Biometrics,
Lip features,
Data classification

Krzysztof WROBEL[1], Rafal DOROZ[1], Piotr PORWIK[1],
Jacek NARUNIEC[2], Marek KOWALSKI[2]

PERSONAL IDENTITY VERIFICATION METHOD BASED ON LIPS PHOTOGRAPHS

The paper presents a personal identification method based on lips photographs. This method uses a new approach to the extraction and classification of characteristic features of the mouth from photographs. It eliminates the drawbacks that occur during the acquisition of lip print images with the use of the forensic method that requires special tools. Geometrical dimensions of the entire mouth as well as of the upper and lower lips were adopted as the features, on the basis of which the verification is performed. An ensemble classifier was used for the classification of the features obtained. The effectiveness of the classifier has been verified experimentally.

1. Introduction

Biometrics is a science that uses methods of mathematical statistics for examining the variability of physical and behavioural features of living organisms. The term "biometrics" is derived from the Greek words *bio* (life) and *metren* (to measure). Important applications of biometrics include identification, i.e. determination of personal identity, and verification, the aim of which is to confirm whether a given person is the one that he or she claims to be [1].

Nowadays, the personal identification and verification based on biometric features is widely used, for example at airports, in work establishments and banks, as well as in forensics for the identification or verification of suspected offenders. The number of available biometric methods is so large that they allow identifying a person with nearly complete certainty [6].

Identification or verification can be carried out based on a number of individual biometric features. Examples of such features include fingerprints, iris and retina, the face or palm shape, and lip prints [1], [3], [6]. This paper presents studies, in which the personal identity is verified through an analysis of the mouth shape. The mouth shape and the lip print left on various objects (e.g. on a glass) are used in forensics for identification or verification of the identity of suspected offenders. Recently published studies have shown that mouth features can also be used for biometric identification and verification [8], [11].

[1]Institute of Computer Science, University of Silesia, ul. Bedzinska 39, 41-200 Sosnowiec, Poland
[2]Faculty of Electronics and Information Technology, Warsaw University of Technology, ul. Nowowiejska 15, 00-665 Warsaw, Poland

2. Methods of acquiring lip prints and mouth shape

In forensic applications, a lip print from a suspected offender is generally acquired with the use of special tools such as a cheiloscopic roller, ferromagnetic powder, a magnetic applicator, as well as cosmetic cream used to moisten the mouth during the acquisition. The acquisition performed with the use of the aforementioned tools has been described in detail in [12]. Many external factors, such as the pressure of the roller on the lips, the speed of rolling or the selection of the place of starting the collection of the lip print have the impact on the result of the acquisition with the use of a cheiloscopic roller. All these factors cause that lip prints obtained from the same person may differ. The difference may concern all the prints or the location of their certain fragments. These are undoubtedly disadvantages of this method. An example of differences between two lip prints obtained from the same person is shown in Figure 1.

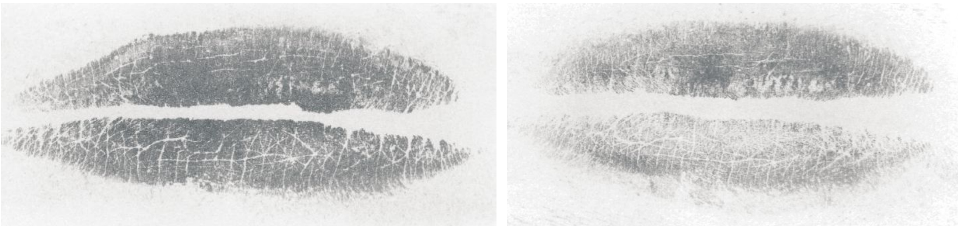

Fig. 1. Two lip prints obtained from the same person.

The aforementioned method of acquisition is used only in forensics. The time consumption and drawbacks mentioned above cause that it is difficult to use this acquisition method for the biometric personal identification or verification, e.g. at airports or in banks.

Therefore, a different approach was used in this study. It consisted in the use of a digital camera for the mouth image acquisition. Thanks to this approach, our method does not require a direct contact of the person carrying out the acquisition with the mouth of the person examined.

Based on the photograph obtained, the extraction and analysis of many features of the mouth is performed. Such features may include for example the shape of the mouth, its size, as well as the height and width of the lips. The studies have shown that the features used for the analysis allow obtaining a high efficiency of the personal identity verification.

3. The mouth detection based on face detection algorithm

In presented work, to extract the shape of the lips, a face detection algorithm was used. Our face detection algorithm is based on the sliding window approach. We use two types of features along with two types of classifiers:

- Extended set of Haar features as in [7], along with simple probabilistic classification,
- HOG features [4] classified by a logistic regression method.

A set of features is then boosted into a strong classifier using the GentleBoost algorithm. Every possible window scale and position across the image is classified as face or background. Since we are interested only in faces with sufficient pixel size, the smallest possible face rectangle size (in pixels) is 80 pixels. In the last step close rectangles are merged into a final decision.

Our face alignment method is based on a recently popular cascaded regression scheme [2], [10], [13], where starting from an initial face shape estimate S^0 the shape is refined in a fixed number of iterations. At each iteration t an increment that will refine the previous pose estimate S^{t-1} is found by regressing a set of features:

$$S^t = S^{t-1} + R^t \Phi^t(I, S^{t-1}), \tag{1}$$

where R^t is the regression matrix at iteration t and $\Phi^t(I, S^{t-1})$ is a vector of features extracted at landmarks S^{t-1} from image I.

The key difference between different methods based on the cascaded regression scheme is the choice of features $\Phi^t(I, S^{t-1})$. In our method, similarly to [10], we use sparse binary features generated by regression forests. Each of the 68 landmarks that constitute the face shape has a corresponding regression forest that was trained to minimize the error of that landmark in the training set. Once each tree in the forest has been traversed, the binary feature vector is created by outputting 1 for each tree leaf that was reached and 0 for each leaf that was not reached.

Each of the forests in our method consists of 3 trees, one for minimizing the error in x axis, one for y axis and one that minimizes the errors in both axes simultaneously. The trees are constructed based on HOG [4] features extracted at each of the landmarks. As opposed to [10] where the trees only use locally sampled features, each of our trees uses features extracted at many landmarks. This allows the trees to exploit information gathered on different parts of the face.

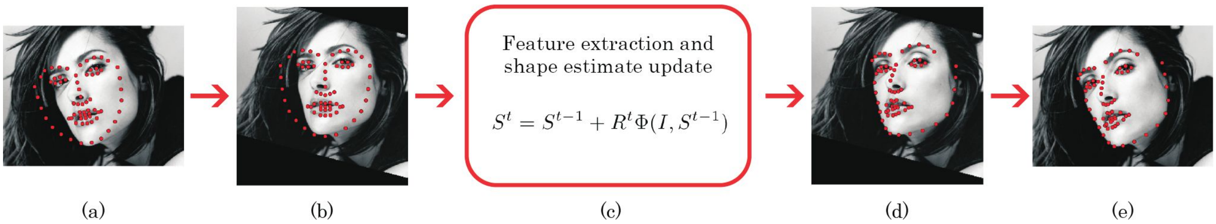

Fig. 2. The process of face alignment. (a) The mean shape is aligned with the landmarks from the previous video frame. (b) The shape and image are transformed to a canonical scale and orientation. (c) Cascaded regression is performed for a predefined number of iterations. (d) The resulting shape. (e) An inverse transformation is applied to the shape to obtain its location in the original image.

Figure 2 summarizes the whole process of face alignment. Our method was trained using parts of the 300-W [9] and Multi-PIE [5] datasets.

As a result of face recognition algorithm we obtained points which describe shape of various elements of face e.g. eyes, nose, mouth. For the further analysis we are using points that refer only to shape of the mouth.

4. Lip feature extraction

As the result of the mouth detection on the image there was obtained the set $M = \{p_1, ..., p_n\}$, where $p_i = (x_i, y_i)$, which includes the coordinates of n points located on the contour of the mouth. An example of mouth with points marked on it is shown in Fig. 3.

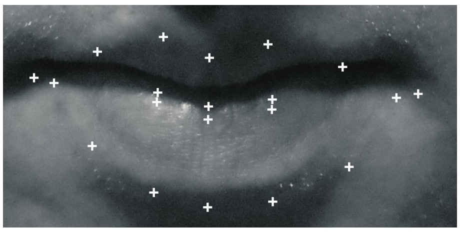

Fig. 3. Points located on the contour of the mouth.

When taking photos, the persons being photographed may be located at different distances from the camera and their heads may be bent. This causes that the size, position and inclination of the mouth may be different in each image, which adversely affects the efficiency of the classification. In order to make the position, inclination and size of the mouth uniform, points from the set M were subjected to processes of rotation, translation and scaling. The points are rotated by the $Alpha$ angle, which is determined between the section A and the X axis. The section A is the section between the points indicating the position of corners of the mouth (Fig. 4).

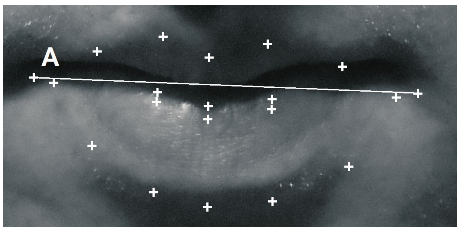

Fig. 4. Section A determined between the points indicating the position of corners of the mouth.

After the image has been rotated in such a way that the section A coincides with the X axis, the points of the mouth were subjected to the process of translation to the origin of the coordinate system and scaling to the same interval.

On the basis of coordinates of the points, the features describing the analysed mouth are determined. Distances between selected points were adopted as features of the mouth. The points were selected in such a way so that it was possible to calculate the width of the entire mouth, the widths of the lower and upper lips and the heights of both lips. For this purpose a letter (sequentially from "a" to "s") is assigned to each point according to the diagram shown in Fig. 5. Then the distances between the points determined earlier are calculated based on the following formula:

$$d(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{2}$$

where $d(p_i, p_j)$ is the distance between points $p_i = (x_i, y_i)$ and $p_j = (x_j, y_j)$.

Taking into account the above assumptions, sections between the following points were determined: $d(p_a, p_b)$, $d(p_c, p_d)$, $d(p_e, p_f)$, $d(p_g, p_h)$, $d(p_i, p_j)$, $d(p_k, p_l)$, $d(p_m, p_n)$, $d(p_o, p_p)$, $d(p_r, p_s)$.

A mouth image with points marked on is shown in Fig. 5. In addition, the figure shows the sections connecting the points, the distances between which were calculated.

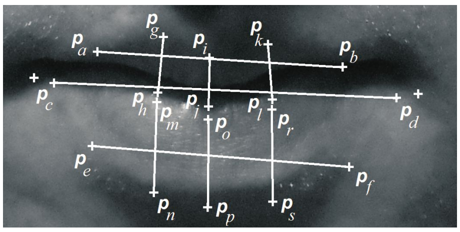

Fig. 5. Distances between points subject to the analysis.

All the distances between paired points are related to the features of a given k–th person. Finally, these data form a vector $\mathbf{v}^k$:

$$\begin{aligned} \mathbf{v}^k = [&d(p_a, p_b), d(p_c, p_d), d(p_e, p_f), d(p_g, p_h), d(p_i, p_j), ... \\ &..., d(p_k, p_l), d(p_m, p_n), d(p_o, p_p), d(p_r, p_s)]. \end{aligned} \tag{3}$$

5. Classification method

The proposed human's mouth recognition method is based on ensemble of classifier. The ensemble consists of the three sub-classifiers $C_i, i = 1, ..., 3$, which work parallelly and all sub-classifiers are the same type. The general structure of the classifier shown in Fig. 6.

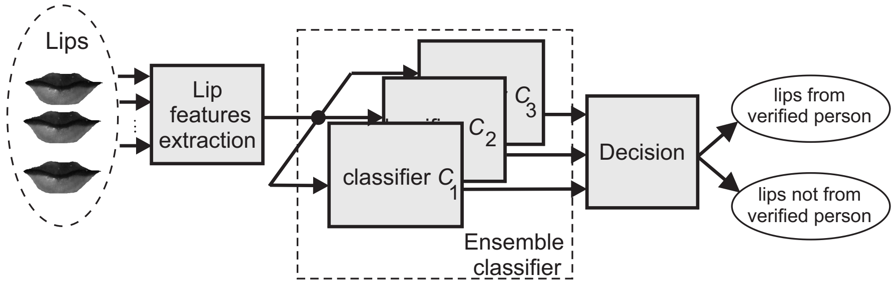

Fig. 6. The general structure of proposed classifier.

The proposed classifier works in the two modes: the learning and testing mode. Training of the classifiers is preformed separately for each user (say O) on the accessible database D. The database includes m users $D = \{O^k\}_{k=1}^m$. Each user O^k is represented by the features vector $\mathbf{V}^k = [\mathbf{v}_1^k, \mathbf{v}_2^k, ..., \mathbf{v}_r^k]$, which consists of r vectors. The indexes r and k of the vector $\mathbf{v}_r^k$ denote r–th mouth image of the given person k–th.

Sub-classifiers, $C_i, i = 1, ..., 3$, are learned separately by means of the its own learning set DS_i^k. The learning set DS_i^k is created on the basis of the vector $\mathbf{V}^k$ for k–th user and randomly selected vector $\mathbf{V}^l$ of other person, where $l \in \{1, ..., k-1, k+1, ..., m\}$. For the each set DS_i^k, the vector $\mathbf{V}^l$ is formed. The three learning sets for the k–th persons are defined as follows:

$$\begin{array}{l} DS_1^k = \mathbf{V}^k \cup \mathbf{V}^a, DS_2^k = \mathbf{V}^k \cup \mathbf{V}^b, DS_3^k = \mathbf{V}^k \cup \mathbf{V}^c, \\ a, b, c, k \in \{1, ..., m\} \text{ and } a \neq b \neq c \neq k. \end{array} \quad (4)$$

After training phase, classifier can be switched into verify mode. During verification the i–th classifier C_i calculates probability P_i, that a given person is legitimated or not. Depending on probability value classifier creates decision, whether lip belongs to verified person or not. The rule, which implements these assumptions is formulated as follow:

$$\begin{array}{lll} if \quad P_i > 0.5 & then & a_i = 1, \quad i = 1, ..., 3, \\ otherwise & & a_i = 0, \end{array} \quad (5)$$

where a_i is an auxiliary variable of a given, single i–th classifier. In the voting process we verify whether lip belongs to verified person. It can be done on the basis of the following formula:

$$\begin{array}{ll} if \quad \sum_{i=1}^{3} a_i \geq 2 & \text{lip belongs to verified person,} \\ otherwise & \text{lip does not belong to verified person.} \end{array} \quad (6)$$

6. Experiments and results

The studies were conducted with the use of a database containing 96 lips photographs obtained from 24 people. Each person was photographed 4 times. In each of the four photographs the face of a given person was shown in different scales, positions in the photograph and against different backgrounds. The photographs were taken from the front, so that the entire mouth was visible in them. People in the photographs had glasses, facial hair or headgear. Images were recorded only in high resolution (4320×3240 px). The design of the database and the selection of photographs allowed verifying the method reliably.

Then, using the techniques described in previous sections of this paper, features of the mouth were extracted from each photograph. The classification was performed on the basis of the data obtained. The effectiveness of ten popular classifiers was tested as a part of the study. The results obtained for the classifiers tested are presented in Table 1.

Table 1. Obtained results.

Classifiers	**Effectiveness of verification [%]**
Ridor	79.2
J48	86.5
Naive Bayes	84.4
Hoeffding Tree	84.4
KStar	91.1
RBFClassifiers	86.5
NBTree	85.4
PART	86.5
Random Forest	85.9
Furia	77.6

These results show that the highest effectiveness of verification was obtained with the use of the KStar classifier. High values of the effectiveness were obtained also with the use of other classifiers. The results for these classifiers are within the range from 80 to 90% and only the Ridor and Furia classifiers showed a slightly lower effectiveness that did not exceed 80%.

7. Conclusions

This study presents a personal identification method based on lips photographs, which consists in an analysis of the mouth images obtained from the photographs. This approach differs from the approach known from forensics, in which mouth image acquisition consists in obtaining a mouth print on a piece of paper. Thanks to this approach, the image acquisition is faster and easier and thus can be used in biometric solutions.

The studies have shown that it is possible to effectively identify people using the features derived from lips photographs. Among the classifiers tested, the highest effectiveness of classification was obtained with the use of the KStar classifier. The result of 91.1% can be regarded as good considering the effectiveness of forensic acquisition methods. The achieved result encourages for further work.

Future studies will be conducted on a larger number of images in the database. It is also planned to develop new features of the mouth and modify the committee of classifiers in order to increase the effectiveness of the classification.

Bibliography

[1] Bolle R., Connell J., Pankanti S., Ratha N., Senior A. Guide to biometrics. 2004. Springer.
[2] Cao X., Wei Y., Wen F., Sun J. Face alignment by explicit shape regression. International Journal of Computer Vision, 2014, Vol. 107. pp. 177–90.
[3] Choras M. The lip as a biometric. Pattern Analysis and Applications, 2010, Vol. 13. pp. 105–112.
[4] Dalal N., Triggs B. Histograms of oriented gradients for human detection. IEEE Int. Conference of Computer Vision and Pattern Recognition (CVPR), 2005. pp. 886–893.
[5] Gross R., Matthews I., Cohn J., Kanade T., Baker S. Multi–pie. 8th IEEE international conference on Automatic Face & Gesture Recognition, 2008. pp. 1–8.
[6] Jain A. K., Flynn P., Ross A. A. Handbook of biometrics. 2007. Springer, New York.

[7] LIENHART R., MAYDT J. An extended set of haar–like features for rapid object detection. Proceedings of International Conference on Image Processing, 2002, Vol. 1. pp. 900–903.
[8] NEWTON M. The encyclopedia of crime scene investigation. 2008. Checkmark Books.
[9] SAGONAS C., TZIMIROPOULOS G., ZAFEIRIOU S., PANTIC M. A semi–automatic methodology for facial landmark annotation. IEEE Int. Conference of Computer Vision and Pattern Recognition (Workshops) (CVPRW), 2013. pp. 896–903.
[10] SHAOQING R., XUDONG C., YICHEN W., JIAN S. Face alignment at 3000 fps via regressing local binary features. IEEE Int. Conference of Computer Vision and Pattern Recognition (CVPR), 2014. pp. 1685–92.
[11] SIEGEL J., SAUKKO P., KNUPFER G. Encyclopedia of forensic science. 2013. Academic Press.
[12] WROBEL K., DOROZ R., PALYS M. A method of lip print recognition based on sections comparison. IEEE Int. Conference on Biometrics and Kansei Engineering (ICBAKE), 2013. pp. 47–52.
[13] XUEHAN X., DE LA TORRE F. Supervised descent method and its applications to face alignment. IEEE Int. Conference of Computer Vision and Pattern Recognition (CVPR), 2013. pp. 532–9.