




# Variable length sliding models for banking clients face biometry

P. Szczuko<sup>1</sup>  · A. Czyżewski<sup>1</sup> · M. Szczodrak<sup>1</sup>

Received: 16 November 2017 / Revised: 6 July 2018 / Accepted: 20 July 2018 /  
Published online: 16 August 2018  
© The Author(s) 2018

## Abstract

An experiment was organized in 100 bank branches to acquire biometric samples from nearly 5000 clients including face images. A procedure for creating face verification models based on continuously expanding database of biometric samples is proposed, implemented, and tested. The presented model applies to circumstances where it is possible to collect and to take into account new biometric samples after each positive verification of the user. Thus the model can evolve in time, and it can follow changes of user face characteristics, e.g. changes in complexion, variable amount of facial hair, arriving wrinkles, cheeks chubbiness appearance, etc., introduced as effects of changing lifestyle, sunbathing, gaining weight, aging or other processes. The variable length sliding models derived from the gathered experimental data are described in the paper.

**Keywords** Face biometry · Verification · Moving average

## 1 Introduction

The most popular means of identity verification require users to remember a password or PIN or apply special identifiers (access cards or electronic tokens). These solutions fail to guarantee satisfying security because they are sensitive to various types of attacks, such as forcing a person to enter a password or stealing the password. Also It is inconvenient to remember many passwords, which often results in writing them on a paper or using the same password for

---

✉ P. Szczuko  
szczuko@multimed.org

A. Czyżewski  
andcz@multimed.org

M. Szczodrak  
szczodry@multimed.org

<sup>1</sup> Faculty of Electronics, Telecommunications and Informatics, Multimedia Systems Department, Gdańsk University of Technology, Gdansk, Poland

different actions, thus decreasing the security. Additionally, a risk of the identity theft is crucial for many everyday activities, such as e.g. banking. Methods based on biometric natural features of clients are able to ensure both: data security and ease of use. Biometric solutions became increasingly attractive, however they still face several challenges:

- there are no universal methods of biometrics covering multiple areas of use. In many cases the convenience of use and reliability are far from perfect. For this reason, the fusion of various technologies, leading towards the multimodal approach to identity verification, has to be developed in order to respond to market demands.
- already known and used solutions may fail to be effective in the long term. According to studies conducted in 121 banks around the world, currently, the most popular biometric solutions for verification in banking industry are: fingerprint (48%), verification using the scan of blood vessels patterns in a finger (12%) and voice or face verification (15%) [12]. These solutions are burdened with numerous imperfections. Verification by fingerprint, can serve as an example, being moderately cheap, but having too low efficiency by the standards of the banking industry and it is not stable in the long term. Characteristics of the fingerprint are changing with age, and dirty fingertips or humidity influencing results of authentication. Verification with veins distribution in fingers is often flawed by changes in body temperature. Verification by voice or by face image is a relatively cheap solution. However, both voice timbre and face image may change over time, causing a necessity to update the biometric reference patterns stored in banks' databases,
- the need to prevent digital exclusion of elderly or impaired persons should be considered more broadly. Authentication methods and solutions may pose practical problems to elderly or disabled people. Though biometry is in principle more easy to use than memorizing logins and passwords, the manual skills of users should be considered for each modality. Hoarse voice, face contour changes caused by ageing, pathologic heartbeat fluctuations, hand tremor while authenticating by signature or hand vein mapping are examples of many difficulties that should be analyzed and considered.

Consequently, the authors of this paper have identified an urgent need to design an integrated, groundbreaking, secure and at the same time a convenient multimodal technology, as an answer to the current, limited, and fragmentary approaches. Highly regulated industries such as financial services must deal with a risks of extremely costly data breaches, or losses in business and in customer loyalty. The authors of the paper, supported by the results of previous research, were conducting a project aimed at biometrics-based identity verification improvement, that was implemented in practice in real banking environments [10].

The previous work of the authors team shows that face biometry is one of the most convenient and natural method of verification, as it is contactless and involves only a highly popular sensor, namely a digital camera, what positively influences comfort and subjective reliability of the whole biometric identity verification process [25].

The reminder of the paper is organized as follows: Section 2 discusses relevant prior work, Section 3 describes the used database and face description method, Section 4 explains data preprocessing procedure, Section 5 describes our method for creating sliding models, Section 6 contains evaluation of the proposed approach, and Section 7 concludes the paper with a discussion of advantages and strong points of this method.



## 2 Background

Face verification is the process of decision making by the analysis of face images which result is checking whether the person being verified is the one that he or she claims to be. Because of various factors, such as image lighting variations, head pose, facial expressions, and aging this can make a difficult task. Numerous approaches have been developed to verify the face image.

An attempt was made to derive a dedicated transformation matrix improving performance of Euclidean distance in the new feature space. Learning such a transformation matrix is equivalent to learning a Mahalanobis metric in the original space [11].

Other method, proposed by Xing et al., is based on learning the Mahalanobis distance metric for the data clustering [30]. The algorithm aims to minimize the sum of squared distances between similarly labeled inputs, while maintaining a lower bound on the sum of distances between differently labeled inputs.

Weinberger et al. proposed a method that learns a matrix designed to improve the performance of k-NN classification [29]. The metric is trained with the goal that the k-nearest neighbors always belong to the same class while examples from different classes are separated by a large distance margin.

Other verification solution was proposed by Cao et al., employing the Bayesian model for face image verification [7]. The approach is based on a simple generative Bayesian model, combining a KL-divergence based regularizer with a robust likelihood function leading to a scalable implementation via the expectation–maximization algorithm. It was shown that this method outperforms two previous state of the art solutions [3, 9].

A discriminative deep metric learning (DDML) was applied to face verification, using deep neural network automatically deriving from examples a set of hierarchical nonlinear transformations to project face pairs into the same feature subspace, under which the distance of each positive face pair is less than a smaller threshold, and that of each negative pair is higher than a larger threshold [13].

A new tensor-based feature extraction algorithm for face recognition was proposed [26], operating on a low-dimensional tensor subspace and on a discriminative locality alignment. This algorithm utilizes the natural geometric structure of the input samples, i.e. spatial constraints of each pixel, and it applies an optimization criterion for tensor spectral analysis.

As compared to static images, time-varying data are much more complex in processing, classification and recognition. Meanwhile, biometric characteristics change over time, so they must be processed with regard to their dynamic nature. Numerous approaches for dealing with dynamic biometric data are described in literature, including authors work on multimodal behavioral biometry, such as identification based on voice and signature samples [17, 25]. One of new interesting approaches is an algorithm proposed by Liu et al. [18] for the recognition of human activities from multimodal sensors, based on automatic identification of temporal patterns within actions and utilizing this kind of atomic components to represent more specific activities.

Dynamic nature of face appearance was addressed in several attempts to accurately verify the person based on their distant past photo. Panis et al. [21] present an extensive overview of the research on facial ageing that was based on FG-NET database of 82 subjects, but only a small fraction of it was dedicated to the face verification. Generally the approaches can be divided into texture-based and feature-based. Accordingly, we present below our own selection of approaches. The texture-based model was presented by Lanitis et al. [16], aimed at explaining effects of ageing on face appearance using a parameterized statistical model based



on Principal Component Analysis (PCA). The simulation of age effects is automated by using a system trained to apply an appropriate type of ageing deformations to new images. The variation explained by a series of local models and non-linear relationships in Eigenspaces were modeled using polynomial regression. This method was based on individual pixels, therefore it was strongly hampered by a scale changes, head pose, and light conditions, and it required a manual extraction of the face region (in terms of removal of background, hair, torso, etc.). The method was tested on 12 individuals with photos separated by 15 years, achieving an increase from 65 to 81% of the recognition accuracy in case of applying the age simulation.

Other texture-based method used Embedded Hidden Markov Model to recognize face images within a long time span, with a prediction of texture of the ageing faces [23]. New images were generated by a series of shape and texture changes of younger faces. Actual images of older faces were compared to predictions. The method achieved up to 68% accuracy on a set of 18 persons, for the most convenient scenario of recognition employing samples aged by 10–20 years.

A feature-based face recognition method was proposed [28], that is biologically inspired by the HMAX model, with image parameterization motivated by a structure of human visual cortex exhibiting a temporal invariance. Dedicated features are extracted, aimed at texture and shape information, then reduced by PCA, and then finally classified by k-NN. The method improved the accuracy of aging face recognition in a selected group of adults. Depending on the number of processed subjects and on the score rank  $k$  the recognition rate varied in the range of 30%–84%. The well-known SIFT local features were used [20] to identify the verification based on the processing of positions and scales of keypoints, that proved to be robust to age-related appearance changes. For FAR set at 16.66% the 81.81% accuracy was achieved on a database of 82 subjects.

Smith et al. proposed generative adversarial neural networks combined with the Local Manifold Adaptation method for synthesizing realistic faces of different age, based on previous images, for the face verification [1]. The age-normalization algorithm was able to rejuvenate or to age the unknown image, and was used to increase the face verification results by 12%.

Other method is close to our approach, as it focused on self-update procedure, through re-learning the user appearance during every interaction [22]. The evaluated systems used Eigenfaces, Fisherfaces, and Similarity-Based Fisherfaces, with personalized face models for each user. The identification approach revealed a corruption of models after an incorporation of the impostor face image.

One of the most interesting findings is the fact that the intensity of ageing effects on different facial areas may differ: the upper part of a face has a higher ageing invariance than the lower part. For example Juefei-Xu et al. [14] focused on the processing a periocular region only to extract age-invariant features. Those regions were further processed by Walsh-Hadamard transform encoded local binary patterns and unsupervised discriminant projection classifier, achieving 98% verification rate at 0.1% FAR.

Available approaches usually deal with small (up to 82 subjects) databases, and they are not extensively documented, lacking details on the false acceptance and on the false rejection rates. Our approach differ from them by taking into account a very large image database, and by introducing a process for maintaining several models of face appearance at the same time. It can be beneficial in case of frequent appearance changes due to hair styles, make-up, variations in light conditions, and face expressions.

The method described in the paper aims at generating numerous models for every person, and at assessing models performance. Each model reflects average appearance derived from



enrolment images (by means of visual features extracted from the image – see Section 3). Models vary in number and time span of samples being included in them. Therefore several modes, i.e. different face appearances for a single person, are stored and then they are recalled and matched against the new image during the verification phase.

### 3 Face description method

Currently existing datasets are not suitable for biometric identity verification experiments, as usually not enough images are available. For the purpose of this work a new dataset was created, comprising data for 4854 persons including: 5 training images, 5 validation images collected during the enrolment, and 5 testing images taken at random time interval after the enrolment, acquired in the time span of a half of the year. Consequently 72,810 images were submitted to the analysis. Those persons were selected for the purpose of the presented research from the large database of biometric samples gathered by 100 biometric stands installed in branches of the largest Polish Bank, PKO Bank Polski [6]. The original images stored in the experimental biometric system database were parameterized after a facial keypoints detection employing the feature extraction algorithm presented below, resulting in 768 features for each face image. Face images were then compared by means of calculating the Euclidean distance between a feature vector extracted from a given image and a model feature vector including the mean value of features of previous images related to the same person. The distance obtained by the comparison of features served the verification of a new face.

The face image-based verification is one of the most convenient modality with regard to person physical interaction with an acquisition device. It assures the client comfort, and it still leaves some degree of freedom, especially because of the tolerance to face position fluctuations.

The face parameterization applied in this research includes several processing stages [24]. For the frontal face detection, Viola-Jones object detector [27] is employed. For the eye detection, the Average of Synthetic Exact Filters (ASEF) method is used. The ASEF method is known to outperform most of correlation filters used for template matching, as well as Haar-based Cascade Classifiers [4]. After the face in the image is detected, face landmarks are to be found using Standard Active Shape Models (STASM) [19]. According to a comprehensive study that was undertaken by Celiktutan, STASM outperformed other face landmarking algorithms in three of the four tests [8]. Figure 1 shows the detection accuracy of the STASM algorithm compared to the others, achieved for m17 landmark points set, which literature

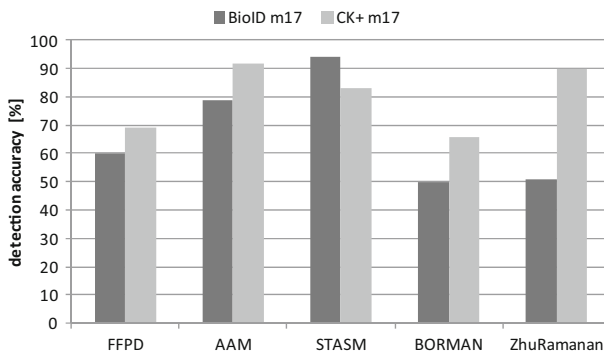


Fig. 1 Detection efficiency of the algorithms tested on BioID and CK+ databases [8]



defines as 17 landmark points within the face contour (4 points for eyebrows, 6 for eyes, 3 for nose, and 4 for mouth) [8]. The m17 landmarks set is presented in Fig. 2 as black dots.

Landmarks positions ( $N = 77$ ) of  $i$ -th face, are described as vector  $\mathbf{x}_i$ :

$$\mathbf{x}_i = (x_{i0}, y_{i0}, x_{i1}, y_{i1}, \dots, x_{ik}, y_{ik}, \dots, x_{iN-1}, y_{iN-1})^T. \quad (1)$$

It is commonly agreed that 5 subsets of landmark points may be used to define borders of regions of interest  $\mathbf{R}_E, \mathbf{R}_B, \mathbf{R}_N, \mathbf{R}_M, \mathbf{R}_F$ , corresponding to eyes, eyebrows, nose, mouth, face, e.g. for mouths in  $i$ -th image:

$$\mathbf{R}_{iM} = ((\min_{59 \leq n \leq 76}(x_{in}, y_{in})), (\max_{59 \leq n \leq 76}(x_{in}, y_{in}))). \quad (2)$$

The parameter space vector of  $i$ -th face is expressed as  $\mathbf{P}_i = F(\mathbf{R}_E, \mathbf{R}_B, \mathbf{R}_N, \mathbf{R}_M, \mathbf{R}_F) = (p_{i0}, p_{i1}, \dots, p_{ik}, \dots, p_{iP-1})^T$ .

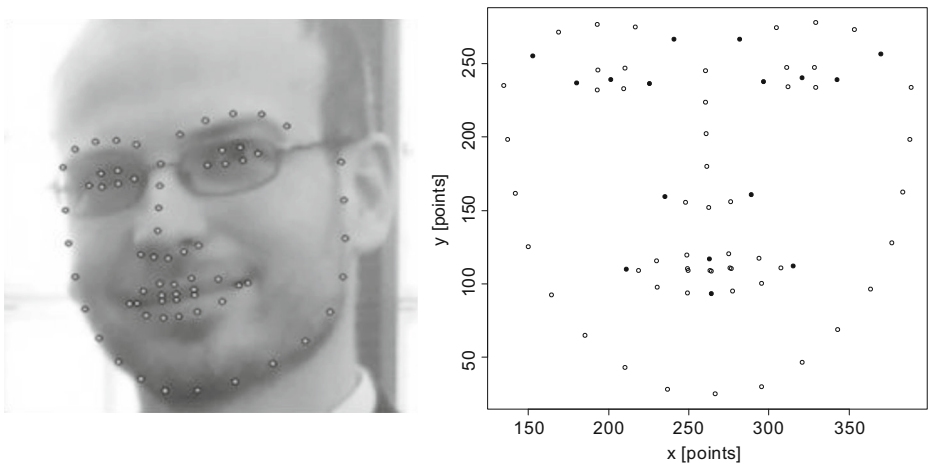
Within the function  $F$ , for each  $\mathbf{R}_c$ , the Histogram of Oriented Gradients and the Local Binary Pattern are calculated [2]. Parameters from each image fragments are truncated using Linear Discriminant Analysis [5]. Subsequently all parameter sets calculated from regions  $\mathbf{R}_E, \mathbf{R}_B, \mathbf{R}_N, \mathbf{R}_M, \mathbf{R}_F$  are concatenated. Then the LDA analysis is performed in order to create the feature vector of the size equal to  $P = 768$  elements. In the last step the parameters are normalized using L2-norm.

Whenever the verification of new face occurs, the described image processing is applied to obtain  $\mathbf{P}_a$  for the “attempter” image, and then the vector is compared with the enrolled model of the face ( $\mathbf{P}_m$ ) using the Euclidian distance:

$$d(\mathbf{P}_a, \mathbf{P}_m) = \sqrt{\sum_{j=1}^P (\mathbf{P}_{a_j} - \mathbf{P}_{m_j})^2}, \quad (3)$$

where:  $a$  is the vector of features extracted from the new attempted face,  $m$  represents the vector extracted from the face model, and  $P$  is the size of the vector.

The face image parameterization algorithms were implemented in C++ programming language with use of OpenCV and OpenBR libraries [15]. Images were acquired with SoftKinetic DS325 RGB sensor, with  $1920 \times 1080$  pixels spatial resolution and 30 frames



**Fig. 2** Face landmarks in processed video frame, face rotated (left) and face automatically aligned in front of the camera (right)

per second rate. The hardware platform was Kontron industrial PC with Intel Atom E3845 1.91GHz CPU and 8 GB RAM. The measured performance of the feature extraction algorithm was around 70 ms per one image. Typically up to 10 images are taken for the verification, so the processing delay is nearly unnoticeable for the user. Sample results of face landmark detection are presented in Fig. 2.

#### 4 Data pre-processing

Every feature extracted from the face image has generally a different range of values, thus using unprocessed data for measuring distance between samples would entail undesired effects. Namely, features having values within a wide range are by principle falling by the absolute distance farther apart, compared to features having values within narrow ranges. Therefore, the total distance, e.g. expressed by Euclidean distance, between various biometric samples is greatly influenced by those features (see a graphical illustration in Fig. 3). Since raw input features are not normalized over all classes (identities), therefore the first stage of data pre-processing is to analyze means and standard deviations over randomly selected set of 50% of cases. Subsequently, the second stage consists of standardization of values of each feature by subtracting the mean of the feature and dividing the result by the feature values standard deviation (Eq. 4).

Consequently, each value of every feature  $a_{i,j}$  is normalized to a new value  $a_{i,j}^*$ :

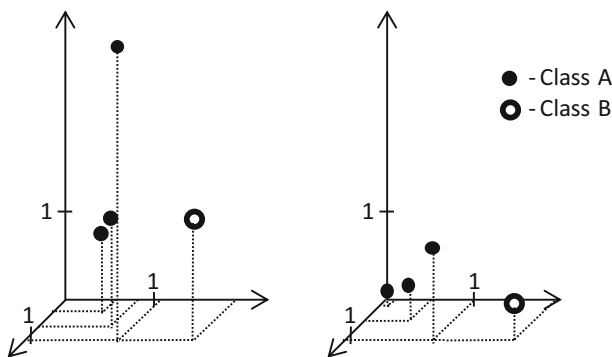
$$a_{i,j}^* = (a_{i,j} - a_i) / \sigma_i \quad (4)$$

based on mean  $a_i$  and standard deviation  $\sigma_i$  determined over  $N$  randomly selected samples:

$$a_i = 1/N \cdot \sum_{n=1 \dots N} (a_{i,n}) \quad (5)$$

$$\sigma_i = \left( 1/N \cdot \sum_{n=1 \dots N} (a_{i,n} - a_i)^2 \right)^{1/2} \quad (6)$$

where  $i = 1 \dots 768$  is the feature number,  $j = 1 \dots J$  and  $n = 1 \dots N$  are face images,  $J = 72,810$  is the number of all selected face images, and  $N = 36,405$  is the number of images randomly taken



**Fig. 3** Graphical explanation of distances for samples before and after normalization made: a) by means of Euclidean distance in 3D space the class B sample is closer to two samples of class A than to the outlier sample of class A, b) as an effect of normalization the influence of feature on vertical axis is reduced, thus the class A cluster gets more condensed



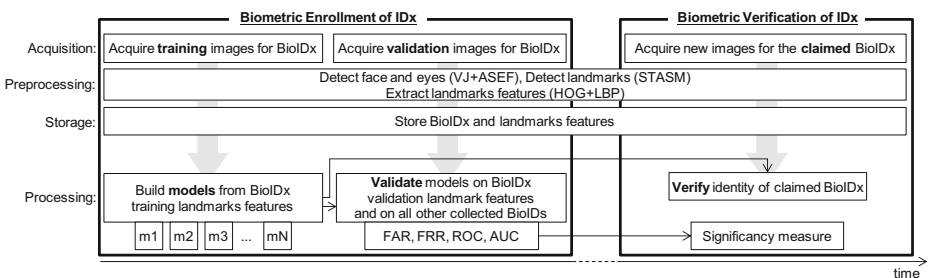
for normalization. After the normalization each feature  $a_{ij}^*$  is ensured to have its mean value equal to 0, and its standard deviation is equal to 1.

Employment of only randomly selected subset of samples allows one to emulate adding of a new, previously unobserved samples, having unknown values, that are supposed to not influence the calculation of normalization factors at the pre-processing stage.

## 5 Variable length sliding model approach

The main purpose of applying the proposed model is to take into account fluctuations of features over time. Generally in video frames or in consecutive photos the face is not motionless, but it performs involuntary microexpressions, nods, and eye blinks. Thus, a long time interval between samples collection can result in introduction of various visual changes. Even a slight change in face geometry influences the amount of light reflected and scattered by the skin, therefore each local feature of the face keypoint is getting prone to changes. Moreover, the face is an elastic structure, thus the local surrounding of a keypoint usually changes due to facial expressions. An efficient face descriptor was applied, described briefly in the previous paragraph accounting for those observations, thus it exhibits some level of invariance to rotation, deformation, and lighting.

The presented approach is motivated by practical scenarios of enrolment and consecutive verification of bank clients (Fig. 4). First, it is always known what is the claimed identity of the client, thus the biometric system is expected to perform a verification of the supposed identity, by measuring similarity between previously collected samples and the currently provided image. Since the client may visit the bank in long intervals, thus a change in appearance is expected and highly probable. Finally, since the client is not accustomed to the usage of a biometric system, therefore it is hard for him or her to follow any strict requirements for face acquisition, such as maintaining a head pose or a face expression. Therefore, the presented approach is focused on proposing several concurrent models based on visual features extracted from numerous face images, representing an averaged appearance of the face taken at various moments in time. The verification relies on the comparison made among all models for the biometric ID being claimed and the new face image acquired.



**Fig. 4** Flowchart explaining elements of the approach: two phases are presented, enrolment and verification, both comprising image acquisition, preprocessing, parameterization and storage of extracted features. Next, during the enrolment several models are derived from training data, and their performance is validated against all other identities. Finally, during the verification the image is compared to all models for the particular BioIDx, and significance of results is assessed



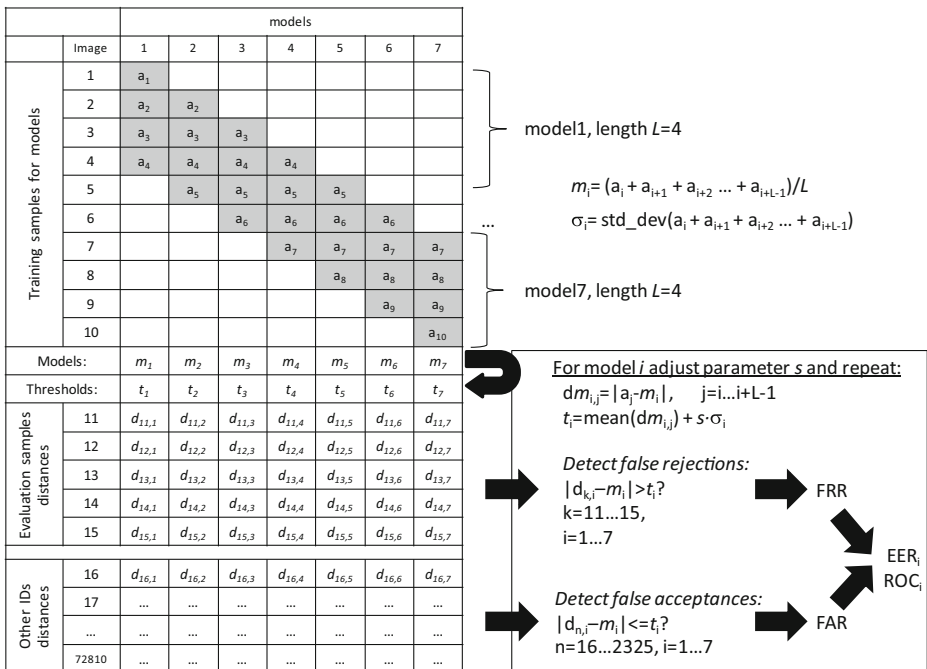
The presented methodology allows for the estimation of significance of the face verification, by expressing the performance of all models for the given person. Performance metrics are based on FAR, FRR, and AUC values, determined during the enrolment stage (Fig. 4).

The proposed face verification approach relies on a comparison of a new sample to each of sliding averages of the length  $L$  calculated over enrolment samples, preordered chronologically. Depending on the number of collected enrolment samples, several different models could be created, varied by the starting sample selection and by the model length. The first model takes into account first  $L$  samples (video frames or photos) numbered as 1 to  $L$ , the second model uses samples numbered as 2 to  $L + 1$ , and the  $n$ -th model is calculated over last samples numbered as  $n$  to  $n + L - 1$  (illustrated in Fig. 5). The maximum number of models  $i_{max}$  depends on the given length  $L$ , and on the number of collected samples  $I$  (Eq. 7):

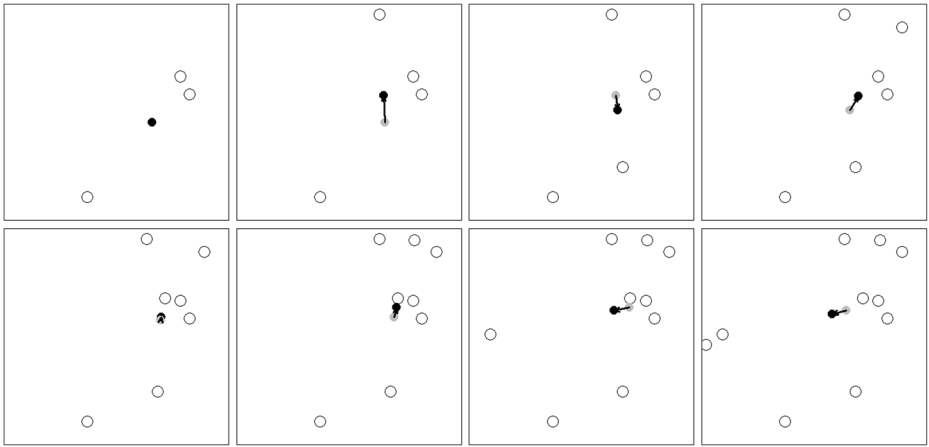
$$i_{max} = I - L + 1 \tag{7}$$

where:  $I$  – number of collected samples,  $L$  – model length.

The sliding model, while compared to a cumulative model (Figs. 6 and 7) tends to change in time significantly, in effect of discarding samples  $a_i$  where  $i < L$  and of adding samples with  $i > L$ . The cumulative model converges to a stable mean value, that is less influenced by adding new samples comparing to the sliding model, as the large set of previous samples has a stronger impact than the one new sample.



**Fig. 5** Variable length sliding models creation and its performance evaluation process. Seven possible models of the length  $L = 4$  are visualized. The order of samples reflects that samples No. 1 to No. 15 are taken from a single person, while samples No. 16 to No. 72810 are taken from all other 4853 persons selected from the biometric database



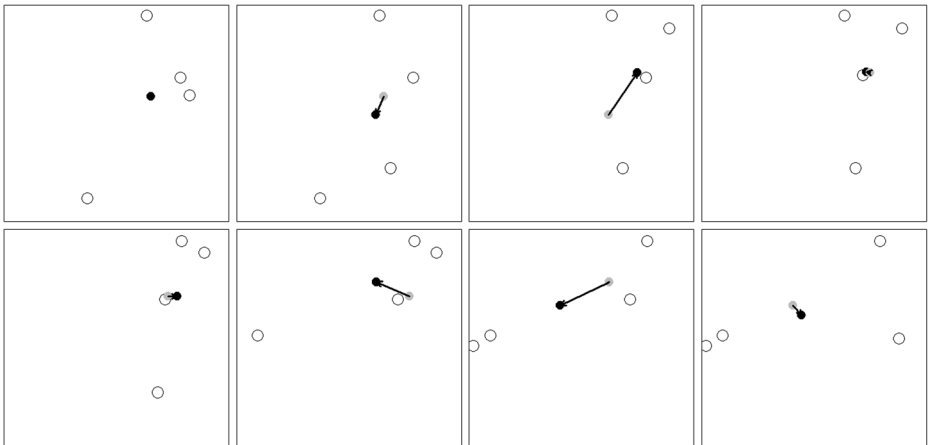
**Fig. 6** Changes of mean value (black dot) in a cumulative model comprising consecutively added new samples (open circles), starting from 3 up to 10 samples. Arrows indicate shift of mean value, while compared to the previous model. Example is based on one person's biometric dataset, two features were selected randomly out of 768 to make this illustration

For each biometric identity  $i$ -th model  $m_{ID,L,i}$  of the length  $L$  is calculated as a mean of  $L$  768-dimensional vectors, thus  $i$ -th standard deviation of  $L$  samples  $\sigma_{ID,L,i}$  is obtained (Eqs. 8, 9):

$$m_{ID,L,i} = (a_{ID,i} + a_{ID,i+1} + a_{ID,i+2} \dots + a_{ID,i+L-1})/L \quad (8)$$

$$\sigma_{ID,L,i} = \left( \frac{1}{L} \cdot \left( (a_{ID,i} - m_{ID,i})^2 + (a_{ID,i+1} - m_{ID,i})^2 + (a_{ID,i+2} - m_{ID,i})^2 \dots + (a_{ID,i+L-1} - m_{ID,i})^2 \right) \right)^{1/2} \quad (9)$$

A separate set of models  $M_{ID,L,i} = \{m_{ID,i}, i = 1 \dots i_{max}\}$  is created for each bank client and for given  $L$  value, but all models are governed by the single global hyperparameter  $s$  being independent from any personal characteristics. The decision thresholds  $t_i$  defining maximal



**Fig. 7** Changes of mean value (black dot) in the sliding model comprising  $L = 3$  most recent samples (the same data used as in Fig. 6)

allowed distance between true positive samples and the model center is set based on distances within the model  $dm_{ID,L,i,j}$ , standard deviation of the model  $\sigma_{ID,L,i}$ , and the parameter  $s$ :

$$dm_{ID,L,i,j} = |a_{ID,j} - m_{ID,L,i}|, \quad j = i \dots i + L - 1 \tag{10}$$

$$\text{mean}(dm_{ID,L,i,j}) = 1/L \cdot \sum_{j=i \dots i+L-1} (dm_{ID,L,i,j}) \tag{11}$$

$$t_i = \text{mean}(dm_{ID,L,i,j}) + s \cdot \sigma_{ID,L,i} \tag{12}$$

The graphical interpretation for the model and the threshold dependence on samples, their mean values, and standard deviation are presented in Fig. 8.

As  $m_{ID,L,i}$  and  $\sigma_{ID,L,i}$  are determined by the true positive samples, so for the particular person, and for the number of samples  $L$  the factor  $s$  is adjusted automatically to produce the lowest EER value.

Factor  $s$  influences the decision threshold and it changes observed rates of false acceptances – FAR (high  $s$  value makes the threshold too high, while negative samples tend to be accepted). It also influences false rejection rates – FRR (small  $s$  value makes the threshold too low, while positive samples are rejected). The performance of the engineered biometric verification system is expressed by the amount of such errors, often characterized by providing EER (equal error rate) as in Fig. 9a. Following ROC (receiver operating characteristic) curves (Fig. 9b) illustrate values of TPR (true positive rate) plotted against FPR (false positive rate) obtained by applying various levels of the threshold factor  $s$ . The point of ROC intersection with a diagonal ( $TPR = 1 - FPR$ ) determines the EER value, where false acceptance and false rejection rates are equal.

It should be stressed that the system is treated as a binary classifier, trained on enrolment samples, and evaluated on validation samples plus on all false images. Therefore, the performance metrics used for the biometry can be evaluated, but in practice the system is expected to perform only confirmation of the person’s identity, but not to make the identification of the person among others, thus false acceptance ratio does not reflect a real risk of accepting an impostor.

The scenario for potential identity fraud would require the impostor to know personal details and to be aware of his or her visual similarity to the actual person. One must first provide personal details during the verification such as name or login, and then an attempt to validate one’s identity by using face biometry is made. Therefore, evaluating false acceptance

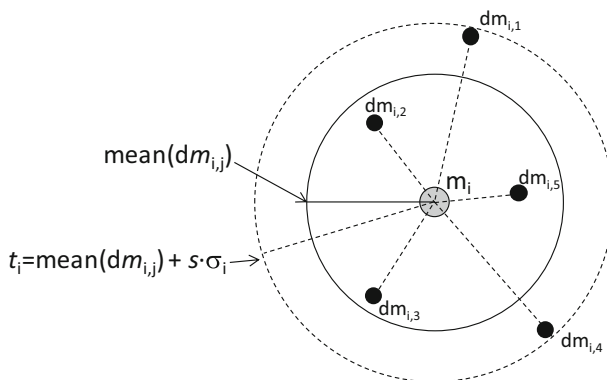
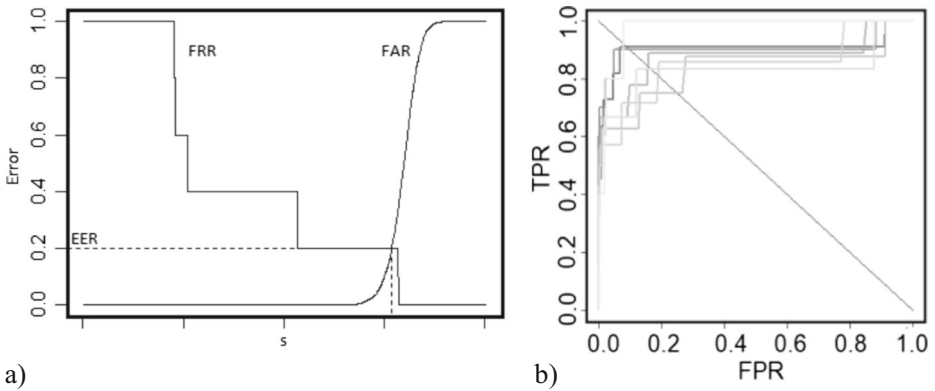


Fig. 8 Graphical interpretation of threshold  $t_i$  definition: black dots – training samples,  $m_i$  – mean of samples,  $d_{mi,j}$  – distance between  $m_i$  and  $j$ -th sample, inner circle – mean distance  $\text{mean}(d_{mi,j})$ , outer dotted circle – threshold  $t_i$



**Fig. 9** Performance metrics: a) FRR, FAR, and EER for the person No. 1, model  $m_i$ , depending on decision factor  $s$ ; b) Receiver Operating Characteristics curves for the person No. 1,  $L = 4$ . Models  $m_i$  for  $i = 1 \dots 7$  represented as gradually lighter shades of gray. Intersection of curves with the diagonal denote EER

ratio by testing the method on all possible images in the database is considered as not suitable for expressing the system performance. It can be compared to a brute force applied to guessing a password, by providing all possible variants, whereas the identity is expected to be accepted or rejected in few attempts, but not in thousands of them.

Nevertheless, in the presented research the FPR was determined based on the whole set of collected negative samples, as it is common practice in the biometric identification.

## 6 Evaluation of model performance

For each of 4854 biometric sets of data 8 models  $M_{ID,L,I}$  were created, with lengths  $L = 3 \dots 10$ , employing  $I = 10$  enrolment samples, thus resulting in  $i_{\max} = 8$  to  $i_{\max} = 1$  models for each person.

It should be stressed that the number of models for a single person, that are possible with regard to this method is large, and it can be expressed as (13):

$$I_{\text{Total}} = 0.5 \cdot (1 + (I - L_L + 1)) \cdot (I - L_L + 1) \quad (13)$$

where:  $I$  is the number of enrolment samples of the given person.

$L_L$  is the assumed shortest length of the sliding model (longest model is encompassing all  $I$  samples).

For example for 10 samples and models with lengths from  $L_L = 3$  to maximal length of 10 there will be 36 models. Each model requires evaluation of performance metrics, including FAR and FRR. This involves a comparison of all collected biometric features with all models. The total amount of 2,621,160 ( $36 \cdot 15 \cdot 4854$ ) distances must be calculated between vectors of 768 features. The speed of generation of all models and comparison with all samples was measured and averaged over all persons. It was determined that the whole process of: adding new identity, collecting 10 enrolment samples, generating new 36 models, comparing with all other images, calculating FFR, FAR and AUC, takes 12 s on 3.2GHz single core CPU, with face capture, preprocessing and feature extraction implemented in C++, and for all calculations related to models and verification implemented in R environment. This processing time can be regarded as a drawback in case of

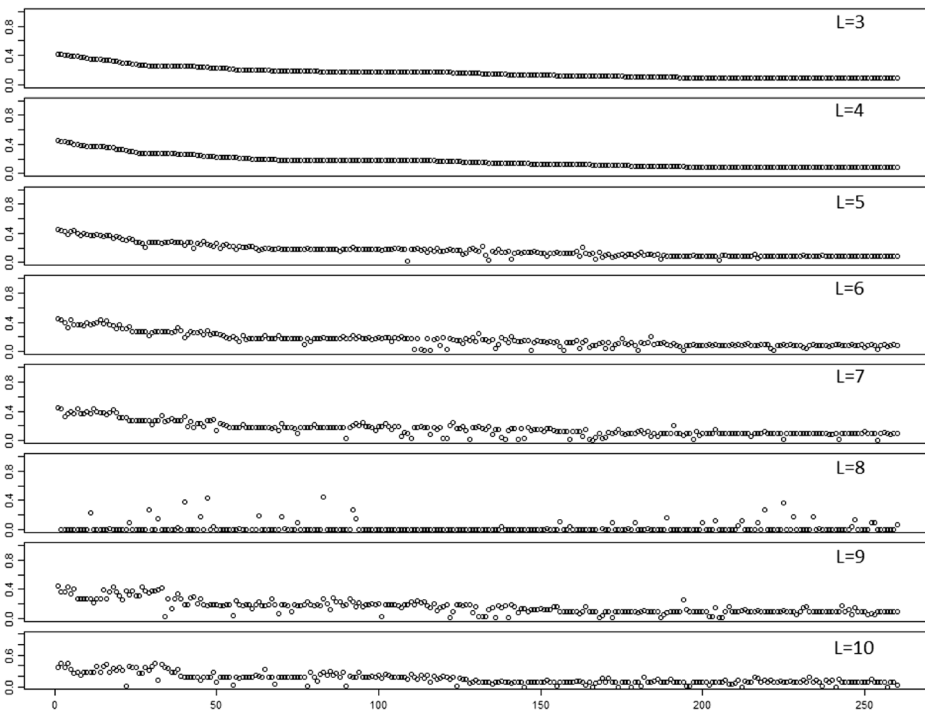


adding a new identity, because a twofold validation must be performed. First, new models must be validated against old samples, and then models for all identities should be validated against all new features. This assures that a new person would not be mistaken with any other who is already represented in the database.

For each of evaluated models the threshold  $t_i$  was adjusted by changing values of  $s$  factor, while the number of positive samples exceeding the threshold and of negative samples remaining below the threshold was determined to calculate ROC characteristics.

A significant group of biometric identities (4595 persons) was verified without any errors, therefore ROC characteristics proved to be optimal, therefore they are not discussed here. Instead, the results are characterized by the analysis of obtained Equal Error Rates (Fig. 10). In this section only 259 worst cases were analyzed, and then verified against 72,810 biometric samples of all 4854 persons. It can be observed that EER for the worst 133 persons is higher than 0.2, what is unacceptable in case of practical biometric verification systems. The interpretation of such high EER is that if it was allowed for all registered persons to repeatedly attempt a counterfeit without any additional verification, and the attack would be targeted against one of those 133 persons it would succeed with 20% chance. A chance to randomly select a single person and to succeed is therefore equal to:  $0.2/4853 \cdot 133/4853 = 1,13 \cdot 10^{-6}$ .

Beside the EER the ROC curve data were collected as well and reduced to AUC (area under curve) measures, as an efficient mean of expressing model performance by a single value.



**Fig. 10** Worst EER values obtained for given model lengths. Each point denote values for 8 to 1 for sliding models generated from  $L$  samples, with identities sorted by decreasing mean error. Only 259 worst results are shown, as remaining 4595 have errors equal to zero in all available models



AUC close to 0.5 is a random classification, and AUC equal to 1.0 reflects perfectly accurate classification.

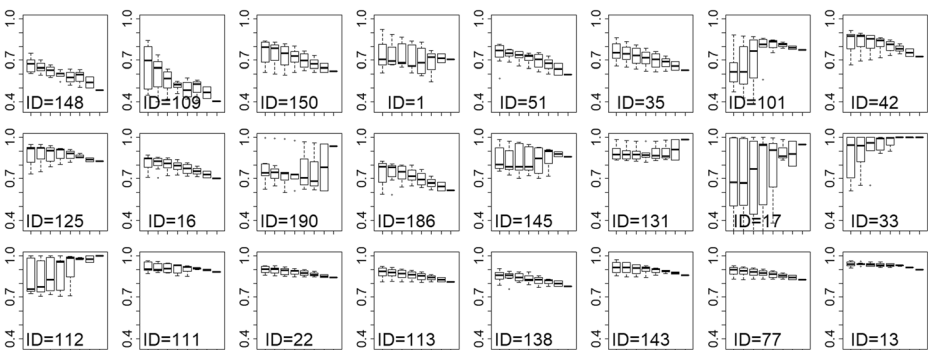
The correctly verified 4595 identities have AUC 1.0 for any model length  $L$ . For the remaining 259 identities errors of varying nature were observed (Fig. 11). Some identities are often mistaken with others, e.g. Person No. 148 (as in the first plot shown in Fig. 11), and their AUC, regardless of used model length is low. Some persons are verified incorrectly by some models and correctly by others, while AUC values are scattered in ranges of 0.5 to 0.9. Few cases are verified with a satisfactory performance by many models, their AUC remaining close to 0.9 or higher.

The set of models incorporating the incorrectly obtained face image tends to have a low first quartile, but a high median. This indicates that one of the models in the set decreases the performance, but other models omit the incorrect image, thus they rely on newer samples, therefore they tend to verify new incoming images better.

Total EER calculated as an average over all errors observed for 4854 identities being verified by particular models  $M_{L,I}$  are presented in Table 1. It can be observed that models involving a lower number of images, e.g.  $L = 3$  and  $L = 4$ , have generally lower equal error rate than the ones with more images. Among all models of given length, shown as columns in Tab. 1, usually first ones with  $n = 1$  and  $n = 2$  have the highest EER, decreasing for larger  $n$ . It is related to the fact, that models initialized on older samples take into consideration the face images, that not necessarily represent accurately images collected at the later stage of identity verification. It should be noted here that in general, errors increase as more samples are added to models, entailing results diverge instead of converging. In fact it was expected, as the method is designed to create and to use partial models based on fraction of images. Large models e.g. with  $L = 10$ , are based on all samples, so they will not reflect all possible appearances of user's face.

## 7 Summary

The presented research is dedicated to the evaluation of the sliding model method, involving many models for every single biometric identity verification, varied by length, i.e. taking into account  $L = 3 \dots 10$  images at once, ordered chronologically. The proposed model is based on



**Fig. 11** AUC for some identities with EER larger than 0. Each plot shows boxplots of AUC values obtained by applying models with length of  $L = 3 \dots 10$ , the higher values are preferred. Each plot shows results for one person, vertical axes are AUC values, horizontally are model lengths from 3 to 10



**Table 1** Total EER of models of given length  $L$ , initiated by samples  $n$  to  $n+L-1$ . Two lowest values in each model length are marked in bold

	$L = 3$	$L = 4$	$L = 5$	$L = 6$	$L = 7$	$L = 8$	$L = 9$	$L = 10$
$n = 1$	0.034	0.035	0.034	0.033	0.034	0.036	<b>0.037</b>	<b>0.036</b>
$n = 2$	0.031	0.032	0.031	0.031	0.031	<b>0.033</b>	<b>0.034</b>	
$n = 3$	0.030	0.029	0.029	0.029	<b>0.029</b>	<b>0.031</b>		
$n = 4$	0.028	0.028	0.027	<b>0.028</b>	<b>0.028</b>			
$n = 5$	0.026	0.026	<b>0.026</b>	<b>0.026</b>				
$n = 6$	<b>0.024</b>	<b>0.025</b>	<b>0.024</b>					
$n = 7$	0.025	<b>0.024</b>						
$n = 8$	<b>0.023</b>							

calculating Euclidean mean in multidimensional (768-dimensional) feature space, and then on applying the proposed decision threshold to a distance between feature vector of a new sample and the mean value.

The most important advantage and strong point of this method is availability of many models reflecting face features in the past, allowing to incorporate the knowledge of previously seen variants of the same face (e.g. differing in appearance, face expression, head pose). The method does not require any hand processing of collected images, as erroneous images are rejected by face and keypoints detection algorithms. Moreover, a particular image is used only for a fraction of models. Therefore, in case of a faulty image the models will score low in performance (e.g. they will cause a high false-rejection, or a false-acceptances). The assumption of correct images composing highly performing models is positively validated during each enrolment by determining FAR, FRR, and AUC for each model. The system administrator is informed of cases where too low number of models achieved satisfactory performance, since it is required to reinitiate the enrolment process. Usually responses for a new verification image differ, as some models may correspond more closely to the image than others. Finally, a single reliable match is produced through scaling responses of every model by their respective performance metrics, what diminishes the impact of models with false-acceptances.

On the basis of the selected sample set of images it was shown that the verification was successful for 4595 selected persons validated against 4854 identities comprising a set of 72,810 faces, for which their models achieved EER = 0.0, regardless of the used length of model.

In the remaining 50 identities error rates were larger, for the worst 15 person cases being higher than 0.2, what would be unacceptable result in terms of practical verification system applications. On average the EER for the whole set of 4854 persons remained in the range 0.023 to 0.037. The practical application of the verification system should employ small and restrictive decision thresholds, ensuring FAR remaining lower than EER. In such a system FRR is expected to surpass the value of EER in result of decreasing the threshold. As the system implemented in bank branches operates in interaction with the user, thus it allows for rejecting some of the collected frames of user face (e.g. exposed in adverse lighting conditions, with closed eyes, rotated head), as long as the correctly exposed images with proper face pose are accepted.

The solution was proven to operate correctly in large number of cases, since only data related to 259 persons out of 4854 introduced low verification rates. After the presented data analysis, the original images for the worst cases were reviewed, and it was observed that



among training and validating samples were photos presenting marginally different head poses, e.g. frames with closed eyes, with large rotation, or half-profile instead of en-face image. In practical applications of banking client face recognition systems it would be recommended to verify subjectively the quality of newly acquired samples during the enrollment and the verification, in order to reject samples not fulfilling pose requirements. The banking teller assisting the process of face image acquisition may adopt this role.

**Acknowledgements** This work was supported by the grant No. PBS3/B3/26/2015 entitled “Multimodal biometric system for bank client identity verification” co-financed by the Polish National Center for Research and Development.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Antipov G, Baccouche M, Dugelay JL (2017) Boosting cross-age face verification via generative age normalization, Intl Conf Biometrics (IJCB), pp 191–199. <https://doi.org/10.1109/BTAS.2017.8272698>
2. Benlamoudi A, Samai D, Ouafi A, Taleb-Ahmed A, Bekhouche SE, Hadid A (2015) Face spoofing detection from single images using active shape models with stasm and LBP. In: Proc Troisième Conference Internationale Sur La Vision Artificielle CVA
3. Berg T, Belhumeur PN (2012) Tom-vs-pete classifiers and identity-preserving alignment for face verification. In BMVC
4. Bolme DS, Draper BA, Beveridge JR (2009) Average of synthetic exact filters. IEEE Conf Comput Vis Pattern Recognit 2009:2105–2112
5. Borade SN, Deshmukh RR, Ramu S (2016) Face recognition using fusion of PCA and LDA: Borda count approach, 24th Mediterranean Conf. Control and Automation (MED), Athens, 1164–1167
6. Bratoszewski P, Czyżewski A, Hoffmann P (2017) Pilot testing of developed multimodal biometric identity verification system. In: 2017 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), pp 184–189. <https://doi.org/10.23919/SPA.2017.8166861>
7. Cao X, Wipf D, Wen F, Duan G, Sun J (2013) A practical transfer learning algorithm for face verification. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp 3208–3215
8. Celiktutan O, Ulukaya S, Sankur B (2013) A comparative study of face Landmarking techniques. EURASIP J Image Video Process 13:2013
9. Chen D, Cao X, Wen F, Sun J (2013) Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification. In CVPR
10. Czyżewski A, Bratoszewski P, Hoffmann P, Lech M, Szczodrak M (2017) The project IDENT: multimodal biometric system for bank client identity verification. In: Dziech A, Czyżewski A (eds) Multimedia communications, services and security. Communications in Computer and Information Science, vol. 785. Springer. [https://doi.org/10.1007/978-3-319-69911-0\\_2](https://doi.org/10.1007/978-3-319-69911-0_2)
11. Goldberger J, Hinton G, Roweis S, Salakhutdinov R (2005) Neighbourhood components analysis. In Advances in neural information processing systems, pp 513–520
12. Hosseini SS, Mohammadi S (2012) Review banking on biometric in the World’s banks and introducing a biometric model for Iran’s banking system. J Basic Appl Sci Res 2(9):9152–9160
13. Hu J, Lu J, Tan YP (2014) Discriminative deep metric learning for face verification in the wild, In IEEE CCVPR pp 1875–1882. <https://doi.org/10.1109/CVPR.2014.242>
14. Juefei-Xu F, Luu K, Savvides M, Bui TD, Suen CY (2011) Investigating age invariant face recognition based on periocular biometrics. Intl Conf Biometrics (IJCB) pp 1–7. <https://doi.org/10.1109/IJCB.2011.6117600>



15. Klontz J, Klare B, Klum S, Burge M, Jain A (2013) Open source biometric recognition, biometrics: theory, applications and systems, Washington DC
16. Lanitis A, Taylor C (2000) Towards automatic face identification robust to ageing variation. *Proc IEEE Automatic Face and Gesture Recognition*, pp 391–396
17. Lech M, Czyżewski A (2017) Modified dynamic time warping method applied to handwritten signature authenticity verification. *Elektronika* 58:18–25. <https://doi.org/10.15199/13.2017.4.4>
18. Liu Y, Nie L, Liu L, Rosenblum DS (2016) From action to activity: sensor-based activity recognition. *Neurocomputing* 181:108–115. <https://doi.org/10.1016/j.neucom.2015.08.096>
19. Milborrow S, Nicolls F (2014) Active shape models with SIFT descriptors and MARS, *Int Conf Computer Vision Theory and Applications (VISAPP)*, Lisbon, Portugal, pp 380–387
20. Ouloul IM, Afdel K, Amghar A, Moutakki Z (2015) Automatic face recognition with aging using the invariant features, *5th Intl Conf Information & Communication Technology and Accessibility (ICTA)*, pp 1–5. <https://doi.org/10.1109/ICTA.2015.7426876>
21. Panis G, Lanitis A, Tsapatsoulis N, Cootes T (2016) Overview of research on facial ageing using the FG-NET ageing database. *IET Biom* 5(2):37–46
22. Pavani SK, Sukno FM, Delgado-Gomez D, Butakoff C, Planes X, Frangi AF (2012) An experimental evaluation of three classifiers for use in self-updating face recognition systems. *IEEE Trans Inf Forensics Secur* 7(3):932–943. <https://doi.org/10.1109/TIFS.2012.2186292>
23. Sun Y, Zhang J-M, Wang L-M, Zhan Y-Z, Song S-L (2005) A novel method of recognizing ageing face based on EHMM. *Intl Conf Mach Learn Cybern* 8:4599–4604. <https://doi.org/10.1109/ICMLC.2005.1527749>
24. Szczodrak M, Czyżewski A (2017) Evaluation of face detection algorithms for the bank client identity verification. *Found Comput Decis Sci* 42(2):137–148
25. Szczuko P, Czyżewski A, Hoffmann P, Bratoszewski P, Lech M (2017) Validating data acquired in bank branches with experimental multimodal biometric system. *J Intell Inf Syst*. <https://doi.org/10.1007/s10844-017-0491-2>
26. Tao D, Guo Y, Li Y, Gao X (2018) Tensor rank preserving discriminant analysis for facial recognition. *IEEE Trans Image Process* 27(1):325–334. <https://doi.org/10.1109/TIP.2017.2762588>
27. Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vis* 57(2):137–154
28. Wang S, Xia X, Huang Y, Le J (2013) Biologically-inspired ageing face recognition using C1 and shape features, *5th Intl Conf Intelligent Human-Machine Systems and Cybernetics* pp 574–577. <https://doi.org/10.1109/IHMSC.2013.285>
29. Weinberger K, Blitzer J, Saul L (2006) Distance metric learning for large margin nearest neighbor classification. *Adv Neural Inf Proces Syst* 18:1473–1480
30. Xing EP, Ng AY, Jordan MI, Russell S (2003) Distance metric learning, with application to clustering with side-information. *Adv Neural Inf Proces Syst* 15:505–512



**Piotr Szczuko** received the M.S. degree in telecommunication in 2000, and Ph.D degree in 2008 from the Gdansk University of Technology. Since 2008 he has been an Assistant Professor in the Multimedia Systems Department of Gdansk University of Technology. His research focuses on applications of classification methods in multimedia and human-computer interaction applications, especially in images and video. He is also interested in: machine vision, artificial intelligence and automatic inference methods, classification and perception of

acoustic and visual signals. He is an author of over 80 conference papers and journal articles. Dr. Szczuko was a recipient of Best Paper Awards of IEEE SPA 2015, MCSS 2014, and MCSS 2011 and Prize of Prime Minister for his PhD dissertation in 2009.



**Prof. Andrzej Czyżewski** - Head of the Multimedia Systems Department is author of more than 400 scientific papers in international journals and conference proceedings. He has led more than 30 R&D projects funded by the Polish Government and participated in 5 European projects. He is also author of 20 Polish patents and 4 international patents. He has extensive experience in soft computing algorithms and sound & image processing for applications in multimedia technology.



**Maciej Szczodrak** - Received his M.Sc. degree from the Gdansk University of Technology in 2006. His thesis concerned implementation of algorithms for open air sound propagation analysis. He is interested in audio and video processing, environmental acoustics. He is also concerned about programming. So far he has published, as author or co-author, 70 publications including international journal conference papers and journal articles.

