



# Using a Probabilistic Neural Network for lip-based biometric verification



Krzysztof Wrobel<sup>a,\*</sup>, Rafal Doroz<sup>a</sup>, Piotr Porwik<sup>a</sup>, Jacek Naruniec<sup>b</sup>, Marek Kowalski<sup>b</sup>

<sup>a</sup> University of Silesia, Institute of Computer Science, ul. Bedzinska 39, 41-200 Sosnowiec, Poland

<sup>b</sup> Warsaw University of Technology, Faculty of Electronics and Information Technology, ul. Nowowiejska 15, 00-665 Warsaw, Poland

## ARTICLE INFO

### Keywords:

Biometrics

Lip

Image processing

Probabilistic Neural Network

Particle swarm optimization

## ABSTRACT

In classical recognition techniques only raw features of objects are employed. Our approach allows use the composed features — so called *Sim* coefficients and landmarks which determine the area where biometric features should be searched. Biometric composed features are associated with appropriate similarity coefficients. Such approach brings significant advantages — recognition level of objects is higher compared to method based on the raw data. In this paper, a novel and effective lip-based biometric recognition approach with the Probabilistic Neural Network (PNN) is proposed. Lip based recognition has been less developed than the recognition of other human physical attributes such as the fingerprint, voice patterns, blood vessel patterns, or the face. For this reason, achieved results on this field are still improved and new recognition techniques are searched. Results achieved by PNN were improved by the Particle Swarm Optimization (PSO) technique.

In the first step, lip area is restricted to a Region Of Interest (ROI) and in the second step, features extracted from ROI are specifically modeled by dedicated image processing algorithms. Extracted lip features are then an input data of neural network. All experiments were confirmed in the ten-fold cross validation fashion on three diverse datasets, Multi-PIE Face Dataset, PUT database and our own faces dataset. Obtained in researches result show that proposed approach achieves an average classification accuracy of 86.95%, 87.14%, and 87.26%, on these three datasets, respectively. Announced results were verified in the comparative studies and confirm the efficacy of the proposed lip based biometrics learned by PSO technique.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

Technologies for cyber-physical security have achieved rapid growth over the past two decades. There is mounting evidence that they are important in biometric-based access control. Today, fingerprints, footprints, pistol bullets with their ballistic traits, but also hair, voice, blood, semen, DNA and the fibers from clothes are widely employed in crime detection and some of them can be used in biometrics. The biometric fields are increasing each year and this tendency is well visible due to need for security at borders, buildings, airports, corridors, banking transaction, etc., as a method of screening test and help for security service (Doroz et al., 2016; Porwik et al., 2014; Doroz et al., 2014; Dollár et al., 2010; Bedagkar-Gala and Shah, 2014). There is still much room for improvement with respect to recognition the new biometric features. One of these domains is human recognition through distinctive facial features. Recognition algorithms can be divided into two main approaches: photometric or geometric. Photometric algorithms try to overlay the pictures for a match, whereas geometric algorithms extract landmarks or other features from an image. Our strategy concerns the

second type of such algorithms. We will focus on lip-based biometric, where digital photos including images of lip are analyzed (Aarabi and Mungamuru, 2004; Cao et al., 2014; Raheja et al., 2010; Howell et al., 2016). It has a great practical meaning because biometric modalities can be deployed in many areas including remote access control, secure special areas in airports, banks, to identify the cell phone users or people search. For this reason, analysis of various regions of images including human faces is still an important task and can improve identification effectiveness level. We noticed that features selection procedure can be linked with various similarity measures and classifiers (Doroz et al., 2016, 2014; Czabanski et al., 2012; Krawczyk and Wozniak, 2016). It has improved a machine-learning method of data classification, what is presented in this paper. Recognition of lip images can be conducted on the basis of groups of features — global and local lip features which are located inside of the contour of the lips.

Global features treat patterns as a whole, while local features are extracted from a limited area of pattern. A global feature is a contour of the lip image (latin: *rubor labiorum*). This contour comprises upper and

\* Corresponding author.

E-mail address: [krzysztof.wrobel@us.edu.pl](mailto:krzysztof.wrobel@us.edu.pl) (K. Wrobel).

lower lip. Local features are determined on the basis of measurements of angles and distances between characteristic points of the lip. The lip contour can be considered as a unique signature of face, and we assumed that some characteristic points of the lip contour can be exploited as biometric features. In this paper all measurements on the shape of lip image contour and inside of contour area have been considered in the Cartesian coordinate plane.

## 2. Related works

Lip visual features are generally grouped into three categories: (a) appearance-based features; (b) shape-based features and (c) a combination of both appearance and shape features (Howell et al., 2016; Ibrahim and Mulvaney, 2015). Depending on the method of lip features extraction, researches are focused on the various domains. Part of the works concentrates around automatic speech recognition in human computer interfaces (Aarabi and Mungamuru, 2004; Ibrahim and Mulvaney, 2015), where image containing the lips region is processed and appropriate ROI is established. It was achieved by the Viola–Jones object detection framework. Later, the geometrical information obtained from the lips contour was used for the lips reading process. In the work (Newman and Cox, 2009) authors try to determine the language in a given person speaks. For this task the lip movements is registered during utterance of text. The authors use Active Appearance Models to locate the face and mouth, and build a vector that represents the lip shape for video frame.

Another field of investigations is recognition of facial expression (Raheja et al., 2010). In this method the lip contour is extracted from the object. In the next stage histogram of the lip contour is built and classified. It allows to achieve a lip gesture recognition rate of up to 95%.

Nowadays, global and local lip features are also used as genetic or morphological marker of ethnic differentiation in people populations in many countries (Domiaty et al., 2010; Kapoor et al., 2015; Rastogi and Parida, 2012). Separate researches are focused on medical treatments, where dental and surgical as well as post mortem analysis are preformed (Sharma et al., 2009; Utsuno et al., 2005). Some researches describe also biometric-based identification methods centered on static mouth or face images. It brings a new impulse to improve previously announced results.

There have been various investigations into recognizing a person directly from the lips shape and contour, where lip ROI is determined on the basis of colors distribution around the lip area (Choraś, 2010; Jin et al., 2007). It follows from the fact, that lip color usually has a stronger red component than other parts of the face. These principles were applied in Choraś (2010).

In the biometric system in the first step appropriate set of discriminative biometric features have to be determined and then these features can be classified (Bolle, 2004). The mouth area position on the face image can be sought by means of various methods (Ibrahim and Mulvaney, 2015): (a) matching technique; (b) feature invariant; (c) machine learning and (d) knowledge based.

Matching technique is known method used in image processing where correlation between a template image and an unknown image have to be determined and then classified. Such a technique has been used to detect the face (Ibrahim and Mulvaney, 2015; Jin et al., 2007; Marzec et al., 2015), as well as the eyes and mouth (Bartlett et al., 2006; Naruniec, 2014).

Invariant-based methods employed features which are stable, although some appreciable changes can be observed in the whole image such as its brightness or the pose of the subject, for example. For face, invariant features can be found by color analysis in the mouth ROI (Ibrahim and Mulvaney, 2015; Choraś, 2010; Jin et al., 2007). Machine learning methods are applied to recognizing the different face poses and head rotation registered in image video sequences. The most popular machine learning approaches are Artificial or Convolutional

Neural Networks (Tadeusiewicz, 2015; Specht, 1990; Li et al., 2015), decision trees (Markus et al., 2013), the Viola–Jones object recognizer (Viola and Jones, 2004; Murphy et al., 2016) and HyperFace methods (Ranjan et al., 2016) for face detection and fiducial point extraction. The HyperFace detector automatically extracts all the faces from a given image.

Knowledge-based technique employs knowledge about human face and mouth location — the mouth is positioned in the lower part of the head below the nose and on a line of vertical symmetry and skin color (Ibrahim and Mulvaney, 2015). It is sufficient to determine the mouth ROI. The set of biometric features of a given object is mostly formed as a multidimensional normalized vector. It can be reduced (Porwik et al., 2014; Wang and Liew, 2012), and classified by means of the various methods such as dynamic time warping (DTW) (Ibrahim and Mulvaney, 2015), as well as combination of classifiers (Doroz et al., 2014; Kurzynski et al., 2016; Tadeusiewicz and Horzyk, 2014; Ooi et al., 2016). To the best of our knowledge lips biometrics for person verification is still developed domain. In this paper we show that combining face together with lip landmarks descriptors improving announced in the literature achievements and we demonstrate the validity of this arguments in performed experiments.

## 3. Proposed innovation

The surface of lips is covered by contours and numerous reticular depressions (grooves) that form system of lines. The lines of a lip pattern can be illustrated as shapes and can have various representations. As was previously stated, to date most of the existing work concentrates on lip movement of speaker or language recognition, where combination of lip movement and static lip features are analyzed. In such systems lip dynamics is sufficient to limited number of spoken words recognition (Aarabi and Mungamuru, 2004; Ibrahim and Mulvaney, 2015; Newman and Cox, 2009). This follows from the fact that this domain is still relatively new, being explored and developed only over the recent years. Popularity of lip print sensors is not yet high and investigations in this area are still being developed. It is the second reason why biometric systems based on lip analysis are not widespread.

It should be noted that also multimodal biometric systems are applied because high performance of the system can be easier to get, when various modalities will be employed. Designing of the biometric system with high accuracy is more difficult when only single modality is available. It is usually a more challenging task. In this work, we focus on the single modality approach, based on static shape information. These data can be collected either from photographs or video cameras as single video frames.

We present a novel biometric system based only on lip contours and new lip geometrical measurements. It means that in contrast to other methods texture of the lip surface is ignored. It simplifies computations without deterioration of the recognition quality. Lip extraction is possible through the two new independently working classifiers which operate on sets of the different features extracted from the image. Proposed verification model is trained by Probabilistic Neural Network (PNN) and optimized by Particle Swarm Optimizer (PSO) (Hashemi and Meybodi, 2011; Garg, 2016; Ali et al., 2015). It should be emphasize that the proposed technique achieves a classification accuracy of up to 87.26%. To the best of our knowledge, a study in lip-based biometrics using PNN + PSO methods has not been investigated before.

It is well known, that there are dozens of biometric modalities because any measurable biological or behavioral characteristic can be a biometric modality. In our case instead of face-based modality we propose lip-based modality. It causes that some problems which occur in the biometric systems based on face analysis can be eliminated. For example, if a given person wears glasses or beard, face recognition process can be unsettled in cases when other face image of the same person will be taken without glasses or facial hair. Our approach allows also correctly to designate all landmarks for ethnic faces because, in

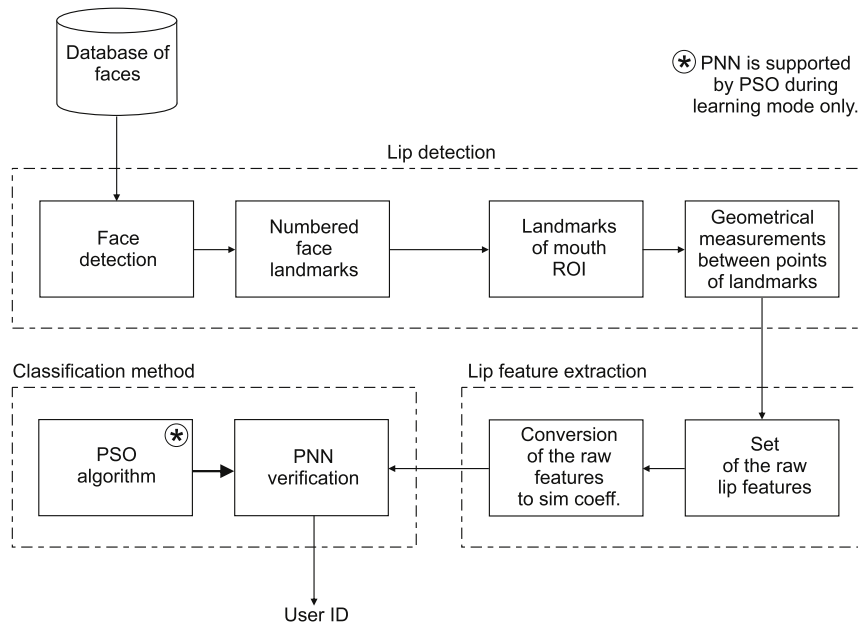


Fig. 1. Block diagram of the proposed approach.

contrast to other methods, processing phase neglects color of the image. It will be shown that lip-based biometrics gives also high recognition accuracy. The system aims to provide efficient and reliable detection of lip landmarks on the face ROI. It means that any additional searching of lip zone is not needed. It would be easy to get a high accuracy via the multimodal approach with many biometrical data, however, developing a high performance system based on solely one modality is usually a more challenging task.

#### 4. Preprocessing – face and mouth detection

Our biometric model consists of the three parts — face and mouth detection, mouth and lip features extraction and classification. Dataset can come from appropriate databases consist of photographs or single video frames. In our experiments some static image samples were captured and taken with various resolutions. From practical point of view, there are three important well known factors that need to be considered when designing and implementing a lip-based biometric verification system. The first is the choice of the method to locate the position of the mouth in the image of a given person; the second is the choice of appropriate lip features extracted from the mouth ROI and the third is choice a classification method. Such idea was presented by means of block diagram scheme (Fig. 1).

In the first stage, both the face and mouth ROI are designated. In the next stage in the mouth and lip area, the various landmarks are located. Between these points, specific Euclidean distance-based measurements are then performed. The data are then converted to so-called *Sim* coefficients, which are input data of biometric classifier. Principles of the *Sim* coefficient calculation will be presented in details in the next paragraphs. Computed values form a vector of features and then features are used in the classification process. Person images as acquired from the single frames of the video camera sequence in real time, so face and mouth detection algorithm has to work fast. In our approach we use two classifiers which increase the accuracy of face localization and provide robustness against majority small geometrical transformations of faces. We propose to use two types of features along with two types of classifiers (Naruniec, 2014):

- Histograms of Oriented Gradient (HOG) features (Dalal and Triggs, 2005; Surinta et al., 2015) classified by a logistic regression method (Dollár et al., 2010; Cao et al., 2014).

- Extended set of the Haar-like features based on Viola–Jones framework (Lienhart and Maydt, 2002; Murphy et al., 2016). Obtained features are classified by simple probabilistic classifier.

At first, a set of rectangular analysis windows is formed. Analysis windows define the possible face scales and locations, spread greedily across the whole image. Their height and width are multiplies of  $24 \times 24$  window. Images formed in the region of each analysis window are then classified by the ensemble of HOG-based and Haar-based classifiers. In HOG-based classification method, image is divided into blocks and cells, in which the histograms of gradient orientations are evaluated. The descriptor of analysis window consists of a concatenation of the histograms of chosen blocks. Image descriptors can be then classified by the various methods (Naruniec, 2014; Dalal and Triggs, 2005).

The Haar-like features are adjacent, rectangular regions, divided into smaller positive and negative rectangles. In each rectangle, intensities of pixels are added up. Subsequently, the difference between the sums of positive and negative regions is calculated. Such simple features itself cannot be used to recognize complex objects like, for example, human faces. For this reason the strong GentleBoost classifier is used to boost the classification performance of a simple (so-called weak) cascade classifiers. We chose the GentleBoost classifier because it is resistant to noisy data (Dollár et al., 2010). The both HOG and Haar-based algorithms are implemented in many programming languages like Matlab, Java, C++ and others. Discussed algorithms are also well represented in the literature (Baş, 2015; Sagonas et al., 2013), for this reason details of these procedures have been omitted.

After all previously mentioned stages, we obtain image, where face zone is clearly separated. This zone in the rectangle form is depicted in Fig. 4(a). In the next step semantic facial landmarks such as eyes, nose, mouth and chin have to be identified. The detection of facial landmarks is an essential part of face recognition systems and plays a key role in our biometric system therefore idea of the landmarks localization will be more precisely presented in this paper.

In our approach we use the predetermined shape  $S^0$ , which has a following geometrical construction (Naruniec, 2014).

A face ground initial shape model  $S^0 = [(x_1, y_1), \dots, (x_{68}, y_{68})]$  consists of the 68 facial landmarks represented by means of the Cartesian coordinates (Fig. 2). The landmarks can be independently moved which allows fitting individual landmarks to the given face elements. The goal of face alignment is to transform a shape  $S^0$  to the shape of a real face

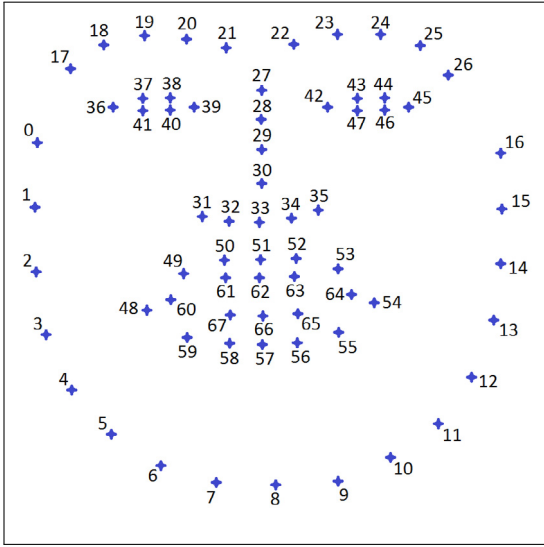


Fig. 2. The initial geometrical face (ground shape) model  $S^0$  with numbered details.

$S^I$  on the image  $I$ . It can be done by minimizing alignment error. At the beginning the shape  $S^I$  is unknown, therefore the alignment procedure have to be derived in the multistage learning process. Realization of this task requires pool of the  $M$  training images. The search region is limited to the face ROI. Individual landmarks from Fig. 2 are manually imposed on a given image  $I$  and then moved and fitted to the all face elements, so the set  $\{I_i, S_i^0\}_{i=1}^M$  of training examples is formed. As was previously stated, every landmark  $(x, y) \in S^0$  on the template image  $I$  is manually matched, hence for this landmark its HOG feature  $h$  is also known. It means that the learning set  $D = \{\{I_i, h_i\}_{i=1}^M, \{(x_j, y_j)\}_{j=1}^{68}\}$  is established. Each landmark is accurately recognized by a fast regression tree classifier, which is trained from the appearance HOG feature  $h$  around this landmark. Hence, regression forest is an ensemble of different regression trees. Learning process is stopped when alignment error  $\epsilon$  no longer decreases — differences between ground shape and last estimation is stabilized. It means that we search the landmarks mapping function  $f(I, S^0) \rightarrow \delta S$  at the facial image  $I$  that minimize landmarks alignment of the Euclidean norm error:

$$\|S^0 + \delta S - S^I\| = \epsilon, \quad (1)$$

where  $S^I$  is the ground truth shape of the target image  $I$ . Above formula shows that difference between the ground shape and current estimated shape is computed. The best shapes adjustment can be performed in the two well known strategies (Dollár et al., 2010; Naruniec, 2014):

- optimization-based,
- cascade-regression (CR)-based framework.

We prefer CR-based alignment method, because optimization-based approach has limitation for faces captured in various poses, expressions and image illumination (Dollár et al., 2010). Regression based methods, learn a regression function from image appearance (features) to the target shapes (Dollár et al., 2010; Cao et al., 2014). Cascade architecture is usually employed and explored in regression based models. In each stage of the cascade architecture, shape-index features (Cao et al., 2014) are extracted to predict the shape increment with linear regression.

The proposed CR-based classifier was trained using randomly selected images from the HELEN (Le et al., 2012) and LFPW (Belhumeur et al., 2013) datasets. Since our study focuses on static lip features, only limited range of poses presented in the mentioned above databases (neutral, smiling and indifference) were used in learning and then testing experiments. Furthermore, images containing non-frontal poses have been ignored.

Each training sample consists of a training image, an initial shape and a ground truth shape. Starting from an initial face shape more fitted shapes are estimated in the fixed number of iterations:

$$S^t = S^{t-1} + A^t \cdot \varphi^t(I, S^{t-1}), \quad (2)$$

where  $A^t$  is a trained matrix projection, and  $\varphi^t(I, S^{t-1})$  is a shape-related feature mapping function.

A few steps of landmarks alignment present Fig. 3. After all preprocessing stages, real images can be tested. One of them is presented in Fig. 4a, whereas the same face image with all landmarks is depicted in Fig. 4b. It is important that in our method the mouth ROI is automatically detected and any additional analysis techniques are not needed. During face detection phase, mouth ROI is determined by numbered landmarks (Fig. 4c and Fig. 4d), separately for upper and lower lip. Numbered points will be used in the mouth geometrical measurements what simplify organization of automatic measurements. Numbering principles are always the same for every face-based image. It is worth to notice that in our approach, number of generated landmarks is independent of size of image. It is an advantage of the method and it simplify both preprocessing and verification processes.

Proposed in this paper lip extraction technique is robust to the change of skin colors as well as background color. This was demonstrated by the lip contour extraction results on the Multi-PIE dataset illustrated in Fig. 5. Lip contour extraction as well as landmarks localization is automatically performed. On the presented images faces skin is from light to much darker skin color. These examples also show clearly that proposed lip extraction approach correctly determine lip ROI, even on faces with facial hair. It follows from the fact that the face detector (Fig. 4) is matched to the all face's elements, so even partially hidden mouth elements are correctly marked out. It is clearly visible

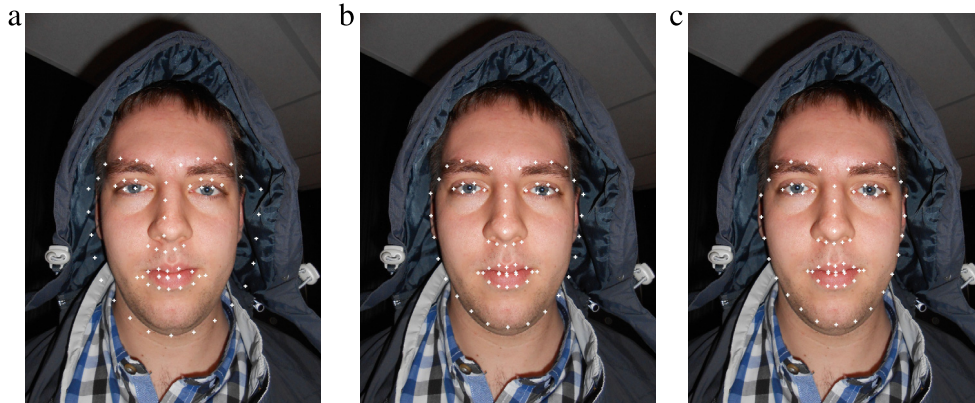


Fig. 3. A few iterative steps of landmarks localization: (a) the first image, (b) one step of alignment and (c) the last image with the best landmarks alignment.



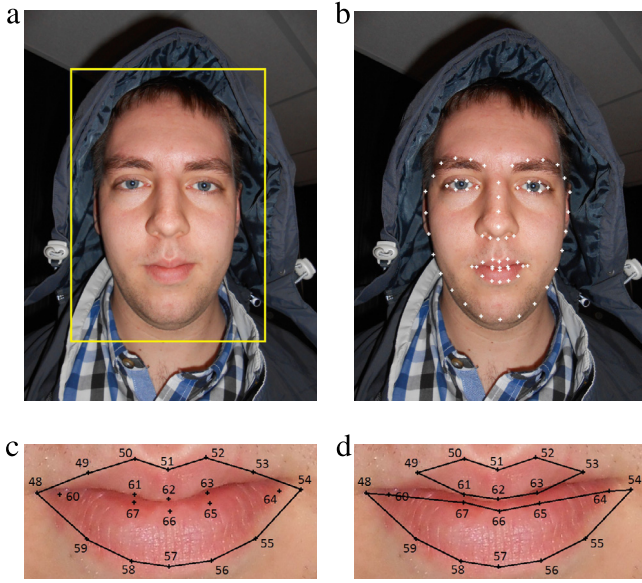


Fig. 4. (a) Input image with face detector (ROI), (b) the same image with all landmarks alignment by using the Haar and HOG features, (c) contour of the mouth ROI, (d) upper and lower numbered landmarks for lip characterization.

on Fig. 5 for the persons with mustache and beard. It is worth pointing out that, it is not necessary to conduct images rotation correction since there are no rotated lips presented.

##### 5. Biometric lip parameters determination

For the best classification accuracy the following distances, curvatures and shapes have been measured on the lip contour and on the lip area. These distances form a features vector for every individual. Principles of these measurements can be shown graphically — it is a simplest and well understandable description. For simplification the landmarks are not numbered.

Fig. 6 shows that 11 different feature vectors were formed and in total the 132 various parameters were measured for every lip-based image. Number of these measurements results from the fact, that human face is asymmetrical from anthropometry point of view. It enforces various types of measurements — predominantly from left and right corner of mouth. Additional, what is presented in Fig. 6g, whole lip

area is also calculated as well as the upper and lower lip area. In the areas presented in Fig. 6j and Fig. 6k contour curvature by means of the IPAN 99 algorithm has been determined. Mentioned algorithm is described in Doroz et al. (2015) and Chetverikov (2003), therefore implementation details were here omitted.

In a nutshell the main principles of the IPAN 99 algorithm are as follows. Let  $p_i$  be an actually analyzed point of a given contour  $l$  then between two adjacent selected points the triangle can be inscribed (Fig. 7). For every triangle the angle  $\alpha$  is calculated:  $\alpha = \arccos \frac{a^2+b^2-c^2}{2ab}$ . Finally, for a given contour  $l$  the set  $G_l = \{\alpha_i^l, (x_i, y_i)^l\}_{i=1}^{20}$ , is formed, where  $\alpha_i^l$  denotes curvature of the contour  $l$  at the point coordinates  $(x_i, y_i)$  and 20 denotes number of all landmarks in the analyzed lip contour.

##### 6. The lip features preparation

The main goal of this paper is analysis of the efficiency of the recognition algorithms and verify images of faces which are stored in the database. For this reason the two-class verification problem will be discussed. In the first step, for each person, two sets of images are formed. Let the set containing images of original faces of a given person be denoted as  $\pi_1 = \{O_1, O_2, \dots, O_c\}$  and the set containing faces of other users randomly selected from the same database be denoted as  $\pi_2 = \{O_1^A, O_2^A, \dots, O_d^A\}$ . Instead of face images their lip-based features are analyzed. For all images the same biometric features have been determined. It was shown in previous paragraph (see Fig. 6). Let the set of lip-based features registered in the same preprocessing stage be denoted as  $F = \{f_1, f_2, \dots, f_u\}$ .

Objects can be compared by means of a range of different methods. The most popular coefficients and similarity measures are reported in Porwik et al. (2009) and Cha (2007). Instead of “similarity measure”, we will also use the alternative name “method” frequently.

Let  $p_i$  represent an  $i$ th value of a given feature  $f$  of the face  $P$ , while  $w_i$  represents an  $i$ th value of the same feature on the face  $W$  and  $z$  is a total number of samples of a given feature. Values  $p_i$  and  $w_i$  represent the same feature  $f$  measured for both  $P$  and  $W$  facial image. Hence, the similarity  $\omega$  of the feature  $f$  in the images  $P$  and  $W$ , can be expressed in the space  $\mathbb{R}^z$ , for example, as follows:

- for the Euclidean distance measure:

$$\omega_E(f) = \left( \sum_{i=1}^z |p_i - w_i|^2 \right)^{1/2}, \quad (3)$$

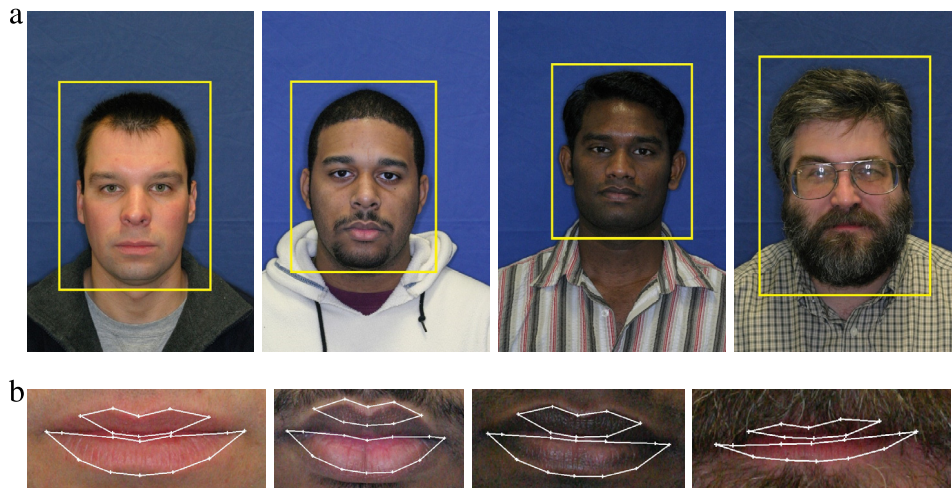


Fig. 5. The lip detection for the Multi-PIE face dataset: (a) example of original images, (b) magnification of corresponding extracted contours.

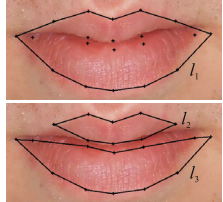
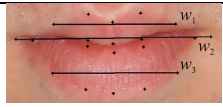
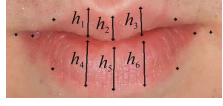
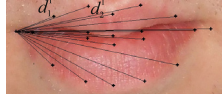
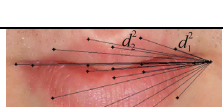

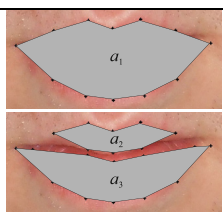
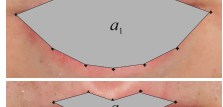
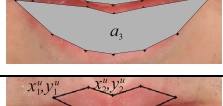
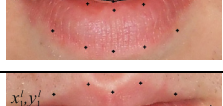
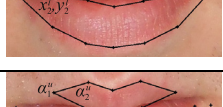
| No | Type of measuring                                                       | Feature                                                    | Feature visualization                                                                 |
|----|-------------------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------------------------------------|
| a) | Lengths of the mouth contour around ROI and upper and lower lip contour | $\mathbf{f}_1 = [l_1, l_2, l_3]$                           |    |
| b) | Widths of the lips in the selected points                               | $\mathbf{f}_2 = [w_1, w_2, w_3]$                           |    |
| c) | Heights of the lips in the selected points                              | $\mathbf{f}_3 = [h_1, \dots, h_6]$                         |    |
| d) | Distances from the left mouth corner to the remaining contour points    | $\mathbf{f}_4 = [d_1^1, \dots, d_{19}^1]$                  |    |
| e) | Distances from the right mouth corner to the remaining contour points   | $\mathbf{f}_5 = [d_1^2, \dots, d_{19}^2]$                  |    |
| f) | Distances from the center of the mouth to the remaining contour points  | $\mathbf{f}_6 = [d_1^3, \dots, d_{19}^3]$                  |    |
| g) | Areas of the mouth ROI and the upper and lower lip                      | $\mathbf{f}_7 = [a_1, a_2, a_3]$                           |   |
| h) | Coordinates $x^u, y^u$ of the upper lip contour                         | $\mathbf{f}_8 = [x_1^u, y_1^u, \dots, x_8^u, y_8^u]$       |  |
| i) | Coordinates $x^l, y^l$ of the lower lip contour                         | $\mathbf{f}_9 = [x_1^l, y_1^l, \dots, x_{12}^l, y_{12}^l]$ |  |
| j) | Curvatures $\alpha^u$ of the upper lip contour                          | $\mathbf{f}_{10} = [\alpha_1^u, \dots, \alpha_8^u]$        |  |
| k) | Curvatures $\alpha^l$ of the lower lip contour                          | $\mathbf{f}_{11} = [\alpha_1^l, \dots, \alpha_{12}^l]$     |  |

Fig. 6. Measurement principles of the lip-based features.

- for the Pearson coefficient:

$$\omega_P(\mathbf{f}) = \frac{\sum_{i=1}^z (p_i - \bar{p})(w_i - \bar{w})}{\left[ \sum_{i=1}^z (p_i - \bar{p})^2 \right]^{1/2} \left[ \sum_{i=1}^z (w_i - \bar{w})^2 \right]^{1/2}}, \quad (4)$$

where  $\bar{p}$  and  $\bar{w}$  are mean values of the appropriate lip feature  $\mathbf{f}$ .

In the next part of the paper, similarity coefficients (3), (4) and others will be denoted as *Sim*. It will be explained below. Through the use of the same feature  $\mathbf{f}_m \in F$  occurring in two lip-based images, the similarity of

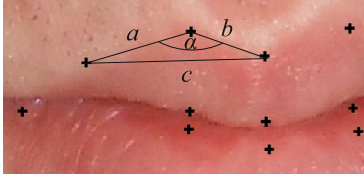


Fig. 7. Lip landmarks and typical triangle localization.

these images can be computed. The prepared data can also be collated in the form of a matrix. The matrix  $X$  is built on the basis of the set  $\pi_1$ . The matrix  $X$  contains values of the similarity coefficients  $Sim$  calculated between the pairs of images from the set  $\pi_1$ .

Let  $[O_i \leftrightarrow O_j]$  denotes pair of the images  $O_i$  and  $O_j$ . The matrix  $X$  comprises of number of the one-columnar vectors  $[O_i \leftrightarrow O_j]$  and, in general, has the following structure:

$$X = [[O_1 \leftrightarrow O_2], \dots, [O_1 \leftrightarrow O_c], \dots, [O_{c-1} \leftrightarrow O_c]]_{u \times C_c^2}, \quad (5)$$

where:  $O_i, O_j$  is the  $i$ th and  $j$ th lip image of a given person,  $c$  is the number of all images of a given person,  $C_c^a$  is the binomial coefficient (the Newton's symbol)  $\binom{b}{a}$ .

Each columnar vector contains values of the similarity-composed features between all images from the set  $\pi_1$ . The values of similarity coefficients are computed using all combinations of “feature-method” pairs which will be called the  $Sim$  coefficients. An example of the first columnar vector of the matrix  $X$  is shown below:

$$[O_1 \leftrightarrow O_2] = \begin{bmatrix} Sim(O_1, O_2)^{f_1} \\ Sim(O_1, O_2)^{f_2} \\ \vdots \\ Sim(O_1, O_2)^{f_u} \end{bmatrix}_{u \times 1}, \quad (6)$$

where:  $Sim(O_a, O_b)^{f_m}$  is the similarity coefficient of the feature  $f_m$ , measured for the lip images  $O_a, O_b \in \pi_1$ . Finally, the fully populated matrix  $X$  has  $u \times C_c^2$  elements.

The second matrix  $Y$  is created on the basis of the both  $\pi_1$  and  $\pi_2$  sets. It can be observed that the matrix  $X$  was constructed on the basis of the original images of a given person whereas the matrix  $Y$  consists of data from original images of a given person and face images of other users. The matrix  $Y$  is built as follows:

$$Y = [[O_1 \leftrightarrow O_1^d], \dots, [O_1 \leftrightarrow O_d^d], \dots, [O_c \leftrightarrow O_d^d]]_{u \times (c \cdot d)}, \quad (7)$$

where  $O_i, O_j^d$  are the  $i$ th original lip image of a given person and the  $j$ th lip image of other person from the same database,  $d$  is the number of all other images randomly selected in the learning mode.

Columns of the matrix  $Y$  are constructed similarly to the columns of the matrix  $X$ . The first columnar vector has the following structure:

$$[O_1 \leftrightarrow O_1^d] = \begin{bmatrix} Sim(O_1, O_1^d)^{f_1} \\ Sim(O_1, O_1^d)^{f_2} \\ \vdots \\ Sim(O_1, O_1^d)^{f_u} \end{bmatrix}_{u \times 1} \quad (8)$$

where  $Sim(O_a, O_b^d)^{f_m}$  is the similarity coefficient of the feature  $f_m$  between the lip images  $O_a \in \pi_1$  and  $O_b^d \in \pi_2$ .

The matrix  $Y$  includes only similarities  $Sim$  between the original images of a given person and images other users from the same database. Hence the fully populated matrix  $Y$  contains  $u \times (c \cdot d)$  elements. Matrices  $X$  and  $Y$  always have the same number  $u$  of rows. In both matrices  $X$  and  $Y$ , each  $Sim$  value represents the value of composed feature. These matrices form the global matrix  $J = [X; Y]$ . The global matrix will be used in the classification of the lip-based faces.

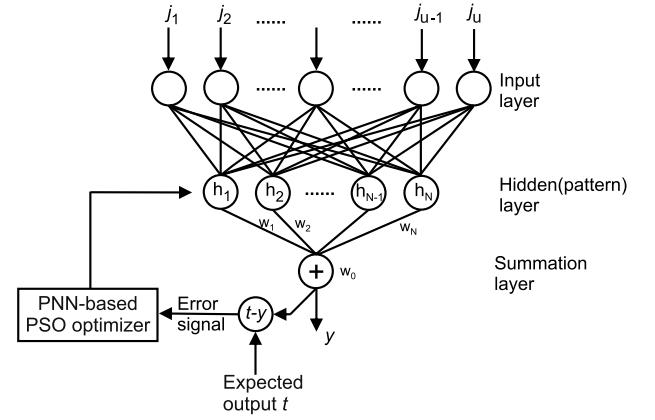


Fig. 8. General structure of the PNN aided by the PSO optimizer in the biometric lip-based classification.

## 7. Probabilistic neural network as a lip-based classifier

In the literature, most of the studies implement the back-propagation neural network (BPNN) as a decision-making model for the monitoring system (Tadeusiewicz, 2015). However, the BPNN needs a large amount of data and time to train the network. Furthermore, when weights between each neuron have been set up, the model cannot be modified. To solve this issue, a Probabilistic Neural Network (PNN) we propose as a self-learning tool. It should be noted that there are many types of classifiers; the most representative of them will be tested in this paper. From investigations carried out, it follows that classifiers based on Probabilistic Neural Networks (PNN) (Tadeusiewicz, 2015; Specht, 1990) give significantly better lip-based verification results compared to other classifiers. For this reason, the proposed structure of the PNN has been more precisely presented. The tasks that artificial neural networks are applied to tend to fall within broad categories such as function approximation, classification including handwriting recognition (Porwik et al., 2016), data processing, etc. (Gaines and Compton, 1995).

The PNNs belong to family radial basis networks (RBNs), where radial basis functions (RBFs) are applied (Tadeusiewicz, 2015). Basic functions are known as kernel functions. Probabilistic Neural Network (PNN) often learns more quickly than many neural network models such as back propagation networks and has been successfully used in variety of applications. Proposed architecture was designed for recognition of the lip-based biometric data of a given person. Therefore, such classifier works in a verification mode. The PNN architecture is composed of many interconnected neurons organized in successive layers. A general structure of the proposed PNN consists of four layers, including input layer, hidden (pattern) layer, summation layer and output layer. The proposed structure of PNN working in the verification mode is shown in Fig. 8.

In experiments, the previously presented matrices  $X$  and  $Y$  have been used. These matrices form a global matrix  $J = [X; Y]$ . The matrix  $J$  includes similarities between all original face images and similarities between original and random face images stored in the database. These similarities have been computed on the basis of lip-based features and selected similarity measures. Columns of the matrix  $J$  can be treated as separate vectors:

$$J = [J^1, J^2, \dots, J^N], \quad J^n = [j_1^n, j_2^n, \dots, j_u^n]^T \in \mathbb{R}^u, \quad n = 1, \dots, N, \quad (9)$$

where  $N = C_c^2 + (c \cdot d)$ , what follows from the dimensions of the matrices  $X$  and  $Y$ . The elements of the vectors  $J^n$  are directly distributed to the input layer the PNN.

The input layer unit does not perform any computation and simply distributes the input to the neurons in the pattern layer. The input layer receives the elements of the vectors  $J^n$ . Fan-out input nodes branch each



input node to all nodes in the hidden layer, so each hidden node receives the complete input features of the vector  $\mathbf{J}^n$ . This means that the input layer contains  $u$  nodes — one node for each feature. The selection of the number of neurons in the hidden layer still remains an open problem and only general recommendations dictated by intuition and practical experience can be formulated (Specht, 1990). In a proposed approach, the pattern layer consists of  $N$  neurons. This number is equal to the number of input training vectors. Input vectors form the  $N$  centroids with the same number of centers  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N \in \mathbb{R}^S$ . In the next stage, for the each pattern node its output activation function  $\mathbf{h}^n = G(\mathbf{J}^n, \mathbf{c}_i)$ , is formed, where  $G$  is a radial Gauss function:

$$G(\mathbf{J}^n, \mathbf{c}_i) = \exp \left( -[(\mathbf{J}^n - \mathbf{c}_i)^T (\mathbf{J}^n - \mathbf{c}_i)]^{1/2} / 2\sigma^2 \right), \quad i = 1, \dots, N, \quad (10)$$

where  $\sigma$  is a free parameter and distance between vectors  $\mathbf{J}^n$  and  $\mathbf{c}_i$  is determined by an Euclidean metrics. It should be noted that  $\sigma$  coefficient will be selected by PSO optimizer.

For a given input training vector  $\mathbf{J}^n$ , in the pattern layer activation functions are calculated:

$$\mathbf{h}^n = [G(\mathbf{J}^n, \mathbf{c}_1), G(\mathbf{J}^n, \mathbf{c}_2), \dots, G(\mathbf{J}^n, \mathbf{c}_N)]. \quad (11)$$

These formulas are reflected in the connections between neurons of the two first layers: the input and pattern layers, where all input neurons have a connection to all of the neurons of the pattern layer neurons (Fig. 8). On the basis of the vectors  $\mathbf{J}^n$ ,  $n = 1, \dots, N$ , the all neurons activation matrix  $\mathbf{H}$  in the pattern layer is formed:

$$\mathbf{H} = \begin{bmatrix} j_1^1 & \dots & j_u^1 \\ \vdots & \dots & \vdots \\ j_1^N & \dots & j_u^N \end{bmatrix} \rightarrow \mathbf{H} = \begin{bmatrix} G(\mathbf{J}^1, \mathbf{c}_1) & \dots & G(\mathbf{J}^1, \mathbf{c}_N) \\ \vdots & \dots & \vdots \\ G(\mathbf{J}^N, \mathbf{c}_1) & \dots & G(\mathbf{J}^N, \mathbf{c}_N) \end{bmatrix}. \quad (12)$$

Elements of the matrix  $\mathbf{H}_{N \times N}$  are input data for the output layer of the PNN. In the proposed approach, the output layer consists of one neuron only. For the input vector  $\mathbf{J}^n$ , on the PNN output  $y^n$ , the following signal is produced:

$$y^n = G(\mathbf{J}^n, \mathbf{c}_1)w_1 + G(\mathbf{J}^n, \mathbf{c}_2)w_2 + \dots + G(\mathbf{J}^n, \mathbf{c}_N)w_N + w_0, \quad (13)$$

where weights  $w$  are created during training process and  $y^n \in \{-1, +1\}$ .

After performing mentioned above stages, the PNN is trained. The PNN error is defined as the mean squared error (MSE):

$$MSE = \sum_{n=1}^N (t^n - y^n)^2, \quad (14)$$

where  $t^n \in \{+1, -1\}$  is the desired target for the  $n$ -th training pattern vector.

If instead of  $y^n$  in the Eq. (14), substitute the right side of the Eq. (13) then the  $N$  linear equations will be constructed to minimize the MSE value. This can be done by appropriate selection of the weights  $w$  in the Eq. (13). The value  $t^n$  is an expected answer of the PNN during learning stage. The parameter  $\sigma$  which appears in the Eq. (10) is so-called smoothing parameter (Berthold and Diamond, 1998). In our method, metaheuristic PSO optimization method has been proposed. The PSO gets better results in a faster and cheaper way compared with other methods (Berthold and Diamond, 1998). Another reason that PSO is attractive is that there are few parameters to adjust.

## 8. PSO-based optimizer for the PNN error minimization

The Particle Swarm Optimization (PSO) algorithm can be applied in function optimization, artificial neural network training, fuzzy systems control, and other areas (Hashemi and Meybodi, 2011; Garg, 2016; Ge et al., 2008). In this approach, global optimums can be searched as well as local optimal solutions. For PSO, the appropriate particle has a dominant influence over the whole swarm; therefore, the solution can be found in the shorter time. For these reasons, the PSO algorithm has been selected for the optimizing of the  $\sigma$  value in the PNN classifier. The general structure of the PNN tuned by the PSO procedure presents

Fig. 8. It means that during optimization process, the values of both the smoothing parameter  $\sigma$  and the weights  $w_i$  in ((12), (13)) will be computed again. The  $\sigma$  parameter estimation for minimizing the MSE of the PNN can be carried out by PSO. The PSO algorithm allows to find the optimal  $\sigma$  value fast and in a wider range of changes. The particle (agent) at iteration  $t$  has two attributes: a current position  $\sigma_g^t$  and a current velocity  $\lambda \sigma_g^t$ .

In every iteration, each particle  $P$  is updated by the two values  $P_i$  and  $P_g$ . The value  $P_i$  is the best previous position of the particle, while the  $P_g$  value is the best position among all the particles in the swarm. It is the global most relevant position. After finding these two values, the particle updates its velocity and position according to the following formulas:

$$\lambda \sigma_g^{t+1} = \lambda \cdot \lambda \sigma_g^t + c_1 \cdot r_{1,t} \cdot (P_i - \sigma_g^t) + c_2 \cdot r_{2,t} \cdot (P_g - \sigma_g^t). \quad (15)$$

Updating of the each particle position:

$$\sigma_g^{t+1} = \sigma_g^t + \lambda \sigma_g^{t+1}, \quad (16)$$

where  $\lambda$  is the internal weight controlling the impact of the current velocity on the next velocity,  $c_1, c_2 \in \mathbb{N}$  are the acceleration parameters pulling each particle toward the best positions,  $r_1$  and  $r_2$  are random numbers uniformly distributed in the range  $[0, 1]$ ,  $\sigma_g^t$  is the current position of the  $g$ -th particle at the epoch  $e$ . The convergent condition is a set of predefined number of epochs  $e = 1, \dots, e_{\max}$  which should be performed. However, to reduce computational time, it is often common practice for the algorithm to be stopped before complete convergence using heuristic approaches. In our case, PSO algorithm is stopped when estimated parameter  $\sigma$  become stable. In our experiments, the parameter  $e_{\max}$  was established to 20 and the number of randomly located particles in the PSO algorithm was established to 30. The searching domain of the  $\sigma$  parameter has been restricted to the range  $\sigma \in [0.1, 5]$ . In the each PSO epoch, separately for every particle  $g = 1, 2, \dots, 30$ , the PNN smoothing parameter  $\sigma$  and network accuracy have been established on the basis of the same training set.

## 9. Characteristic of the datasets

All experiments conducted were carried out on the basis of three face-content databases. Databases include images of various sizes and different facial expressions. Since our study focuses on static lip features, only neutral, smiling and indifference expressions present in the databases have been selected because it suit our purpose. Images containing non-frontal poses or subjects with their mouth open have been omitted in experiments.

The *Multi-PIE database* (Sim et al., 2003; The MultiPie database, 2009) comprises of color images of size  $2048 \times 3072$ . From this database 337 objects were taken into consideration, 5 images per object. This database consists of systematically captured subjects with varying poses and illuminations. Images were recorded in up to four sessions over the span of five months.

The *PUT face database* (Kasinski et al., 2008; The Put database, 2008) contains almost 10,000 images of 100 people acquired in partially controlled conditions and stored in  $2048 \times 1536$  pixels color images. In experiments 100 objects, 5 samples per object were analyzed. Those images were taken while people were moving heads without any constraint to the pose or expression. Images of each person are marked and stored in separate folders.

Our *own faces database* contains 50 both color and grayscale images of objects, 5 images per object. Images have various sizes – from range  $3096 \times 3456$  to  $4128 \times 4608$  pixels. Mentioned own faces dataset is available on the website (Our database, 2016) and can be used in other experiments. It should be noticed that our facial database is not a homogeneous, so it is more realistic in contrast to other databases, where only color images with the same resolution are stored. Proposed in our approaches lip features determination is independent of image color, as a result, some preprocessing steps may be skipped – in contrary to other analysis methods (Choraś, 2010; Marzec et al., 2015).



**Table 1**

Lip features influence on the best accuracy of the PNN classifier (Multi-PIE database).

| Lip features |       |       |       |       |       |       |       |       |          |          | Accuracy               | FAR                    | FRR                    | ERR                    |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|------------------------|------------------------|------------------------|------------------------|
| $f_1$        | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | [%]                    | [%]                    | [%]                    | [%]                    |
| x            | x     | x     | x     | x     | x     |       | x     | x     |          | x        | <b>86.95</b><br>± 1.21 | <b>15.32</b><br>± 0.35 | <b>12.63</b><br>± 0.56 | <b>12.71</b><br>± 0.53 |
| x            | x     | x     | x     | x     | x     | x     |       |       | x        | x        | 84.73<br>± 1.33        | 17.53<br>± 0.57        | 13.58<br>± 0.46        | 14.21<br>± 0.46        |
| x            | x     | x     |       |       |       |       | x     |       | x        | x        | 83.89<br>± 0.89        | 17.77<br>± 0.49        | 15.46<br>± 0.42        | 15.97<br>± 0.64        |
| x            | x     | x     | x     | x     | x     | x     | x     | x     | x        | x        | 82.31<br>± 1.02        | 19.32<br>± 0.66        | 16.21<br>± 0.61        | 17.23<br>± 0.48        |
| x            | x     | x     |       |       |       |       | x     |       |          |          | 82.12<br>± 0.94        | 19.45<br>± 0.37        | 16.61<br>± 0.55        | 17.94<br>± 0.51        |
|              |       |       | x     | x     | x     |       |       |       |          |          | 81.58<br>± 1.18        | 21.12<br>± 0.65        | 17.31<br>± 0.49        | 18.37<br>± 0.52        |

## 10. Results of the experiment

The aim of image classification is to automatically build computerized models able to predict accurately the class of new images, previously trained from a set of labeled images. This strategy is well-known in automatic recognition of handwritten characters, faces images, cells, to name but a few. In this section experimental investigations are presented in order to examine the quality of proposed the new lip-based biometric approach. The main idea of the biometric verification is based on the mentioned above *Sim* coefficients gathered in the matrices  $\mathbf{X}$ ,  $\mathbf{Y}$ . These data have been recognized by various classifiers, especially by the PSO-based technique. In order to compare our method with other classifiers reported in this paper, the same scenario was considered. The available data were divided into training, validation and test set. Classification was performed using the Leave-One-Out cross validation strategy (Berthold and Diamond, 1998). It ensures that the risk of overoptimistic results is significantly reduced. All results are reported in terms of classification accuracy based on different runs prepared separately for each dataset. In experiments, as a similarity measure the Pearson coefficient has been applied. In the first stage the matrix  $\mathbf{X}$  is formed. This matrix contains the Pearson's similarity coefficients determined between the  $c = 5$  face-based images of the same person. Thus, the matrix  $\mathbf{X}$  consists of  $u \times u \times C_c^2 = u \times 10$  elements, where  $u$  denotes the number of analyzed lip features. In the next stage the matrix  $\mathbf{Y}$  is constructed. This matrix includes similarities between objects from the legitimate images of a given person and other ( $d = 2$ ) randomly selected object from a given database. Thus, the matrix  $\mathbf{Y}$  comprises  $u \times (c \cdot d) = u \times (5 \times 2) = u \times 10$  elements. In the last stage, both  $\mathbf{X}$  and  $\mathbf{Y}$  matrices are joined and form the global matrix  $\mathbf{J} = [\mathbf{X}; \mathbf{Y}]$  but information about the two parts of the global matrix is also stored. In classifiers learning mode, elements of the matrix  $\mathbf{X}$  are fixed and the matrix  $\mathbf{Y}$  is formed 10 times – always for other images than those in the matrix  $\mathbf{X}$ , so we construct 10 various matrices  $\mathbf{J}$ . In the future, based on the  $\mathbf{J}$  matrix, the models of the various classifiers will be tested in the Leave-One-Out strategy. Finally, the average classifier's accuracy is calculated.

### Experiment 1

*Set up:* Described previously the composed similarity measure  $Sim(O_a, O_b)^{f_m}$  can be treated as a new type of data. The aim of the first experiment was to investigate the influence of different lip features on the overall accuracy of the proposed PNN classifier. In this experiment, usefulness of the features  $f_i$ ,  $i = 1, \dots, 11$  defined at the Fig. 6 has been checked.

*Results:* The summary of the results of the first experiment has been shown in Tables 1–3 separately for each database. These tables show various combinations of the features and only best selections are only performed. The mean and standard deviation of the classification accuracy, FAR, FRR and ERR factors of each run over all calculations

are reported in all Tables in the paper. If the standard deviation is low it means better stability of the mean value.

In Tables the bold typeface numbers indicate the best accuracy, FAR, FRR and EER under the mean and standard deviation points of view. Tables show different lip features combinations taken into consideration during classification process. In these tables all possible lip–mouth features as well as selected combinations have been shown in the objects verification process. Experiment was conducted on the three, previously presented, facial databases. It is clearly shown that only one combination of lip features ( $f_1, f_2, f_3, f_4, f_5, f_6, f_8, f_9, f_{11}$ ) allows achieving the highest PNN accuracy from 86.95% to 87.26%. It can be observed for all tested datasets. In the next experiments the same features was utilized as well as type of features. It means that in the subsequent experiments selected collection of the lip–mouth features will be employed. Thus, from this moment the global matrix  $\mathbf{J} = [\mathbf{X}; \mathbf{Y}]$  will be always of size  $9 \times 20$  because the highest classifier's accuracy level was reached for the  $u = 9$  carefully selected lip features. Mentioned set of the selected lip features will be input data for other canonical classifiers. So, the accuracy of the proposed the PNN-based classifier can be reliably compared with many others classifiers.

### Experiment 2

*Set up:* The aim of the second experiment was to investigate the behavior of various classifiers in the lip recognition process, where the quality of classification has been measured by classifiers accuracy level. In the comparative study, canonical classifiers proven in literature and implemented in Matlab, R, or WEKA systems have been evaluated and all of them worked on the *Sim*-type data:

- Bayes NET (NBa) (Muramatsu et al., 2006),
- Hoeffding Tree (HoTr) (Kirkby, 2007),
- $k$ -Nearest Neighbors Classifier (kNN) (Kirkwood and Sterne, 2003; Shakhnarovich et al., 2005),
- J48 — C4.5 decision trees (J48) (Quinlan, 1993),
- Random Forests — forest of random trees (RaFo) (Breiman, 2001),
- Random Tree — tree that considers  $K$  randomly chosen attributes at each node (RaTr) (Breiman, 2001),
- RIDOR RIpplE-DOWn Rule learner (Rid) (Gaines and Compton, 1995).

The Bayes Nets (NBa) are a compact way to represent the joint distribution of a set of random variables. The nodes represent Random Variables. Random variables provide a mapping from values to probabilities. Other attractive BN characteristics include a capability for incorporating expert knowledge, and automated learning of relationship structures and conditional probabilities from databases which may include missing values.

The Hoeffding Tree (HoTr) is an incremental, anytime decision tree algorithm that is capable of learning from massive data streams, assuming that the distribution generating examples does not change over

**Table 2**

Lip features influence on the best accuracy of the PNN classifier (PUT database).

| Lip features |       |       |       |       |       |       |       |       |          |          | Accuracy               | FAR                    | FRR                    | ERR                    |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|------------------------|------------------------|------------------------|------------------------|
| $f_1$        | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | [%]                    | [%]                    | [%]                    | [%]                    |
| x            | x     | x     | x     | x     | x     |       | x     | x     |          | x        | <b>87.14</b><br>± 1.06 | <b>15.53</b><br>± 0.43 | <b>11.67</b><br>± 0.51 | <b>12.45</b><br>± 0.52 |
| x            | x     | x     | x     | x     | x     | x     |       |       | x        | x        | 84.54<br>± 0.97        | 18.08<br>± 0.43        | 12.31<br>± 0.37        | 13.78<br>± 0.47        |
| x            | x     | x     |       |       |       | x     |       |       | x        | x        | 83.27<br>± 1.31        | 18.74<br>± 0.63        | 14.36<br>± 0.42        | 15.41<br>± 0.46        |
| x            | x     | x     | x     | x     | x     | x     | x     | x     | x        | x        | 83.78<br>± 0.94        | 19.44<br>± 0.56        | 14.61<br>± 0.47        | 15.75<br>± 0.50        |
| x            | x     | x     |       |       |       | x     |       |       |          |          | 82.02<br>± 1.12        | 21.23<br>± 0.53        | 15.78<br>± 0.67        | 16.83<br>± 0.45        |
|              |       |       | x     | x     | x     |       |       |       |          |          | 81.89<br>± 1.23        | 23.33<br>± 0.75        | 17.65<br>± 0.62        | 18.84<br>± 0.64        |

**Table 3**

Lip features influence on the best accuracy of the PNN classifier (Our database).

| Lip features |       |       |       |       |       |       |       |       |          |          | Accuracy               | FAR                    | FRR                    | ERR                    |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|------------------------|------------------------|------------------------|------------------------|
| $f_1$        | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | [%]                    | [%]                    | [%]                    | [%]                    |
| x            | x     | x     | x     | x     | x     |       | x     | x     |          | x        | <b>87.26</b><br>± 0.91 | <b>15.99</b><br>± 0.57 | <b>11.63</b><br>± 0.47 | <b>12.57</b><br>± 0.41 |
| x            | x     | x     | x     | x     | x     | x     |       |       | x        | x        | 85.13<br>± 1.03        | 17.66<br>± 0.62        | 11.24<br>± 0.64        | 12.44<br>± 0.59        |
| x            | x     | x     |       |       |       | x     |       |       | x        | x        | 83.54<br>± 1.14        | 19.54<br>± 0.52        | 13.22<br>± 0.68        | 14.16<br>± 0.55        |
| x            | x     | x     | x     | x     | x     | x     | x     | x     | x        | x        | 83.97<br>± 0.78        | 18.54<br>± 0.82        | 13.91<br>± 0.48        | 15.18<br>± 0.60        |
| x            | x     | x     |       |       |       | x     |       |       |          |          | 82.76<br>± 1.45        | 20.32<br>± 0.65        | 14.90<br>± 0.52        | 16.10<br>± 0.56        |
|              |       |       | x     | x     | x     |       |       |       |          |          | 82.34<br>± 1.16        | 19.64<br>± 0.52        | 16.02<br>± 0.75        | 16.89<br>± 0.63        |

time. Hoeffding trees exploit the fact that a small sample can often be enough to choose an optimal splitting attribute. This idea is supported mathematically by the Hoeffding bound, which quantifies the number of observations (examples), needed to estimate some statistics within a prescribed precision (goodness of an attribute).

The *k*-Nearest Neighbors (kNN) algorithm is a non-parametric method used for classification. The input consists of the *k* closest training examples in the feature space. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors.

The *C4.5 algorithm* (J48) based J48 decision tree classifier uses a separate-and-conquer approach, builds a partial C4.5 decision tree in each iteration, and makes the “best” leaf into a rule, have been tested.

The *Random Forests* (RaFo) is supervised a machine learning algorithm for classification. In contrary to classic decision trees, Random Forests are built using randomly selected subsets of features for each node. In this method, the final class is assigned to an object based on the majority vote of all trees in the forest.

The *Random Tree* (RaTr) is a tree that is formed by a stochastic process. The random trees classifier takes the input feature vector, classifies it with every tree. All the trees are trained with the same parameters but on different training sets. These sets are generated from the original training set.

The *Ripple-Down Rule* (Rid) strategy consists of the following stages. First, it generates a default rule and then the exceptions for the default rule with the least (weighted) error rate. Then it generates the “best” exceptions for each exception and iterates until pure. Thus, it performs a tree-like expansion of exceptions. The exceptions are a set of rules that predict classes other than the default.

**Results:** The summary of the results of the second experiment, presented in Tables 4–6.

According to the Leave-One-Out strategy, presented in Tables 4–6 values represent the average values. These values together with

minimum and maximum data and outliers can conveniently also present by means of the box-plots. Outliers are plotted as individual points. It was collectively depicted in Fig. 9. Similarly as data from Tables 4–6, descriptive statistics in the box-plot form confirm that the proposed PNN classification approach give the better results compared to other classifiers. The better results have been reached in the FAR/FRR as well as in the ERR factor. It shows that robust face-based verification systems and algorithms, where machine learning techniques are employed, allow to improve quality of biometric systems.

### Experiment 3

The one of the most important feature of the analyzed images is their pixel resolution which is defined as capability to measure the smallest object with distinct boundaries. In pixel resolution, the term resolution refers to the total number of count of pixels in an digital image. In our case we used images gathered in specialized databases (for example Multi-PIE face database) and these images were taken with various pixel resolution. In proposed approach, landmarks of the lip contour consist of 20 points (see Figs. 2 and 4). Regardless of the image pixel resolution, mentioned landmarks have to be always visible and their number constant. We start from the Multi-PIE database where images are taken in size Width × Height = 2048 × 3072 pixels. These images are then rescaled according to the formula:  $\text{Height}_{\text{new}} = \text{round}((\text{Width}_{\text{new}} \cdot 3072) / 2048)$ , where  $\text{Width}_{\text{new}} = 100, 200, \dots, 2048$ . After each rescaling phase, landmarks were imposed on the image and number of visible landmarks have been counted. It was assumed that on the every rescaled image number of landmarks shall be constant and always equal to 20. During reduction of pixel resolution quality of the image decreases and landmarks are inappropriately imposed. From experiments conducted, starting from size 600 × 900 of images, position of the landmarks is stable. These observations are similar for images from other databases. It will also be demonstrated in numerical

**Table 4**Accuracy, FAR, FRR, EER and  $\pm$  standard deviation for various classification methods (Multi-PIE database).

| Classifier       | Accuracy<br>[%]                    | FAR<br>[%]                         | FRR<br>[%]                         | ERR<br>[%]                         |
|------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| <b>PNN + PSO</b> | <b>86.95 <math>\pm</math> 1.21</b> | <b>15.32 <math>\pm</math> 0.35</b> | <b>12.63 <math>\pm</math> 0.56</b> | <b>12.71 <math>\pm</math> 0.53</b> |
| Bayes NET        | 84.67 $\pm$ 0.91                   | 17.45 $\pm$ 0.54                   | 13.43 $\pm$ 0.64                   | 14.29 $\pm$ 0.64                   |
| Hoeffding Tree   | 84.11 $\pm$ 0.86                   | 17.64 $\pm$ 0.64                   | 13.45 $\pm$ 0.53                   | 14.16 $\pm$ 0.53                   |
| k-NN             | 83.58 $\pm$ 1.12                   | 19.33 $\pm$ 0.46                   | 13.22 $\pm$ 0.56                   | 15.04 $\pm$ 0.56                   |
| J48              | 81.79 $\pm$ 2.31                   | 21.23 $\pm$ 0.52                   | 17.61 $\pm$ 0.67                   | 18.34 $\pm$ 0.67                   |
| Random Forest    | 84.23 $\pm$ 1.42                   | 20.23 $\pm$ 0.52                   | 14.78 $\pm$ 0.56                   | 16.68 $\pm$ 0.56                   |
| Random Tree      | 83.07 $\pm$ 1.83                   | 19.21 $\pm$ 0.46                   | 15.39 $\pm$ 0.34                   | 17.03 $\pm$ 0.34                   |
| RIDOR            | 78.46 $\pm$ 1.67                   | 26.54 $\pm$ 0.45                   | 19.62 $\pm$ 0.52                   | 20.37 $\pm$ 0.52                   |

**Table 5**Accuracy, FAR, FRR, EER and  $\pm$  standard deviation for various classification methods (PUT database).

| Classifier       | Accuracy<br>[%]                    | FAR<br>[%]                         | FRR<br>[%]                         | ERR<br>[%]                         |
|------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| <b>PNN + PSO</b> | <b>87.14 <math>\pm</math> 1.06</b> | <b>15.53 <math>\pm</math> 0.43</b> | <b>11.67 <math>\pm</math> 0.51</b> | <b>12.45 <math>\pm</math> 0.52</b> |
| Bayes NET        | 83.72 $\pm$ 1.02                   | 18.65 $\pm$ 0.74                   | 13.88 $\pm$ 0.68                   | 14.96 $\pm$ 0.53                   |
| Hoeffding Tree   | 83.55 $\pm$ 1.28                   | 18.54 $\pm$ 0.57                   | 12.34 $\pm$ 0.86                   | 13.66 $\pm$ 0.71                   |
| k-NN             | 82.15 $\pm$ 0.77                   | 21.12 $\pm$ 0.76                   | 15.63 $\pm$ 0.45                   | 16.06 $\pm$ 0.58                   |
| J48              | 81.43 $\pm$ 1.61                   | 22.75 $\pm$ 0.48                   | 16.32 $\pm$ 0.76                   | 17.34 $\pm$ 0.84                   |
| Random Forest    | 82.62 $\pm$ 0.97                   | 21.65 $\pm$ 0.55                   | 15.24 $\pm$ 0.37                   | 16.43 $\pm$ 0.57                   |
| Random Tree      | 82.54 $\pm$ 1.52                   | 20.54 $\pm$ 0.45                   | 16.46 $\pm$ 0.61                   | 17.21 $\pm$ 0.75                   |
| RIDOR            | 77.36 $\pm$ 2.17                   | 25.43 $\pm$ 0.47                   | 20.36 $\pm$ 0.65                   | 21.46 $\pm$ 0.58                   |

**Table 6**Accuracy, FAR, FRR, EER and  $\pm$  standard deviation for various classification methods (Our database).

| Classifier       | Accuracy<br>[%]                    | FAR<br>[%]                         | FRR<br>[%]                         | ERR<br>[%]                         |
|------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| <b>PNN + PSO</b> | <b>87.26 <math>\pm</math> 0.91</b> | <b>15.99 <math>\pm</math> 0.57</b> | <b>11.63 <math>\pm</math> 0.47</b> | <b>12.57 <math>\pm</math> 0.41</b> |
| Bayes NET        | 85.02 $\pm$ 1.23                   | 17.76 $\pm$ 0.43                   | 12.36 $\pm$ 0.65                   | 13.71 $\pm$ 0.43                   |
| Hoeffding Tree   | 85.49 $\pm$ 0.98                   | 17.21 $\pm$ 0.53                   | 13.99 $\pm$ 0.76                   | 14.21 $\pm$ 0.67                   |
| k-NN             | 82.97 $\pm$ 1.36                   | 20.31 $\pm$ 0.47                   | 18.02 $\pm$ 0.65                   | 19.32 $\pm$ 0.65                   |
| J48              | 82.51 $\pm$ 2.15                   | 20.31 $\pm$ 0.54                   | 16.11 $\pm$ 0.56                   | 17.18 $\pm$ 0.77                   |
| Random Forest    | 83.37 $\pm$ 1.58                   | 19.56 $\pm$ 0.77                   | 14.98 $\pm$ 0.53                   | 15.85 $\pm$ 0.76                   |
| Random Tree      | 83.72 $\pm$ 1.30                   | 20.56 $\pm$ 0.34                   | 15.98 $\pm$ 0.43                   | 16.64 $\pm$ 0.48                   |
| RIDOR            | 77.14 $\pm$ 1.96                   | 26.32 $\pm$ 0.32                   | 19.99 $\pm$ 0.47                   | 21.93 $\pm$ 0.55                   |

**Table 7**Classification accuracy depending on pixel resolution Width  $\times$  Height of the image from the Multi-PIE database.

| Input image |        | Lip area |        | Accuracy         |
|-------------|--------|----------|--------|------------------|
| Width       | Height | Width    | Height |                  |
| 100         | 150    | 20       | 10     | 22.11 $\pm$ 6.43 |
| 200         | 300    | 40       | 20     | 57.29 $\pm$ 4.17 |
| 400         | 600    | 80       | 35     | 81.89 $\pm$ 2.86 |
| 600         | 900    | 120      | 50     | 86.71 $\pm$ 2.05 |
| 800         | 1200   | 160      | 70     | 86.74 $\pm$ 1.44 |
| 1600        | 2400   | 320      | 135    | 86.87 $\pm$ 1.56 |
| 2048        | 3072   | 410      | 170    | 86.95 $\pm$ 1.21 |

experiments, for various classifiers, what will be depicted in consecutive studies.

When pixel resolution is too small, then classification accuracy will be low. These dependences demonstrate [Table 7](#) for the PNN classifier boosted by the PSO procedure. The same table presents not only various resolution of images. For each rescaled image, the mouth ROI dimension is shown in the column “Lip area”. In the mentioned ROI all mouth features have been measured.

From [Table 7](#) follows that images with resolution lower than  $600 \times 900$  are unsuitable for biometric recognition of human lips. It follows from the fact, that for low pixel resolutions of image some landmarks will be improperly imposed — due to image blurring. In proposed approach biometric recognition process depends only on number of lip-based landmarks. In our case number of lip landmarks is constant and equal to 20. For this reason, recognition procedures are realized in constant time 21ms per image whereas classifier’s training

time is proximately equal to 4.51min, irrespective of image resolution and type of database. Learning procedure was performed by means of the *Sim* coefficients. The idea of using the *Sim* coefficients has been introduced in [Section 6](#).

In all experiments, the measurement time was calculated by a PC class computer equipped with Intel i5-2500 processor, 3.3 GHz, 16 GB RAM, and Windows 7 x64 operating system.

All face-based images comprised in datasets were taken in the JPEG format. In [Table 8](#) we show how the image compression ratio changes classification accuracy. Image compression process was automatically performed in the external computer program. We started from 60% compression ratio because for smaller ratios fluctuations of the accuracy values were insignificant. We can conclude that influence of image compression on the verification accuracy can be always observed but for images with resolution smaller than  $600 \times 900$  and compression ratio greater than 70% this influence is the greatest. Mentioned values were reached for Multi-PIE face-based images. In the case of the two remaining databases obtained results of accuracy measurements were similar.

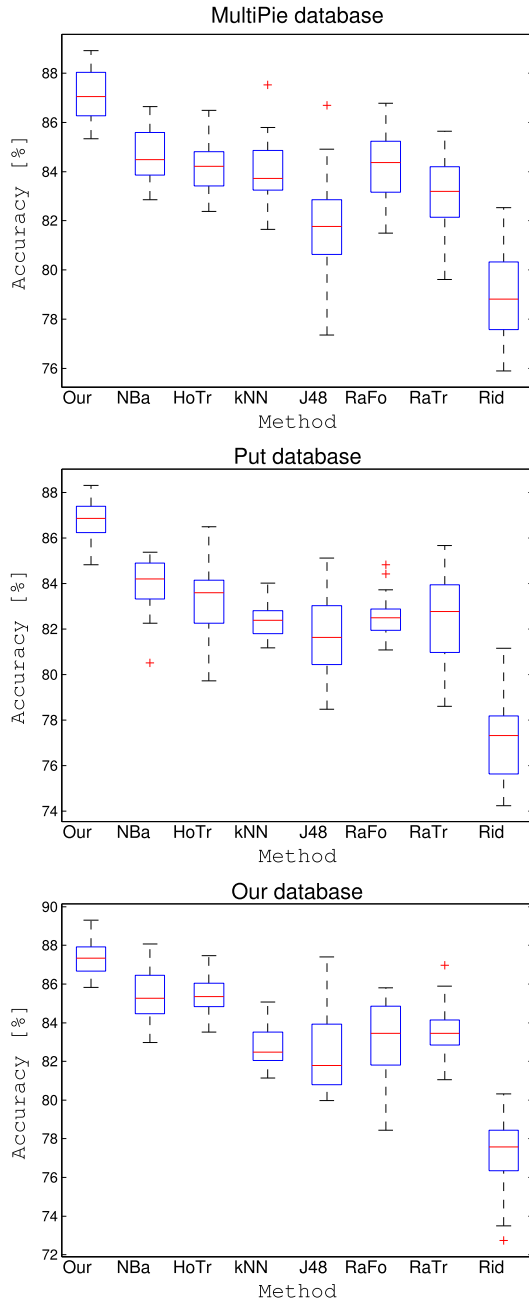
#### Experiment 4

In this experiment we evaluated above described various classification methods for used verification. [Fig. 10](#) shows the receiver operating characteristic (ROC) curves of various classifiers using five training images per class on the Multi-PIE, PUT and our databases, respectively. From the comparison of these charts follows that PNN approach significantly outperforms other approaches, especially on the challenging in the paper three datasets. For the baseline approach using our method, verification errors of other classifiers are higher in FAR as well in FRR achievements.

**Table 8**

Classification accuracy [%] depending on image compression rate for the Multi-PIE database.

| Input image |        | The JPEG-type compression ratio |              |              |              |              |
|-------------|--------|---------------------------------|--------------|--------------|--------------|--------------|
| Width       | Height | 60%                             | 70%          | 80%          | 90%          | 95%          |
| 100         | 150    | 8.50 ± 2.67                     | 7.24 ± 2.94  | 5.31 ± 1.17  | 3.46 ± 1.76  | 1.67 ± 0.93  |
| 200         | 300    | 28.53 ± 3.52                    | 16.94 ± 2.88 | 10.38 ± 2.23 | 6.16 ± 2.02  | 3.42 ± 1.62  |
| 400         | 600    | 64.43 ± 4.51                    | 46.26 ± 3.73 | 35.72 ± 4.45 | 27.30 ± 5.24 | 14.43 ± 5.89 |
| 600         | 900    | 86.71 ± 5.05                    | 81.69 ± 3.89 | 67.41 ± 5.37 | 53.11 ± 5.78 | 32.16 ± 6.29 |
| 800         | 1200   | 86.74 ± 5.44                    | 83.03 ± 6.62 | 74.74 ± 4.01 | 60.47 ± 6.49 | 51.18 ± 6.01 |
| 1600        | 2400   | 86.87 ± 4.56                    | 86.87 ± 6.56 | 85.61 ± 5.40 | 79.22 ± 5.91 | 71.66 ± 6.28 |
| 2048        | 3072   | 86.95 ± 5.21                    | 86.95 ± 5.21 | 86.95 ± 5.21 | 84.38 ± 4.43 | 75.21 ± 4.03 |

**Fig. 9.** Accuracy of various classification methods for different databases.

The analysis of the achieved results shows that proposed composed similarity measure *Sim*, generates the similar accuracy level for all classifiers. Regardless of the selected classifier and biometric face-based

datasets, the higher classification accuracy was always reached for the PNN classifier.

### Experiment 5

**Set up:** The aim of this experiment was to check how the raw data, calculated directly on the input image (see Fig. 6), will be assigned by the PNN classifier. Experiment was performed similarly as previous – separately for each face-based dataset. Measured previously raw features ( $f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}$ ) were designated as an input data of the PNN classifier (Fig. 8). All possible combinations of the raw data have been checked. In our case we have the set of  $n = 11$  elements, thus number of the  $k$ -combinations ( $k = 1, \dots, 11$ ) is equal to the summed binomial coefficient:

$$\sum_{k=1}^{11} \frac{11!}{k!(11-k)!} \quad (17)$$

In the tables only most promising results have been depicted. The rest of the thousands of results are worse and are not interesting from our point of view.

**Results:** The summary of the results of the fourth experiment presented in Tables 9–11.

Observation of the Tables 9–11 allows to conclude that classification based on raw data gives significantly worse recognition level compared to classification based on *Sim* coefficients. It applies to the all classifiers and face-based databases. It is also worth to note, that unlike the *Sim*-based approach, raw data best classification is performed on the basis of another combination of the lip features.

### Experiment 6

The results have been also statistically evaluated by the Wilcoxon test. It can be used as an alternative to the *t*-paired Student's test. In contrast to the Student test, the Wilcoxon approach allows to compute statistic for data which do not have the normal distribution. Obtained results are sufficient to conclude that there are statistically significant differences between accuracy in groups of classifiers. The results obtained by the Wilcoxon test for the level of significance  $\alpha = 0.05$  are summarized in Table 12. In this table,  $R^+$  denotes the sum of the ranks in favor of our method, whereas  $R^-$  sum of ranks of the other methods.

In the case of the confidence measure, the PNN approach has presented better results, the  $R^+$  value is greater than the  $R^-$  and the  $p$ -value obtained is lower than the level of significance considered. Therefore, the test confirms the hypothesis concluding that our classifier achieves the better accuracy level than the other classifiers working on the same datasets. Results of Table 12 show the good behavior of the proposed approach compared to other related classifiers because for all rows of table the  $p$ -value is less than 0.05 and  $R^+ > R^-$  – so the PNN-based classifier gives better and statistically significant differences from accuracy point of view in respect to all other classifiers mentioned in Table 12.

## 11. Comparison with state-of-the-art methods

Proposed in this paper biometric recognition strategy was widely confronted with recently published classification methods, where various lip-based biometric systems were announced. It is a well-known



**Table 9**

Lip raw features influence on the best accuracy of the PNN classifier (Multi-PIE database).

| Lip features |       |       |       |       |       |       |       |       |          |          | Accuracy               | FAR                    | FRR                    | ERR                    |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|------------------------|------------------------|------------------------|------------------------|
| $f_1$        | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | [%]                    | [%]                    | [%]                    | [%]                    |
| x            | x     | x     | x     |       | x     |       | x     | x     | x        | x        | <b>68.22</b><br>± 2.01 | <b>38.45</b><br>± 0.42 | <b>26.24</b><br>± 0.36 | <b>28.43</b><br>± 0.47 |
| x            | x     | x     | x     | x     | x     | x     |       |       | x        | x        | 64.31<br>± 1.16        | 42.34<br>± 0.63        | 28.52<br>± 0.44        | 29.41<br>± 0.39        |
| x            | x     | x     | x     |       |       | x     |       |       | x        | x        | 63.48<br>± 1.28        | 40.55<br>± 0.42        | 33.35<br>± 0.53        | 35.72<br>± 0.49        |
| x            | x     | x     | x     | x     | x     | x     | x     | x     | x        | x        | 60.59<br>± 1.47        | 43.36<br>± 0.36        | 36.63<br>± 0.42        | 38.63<br>± 0.44        |
|              |       |       | x     | x     | x     |       |       |       |          |          | 57.79<br>± 1.94        | 49.74<br>± 0.43        | 35.25<br>± 0.56        | 37.36<br>± 0.48        |
| x            | x     | x     | x     |       |       | x     |       |       |          |          | 52.51<br>± 2.25        | 56.25<br>± 0.36        | 42.46<br>± 0.53        | 44.59<br>± 0.55        |

**Table 10**

Lip raw features influence on the best accuracy of the PNN classifier (PUT database).

| Lip features |       |       |       |       |       |       |       |       |          |          | Accuracy               | FAR                    | FRR                    | ERR                    |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|------------------------|------------------------|------------------------|------------------------|
| $f_1$        | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | [%]                    | [%]                    | [%]                    | [%]                    |
| x            | x     | x     | x     |       | x     |       | x     | x     | x        | x        | <b>67.14</b><br>± 1.06 | <b>38.65</b><br>± 0.69 | <b>29.76</b><br>± 0.75 | <b>30.95</b><br>± 0.61 |
| x            | x     | x     | x     | x     | x     | x     |       |       | x        | x        | 65.84<br>± 1.72        | 40.58<br>± 0.75        | 31.75<br>± 0.65        | 33.17<br>± 0.53        |
| x            | x     | x     | x     |       |       | x     |       |       | x        | x        | 61.67<br>± 1.93        | 47.86<br>± 0.67        | 35.48<br>± 0.67        | 37.50<br>± 0.61        |
| x            | x     | x     | x     | x     | x     | x     | x     | x     | x        | x        | 60.28<br>± 2.41        | 46.65<br>± 0.66        | 34.99<br>± 0.86        | 36.87<br>± 0.71        |
|              |       |       | x     | x     | x     |       |       |       |          |          | 54.21<br>± 1.36        | 49.74<br>± 0.86        | 39.43<br>± 0.54        | 41.58<br>± 0.62        |
| x            | x     | x     | x     |       |       | x     |       |       |          |          | 49.92<br>± 1.56        | 60.34<br>± 0.54        | 41.46<br>± 0.66        | 42.25<br>± 0.63        |

**Table 11**

Lip raw features influence on the best accuracy of the PNN classifier (Our database).

| Lip features |       |       |       |       |       |       |       |       |          |          | Accuracy               | FAR                    | FRR                    | ERR                    |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|------------------------|------------------------|------------------------|------------------------|
| $f_1$        | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | [%]                    | [%]                    | [%]                    | [%]                    |
| x            | x     | x     | x     |       | x     |       | x     | x     | x        | x        | <b>68.62</b><br>± 0.89 | <b>37.56</b><br>± 0.65 | <b>27.65</b><br>± 0.67 | <b>28.48</b><br>± 0.59 |
| x            | x     | x     | x     | x     | x     | x     |       |       | x        | x        | 64.71<br>± 2.13        | 41.45<br>± 0.68        | 31.63<br>± 0.74        | 32.02<br>± 0.57        |
| x            | x     | x     | x     |       |       | x     |       |       | x        | x        | 62.25<br>± 1.46        | 43.89<br>± 0.75        | 32.47<br>± 0.69        | 33.88<br>± 0.58        |
| x            | x     | x     | x     | x     | x     | x     | x     | x     | x        | x        | 59.79<br>± 1.87        | 50.65<br>± 0.67        | 36.95<br>± 0.74        | 37.92<br>± 0.66        |
|              |       |       | x     | x     | x     |       |       |       |          |          | 57.43<br>± 1.52        | 49.69<br>± 0.85        | 38.74<br>± 0.56        | 40.12<br>± 0.48        |
| x            | x     | x     | x     |       |       | x     |       |       |          |          | 53.64<br>± 1.26        | 55.44<br>± 0.58        | 40.48<br>± 0.84        | 42.16<br>± 0.63        |

**Table 12**The Wilcoxon test to compare our classification method with other classifiers.  $p$ -value and ( $R^+/R^-$ ) ranks are performed.

| Databases | Our method (with PNN classifier) vs: |                     |                     |                     |                     |                     |                     |
|-----------|--------------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|           | Bayes NET                            | Hoeffding Tree      | k-NN                | J48                 | Random Forest       | Random Tree         | RIDOR               |
| Multi-Pie | 2.68e−06<br>(209/1)                  | 2.56e−07<br>(210/0) | 6.91e−07<br>(204/6) | 2.56e−07<br>(210/0) | 9.12e−07<br>(210/0) | 9.17e−08<br>(210/0) | 6.79e−08<br>(210/0) |
| Put       | 1.65e−07<br>(210/0)                  | 3.93e−07<br>(209/1) | 6.79e−08<br>(210/0) | 7.89e−08<br>(210/0) | 6.79e−08<br>(210/0) | 9.17e−08<br>(210/0) | 6.79e−08<br>(210/0) |
| Our       | 3.70e−05<br>(194/16)                 | 8.59e−06<br>(201/9) | 6.79e−08<br>(210/0) | 3.41e−07<br>(209/1) | 6.79e−08<br>(210/0) | 2.56e−07<br>(210/0) | 6.79e−08<br>(210/0) |

fact that recognition performance decreases when number of samples in a database of biometric features is increased. It can be noticed even for small number of database records (Bharadi and Kekre, 2010). Such important remark is often ignored; therefore, we postulate

that presented results should always be normalized and presentation principles should be respected. In presented approach all results were obtained for databases, which are fully available (The Put database, 2008; The MultiPie database, 2009; Our database, 2016). Hence, results

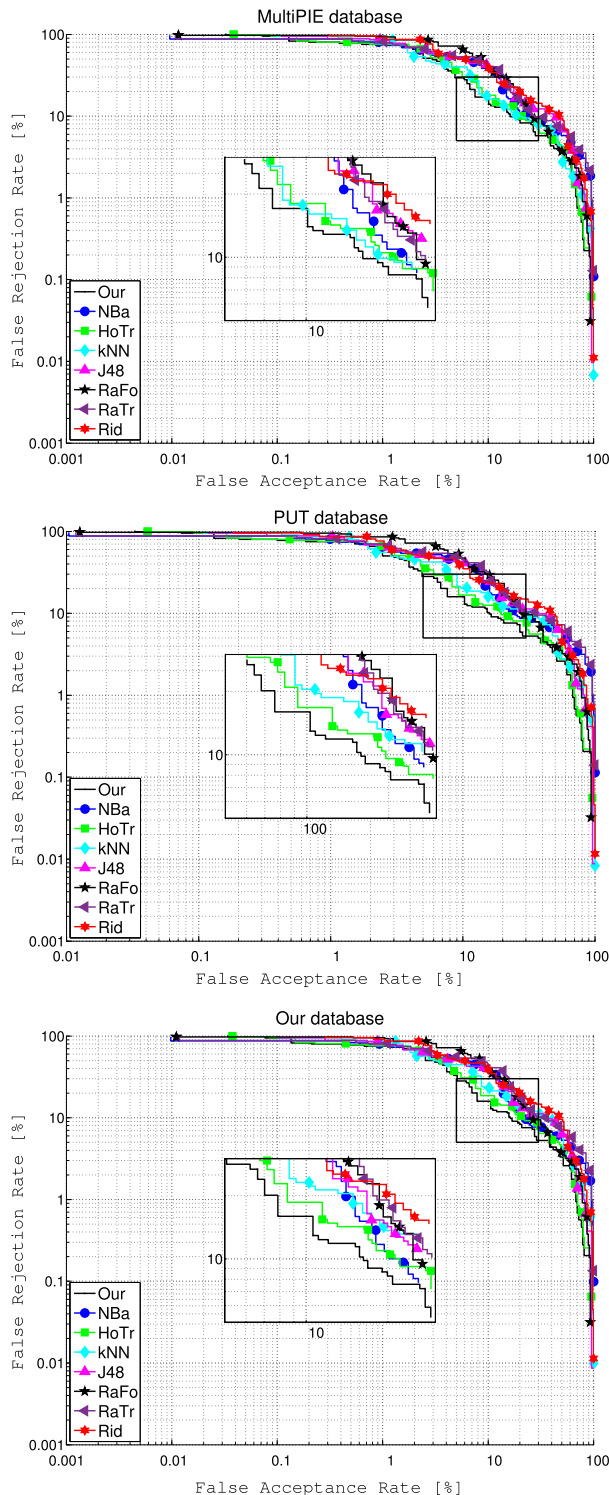


Fig. 10. ROC curves for the various databases using 5 training samples per class.

obtained and proposed algorithms can be always reliably compared with achievements of other authors. On the other hand, it can also be observed that results reported in many papers use different, often private, databases, coefficients (FAR, FRR, EER) and factors from statistics (type of errors), machine learning (accuracy, true positive rate, true negative rate) as well as from medical informatics (sensitivity, specificity). Unfortunately, only one of these parameters is very often treated as a single quality factor of described biometric systems. It

is another obstacle precluding comparison of achieved results. An influence of mentioned difficulties can be observed in the short review of obtained results in the work (Johnson et al., 2003).

The main state-of-the-art directions in the field of lip-based processing have been presented in the Introduction as well as in the Related work Sections. It concerns lip contour extraction for limited speech recognition (Aarabi and Mungamuru, 2004; Ibrahim and Mulvaney, 2015; Le and Savvides, 2016), facial expression analysis (Raheja et al., 2010; Moeini et al., 2017; Kamarol et al., 2017) or statistical lip features distribution for sex and population determination (Domiaty et al., 2010; Kapoor et al., 2015; Rastogi and Parida, 2012). From our point of view, mentioned trends are less interesting, although they use in some steps similar techniques as in our paper. Our approach is focused on biometric recognition of persons, therefore our strategy will be compared with relevant papers. The recent and most interesting investigations on this field depicts Table 13.

In Table 13 all published by the authors recognition factors have been shown. It should be noticed that our verification method was tested in so-called open set-mode by means of the three large and accessible biometric datasets. Verification mode is the most frequently used in practice. In the open-set there is no guarantee that a record of a given individual is contained in the existing database. So, biometric verification system can be deliberately cheated and it should be indicated. Analysis of Table 13 shows that the most important results were presented in the work (Travieso et al., 2014), where recognition accuracy is very high when biometric system works in the identification mode. Additionally, in the works (Travieso et al., 2014, 2011), the databases comprise from 9 to 11 samples per image, whereas we utilized datasets with 5 samples per image only. It means that classifier's learning process in our approach was more restricted and is more realistic from biometric point of view. Results of the lip-based identification rate were achieved for closed set of data (closed-set), so without counterfeits. Such mode is less restricted compared to our strategy where counterfeits have been acceptable. It should be stressed that our method is designated to work in verification mode, therefore direct comparison with identification system is not possible.

## 12. Final remarks

High volume of data can be stored nowadays, what is a trivial observation. These data have to be efficiently analyzed therefore the use of efficient computational techniques has become a task of the utmost importance. We have summarized the main research on the PNN biometric classifier. Key issues related to the problem under consideration are classifier diversity and methods of classifier combination. The diversity is believed to provide improved accuracy and classifier performance. The results of conducted experiments demonstrate that our classification model use a new kind of similarity scores what brings benefits in verification of biometric data. The investigations performed demonstrate the validity of our results. Public datasets retrieved from the biometric repository have been used to test the quality of the proposed approach. All the results show a remarkable performance of the proposed PNN classifier which outperforms other checked classifiers. Our studies make it clear that face-based lip features extraction might become an important tool for biometric identification.

The lip biometrics proposed in this paper could also be used as an additional method to enhance the effectiveness of other well-known biometrics through the implementation of multimodal systems, including fusion or hybrid systems. Such systems are the focus of intense research recently.

Conclusions concerning a new lip-based biometric verification method, can be summarized as follows:

- Our verification method is robust to the change of skin color and can be used for color as well as gray-scale images. Additionally, proposed method is largely resistant to changes of resolution and image compression.

**Table 13**

Comparison of our results with state-of-the-art achievements.

| Method                                    | FAR [%]      | FRR [%]      | EER [%]      | ACC [%]      | Database                                     | Comments                            |
|-------------------------------------------|--------------|--------------|--------------|--------------|----------------------------------------------|-------------------------------------|
| <b>Verification process</b>               |              |              |              |              |                                              |                                     |
| Our approach                              | 15.32 ± 0.35 | 12.63 ± 0.56 | 12.71 ± 0.53 | 86.95 ± 1.21 | Multi-PIE: 337 persons, 5 images per person. | Color image ✓                       |
|                                           | 15.53 ± 0.43 | 11.67 ± 0.51 | 12.45 ± 0.52 | 87.14 ± 1.06 | PUT: 100 persons, 5 images per person.       | Grayscale images ✓                  |
|                                           | 15.99 ± 0.57 | 11.63 ± 0.47 | 12.57 ± 0.41 | 87.26 ± 0.91 | Own: 50 persons, 5 images per person.        |                                     |
| Wang and Liew (2012)                      | –            | –            | 18.50        | –            | Own: 40 persons.                             | Color image ✓<br>Grayscale images ✗ |
| Travieso et al. (2011)                    | –            | –            | 12.75        | –            | PIE: 68 persons, 11 images per person.       | Color image ✓<br>Grayscale images ✗ |
| <b>Identification process<sup>a</sup></b> |              |              |              |              |                                              |                                     |
| Kim et al. (2004)                         | 1.10         | 3.60         | –            | 95.30        | Own: 24 persons.                             | Color image ✓<br>Grayscale images ✗ |
| Choraś (2010)                             | –            | –            | –            | 82.00        | Own: 38 persons, 3 images per person.        | Color image ✓<br>Grayscale images ✗ |
| Travieso et al. (2014)                    | –            | –            | –            | 99.89 ± 0.16 | Own: 50 persons, 10 images per person.       | Color image ✓                       |
|                                           | –            | –            | –            | 97.13 ± 0.56 | PIE: 68 persons, 11 images per person.       | Grayscale images ✗                  |
|                                           | –            | –            | –            | 98.10 ± 0.40 | RaFD: 60 persons, 9 images per person.       |                                     |

<sup>a</sup> Results were achieved for closed set of data (closed-set), so without counterfeits. It should be stressed that our method is designated to work in verification mode, therefore direct comparison with identification system is not possible.

- Proposed PNN + PSO strategy demonstrates the high recognition accuracy. This advantage is seen for Multi-Pie, PUT and our databases. In the all cases our method gives the high accuracy level. Accuracy all of classifiers has been checked in the Leave-One-Out validation procedure (Tables 4–6).
- Obtained results were achieved thanks to use the new composed lip feature. Composed feature was computed on the basis of an accessible similarity coefficients and a lip features extracted from the face-based image.
- Correctness and usefulness of proposed algorithm has been also statistically confirmed by the Wilcoxon test (Table 12).
- The main goal of building a classifier was its high accuracy. Preformed investigations confirm that this goal has been achieved by the PNN method busted by PSO strategy.
- The accuracy of the proposed method was carried out on the wide range of data originating from realistic biometric datasets. The results indicate that in real problems, proposed lip-based biometric verification method is valuable.

## References

- Aarabi, P., Mungamuru, B., 2004. The fusion of visual lip movements and mixed speech signals for robust speech separation. *Inf. Fusion* 5, 103–117.
- Ali, J.B., Saidi, L., Mouelhi, A., Chebel-Morello, B., Fnaiech, F., 2015. Linear feature selection and classification using {PNN} and {SFAM} neural networks for a nearly online diagnosis of bearing naturally progressing degradations. *Eng. Appl. Artif. Intell.* 42, 67–81.
- Bartlett, M.S., Littlewort, G.C., Frank, M.G., Lainscsek, C., Fasel, I.R., Movellan, J.R., 2006. Automatic recognition of facial actions in spontaneous expressions. *J. Multimedia* 1, 22–35.
- Başa, B., 2015. Implementation of Hog Edge Detection Algorithm Onfpga's. *Proc. Soc. Behav. Sci.* 1567–1575.
- Bedagkar-Gala, A., Shah, S.K., 2014. A survey of approaches and trends in person re-identification. *Image Vis. Comput.* 32, 270–286.
- Belhumeur, P.N., Jacobs, D.W., Kriegman, D.J., Kumar, N., 2013. Localizing parts of faces using a consensus of exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2930–2940.
- Berthold, M.R., Diamond, J., 1998. Constructive training of probabilistic neural networks. *Neurocomputing* 19, 167–183.
- Bharadi, V.A., Kekre, H.B., 2010. Off-line signature recognition systems. *Int. J. Comput. Appl.* (27), 48–56.
- Bolle, R., 2004. Guide To Biometrics. In: Springer professional computing, Springer, New York.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Cao, X., Wei, Y., Wen, F., Sun, J., 2014. Face alignment by explicit shape regression. *Int. J. Comput. Vis.* 107, 177–190.
- Cha, S.-H., 2007. Comprehensive survey on distance/similarity measures between probability density functions. *Int. J. Math. Models Methods Appl. Sci.* 1, 300–307.
- Chetverikov, D., 2003. A simple and efficient algorithm for detection of high curvature points in planar curves. In: Computer Analysis of Images and Patterns, 10th International Conference, CAIP 2003, Groningen, The Netherlands, August 25–27, 2003, Proceedings, pp. 746–753.
- Choraś, M., 2010. The lip as a biometric. *PAA Pattern Anal. Appl.* 13, 105–112.
- Czabanski, R., Jezewski, M., Jezewski, J., Wrobel, J., Horoba, K., 2012. Robust extraction of fuzzy rules with artificial neural network based on fuzzy inference system. *Int. J. Intell. Inf. Database Syst.* 6, 77–92.
- Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893.
- Dollár, P., Welinder, P., Perona, P., 2010. Cascaded pose regression, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1078–1085.
- Domiaty, M.A.E., Al-gaidi, S.A., Elayat, A.A., Safwat, M.D.E., Galal, S.A., 2010. Morphological patterns of lip prints in saudi arabia at almadinah almonawarah province. *Forensic Sci. Int.* 200, 179e1–179e9.
- Doroz, R., Porwik, P., Orczyk, T., 2016. Dynamic signature verification method based on association of features with similarity measures. *Neurocomputing* 171, 921–931.
- Doroz, R., Wrobel, K., Palys, M., 2015. Detecting the reference point in fingerprint images with the use of the high curvature points. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9012, pp. 82–91.
- Doroz, R., Wrobel, K., Watroba, M., 2014. A hybrid system of signature recognition using video and similarity measures. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNAI, vol. 8480, pp. 211–220.
- Gaines, B.R., Compton, P., 1995. Induction of ripple-down rules applied to modeling large databases. *J. Intell. Inf. Syst.* 5, 211–228.
- Garg, H., 2016. A hybrid PSO-GA algorithm for constrained optimization problems. *Appl. Math. Comput.* 292–305.
- Ge, H.W., Sun, L., Liang, Y.C., Qian, F., 2008. An effective PSO and AIS-based hybrid intelligent algorithm for job-shop scheduling. *IEEE Trans. Syst. Man Cybern. A* 38, 358–368.
- Hashemi, A., Meybodi, M., 2011. A note on the learning automata based algorithms for adaptive parameter selection in PSO. *Appl. Soft Comput.* 11, 689–705.
- Howell, D., Cox, S., Theobald, B., 2016. Visual units and confusion modelling for automatic lip-reading. *Image Vis. Comput.* 51, 1–12.
- Ibrahim, M.Z., Mulvaney, D.J., 2015. Geometrical-based lip-reading using template probabilistic multi-dimension dynamic time warping. *J. Vis. Commun. Image Represent.* 30, 219–233.
- Jin, Z., Lou, Z., Yang, J., Sun, Q., 2007. Face detection using template matching and skin-color information. *Neurocomputing* 70, 794–800.
- Johnson, A., Sun, J., Bobick, A., 2003. Using similarity scores from a small gallery to estimate recognition performance for larger galleries. In: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, AMFG2003, pp. 100–103.
- Kamarol, S.K.A., Jaward, M.H., Kälviäinen, H., Parkkinen, J., Parthiban, R., 2017. Joint facial expression recognition and intensity estimation based on weighted votes of image sequences. *Pattern Recognit. Lett.* 92, 25–32.
- Kapoor, N., Badiye, A., distribution, A., 2015. sex differences and stability of lip print patterns in an indian population. *Saudi J. Biol. Sci.*
- Kasinski, A., Florek, A., Schmidt, A., 2008. The put face database. *Image Process. Commun.* 13, 59–64.

- Kim, J.O., Lee, W., Hwang, J., Baik, K.S., Chung, C.H., 2004. Lip print recognition for security systems by multi-resolution architecture. *Future Gener. Comput. Syst.* 20, 295–301. Modeling and simulation in supercomputing and telecommunications.
- Kirkby, R., 2007. Improving Hoeffding Trees (Ph.D. thesis). Department of Computer Science, University of Waikato, University of Waikato.
- Kirkwood, B.R., Sterne, J.A., 2003. *Essential Medical Statistics*. Blackwell, Oxford.
- Krawczyk, B., Wozniak, M., 2016. Dynamic classifier selection for one-class classification. *Knowl.-Based Syst.* 107, 43–53.
- Kurzynski, M., Krysmann, M., Trajdos, P., Wolczowski, A., 2016. Multiclassifier system with hybrid learning applied to the control of bioprosthetic hand. *Comput. Biol. Med.* 286–297.
- Le, T.H.N., Savvides, M., 2016. A novel shape constrained feature-based active contour model for lips/mouth segmentation in the wild. *Pattern Recognit.* 54, 23–33.
- Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S., 2012. Interactive facial feature localization. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNCS, vol. 7574, pp. 679–692.
- Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G., 2015. A convolutional neural network cascade for face detection. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 5325–5334.
- Lienhart, R., Maydt, J., 2002. An Extended Set of Haar-Like Features for Rapid Object Detection, Volume 1. p. 1/900.
- Markus, N., Frljak, M., Pandzic, I.S., Ahlberg, J., Forchheimer, R., 2013. A method for object detection based on pixel intensity comparisons. *CoRR abs/1305.4537*.
- Marzec, M., Koprowski, R., Wróbel, Z., 2015. Methods of face localization in thermograms. *Biocybern. Biomed. Eng.* 35, 138–146.
- Moeini, A., Faez, K., Moeini, H., Safai, A.M., 2017. Facial expression recognition using dual dictionary learning. *J. Vis. Commun. Image Represent.* 45, 20–33.
- Muramatsu, D., Kondo, M., Sasaki, M., Tachibana, S., Matsumoto, T., 2006. A markov chain monte carlo algorithm for bayesian dynamic signature verification. *IEEE Trans. Inf. Forensics Secur.* 1, 22–34.
- Murphy, T.M., Broussard, R., Schultz, R., Rakvic, R., Ngo, H., 2016. Face detection with a Viola-Jones based hybrid network. *IET Biometrics* 6, 200–210.
- Naruniec, J., 2014. Discrete area filters in accurate detection of faces and facial features. *Image Vis. Comput.* 32, 979–993.
- Newman, J.L., Cox, S.J., 2009. Automatic visual-only language identification: A preliminary study. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 4345–4348.
- Ooi, S.Y., Teoh, A.B., Pang, Y.H., Hiew, B.Y., 2016. Image-based handwritten signature verification using hybrid methods of discrete radon transform, principal component analysis and probabilistic neural network. *Appl. Soft Comput.* 274–282.
- Our database, 2016. URL: <http://www.biometrics.us.edu.pl/uploads/database1.zip>.
- Porwik, P., Doroz, R., Orczyk, T., 2014. The k-NN classifier and self-adaptive Hotelling data reduction technique in handwritten signatures recognition. *PAA Pattern Anal. Appl.* 18, 983–1001.
- Porwik, P., Doroz, R., Orczyk, T., 2016. Signatures verification based on PNN classifier optimised by PSO algorithm. *Pattern Recognit.* 60, 998–1014.
- Porwik, P., Doroz, R., Wrobel, K., 2009. A new signature similarity measure. In: *2009 World Congress on Nature and Biologically Inspired Computing, NABIC 2009 - Proceedings*, pp. 1022–1027.
- Quinlan, J.R., 1993. C45: Programs for machine learning. In: *The Morgan Kaufmann Series in Machine Learning*, Morgan Kaufmann Publishers, San Mateo, Calif.
- Raheja, J.L., Shyam, R., Gupta, J., Kumar, U., Prasad, P.B., 2010. Facial gesture identification using lip contours. In: *ICMLC 2010 - The 2nd International Conference on Machine Learning and Computing*, pp. 3–7.
- Ranjan, R., Patel, V.M., Chellappa, R., 2016. HyperFace: A Deep Multi-task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition. *CoRR abs/1603.01249*.
- Rastogi, P., Parida, A., 2012. Lip prints - An aid in identification. *Aust. J. Forensic Sci.* 44, 109–116.
- Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., Pantic, M., 2013. A semi-automatic methodology for facial landmark annotation. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 896–903.
- Shakhnarovich, G., Darrell, T., Indyk, P., 2005. *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. In: *Neural Information Processing Series*, MIT Press, Cambridge, Mass.
- Sharma, P., Saxena, S., Rathod, V., 2009. Comparative reliability of cheiloscopy and palatoscopy in human identification. *Indian J. Dent. Res.* 20, 453–457.
- Sim, T., Baker, S., Ssat, M., 2003. The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 1615–1618.
- Specht, D.F., 1990. Probabilistic neural networks. *Neural Netw.* 3, 109–118.
- Surinta, O., Karaaba, M.F., Schomaker, L.R., Wiering, M.A., 2015. Recognition of handwritten characters using local gradient feature descriptors. *Eng. Appl. Artif. Intell.* 45, 405–414.
- Tadeusiewicz, R., 2015. Neural networks as a tool for modeling of biological systems. *Bio-Algorithms Med-Systems* 11.
- Tadeusiewicz, R., Horzyk, A., 2014. Man-machine interaction improvement by means of automatic human personality identification. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8838, pp. 278–289.
- The MultiPie database, 2009. URL: <http://www.multipie.org>.
- The Put database, 2008. URL: <http://biometrics.cie.put.poznan.pl>.
- Travieso, C.M., Zhang, J., Miller, P., Alonso, J.B., 2014. Using a discrete hidden markov model ernel for lip-based biometric identification. *Image Vis. Comput.* 32, 1080–1089.
- Travieso, C.M., Zhang, J., Miller, P., Alonso, J.B., Ferrer, M.A., 2011. Bimodal biometric verification based on face and lips. *Neurocomputing* 74, 2407–2410.
- Utsuno, H., Kanoh, T., Tadokoro, O., Inoue, K., 2005. Preliminary study of post mortem identification using lip prints. *Forensic Sci. Int.* 149, 129–132.
- Viola, P., Jones, M.J., 2004. Robust real-time face detection. *Int. J. Comput. Vis.* 57, 137–154.
- Wang, S.-L., Liew, A.W.-C., 2012. Physiological and behavioral lip biometrics: A comprehensive study of their discriminative power. *Pattern Recognit.* 45, 3328–3335.