SBIQF+NN (Feng et al., 2016)	3.83	6.13	15.5
YCbCr + HSV + LBP (Boulkenafet et al., 2015)	0.4	2.9	6.2
LSTM (Xu et al., 2016)	---	---	5.17
Our Method	0.13	0.63	3.42

The situation is aggravated by the limitedness of the available data sets and, for example, on the notorious Replay-Attack, no one can be surprised by HTER 0%. All this leads to the emergence of very complex architectures, for example, these, with the use of various tricky features, auxiliary algorithms collected in a stack, with several classifiers, the results of which are averaged, and so on … As a result, the authors get HTER = 0.04%!

This suggests that the face anti-spoofing task has been solved within a specific dataset. Let’s summarize in a table various modern methods based on neural networks. As it is easy to see, the "benchmark results" were achieved by a variety of methods that just emerged in the inquisitive minds of developers.

Unfortunately, the good picture of the struggle for tenths of a percent is violated by the same “small” factor. If you try to train a neural network on one dataset, and apply it on another, the results will turn out to be... not so optimistic. Even worse, attempts to apply classifiers in real life leave no hope at all.

For example, let's take the data from 2015, where the metric of its quality was used to determine the authenticity of the presented image.

In other words, the algorithm trained on Idiap data, and applied on MSU, will give a true positive detection rate of 90.5%, and if we do the opposite (train on MSU, and check on Idiap), then only 47.2 will be correctly determined. % (!) For other combinations, the situation worsens even more, and, for example, if you train the algorithm on MSU and check it on CASIA, the TPR will be 10.8%! This means that a huge number of honest users were mistakenly ranked among the attackers, which cannot but depress. Even cross-database training could not change the situation, which seems to be a quite reasonable way out.

In 2017, at the University of Oulu in Finland, a competition was held on its own new dataset with quite interesting protocols focused specifically on the use in the field of mobile applications.

– Protocol 1: There is a difference in lighting and background. The datasets are recorded in different locations and differ in background and lighting.

– Protocol 2: Various models of printers and screens are used for attacks. So, the test dataset uses a technique that is not found in the training dataset.

– Protocol 3: Interchangeability of sensors. Videos of the real user and the attacks are recorded on five different smartphones and used in the training dataset. To check the algorithm, a video from another smartphone is used, which is not included in the training set.

– Protocol 4: Includes all of the above factors.

The results were quite unexpected. As in any competition, there was no time to come up with brilliant ideas, so almost all participants took familiar architectures and modified them by fine-tuning, working with features and trying to somehow use other datasets for training. The prize solution showed an error on the fourth, most difficult protocol, about 10%.

REFERENCE:

1. Bezrukov B.N. Specification of video monitoring of broadcast television images, Materials of the HAT International Congress, Moscow, 2002. – C. 215–216.

2. Vorobel R.A. Image contrast improvement using a modified method of lump stretching. Selection and processing of information / R.A. Vorobel, I.M. Journal. – M.: 2000, – №14 (90), – C. 116–121.

3. Gonzalez R., Woods R. Digital image processing / Pereyev. from English – M.: Technosphere, 2006. – 1070.

4. Beknazarova S., Mukhamadiyev A.Sh. Jaumitbayeva M.K. Processing color images, brightness and color conversion // International Conference on Information Science and Communications Technologies ICISCT 2019 Applications, Trends and Opportunities. Tashkent – 2019.