Computing and Informatics, Vol. 40, 2021, 169–194, doi: 10.31577/cai_2021_1_169

ASSESSMENT OF THE VIABILITY OF A BIOMETRIC CHARACTERISTIC IN THE CONTEXT OF BIOMETRIC AUTHENTICATION ON MOBILE DEVICES

Piotr Nawrocki, Wojciech Kubaty

AGH University of Science and Technology
Faculty of Computer Science, Electronics and Telecommunications
Institute of Computer Science
al. A. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: piotr.nawrocki@agh.edu.pl, wkubaty@gmail.com

Abstract. The issue of safe utilization of mobile devices is becoming an increasingly important problem, among others due to the widespread use of such devices to access sensitive data (such as electronic documents or banking data). In our work we analyze the use of biometric techniques in order to secure a mobile device, with particular emphasis on the viability of selected biometric characteristics. For this purpose, we investigate the possibility of applying machine learning models to assess the authenticity of a biometric characteristic. Results of our tests have shown that the most effective method of assessing the viability of a biometric characteristic involves blink and smile detection.

Keywords: Biometrics, viability of a biometric characteristic, face recognition, biometric authentication, mobile device

1 INTRODUCTION

Due to the widespread use of mobile devices, it is becoming more and more important to ensure their appropriate level of security. In particular, it is important to protect the identity of the mobile device user, including protection of personal information, banking data or other sensitive data. For this purpose, various security methods have been developed over the years. The most common method of securing access to a mobile device is a 4-character PIN password. However, with the ongoing technical progress (including machine learning methods, and, in particular, data processing by deep neural networks), biometric methods are more and more frequently applied to secure access to mobile devices, including, primarily, the facial recognition method. This method must work in real time; moreover, it should work reliably in various environment conditions and despite changes in the appearance of the face itself. Therefore, with the growing popularity of biometric methods that provide high accuracy, scalability and system security, it becomes important to ensure that the biometric characteristic used for authorization remains authentic. The problem of assessing the viability of a biometric characteristic, making sure that the characteristic comes from a real and physically present person, is complex and not yet fully explored. Therefore, in our article, we focus on this issue and examine the effectiveness of various methods of assessing the viability of biometric characteristics.

The structure of the article is as follows: Section 2 presents an overview of research in the field of detecting attacks on the credibility of a biometric characteristic; Section 3 describes evaluation of methods of assessment of the viability of a biometric characteristic; Section 4 presents the results of experiments and Section 5 contains conclusions.

2 RELATED WORK

A biometric system should not only be able to correctly recognize the face, but also be resistant to various types of attacks. The assessment of the viability of a biometric characteristic consists in making sure that the characteristic considered by the system comes from a real person who is physically present at the site of the given activity. Due to the large variety of attacks, the problem of detecting malicious access has not yet been fully resolved. The following types of attacks can be distinguished:

1. static – using a printed or displayed target;
2. dynamic – using facial recording, often with changing perspective or expressions;
3. using 3D masks [5] – ranging from the simplest paper masks to very detailed ones made of resin or silicone.

Given the wide range of potential attacks, many solutions have been developed to detect them. The focus was mostly on detecting only one type of attack. Initially, hand-designed methods were used, such as LBP [18, 4] and its various modifications [6, 24]. Recently, neural networks have been gaining popularity. They perceive features of the image that are difficult to describe using simple algorithm. There are many ways to detect an attack, differing mainly in the degree of interaction with the user. An ideal biometric system would be able to detect an attack attempt in a very short time and without any user interaction. Unfortunately, such methods require additional sensors, which mobile devices are not always equipped with.

The simplest solutions rely on a single photo showing the the user's face and make the appropriate classification by analyzing this image. Another group of solutions is based on the analysis of a sequence of consecutive photos. We can distinguish solutions based on unintentional facial or camera movements, and those that require the user to perform the prescribed motion. The former group is not very invasive, but it does not provide as much information as a forced interaction. By forcing specific movement larger changes are produced between consecutive images; what is more, an attack becomes more difficult to carry out. The following ways of recognizing attacks can be distinguished: analyzing facial expression, analyzing the 3D structure and mimicry of the face, and using image texture analysis.

Utilization of blinking as a simple method of assessing facial vitality was discussed in [11] and [21]. Furthermore, [8] presents a method of assessing vitality by analyzing unintended eye movements and blinking. When eyes are detected, the image is normalized and successive frames of the image sequence are compared. If the differences are large enough, the characteristic is classified as alive.

The possibility of using lip movement analysis while uttering a given phrase was tested in [13], focusing on the correlation between the assumed utterance and the dynamics of the mouth movement. Due to their simplicity and feasibility of implementation on mobile devices, methods based on analysis of the degree of eye closure and smile were tested in this study.

The 3D analysis group includes *Optical Flow* analysis described in [3], where it was noticed that the movement of objects can be divided into four types: rotation, displacement, resizing and changing perspective. The first three types are common to 2D and 3D objects, while the final type is specific only to 3D objects − thus, analysis of such changes may allow us to detect attacks. The paper discussed the idea of using light neural networks to classify Optical Flow maps. This method, in conjunction with facial landmark detection, is also applied in [10]. It can be assumed that as the face moves, different parts of the face move differently. The model uses Gabor decomposition and an SVM classifier. This technique is also used in the method proposed in [12]. However, there is a high probability of rejecting a characteristic if it does not exhibit this kind of movement. In [28] a method has been proposed to detect characteristic points of the face in several photos taken from different angles, by forcing the user's head to move appropriately. The classifier assesses whether the arrangement of points is characteristic of a real face or whether it is an attack. This solution is independent of the user's phone model and environmental conditions, which has been proven by carrying out tests on several databases. However, the proposed approach requires some user engagement.

One of the first publications on image texture and component analysis is [16], where attacks involving a printed photo of a face using the Fourier transform are detected. This solution is based on two observations: the real image contains more high-frequency components than the false image, and even if the face is moving, the standard deviation in successive frames remains small. With the development of printers and improvements in printing precision, this approach has become ineffective. The study of textures is a very popular and widely studied topic. In [4] and [18], the possibility of using LBP in various variants in static images was checked. The method consists in creating local histograms of differences in the values of neighboring pixels, using a window with a predetermined size, for example 3 × 3. It can divide the image into blocks and perform separate operations, and then combine the resulting histograms into one.

The solution proposed in [6] – Dynamic Textures – is an extension of LBP and is based on the analysis of microtextures in time and space. The best results have been obtained using a nonlinear SVM classifier. In [9] both solutions were tested – the discrete two-dimensional Fourier transform and texture detection via LBP, achieving the best results by combining these two methods. The research was conducted on attacks using paper masks, but the method used only one static image at a time. The computed values produced vectors that were used to train the SVM classifier. With the development of machine learning, solutions based on convolutional neural networks have emerged. An example of such a solution is [1], where one photo is required, and the developed method involves non-linear blurring of the image in such a way that the contours of the real face remain visible while the fake face disappears. The prepared images are then classified by the convolutional neural network.

Dataset	**Date**	**Number of Videos**	**Number of People**
NUAA	2010	(photos) 12 000	15
CASIA-FASD	2012	600	50
Replay-Attack	2012	1 200	50
MSU-USSA	2016	9 000	1 000
CASIA-SURF	2018	21 000	1 000
ROSE-Youtu	2018	3 350	20

Table 1. Comparison of datasets used for training and testing new attack detection solutions

There is a group of solutions which combine several methods, often using additional sensors, for example image depth [17], light reflections at different frequencies, or detection of the intensity of skin changes caused by blood flow [19, 17]. The recently created CASIA-SURF [30] database contains images of 1 000 people's faces in visible light, infrared and a depth map. On its basis, several works [22, 29, 26] have been published, with researchers reporting very high effectiveness. Different neural network architectures were compared with the best results for combining the three modalities. Unfortunately, such sensors are not widely available on mobile devices. HOG-based methods (*Histogram of Oriented Gradients*) were also used. In [14] information about the character's surroundings is used and attack detection is performed in a similar way to what an actual person would do – it points out if someone is trying to cheat the system by holding, for example, a photo of the face.

In [7], a comprehensive solution using only RGB images was proposed, using EfficientNet [27] networks and on MobileNetV2 [25], adding the last few layers in such a way as to obtain a binary classification result on the output. It was shown that although the MobileNetV2-based model achieved slightly worse results, it also required a smaller number of network parameters (267 thousand, compared to the the EfficientNet-based model at over 5.5 million parameters), and therefore has greater potential for use on mobile devices. The ROSE-Youtu [15] training set, which was applied in this study, contains attempts of attacks by 20 people using both paper printouts and masks, as well as reconstructed recordings. In [2] the authors propose to combine two solutions. In the former (*patch-based*) random, small, local facial characteristics were examined, increasing the amount of data to train the model. The procedure enables the use of the full resolution of the characteristic image, as opposed to the holistic approach, which often scales the image, and thus forfeits some of its quality. The second aproach (*depth-based*) resulted in a method for creating a comprehensive face depth map, assuming that a real face has more depth than the flat characteristics used in attacks.

Several sets – CASIA-FASD [31], MSU-USSA [23] and Replay-Attack [4] – were used, producing 2.67 %, 0.35 % and 0.79 % EER (*Equal Error Rate*) respectively. Table 1 compares the datasets used to detect attacks.

3 ASSESSMENT OF THE VIABILITY OF A BIOMETRIC CHARACTERISTIC

In order to analyze the means of assessing the viability of a biometric characteristic, various methods were tested in this study, both specific to facial biometrics and enabling detection of attacks regardless of the biometric method used.

3.1 Assessment of Characteristic Viability on the Basis of Facial Movement in Three-Dimensional Space

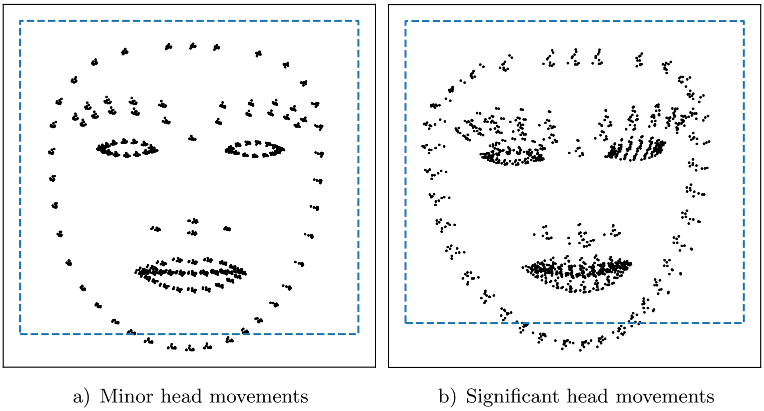
Due to the specificity of face recognition biometrics, methods involving analysis of successive frames of the recording, in particular the movement of characteristic points of the face, were proposed in this paper to assess the viability of the characteristic. The implementation, number and exact location of points may vary, but the most common ones include eyes, mouth and nose. The method proposed in this work bases on the observation that the movement of the face in three-dimensional space differs from the movement of a two-dimensional object, which is a photo of a face printed or displayed on an electronic device. Evaluation of changes in the grid of characteristic points can be performed by analyzing the distance measure. For two matrices, A_1 and A_2, representing the distances between each two characteristic points of the face, the distance is given by the formula:

$$M = \left|\left|\frac{A_1}{||A_1||_F} - \frac{A_2}{||A_2||_F}\right|\right|_F \tag{1}$$

where $||\cdot||_F$ is the Frobenius norm. For the A matrix with dimensions of $m \times n$ it assumes the form:

$$||A||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}. \tag{2}$$

For two identical grids of characteristic points, the value of the M distance will be zero. The value increases along with an increase in differences between the grids. In fact, lack of precision in detecting the characteristic points causes slight changes in distance for meshes which retain similarity[1], For example, by changing the position in the frame, zooming and rotating it, the value will be close to zero. Rotation of the head changes the position of the characteristic points in relation to each other, which implies an increase in distance. Figure 1 shows the position of the characteristic points of the face for two ranges of motion, taking into account 10 successive frames of the recording.


a) Minor head movements

b) Significant head movements

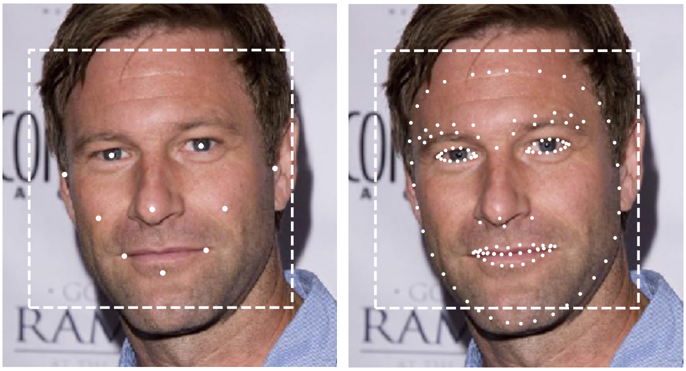
Figure 1. Comparison of the relative coordinates of 131 characteristic points of the face. The edge of the detected face is marked with a dashed line.

The NUAA Imposter DB[2] was used to check the value of the distance for faces that do not change position, as well as faces which exhibit movement. Each recording in the database consists of a sequence of between several dozen and several hundred frames, and within each sequence all frames show the same person. The recordings contain presentation attacks involving various transformations of printed photographs and sequences of real people. The attack set was divided into a part which contains similarity transforms, and a part which involves changes in perspective and bending of the photo.

[1] Similarity – a geometric transformation that maintains the ratio of the distance between points.

[2] NUAA Imposter DB set – `http://parnec.nuaa.edu.cn/xtan/data/NUAAImposterDB.html`

Each photo was subjected to face detection and extraction of characteristic points on a mobile device, resulting in a list of coordinates relative to the upper left corner of the detected face. The *Firebase ML Kit*[3] library was used for this purpose. It enables the use of two characteristic point detection algorithms which return 10 or 131 face characteristic points respectively.



a) b)

Figure 2. Photos of faces with 10 and 131 landmarks respectively

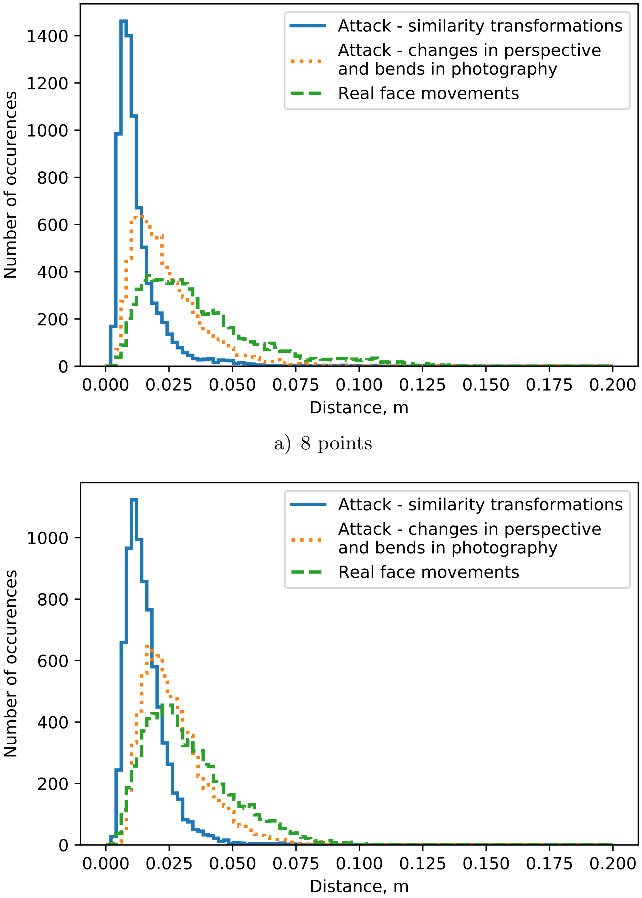
Figure 2 shows the points used in both methods. The positions of some facial features (mainly the ears) which are not directly visible, were approximated. Their positions in subsequent frames differ significantly, so only the remaining eight points were used in further studies. Moreover, in the second drawing (131 points) poor precision of the facial contours can be observed. These contours are probably averaged over existing points rather than independently detected.

To calculate the M distance for data from the NUAA database, both face characteristic point detection algorithms from the Firebase ML Kit library were used. For each frame in the sequence characteristic points of the face were generated. Subsequently, a matrix of mutual distances between each pair of characteristic points was computed. Following final transformations, based on the (1) formula, a distance measure was obtained, expressed as a single numerical value.

Figure 3 presents the distance histograms according to the M distance for both face characteristic point detection algorithms and two types of attacks: the first one based solely on the similarity of still photos and the second one based on distortions and projective transformations. Additionally, the movement distance was included for real subjects. The lowest value of M corresponds to the distortion-free set, while for attacks carried out with distortions and actual subjects very similar values were obtained. Moreover, no significant differences were noticed between the algorithms

[3] Firebase ML Kit library – `https://firebase.google.com/docs/ml-kit/detect-faces`

for detecting characteristic points of the face; thus, further analyses involved the algorithm which returns fewer points.



a) 8 points

b) 131 points

Figure 3. Histogram showing the M distance for the NUAA set for two face characteristic point detection algorithms

Based on the above distance measure, the theoretical classification abilities of the logistic regression model to distinguish attacks from real subjects were tested. Figure 4 a) shows the ROC curve (*Receiver Operating Characteristic*). It takes into account the relationship between FPR^4 and TPR^5 The expected ROC curve should be more convex and the surface area beneath it as large as possible. In this case, the

[4] FPR (*False Positive Rate*) – the percentage of wrong confirmations

[5] TPR (*True Positive Rate*) – sensitivity or true positive percentage

ROC curve shows that only presentation attacks in which there is no facial movement can be detected relatively well. Unfortunately, for this collection, the differences in the movement of the characteristic points of real people's faces in relation to the distortions of photographs are too small to be able to clearly separate the classes from each other.

Due to the lack of unambiguous results assessing the effectiveness of the distance and the shortage of data sets of appropriate size and quality, which would include significant facial movement within one sequence of photos, it was decided to use a data set not specialized for this purpose. The YouTube Faces database[6] containing recordings of 1 595 different people (over 600 000 photos) was used. In order to divide the set into sequences containing significant movement and those containing slight movement, each photo in the sequence was analyzed for facial rotation. The FSA-NET[7] was used to assess the face rotation angle, resulting in a three-dimensional face rotation vector. The threshold for considering facial movement as significant was approximated based on the previously described NUAA dataset, which also assessed the facial rotation angle.

Table 2 presents statistics of the NUAA set for attacks using similarity-based transformations. It has been shown that changes in position in relation to the X and Y axes are relatively small and most do not exceed 7° along any of the axes. The X axis corresponds to the absolute value of face rotation vertically, while the Y axis corresponds to horizontal rotation.

	Face rotation angle for the NUAA set	
	X axis	Y axis
Average	1.97	2.29
Standard deviation	1.76	1.68
50 %	1.29	2.06
75 %	3.36	3.61
99 %	6.55	5.48

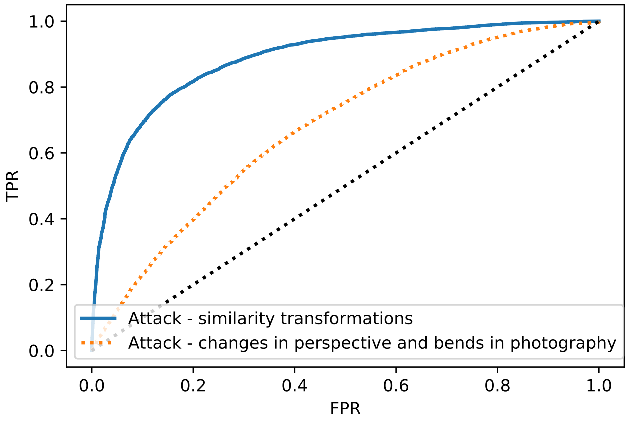
Table 2. Changes in the angle of face rotation for the NUAA dataset

Due to the small characteristic size, frames with a minimum 10° change in the face rotation angle were included in the distinct movement collection. Sequences where movement fell below this value were treated as attack simulations, reflecting the instability and imprecision of facial landmark detection models.

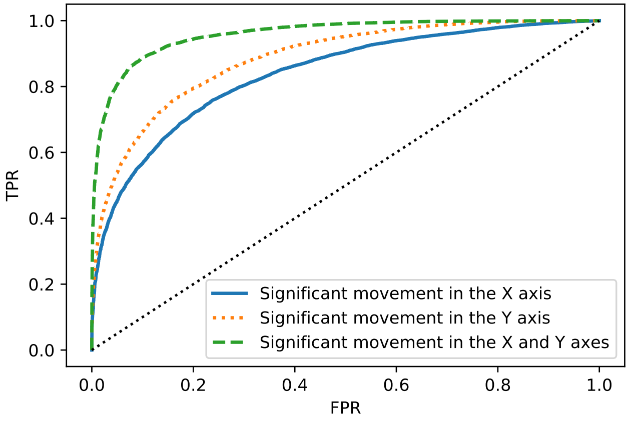
Table 3 summarizes the ROC curve parameters for the logistic regression classifier. Its graph is shown in Figure 4 b). The most significant movement (at least 10° along the X and Y axes) corresponded to the highest average distance values and therefore the best separation between this set and the set where there was no significant movement. For this reason, the EER error was the smallest for this class, equalling 4.99 %. It follows that the combination of facial movement along two axes

[6] YouTube Faces database – `https://www.cs.tau.ac.il/~wolf/ytfaces/`

[7] FSA-NET network – `https://github.com/shamangary/FSA-Net`



a) NUAA dataset



b) YT Faces dataset

Figure 4. ROC curves for NUAA and YT Faces datasets

may be the best indicator of the viability of the characteristic, but may also be the least convenient for the user to perform. Figure 5 presents a histogram showing the M distance values for the YT Faces set.

	X Axis Movements	**Y Axis Movements**	**X & Y Axes Movements**
EER	19.17 %	9.08 %	4.99 %
AUC[8]	0.88	0.97	0.99

Table 3. Efficiency of classification using logistic regression

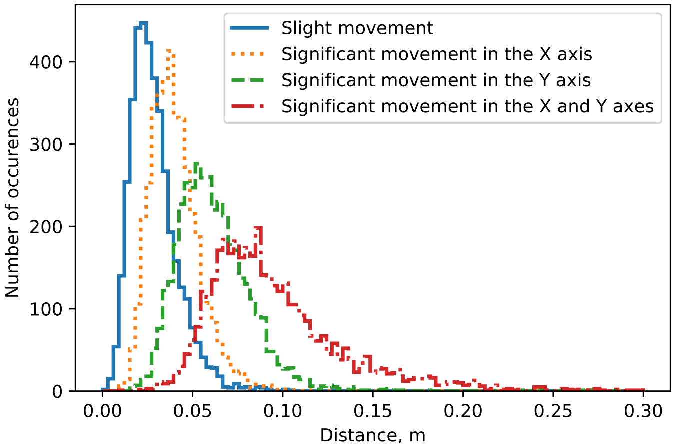


Figure 5. Histogram showing the M distance for the YT Faces set

3.2 Simulating Face Movement by Changing Perspective

Facial movement can also be simulated by changing the perspective under which the 2D characteristic is presented, thus simulating facial rotation. In order to create a model capable of recognizing such attacks a new set of sequences was generated providing it is a faithful representation of real-world attacks. For this purpose, the Facescrub[9] collection was used [20]. For each subject, on the basis of their single photo and projective transformation, new images were created, imitating changes in the angle at which the false characteristic is presented. Each photo was transformed along the X axis, Y axis and both axes simultaneously. The corresponding edges of the photo were enlarged in one of three scales: 120 %, 160 % and 200 %. An example of a Y-axis transformation simulating horizontal face rotation is presented in Figure 6.

The effectiveness of face detection was checked along with its characteristic points depending on the degree of transformation. Results are presented in Table 4. It can be seen that as the degree of transformation increases, the effectiveness of face detection decreases. Moreover, for such transformations the angle of face rotation predicted by the FSA-NET network was also checked. As the degree of transformation increases, the projected angle of rotation increases. The M distance value for the original photo and the generated transformation was analyzed, with the corresponding histograms presented in Figure 7. Results indicate that the distance value may exceed the value obtained when comparing faces under natural movement.

Additionally, the similarity of the transformed photos to the originals was checked by applying the previously described facial recognition model. Figure 8 presents

[9] Facescrub DB set – `http://vintage.winklerbros.net/facescrub.html`

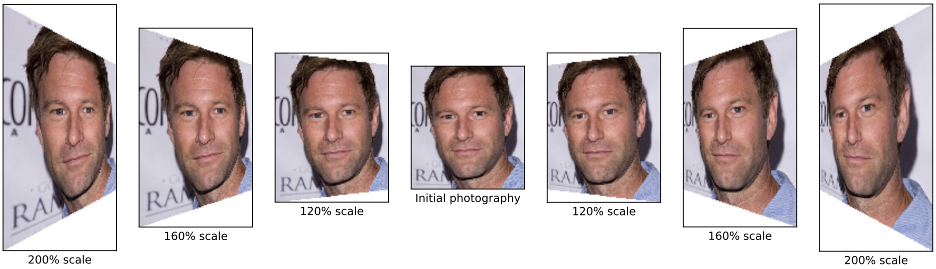
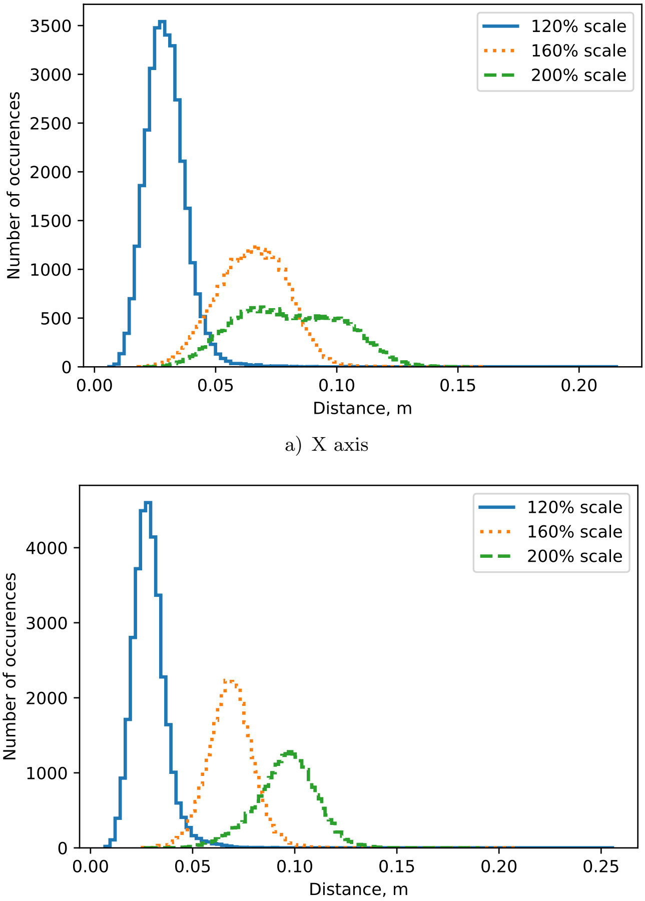
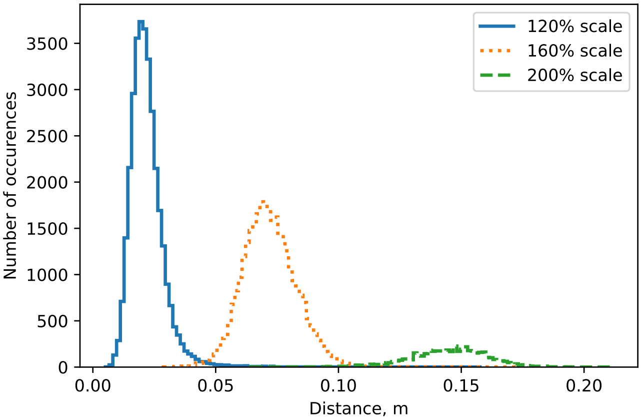


Figure 6. An example of a generated sequence simulating an attack using a face photo perspective change



a) X axis

b) Y axis



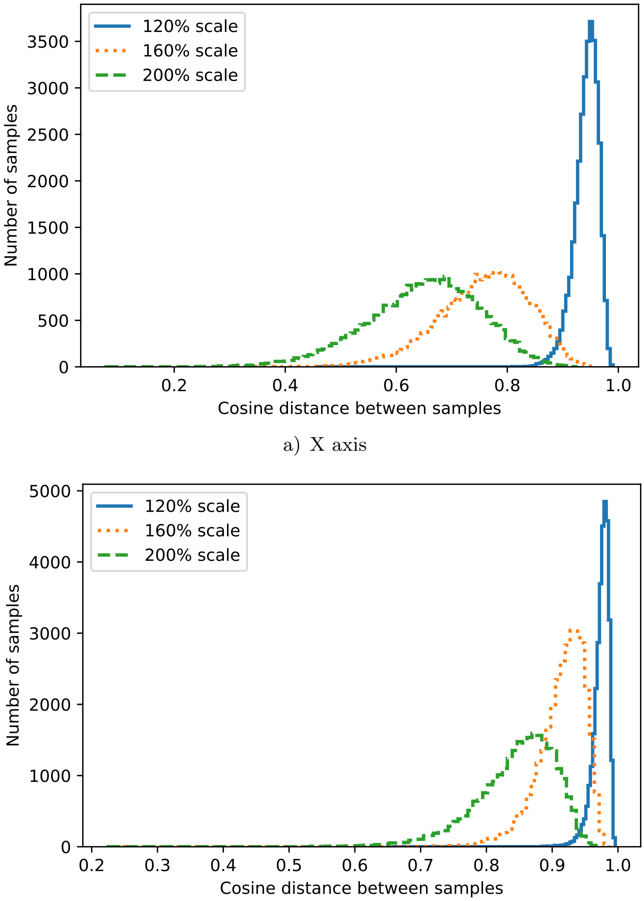
c) X and Y axes

Figure 7. Histograms showing the M distance as a function of the degree of projective transformation

Transform		Average Rotation Angle [°]		Average Distance M Value	Face Detection Efficiency [%]
Axis	Scale [%]	X	Y		
X	120	1.16	3.15	0.029	98.24
	160	4.50	2.52	0.065	96.11
	200	7.16	2.89	0.080	79.53
Y	120	2.22	3.04	0.022	98.86
	160	1.80	6.64	0.068	93.58
	200	2.58	9.51	0.095	78.19
X/Y	120	1.93	1.70	0.029	98.80
	160	7.39	5.96	0.071	92.12
	200	16.75	15.74	0.144	14.64

Table 4. Comparison of rotation angle, average distance value, and face detection accuracy for different projective transforms

the statistical distribution of the cosine distance between the compared characteristics. It is evident that as the transformation scale increases, the distance and thus similarity both decrease. Nevertheless, for the previously determined threshold, the vast majority of characteristics are still positively classified. This means that the facial recognition module is not very sensitive to changes of perspective under which a potential false characteristic is presented, but this has a significant impact on the selection of the method for assessing the viability of the characteristic.



a) X axis

b) Y axis

Figure 8. Histograms showing the value of the cosine distance between the original photograph and the projective transformation at different scales and axes

3.3 Classification of the Movement of Characteristic Points of the Face Using Machine Learning Techniques

A high M distance value is not sufficient as an indicator of the viability of a given characteristic. It only reveals the magnitude of changes in between two sets of facial landmarks. The greater the distance value, the more significant the facial movement, but it is not known whether this is caused by actual changes in the appearance of the subject or merely by distortions and changes in perspective. The corresponding assessment can be performed through more detailed analysis of the movement of characteristic points of the face, using machine learning methods. The previously generated projective transformations and sequences from the YT Faces database were used to classify the face of a real subject and to attack the presentation. Instead of calculating a one-dimensional distance value, a matrix of differences in the distance of characteristic points of the face was used to classify the movement. Four machine learning models from the *scikit-learn*[10] library were tested: two support vector machines (*LinearSVC* and *SVC* with *rbf* kernel), *GaussianNaiveBayes* algorithm, and the multilayer perceptron *MultiLayerPerceptron*. Classification accuracy was checked using the above models for two distance ranges, which may correspond to two scales of projective transformations (120 % and 160 % respectively). Percentage values were determined arbitrarily in the course of experiments and while they were not adjusted to distance ranges, they nevertheless correspond to them with high accuracy. Face detection accuracy dropped dramatically following 200 % projective transformation, and the accuracy of detecting attacks was very high; thus a realistic attack is regarded as unlikely.

Results are presented in Table 5. The best results were obtained for a nonlinear support vector machine. It has been shown that classification accuracy increases along with distance and thus with the scope of transformations. It is therefore easier to detect simulation of motion through perspective changes if the angle at which the characteristics are presented is larger. Therefore, classification should be performed only after exceeding a certain distance value (in this case – 0.05). Movement of the subject's face will therefore have to be significant enough to exceed the predefined threshold, permitting accurate analysis.

Classifier	**Accuracy**	
	$M \in (0.01; 0.05)$	$M \in (0.05; 0.1)$
Linear SVM	85.19 %	90.03 %
Nonlinear SVM	90.76 %	96.90 %
Naive Bayes Classifier	85.14 %	86.61 %
Multilayer Perceptron	85.37 %	95.28 %

Table 5. Classification performance using different machine learning methods for different M distance values

3.4 Classification of Face Movement Using Optical Flow and Mobile Neural Networks

Classification of facial motion by detecting movement of characteristic points of the face requires the user to perform considerable movement with their head. However, for practical reasons, it is advisable to minimize such movement. For this reason, the applicability of a method based on Optical Flow maps was tested. This technique can be used to detect the direction of movement of objects between successive frames

[10] Scikit-learn – `https://scikit-learn.org/stable/`

in a video sequence. In this work, classification is done through a neural network, without assuming any particular motion model.

The data used to train neural networks came from the previously mentioned NUAA Imposter DB database. The Farneback algorithm from the OpenCV library was used to create Optical Flow maps. Among the algorithm's parameters there is the size of the analyzed window within which motion is sought. Window sizes of 10, 15 and 20 pixels were checked. For each pixel, the value and direction of movement are calculated. These values are converted to the HSV color space and then to RGB. Modified EfficientNet neural networks were used for classification. Dropout and Softmax layers were added to the base. Learning accuracy was checked using *Transfer Learning* (trained networks on ImageNet) and teaching the networks from scratch. Color pictures with dimensions of 224 × 224 formed the input to the network. 10-fold cross-validation was used to evaluate the effectiveness of neural network learning. Sequences from the initial set were divided into 10 disjoint parts. Nine of these comprised the training set and one formed the test set. The learning process was repeated ten times, with a different part used as the test set each time. Finally, the average learning success rate for the test set was measured.

Regardless of the value of the window, training the model from scratch was more effective. Moreover, the highest classification efficiency occurs for a window size of 15. Therefore, this value was chosen for further analysis.

In addition, in order to increase the amount and diversity of data, less popular data sets were also used: the BioID[11], which contains 1 521 photos of 23 subjects, and Kaggle DeepFake[12] from which 3 310 10-frame sequences were obtained. The attack collection was also expanded with Optical Flow maps generated on the basis of 40 customized videos displayed on the monitor. Increasing the size of the training sets resulted in increased stability of classification and enabled the authors to forgo cross-validation. Sequences from the initial set were split into 90/10 subsets, with 90 % of data forming the training set and the remaining 10 % used as the test set. The effectiveness of classification was checked both with the use of *Transfer Learning* and by training the network from scratch. As before, classification efficiency turned out to be better for networks trained from scratch.

Table 6 summarizes the effectiveness of classification for different variants of network training. Although training a network from scratch increases the required time, it does not affect the response time of the model on a mobile device. More important is the effectiveness of classification, which turned out to be higher when the network was trained from scratch – which is why this neural network model was used for further tests on mobile devices.

[11] BioId database – `https://www.bioid.com/facedb/`

[12] Kaggle DeepFake DB – `https://www.kaggle.com/c/deepfake-detection-challenge`

Dataset	**Accuracy**	
	Transfer Learning	Without Transfer Learning
NUAA (window = 10)	83.85 %	95.55 %
NUAA (window = 15)	87.46 %	95.55 %
NUAA (window = 20)	84.59 %	93.61 %
Combination of datasets (window = 15)	95.58 %	98.94 %

Table 6. Effectiveness of classification of Optical Flow maps of the EfficientNet network using *Transfer Learning* and training the network from scratch

4 EVALUATION

For practical analysis of the issue of assessing the viability of a biometric characteristic, we have developed a dedicated mobile application used in experiments. Models whose training is described in the previous chapter have been used.

4.1 Implementation

User authentication consists of the facial recognition process and the characteristic viability evaluation process. In a dedicated mobile application the user can choose a specific method of assessing the viability of the characteristic:

- method based on the analysis of the movement of characteristic points of the face (1);
- Optical Flow method using vertical or horizontal face rotation (2);
- Optical Flow method which involves bringing the lens closer to the face (3);
- method based on blink and smile detection (4).

In addition to a real-time preview of the camera image, the user's screen displays a message highlighting the need to perform a given action. Depending on the selected characteristic viability evaluation mode, the screen may also show additional information in graphic form, e.g. detected facial markings or indicators showing the angle of face rotation. The process of acquiring a face photo is automated. After analyzing a given frame, a decision is made whether the requested action has been successfully performed by the user. Upon acceptance, the frame is remembered and the status of the authentication process is updated. A corresponding message is also displayed to the user on the screen. After performing all actions, the collected frames are analyzed by the characteristic viability detection module and the face recognition module.

In the method based on the movement of characteristic points of the face (1), the user must perform any face rotation, e.g. a rotation to the right. Each frame of the recording is analyzed and the face is detected, followed by its characteristic points. A list of the most recent 10 frames is kept. For each pair of consecutive frames

Attempt	**Characteristic Authenticity Prediction** [%]		
	Genuine Characteristic	Characteristic Displayed on Electronic Device Screen	Characteristic Printed on A4 Paper Sheet
1	69.49	11.35	9.44
2	45.91	17.21	2.69
3	51.44	11.31	19.05
4	29.41	5.31	5.79
5	56.81	8.38	2.60
6	78.64	24.15	7.43
7	51.58	7.95	6.67
8	25.05	7.89	22.54
9	21.49	11.81	3.64
10	63.08	18.48	11.08
Average	**49.29**	**12.38**	**9.09**

Table 7. Characteristic authenticity prediction – method based on changing the distance from the tested object

containing facial landmarks, the distance is calculated according to the M distance formula (described in the previous section). If the value of the metric exceeds 0.05 for at least 10 pairs of frames, classification is performed. The decision to accept a characteristic is made by the two classifiers which exhibit the best training efficiency, i.e. the nonlinear classifier *SVM* and the multilayer perceptron *MLP*. The classifiers were trained in Python using the scikit-learn library and converted using the *sklearn-porter*[13]. Appropriate parameters necessary for initialization of classifiers are saved in the *json* format and loaded from the device's internal memory. A characteristic is classified positively if the mean value returned by both classifiers is positive ($> 50\,\%$).

For the method involving rotation of the face (2), the selection of appropriate frames used to create Optical Flow maps is performed by analyzing the face rotation angle. Depending on the selected mode, the *Firebase ML Kit* library is used to analyze the horizontal rotation angle, or *dlib* together with the *OpenCV* library to detect the vertical rotation angle. The rotation angle is analyzed in real time. A point indicator is displayed on the user's screen, showing the current degree of facial rotation, along with four target points. The points are arranged in a straight line vertically or horizontally, depending on the selected mode – two per side. By rotating the face, the user changes the position of the pointer. If it reaches the indicated target point, the frame is saved. After completing the task, Optical Flow maps are computed between frames using the *OpenCV* library. Subsequently, they are classified by the EfficientNet mobile neural network in *tflite* format using the *TensorFlow Lite* library. If the average of all images is above the set threshold, the viability test is accepted.

[13] sklearn-porter library – `https://github.com/nok/sklearn-porter`

Attempt	**Characteristic Authenticity Prediction** [%]		
	Genuine Characteristic	Characteristic Displayed on Electronic Device	Characteristic Printed on A4 Paper Sheet
1	72.85	43.84	27.00
2	43.55	51.30	23.29
3	55.66	13.60	35.91
4	43.12	54.39	9.45
5	77.12	9.47	23.27
6	21.32	24.10	18.95
7	68.72	22.59	38.78
8	68.97	44.94	46.04
9	46.31	43.60	26.69
10	79.49	22.20	40.98
Average	**57.71**	**33.03**	**29.04**

Table 8. Characteristic authenticity prediction – method based on horizontal face rotation

In order to minimize the need to rotate the face while ensuring that movement is registered between each two frames of the recording, the classification capabilities of Optical Flow maps were tested using the natural vibrations of the user's hand, perspective changes in the process of zooming in and out, and involuntary changes in facial expressions (3). On the device screen the currently detected face is marked with a frame, with rectangles depictng two target frames – a smaller one and a larger one. The larger square takes up approx. 80 % of the screen width, while the smaller square takes up approx. 60 %. The user's task is to move the mobile device in such a way that the detected face is inside the smaller and outside the larger rectangle respectively. Once the requirements are met, the frames are saved for further analysis. In total, three frames are obtained in this method - one with the face fitting inside the smaller rectangle, the second when the face protrudes beyond the larger rectangle, and the third in between. An Optical Flow map is computed for each pair of consecutive frames and then classified in the same way as for the facial rotation method.

Contrary to the previous methods, which require face rotation or close-up, method (4) analyzes changes in facial expressions. The user is asked to blink first and then to smile. Detection of these activities is performed using the *Firebase ML Kit* library, which assesses the likelihood of a blink and of a smile for each eye. If an action occurs in the analyzed frame, it is saved for further analysis. Appropriate information about the required next action is displayed at the top of the screen.

4.2 Experiments

For each method of assessing the viability of a characteristic and the type of attack, 10 measurements were performed. Table 7 presents the characteristic authenticity

Attempt	**Characteristic Authenticity Prediction** [%]	
	Genuine Characteristic	Characteristic Printed on A4 Paper Sheet
1	67.1	40.43
2	51.12	28.21
3	53.06	27.83
4	59.17	16.18
5	46.19	8.05
6	40.83	31.19
7	21,53	23.99
8	58.74	14.5
9	76.61	35.14
10	88.71	14.01
Average	**56.31**	**23.95**

Table 9. Characteristic authenticity prediction – method based on the vertical face rotation

assessment for the method based on changes in distance from the tested object. The greater the value, the higher the likelihood that the characteristic is genuine, while lower values suggest an attempted attack. It can be seen that the attacks obtained a much lower mean value than real characteristics, which confirms the statistical effectiveness of the model. However, the tendency of the model to reduce the likelihood that the characteristic is authentic can also be noticed.

Table 8 shows the results of characteristic authenticity prediction for the horizontal face rotation method. The mean value for all three test modes turned out to be greater than for the method based on changes in distance from the test object. In this case, facial movement was more significant, suggesting greater authenticity of the characteristic. In addition, the true characteristic has a greater authenticity estimate than the corresponding attack attempts. Unfortunately, in some cases the authenticity estimate of the counterfeit characteristic is greater than that of the genuine characteristic.

The characteristic authenticity prediction for the vertical face rotation method is shown in Table 9. The test was performed only for the real characteristic and for an attack which exploits a printed photograph. It was not possible to perform tests on an electronic device in this form. The vertical face rotation angle evaluation algorithm used on a mobile device turned out to be unable to detect faces merely via changes in perspective. The flexible nature of paper, however, allowed for more extensive manipulation and enabled a successful initial test. Paradoxically, this turns out to be an effective test of the authenticity of a characteristic displayed on a mobile device.

The characteristic authenticity prediction for the method based on blink and smile detection is shown in Table 10. The attack could not be successfully performed with a single face image, therefore only the true characteristic is included in this

Attempt	**Characteristic Authenticity Prediction** [%]
	Genuine Characteristic
1	84.42
2	66.49
3	77.82
4	72.85
5	67.21
6	70.32
7	82.55
8	62.59
9	95.28
10	76.88
Average	**75.64**

Table 10. Characteristic authenticity prediction – method based on blink and smile detection

table. The reported values turned out to be the highest among all methods, which suggests that changes in facial expressions are the most significant for the method using Optical Flow maps. In all cases, both a face with closed eyes and a smiling face were correctly identified.

Assessment of characteristic authenticity for a method based on analysis of the movement of characteristic points of the face, using the *SVM* classifier and the *MLP* classifier, is presented in Table 11. As can be seen, in most cases the classifiers returned results which were definitive, but divergent from each other.

Attempt	**Characteristic Authenticity Prediction** [%]					
	Genuine Characteristic		Characteristic Displayed on Electronic Device		Characteristic Printed on A4 Paper Sheet	
	SVM	MLP	SVM	MLP	SVM	MLP
1	60	80	0	0	0	0
2	100	100	0	0	0	40
3	100	100	20	20	0	0
4	100	100	0	0	0	0
5	100	100	0	100	10	20
6	0	100	0	0	0	0
7	100	100	10	10	0	0
8	100	100	0	0	0	0
9	100	100	80	60	10	20
10	100	100	40	30	0	0
Average	**86**	**98**	**15**	**22**	**2**	**8**

Table 11. Characteristic authenticity prediction – method based on the analysis of facial landmark movements

Method	**Prediction Accuracy** [%]		
	Genuine Characteristic Acceptance	Rejection of the Characteristic Displayed on Electronic Device	Rejection of the Characteristic Printed on A4 Paper Sheet
Distance change	60	100	100
Horizontal face rotation	60	80	100
Vertical face rotation	70	–	100
Blink and smile	100	–	–
Facial landmark movement analysis	90	100	90

Table 12. Summary of accuracy of characteristic authenticity prediction for all tested methods

A summary of the performance of all tested methods is presented in Table 12. The effectiveness with which the authenticity of a real characteristic is confirmed and false characteristics detected was calculated under the assumption that for methods based on Optical Flow, the input is considered genuine if the average value is at least 50 %. In turn, for the method based on the analysis of the movement of characteristic points of the face, the average value returned by both classifiers, i.e. *SVM* and *MLP*, must be at least 0.5. All tested methods showed good effectiveness. Assuming that only one photo of the face is used for the attack, the most effective method involved detection of blinks and smiles. Presentation attacks cannot be carried out without photo manipulation. In this method, the use of Optical Flow maps does not yield any additional benefits, but can be applied when an attack exploits a paper mask with cutouts for the mouth and eyes.

5 CONCLUSION

In this work we focused on the issue of assessing the viability of a biometric characteristic on a mobile device – focusing on the possibility of using machine learning models to assess the authenticity of such a characteristic. For this purpose, we analyzed the effectiveness of detecting the viability of a biometric characteristic during attacks using a single photograph of the subject's face. Our analysis and experiments indicate that the effectiveness of assessment is largely influenced by the hardware properties of mobile devices. On the one hand, such devices have various types of sensors that can be used to assess the viability of a biometric characteristic; however, on the other hand, limited hardware resources (such as memory, CPU power or screen space) may affect the convenience and efficiency with which biometric characteristic viability assessment is performed, especially using complex machine learning algorithms. As a result of the presented tests, we concluded that the most effective method of assessing the viability of a biometric characteristic is the one which involves blink and smile detection. However, each of the analyzed methods (including the most effective one) was susceptible – to some extent – to attack. Foolproof attack detection and unquestionable assessment of the viability of a biometric characteristic are impossible to achieve; however, any proposed method used should be accurate enough to effectively discourage potential attack attempts.

Acknowledgements

The research presented in this paper was supported by funds from the Polish Ministry of Science and Higher Education allocated to the AGH University of Science and Technology. Wojciech Kubaty's work was supported in part by the National Center for Research and Development (NCBR) under Grant No. CYBERSECIDENT/382354/II/NCBR/2018.

REFERENCES

[1] Alotaibi, A.—Mahmood, A.: Deep Face Liveness Detection Based on Nonlinear Diffusion Using Convolution Neural Network. Signal, Image and Video Processing, Vol. 11, 2017, No. 4, pp. 713–720, doi: 10.1007/s11760-016-1014-2.

[2] Atoum, Y.—Liu, Y.—Jourabloo, A.—Liu, X.: Face Anti-Spoofing Using Patch and Depth-Based CNNs. 2017 IEEE International Joint Conference on Biometrics (IJCB), 2017, pp. 319–328, doi: 10.1109/btas.2017.8272713.

[3] Bao, W.—Li, H.—Li, N.—Jiang, W.: A Liveness Detection Method for Face Recognition Based on Optical Flow Field. 2009 International Conference on Image Analysis and Signal Processing, IEEE, 2009, pp. 233–236, doi: 10.1109/iasp.2009.5054589.

[4] Chingovska, I.—Anjos, A.—Marcel, S.: On the Effectiveness of Local Binary Patterns in Face Anti-Spoofing. Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), IEEE, 2012, pp. 1–7.

[5] Erdogmus, N.—Marcel, S.: Spoofing in 2D Face Recognition with 3D Masks and Anti-Spoofing with Kinect. 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–6, doi: 10.1109/btas.2013.6712688.

[6] de Freitas Pereira, T.—Komulainen, J.—Anjos, A.—De Martino, J. M.—Hadid, A.—Pietikäinen, M.—Marcel, S.: Face Liveness Detection Using Dynamic Texture. EURASIP Journal on Image and Video Processing, Vol. 2014, 2014, No. 1, Art. No. 2, doi: 10.1186/1687-5281-2014-2.

[7] Ghofrani, A.—Toroghi, R. M.—Tabatabaie, S. M.: Attention-Based Face AntiSpoofing of RGB Images, Using a Minimal End-2-End Neural Network. 2019, arXiv: 1912.08870.

[8] Jee, H. K.—Jung, S. U.—Yoo, J. H.: Liveness Detection for Embedded Face Recognition System. World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering, Vol. 2, 2008, No. 6, pp. 2142–2145, doi: 10.5281/zenodo.1060812.

[9] Kim, G.—Eum, S.—Suhr, J. K.—Kim, D. I.—Park, K. R.—Kim, J.: Face Liveness Detection Based on Texture and Frequency Analyses. 2012 5th IAPR International Conference on Biometrics (ICB), IEEE, 2012, pp. 67–72, doi: 10.1109/icb.2012.6199760.

[10] Kollreider, K.—Fronthaler, H.—Bigun, J.: Evaluating Liveness by Face Images and the Structure Tensor. 4th IEEE Workshop on Automatic Identification Advanced Technologies (AutoID '05), 2005, pp. 75–80, doi: 10.1109/autoid.2005.20.

[11] Kollreider, K.—Fronthaler, H.—Bigun, J.: Verifying Liveness by Multiple Experts in Face Biometrics. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–6, doi: 10.1109/cvprw.2008.4563115.

[12] Kollreider, K.—Fronthaler, H.—Bigun, J.: Non-Intrusive Liveness Detection by Face Images. Image and Vision Computing, Vol. 27, 2009, No. 3, pp. 233–244, doi: 10.1016/j.imavis.2007.05.004.

[13] Kollreider, K.—Fronthaler, H.—Faraj, M. I.—Bigun, J.: Real-Time Face Detection and Motion Analysis with Application in "Liveness" Assessment. IEEE Transactions on Information Forensics and Security, Vol. 2, 2007, No. 3, pp. 548–558, doi: 10.1109/tifs.2007.902037.

[14] Komulainen, J.—Hadid, A.—Pietikäinen, M.: Context Based Face Anti-Spoofing. 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2013, pp. 1–8, doi: 10.1109/btas.2013.6712690.

[15] Li, H.—Li, W.—Cao, H.—Wang, S.—Huang, F.—Kot, A. C.: Unsupervised Domain Adaptation for Face Anti-Spoofing. IEEE Transactions on Information Forensics and Security, Vol. 13, 2018, No. 7, pp. 1794–1809, doi: 10.1109/tifs.2018.2801312.

[16] Li, J.—Wang, Y.—Tan, T.—Jain, A. K.: Live Face Detection Based on the Analysis of Fourier Spectra. Biometric Technology for Human Identification, Proceedings of the SPIE, Vol. 5404, 2004, pp. 296–303, doi: 10.1117/12.541955.

[17] Liu, Y.—Jourabloo, A.—Liu, X.: Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 389–398, doi: 10.1109/cvpr.2018.00048.

[18] Määttä, J.—Hadid, A.—Pietikäinen, M.: Face Spoofing Detection from Single Images Using Micro-Texture Analysis. 2011 International Joint Conference on Biometrics (IJCB), IEEE, 2011, pp. 1–7, doi: 10.1109/ijcb.2011.6117510.

[19] Nowara, E. M.—Sabharwal, A.—Veeraraghavan, A.: PPGSecure: Biometric Presentation Attack Detection Using Photopletysmograms. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 56–62, doi: 10.1109/FG.2017.16.

[20] Ng, H. W.—Winkler, S.: A Data-Driven Approach to Cleaning Large Face Datasets. 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 343-347, doi: 10.1109/icip.2014.7025068.

[21] Pan, G.—Sun, L.—Wu, Z.—Lao, S.: Eyeblink-Based Anti-Spoofing in Face Recognition from a Generic Webcamera. 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8, doi: 10.1109/iccv.2007.4409068.

[22] Parkin, A.—Grinchuk, O.: Recognizing Multi-Modal Face Spoofing with Face Recognition Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 1617–1623, doi: 10.1109/cvprw.2019.00204.

[23] Patel, K.—Han, H.—Jain, A. K.: Secure Face Unlock: Spoof Detection on Smartphones. IEEE Transactions on Information Forensics and Security, Vol. 11, 2016, No. 10, pp. 2268–2283, doi: 10.1109/tifs.2016.2578288.

[24] Patel, K.—Han, H.—Jain, A. K.—Ott, G.: Live Face Video vs. Spoof Face Video: Use of Moiré Patterns to Detect Replay Video Attacks. 2015 International Conference on Biometrics (ICB), IEEE, 2015, pp. 98–105, doi: 10.1109/icb.2015.7139082.

[25] Sandler, M.—Howard, A.—Zhu, M.—Zhmoginov, A.—Chen, L. C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520, doi: 10.1109/cvpr.2018.00474.

[26] Shen, T.—Huang, Y.—Tong, Z.: FaceBagNet: Bag-of-Local-Features Model for Multi-Modal Face Anti-Spoofing. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 1611–1616, doi: 10.1109/cvprw.2019.00203.

[27] Tan, M.—Le, Q. V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Chaudhuri, K., Salakhutdinov, R. (Eds.): Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, Proceedings of Machine Learning Research, Vol. 97, 2019, pp. 6105–6114, arXiv: 1905.11946.

[28] Wang, T.—Yang, J.—Lei, Z.—Liao, S.—Li, S. Z.: Face Liveness Detection Using 3D Structure Recovered from a Single Camera. 2013 International Conference on Biometrics (ICB), IEEE, 2013, pp. 1–6, doi: 10.1109/ICB.2013.6612957.

[29] Zhang, P.—Zou, F.—Wu, Z.—Dai, N.—Mark, S.—Fu, M.—Zhao, J.—Li, K.: FeatherNets: Convolutional Neural Networks as Light as Feather for Face Anti-Spoofing. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 1574–1583, doi: 10.1109/cvprw.2019.00199.

[30] Zhang, S.—Wang, X.—Liu, A.—Zhao, C.—Wan, J.—Escalera, S.—Shi, H.—Wang, Z.—Li, S. Z.: A Dataset and Benchmark for Large-Scale Multi-Modal Face Anti-Spoofing. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 919–928, doi: 10.1109/CVPR.2019.00101.

[31] Zhang, Z.—Yan, J.—Liu, S.—Lei, Z.—Yi, D.—Li, S. Z.: A Face Antispoofing Database with Diverse Attacks. 2012 5th IAPR International Conference on Biometrics (ICB), IEEE, 2012, pp. 26–31, doi: 10.1109/icb.2012.6199754.

Piotr Nawrocki is Associate Professor in the Institute of Computer Science at the AGH University of Science and Technology, Krakow, Poland. His research interests include distributed systems, mobile systems, cloud computing, artificial intelligence and service-oriented architectures. He has participated in several EU research projects including MECCANO, 6WINIT and UniversAAL. He is a member of the Polish Information Processing Society (PTI).

Wojciech Kubaty received his M.Sc. in 2020 in computer science from the AGH University of Science and Technology, Kraków, Poland. His interests include practical use of mobile technologies and machine learning algorithms. He is currently working for one of the biggest and fastest growing companies developing mobile applications in Poland.