

Comparative Analysis of Classification Methods for Human Identification by Gait

Lubov Shiripova¹ and Evgeny Myasnikov¹

Samara University, Moskovskoe Shosse 34, Samara, Russia, 443086
mevg@geosamara.ru
<http://ssau.ru>

Abstract. The paper considers the problem of a person identification by gait using a video sequence. It provides a comparative analysis of two methods. The first method is proposed in this paper and consists in the detection of a moving person on a video sequence with the subsequent size normalization, generation of subsequences, and dimensionality reduction using the principal component analysis technique. The person classification is carried out using the support vector machine. In the second method, the known approach based on the hidden Markov model is used. CASIA GAIT dataset is used in this paper to compare the above methods. It was shown that the proposed method outperforms the HMM-based technique and provides high classification accuracy on the considered dataset.

Keywords: gait analysis, person identification, dimensionality reduction, SVM, HMM

1 Introduction

The identification of a person by its biometric parameters is popular and widely used all over the world at present. Such specific features as a face image, voice timbre, fingerprint, iris pattern and even gait are used for identification of a person. Although the use of fingerprints or iris patterns makes it possible to identify a person with little or no error, contactless and remote identification methods are of considerable interest. In this regard, especially important is the problem of recognizing a person using his gait.

Considering a gait as a set of poses and movements, we can distinguish two most common ways of recording (capturing) such information: video [1] (for example, in the optical range) and recording using sensors located on a human body [2]. In addition, there are works, in which gait analysis is performed based on the readings of accelerometers built into a smartphone [3].

Considering that a gait allows to identify a person even in cases where it is not possible to produce it in other ways (the person is at a distance, it is impossible to obtain the high-quality image of the face, etc.), the use of video, for example, from CCTV cameras is of particular interest.

To date, various methods have been used to solve the problem of identification of a person on video by gait.

The approach used in [4] consists in the subsequent segmentation of video frames using the background subtraction algorithm based on a mixture of Gaussian distributions (GMM), dimensionality reduction using the principal component analysis technique (PCA), and classification based on the Fisher linear discriminant (FLDA). Another feature of the work is the combination of movement features with the features of person's trace (footprints).

The first step of the approach proposed in the paper [5] is an improved background subtraction procedure. In this paper, selected motions are described by the descriptors based on the form statistical analysis (Procrustes analysis) technique. The procedure of the supervised classification is constructed using the appropriate measure (Procrustes distance measure).

In the paper [1], the analysis of the efficiency of linear (PCA) and non-linear (ISOMAP, LLE) dimensionality reduction techniques is performed. A Hidden Markov Model (HMM) is used to classify persons using features generated with the above techniques.

In the paper [6], Support Vector Machine (SVM) is used to solve the problem of classification of a person by gait. In particular, the dependence of the classification accuracy on the choice of the type of the SVM kernel is studied in the paper.

In general, it is worth noting that the problem of recognizing a person by gait attracts the attention of an increasing number of researchers. Considerable attention is paid to both feature description techniques and to the choice of effective classification methods.

In this paper, to solve the problem of identification of a person by gait, we follow the general approach used in the above studies [1, 4, 6]. This method is based on the detection and segmentation of a moving person on a video sequence, normalizing the size of frames, generating subsequences and reducing the dimensionality of a subsequences using the principal component analysis technique. The support vector machine is used as a classifier.

The proposed approach differs from the above mentioned papers.

The proposed technique is compared to the HMM-based technique. The paper shows that the proposed method provides higher classification accuracy with a relatively small number of classes.

The paper has the following structure. Section 2 is devoted to the description of the method used in the paper. Section 3 describes the results of experiments. The paper ends up with the conclusion. The list of references is given at the end of the paper.

2 Methods

2.1 Identification of a moving person using principal components analysis and support vector machine

The method developed in this paper consists of the following steps (see Figure 1):

- detection and segmentation of a moving person in the video sequence,
- normalization of the size of detected fragments,
- generation of subsequences of video frames and dimensionality reduction of generated subsequences,
- classification of video sequences.

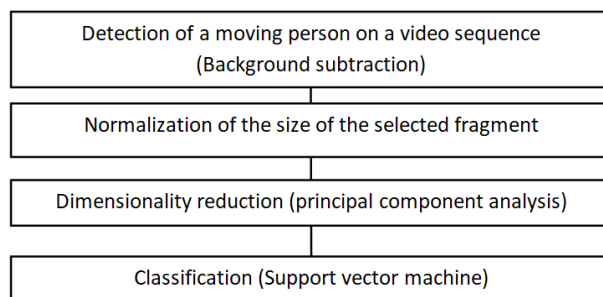


Fig. 1. Main steps of the proposed method

Detection and segmentation of a moving person on a video sequence.

At the first stage of the developed method, a moving person is detected on a video sequence. Background subtraction methods are most frequently used to detect moving objects if the video sequence was obtained using video surveillance camera. The main idea of the methods of this class is to use a certain background model and to decide whether a particular pixel belongs to the background or a moving object. This decision is based on the correspondence of the pixel to the background model. The background model is gradually refined. Although the time-averaged image can be used as a background model in simplest applications, better results in this problem are given by more complex models, for example [7–9].

In this paper, we use the background subtraction algorithm based on a mixture of Gaussian distributions (Gaussian mixture model, GMM) [8]. According to this method, each background pixel is modeled by a weighted sum (mixture) of Gaussians. The weights of Gaussians are determined by time periods, during which the corresponding color is present on the video sequence.

To choose this particular background subtraction technique, we took into account both our preliminary experiments and the experience of using this method by other researchers [4, 5].

Upon completion of the first stage of the method, a sequence of masks corresponding to individual frames of the video sequence is formed. Each mask reflects the result of the segmentation of the corresponding video frame into the foreground region (moving person) and the background.

Normalization of the size of detected fragments. At the second stage of the method, obtained masks are processed as follows. First, the center of mass for each foreground region is calculated. Then the linear dimensions (size) of the region are determined, and a framing (truncation of the mask image) is performed. After that, the cropped image is resized to the specified size. The described scheme is shown in Figure 2.



Fig. 2. Normalization of the size of the selected video sequence fragment

Taking into account the time coordinate, the dimensionality of the sequence of masks, which describes the movement of a person, remains high even after size normalization. In this regard, the third stage reduces the dimensionality of data describing the movement of a person.

Dimensionality reduction using the principal component analysis technique. To reduce the dimensionality of multidimensional data, both linear and nonlinear methods are used. The most commonly used are linear methods, such as the principal component analysis (PCA) [11] and independent component analysis (ICA). Nonlinear dimensionality reduction methods [12] (for example, nonlinear mapping, ISOMAP, LLE) are used less often due to the high computational complexity of such methods. It should be noted that recent attempts have been made to accelerate such methods [13, 14].

In this paper, we use the principal component analysis technique, as the most often used in such cases (see, for example, [1, 4]). This method searches for a linear projection into the subspace of a smaller dimension that maximizes the

variance of data. The PCA method is often considered as a linear dimensionality reduction technique, minimizing the loss of information.

In this paper, before reducing the dimensionality of data, we form a set of subsequences of a fixed length for each sequence of frames. To do this, we successively select subsequences of the predefined length k with the step s starting from the beginning of the whole sequence (see Figure 3).

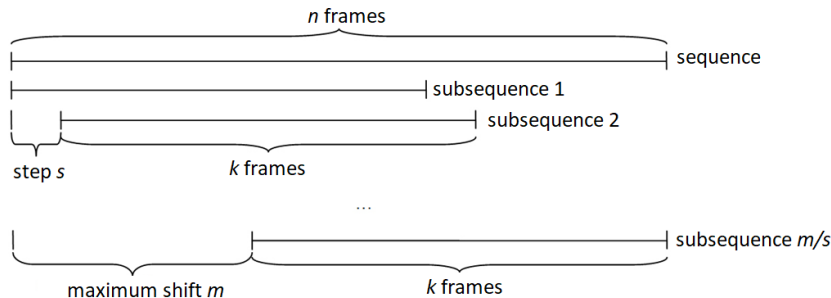


Fig. 3. Parameters of the allocated subsequences

For each selected subsequence, the vector of features is formed as follows: each normalized frame of the subsequence is expanded into a row, and the rows obtained for individual frames are concatenated to each other.

The feature vectors of all sequences for different persons form the input matrix for the principal component analysis technique. When principal components are found, the projection of feature vectors onto the first N principal components is taken as a feature description.

Classification of video sequences. The features obtained as a result of the principal component analysis are used to train the support vector machine (SVM) [15] classifier. In the considered case, the classes correspond to individual persons (individuals), and feature vectors obtained for all the subsequences correspond to individual observations (examples).

The description given above is valid for the training mode, in which the parameters of the dimensionality reduction procedure (PCA) and classifier (SVM) are configured. In the testing mode, the data is processed in the same way, except that the parameters of the linear transformation (which is used to reduce the dimensionality) are fixed to the values obtained in the training mode, and the classification is performed by the trained SVM classifier.

2.2 Identification of a moving person using hidden Markov models

In this paper, we consider the method described in [16] as an alternative approach. In this section, we give a brief description of this approach.

A Hidden Markov Model (HMM) is defined as $\lambda = (\pi, A, B)$, where π is the initial state distribution, A is the state transition probability matrix, and B is the observation probability matrix. As the gait of a person consists of a sequence of cyclic movements, the HMM can be naturally adapted to solve the considered task. To build the HMM model one has to determine the number n of states $S_i, i = 1..n$ of the model, and estimate initial π , transition A , and observation B probability matrices. In according to [16], we define the set of states S_i in such a way that the system successively passes through all these states at equal time intervals in one cycle of a motion. For every state $S_i, i = 1..n$, we estimate corresponding exemplar image (person's stance) $E_i, i = 1..n$ using the sequence of silhouettes (masks) extracted from a video. These exemplar images are used thereafter to estimate observation probabilities.

HMM initialization. The HMM is built for each person separately. The initial state probabilities are set to $\pi_i = 1/n, i = 1..n$, as the first frame of an observed sequence can correspond to any state s_i . The transition probability matrix $A = \{a_{i,j}\}$ is defined as: $a_{i,i} = 0.5$ and $a_{i,i+1} = 0.5, 1 \leq i < n, a_{n,1} = 0.5$. The observation probability matrix $B = \{b_{i,j}\}, i = 1..n, j = 1..m$ for a given set of frames $X = x_j, j = 1..m$ is defined as

$$b_{i,j} = \alpha e^{-\beta D(e_i, x_j)}. \quad (1)$$

Here α and β are some constants, $D()$ is a distance metric. In this paper we used Euclidean distance as a metric. Above mentioned constants are selected so that $\sum_{j=1}^m b_{i,j} = 1, i = 1..n$. The described HMM is shown in the Figure 4.

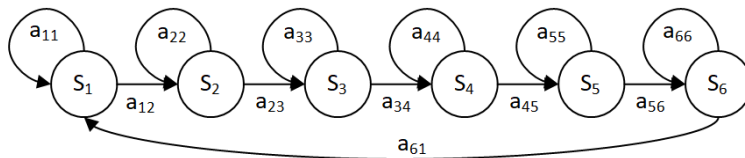


Fig. 4. The hidden Markov model for person identification.

As it can be seen from the above expression, we first need to define exemplar images $e_i, i = 1..n$ to estimate the observation probabilities. To do this, we use the following procedure proposed in [16]. At the first step, for each frame $x_j, j = 1..m$ of a sequence, we compute the mass $g_j, j = 1..m$ of the corresponding silhouette (mask) image. Then mass values g_j are filtered with a median filter, and local minima are estimated. These minima allow us to partition the whole sequence $x_j, j = 1..m$ into some number of gait cycles. After boundaries of gait cycles are determined, we divide each cycle into n intervals of equal (up to one frame) length, correspondingly to HMM states $s_i, i = 1..n$. Finally, ex-

emplar images $e_i, i = 1..n$ are estimated by averaging silhouette (mask) images of corresponding intervals for all gait cycles.

Examples of estimated exemplar images are shown in Figure 5.

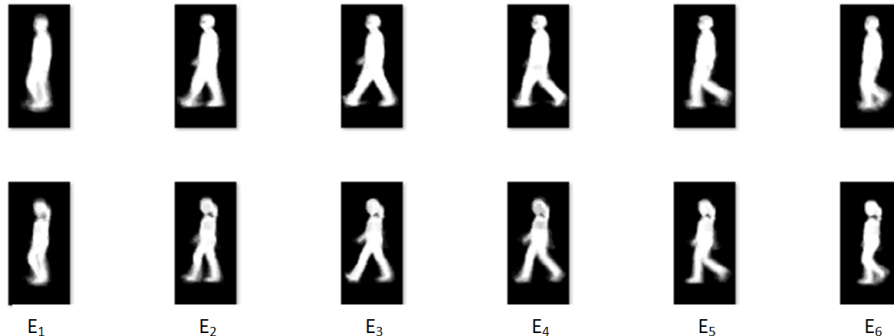


Fig. 5. Exemplar images estimated for two different persons.

HMM training. The initial HMM model described above can be further refined according to [16] using the following algorithm:

1. The most probable sequence of states is estimated for each training sequence of frames using the Viterbi algorithm [17].
2. The exemplar images are refined according to the estimation made at the previous step of the algorithm.
3. The transition probability matrix A is refined using the Baum-Welch algorithm [17] (observation probabilities B are recalculated using re-estimated exemplar images).

Classification of video sequences. Given a number of trained HMM models $\lambda_k, k = 1..K$, and a test frame sequence, $X = \{x_j\}, j = 1..m$, we select the model, which gives maximum probability $p(X|\lambda_k)$ of producing the test sequence. The probability for each trained model is computed using the forward algorithm [17].

3 Experiments

The described above methods were implemented in C++ using the OpenCV library. A PC based on Intel Core i5-3470 CPU 3.2 GHz was used to perform experimental studies.

For the experimental study, the video sequences from the open CASIA GAIT dataset [18] were used. This dataset contains sequences of binary images, which contain silhouettes of moving persons. In this work, we used sequences of 25 persons, in which the shooting angle is 90 degrees, people are depicted in normal

clothes and without bags. There were 6 sequences in each class. The length of each sequence was not less than 60 frames. Classes were divided into training and test samples of 3 sequences each. To estimate the quality of the considered methods, we used the classification accuracy, defined as the proportion of correctly classified sequences.

For the method proposed in Section 2.1, we studied the dependence of the classification accuracy on the dimensionality of feature vectors (output dimensionality of the PCA technique). The results of the experiments are shown in the Figure 6. As it can be seen from the figure, the best values of the classification accuracy are achieved for 64-dimensional feature vectors. The increase in dimensionality is accompanied by the expected increase in processing time, although the changes are not very significant.

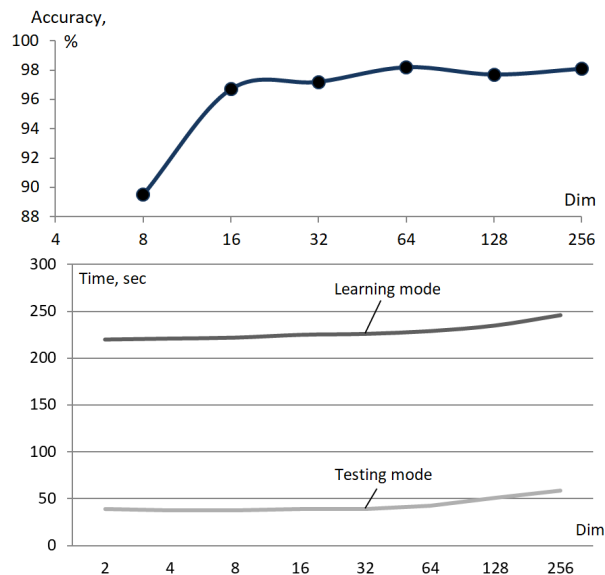


Fig. 6. Dependence of the classification accuracy (top) and processing time (bottom) on the dimensionality of feature vectors.

In the second experiment, we considered the dependence of the classification accuracy on the number of classes (persons). The experiment was carried out for 5, 10, 15, 20 and 25 classes, and other parameters remained fixed. In particular, the step s was equal to 2 frames, the maximum shift m of the beginning of the extracted subsequences was equal to 15 frames, the dimensionality of feature vectors was equal to 64. The results of the experiment are shown in the Figure 7. This experiment was conducted on the first three sequences of random classes with subsequent averaging.

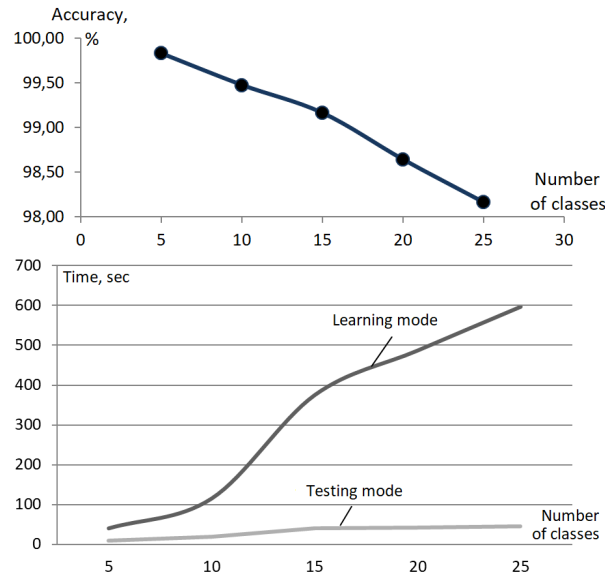


Fig. 7. Dependence of the classification accuracy (top) and processing time (bottom) on the number of classes for the proposed method.

It is worth noting that a direct experimental comparison to other works seems to be quite a challenge in connection with the different data sets used, as well as the potential differences in the experimental conditions. The closest approach to the proposed one is described in the paper [6]. Depending on the classifier configuration, the authors in [6] declared the accuracy from 92.08 to 98.79% for the case with ten objects. In this paper, for 25 classes we achieved 98.80% classification accuracy value, which was validated over twenty possible combinations of three training sequences. Thus, it can be said that the results obtained in this paper correspond to the current state in the considered field of research.

As it can be seen in Figure 7, the processing takes an increasing amount of time as the number of classes increases. Considerable time is taken in the training mode. This fact becomes especially important in scenarios when the number of classes changes dynamically, and it is required to re-train the system regularly.

To compare the proposed method to the known approach, we implemented the HMM-based technique described in Section 2.2. The results of the experiments are shown in the Figure 8. As you can see, the HMM-based approach provided acceptable accuracy for the considered number of classes. But these values are significantly lower than the accuracy demonstrated by the proposed approach.

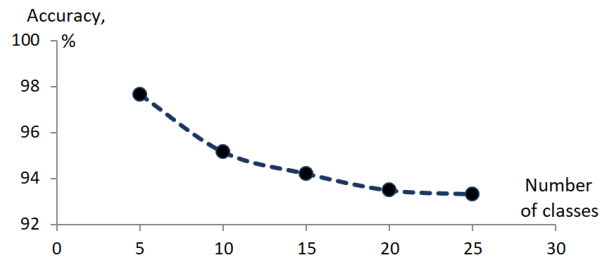


Fig. 8. Dependence of the classification accuracy on the number of classes for the HMM-based method.

4 Conclusion

In this paper, we proposed the method for human identification by gait. The proposed method consists of the detection of a moving person on a video sequence with the subsequent normalization of size, generation of subsequences, dimensionality reduction using the principal component analysis technique, and classification using the support vector machine. The experiments performed on the CASIA GAIT dataset allowed to determine the best values of the parameters of the proposed method and to compare the proposed method to the HMM-based technique. It was shown that the proposed method outperforms implemented HMM-based technique and provides high classification accuracy on the considered dataset. In particular, for 25 classes, the accuracy was 98,8% that corresponds to the current state of research.

The drawbacks of the proposed method include its long operating time. In connection with this, a promising line of research is speeding up this method.

Acknowledgments The reported study was funded by RFBR according to the research project no. 17-29-03190-ofi.

References

1. H. Josinski, A. Switonski, A. Michalczuk, D. Kostrzewa, K. Wojciechowski: Feature Extraction and HMM-Based Classification of Gait Video Sequences for the Purpose of Human Identification. *Vision Based Systems for UAV Applications. Studies in Computational Intelligence*, Vol. 481, pp. 233-245. (2013)
2. J. Suutala, K. Fujinami, J. Roning: Gaussian Process Person Identifier Based on Simple Floor Sensors. *European Conference on Smart Sensing and Context*, pp. 55-68. (2008)
3. D. Dingbo, G. Guangyu, H. Chi, JianMa: Automatic Person Identification in Camera Video by Motion Correlation. *Hindawi Publishing Corporation Journal of Sensors*, 838751. (2014)

4. C. Murukesh, K. Thanushkodi, P. Padmanabhan, Mohamed D. Feroze, Naina: Secured Authentication through Integration of Gait and Footprint for Human Identification. *Journal of Electrical Engineering and Technology*, Vol. 9(6), pp. 2118-2125. (2014)
5. L Wang, T. Tan, W. Hu, H. Ning: Automatic Gait Recognition Based on Statistical Shape Analysis. *IEEE Transactions on image processing*, Vol. 12(9), pp. 1120-1131. (2003)
6. P.B.Shelke, P.R.Deshmukh: Person Identification Using Gait: SVM Classifier Approach. *International Journal of Emerging Technologies and Engineering (IJETE)*, Vol. 1(10). (2014)
7. P. KadewTraKuPong, R. Bowden: An improved adaptive background mixture model for real-time tracking with shadow detection. *Video-Based Surveillance Systems*, pp. 135-144. (2002)
8. Z.Zivkovic: Improved adaptive Gaussian mixture model for background subtraction. *Proc. of the 17th Int. Conf. on Pattern Recognition Cambridge*, Vol. 2, pp. 28-31. (2004)
9. A.B. Godbehere, A. Matsukawa, K. Goldberg: Visual Tracking of Human Visitors under Variable-Lighting Conditions for a Responsive Audio Art Installation. (2012)
10. Background Subtraction. https://docs.opencv.org/3.3.0/db/d5c/tutorial_py_bg_subtraction.html
11. K. Fukunaga: *Introduction to Statistical Pattern Recognition*. London: Academic Press, 2nd edn. (2003)
12. J.A. Lee, M. Verleysen: *Nonlinear Dimensionality Reduction*. New York: Springer-Verlag. (2007)
13. E.V. Myasnikov: A Nonlinear Method for Dimensionality Reduction of Data Using Reference Nodes. *Pattern Recognition and Image Analysis*, Vol. 22 (2), pp. 337-345. (2012)
14. E.V. Myasnikov: Fast Techniques for Nonlinear Mapping of Hyperspectral Data. *Proc. SPIE 10341, Ninth International Conference on Machine Vision (ICMV 2016)*, 103411D. (2017)
15. C. Cortes, V. Vapnik: Support-vector networks. *Machine Learning*, Vol. 20 (3), pp. 273-297. (1995)
16. A. Sundaresan, A. RoyChowdhury, R. Chellappa: A hidden Markov model based framework for recognition of humans from gait sequences. *Proceedings 2003 International Conference on Image Processing*, Vol. 3, pp. II-93-6. (2003)
17. L.R. Rabiner: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, Vol. 77(2), pp. 257-286. (1989)
18. CASIA GAIT dataset. <http://www.cbsr.ia.ac.cn/english/Databases.asp>.