Gait analysis for person recognition using principal component analysis and support vector machines

O V Strukova[1], LV Shiripova[1] and E V Myasnikov[1]

[1]Samara National Research University, Moskovskoe Shosse 34, Samara, Russia, 443086

Abstract. The paper is devoted to the problem of the recognition of a person by gait using a video recorded in the optical range. The method proposedin this paper consists in the detection of a moving person on a video sequence with the subsequent size normalization and dimensionality reduction using the principal component analysis technique. The person classification was carried out using the support vector machine. The experimental studies performed using the CASIA GAIT dataset allowed us to determine the best values of the method parameters. The obtained results showed that with a small number of classes, high classification accuracy canbe achieved.

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 face image, voice timbre, fingerprints, iris pattern and even gait are used for the identification of a person. Although the use of fingerprints or the iris pattern 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 the person using his gait.

Considering the 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 the human body [2]. In addition, there are papers, in which gait analysis is performed based on the readings of the accelerometers built into the smartphone [3].

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

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

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

The first step of the approach proposed in the paper [5] is an improved background subtraction procedure. In this paper, the 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 linear (PCA) and non-linear (ISOMAP, LLE) dimensionality reduction techniques, which are used to form features, is performed. A Hidden Markov Model (HMM) is used to classify the generated features.

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

In general, it can be noted that the problem of recognizing a person by gait attracts the attention of an increasing number of researchers. At the same time, considerable attention is paid to both the methods of the feature description of motion and 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 and reducing the dimensionality of the sequence using the principal component analysis technique. The support vector machine is used as a classifier. Considerable attention is paid to the selection of parameters of a feature description. The study shows the importance of the careful selection of parameters in the solution of the considered problem. This allows to achievehigh quality of the classification 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 used literature is given at the end of the paper.

2. Methods
The method developed in this paper consists of the following steps:
- detection and segmentation of a moving person in the video sequence,
- normalization of the frame size of the selected video sequence fragment,
- dimensionality reduction of the selected video sequence fragment,
- classification of video sequences.

2.1. Detection and segmentation of a moving person on a video sequence
At the first stage of the developed method the moving person is allocated on the video sequence. When the video sequence source is a video surveillance camera, background subtraction methods are used most frequently. The main idea of the methods of this class is to use a certain background model and to decide whether the 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 observation 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 the 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.

As a result of this stage, the set of masks corresponding to individual frames of the video sequence is formed. Each mask reflects the result of the segmentation of a frame into the foreground area corresponding to a moving person and the background. An example obtained using the selected method is shown in figure 1.

2.2. 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 the figure 2.Taking into account the time coordinate, the dimensionality of the sequence of masks, which describes the movement of a

person, remains high even after the size normalization. In this regard, the third stage reduces the dimensionality of data describing the movement of a person.

Figure 1. Detection of a moving person on a video sequence: the frame of the original video sequence (on the left), the formed mask (on the right) [10].

Figure 2. Normalization of the size of detected fragments.

2.3. 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 (for example, nonlinear mapping, ISOMAP, LLE) [12] 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.
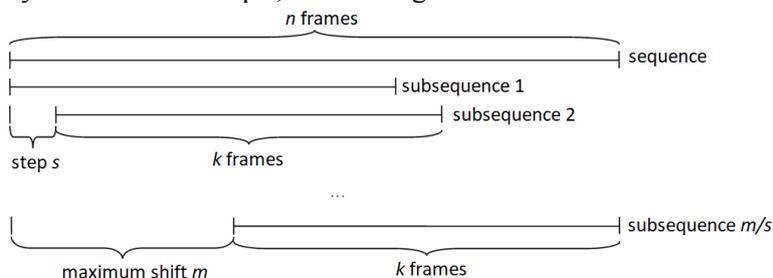

Figure 3. Parameters of the allocated subsequences.

In this paper, before reducing the dimensionality of data, for each sequence of frames we form a set of subsequences of a fixed length. 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).

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.

2.4. 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.

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 [16] were used. This dataset contains the sequences of binary images, which contain the 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.

In the first experiment, the dependence of the classification accuracy on the maximum shift of the subsequences from the beginning of the sequences was investigated (parameter m in figure 3). In this experiment, to reduce the learning time used in the selection of subsequences, the shift step s was 3 frames. Thus, the maximum step of the subsequences m also changed in step 3, taking values from 0 to 15. The last value was determined from the length k of the generated subsequences ($k = 45$ frames), and the minimum length of the sequences selected for the experiments (n=60 frames).

The experimental results are shown in figures 4 and 5. It was experimentally determined that the accuracy of the classification as a whole increases with increasing maximum shift. This observation is quite expected, since the greater the maximum shift is used in the formation of characteristics, the more "complete" is the feature description of the video sequence. This growth is also accompanied by an increase in processing time, as the number of processed subsequences increases.
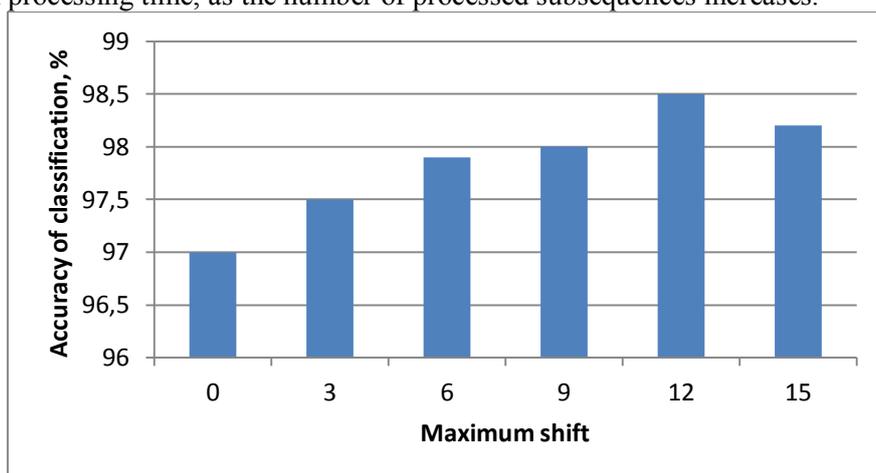

Figure 4. Dependence of the classification accuracy on the maximum shift.

To choose the value of the step s (see figure 3) we considered rather small values from 1 to 4, which allow us to obtain a more "dense" feature description of the video sequence. Our preliminary experiments showed that the best classification accuracy is achieved at small values of the shift s. Taking into account also the computation time, we ended up with the value s=2.For the method proposed in Section 2.3, 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 figures 6 and 7.
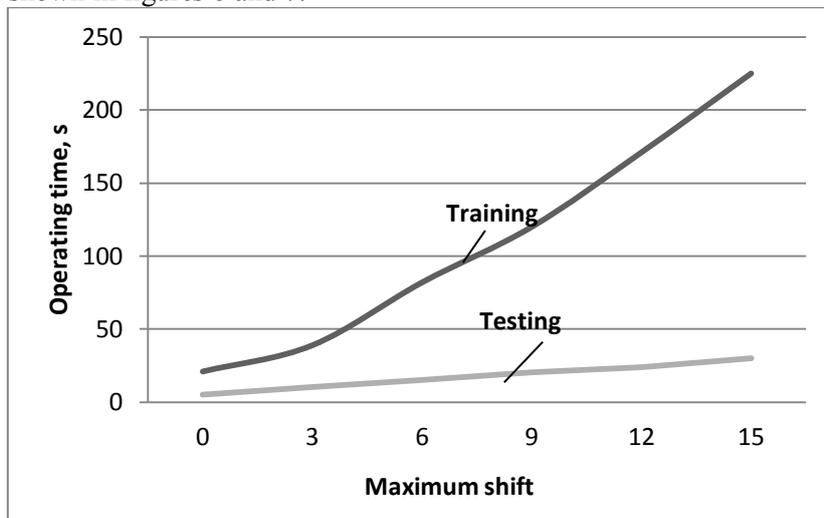

Figure 5. Dependence of training and testing time on the maximum shift.

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.
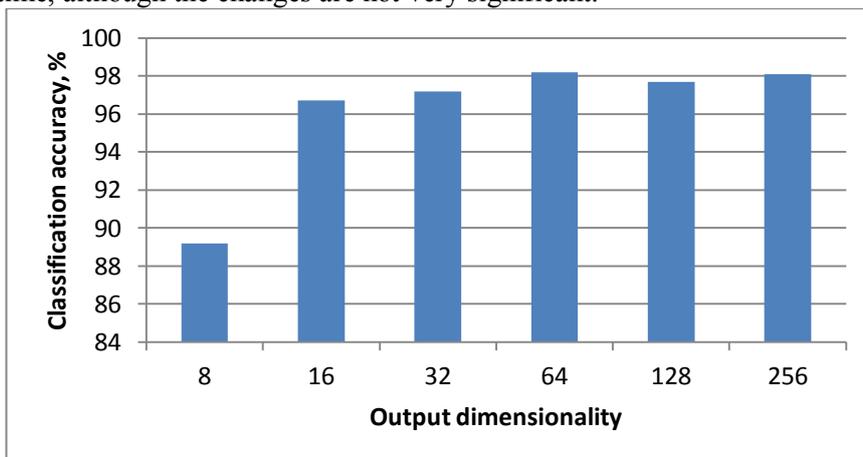

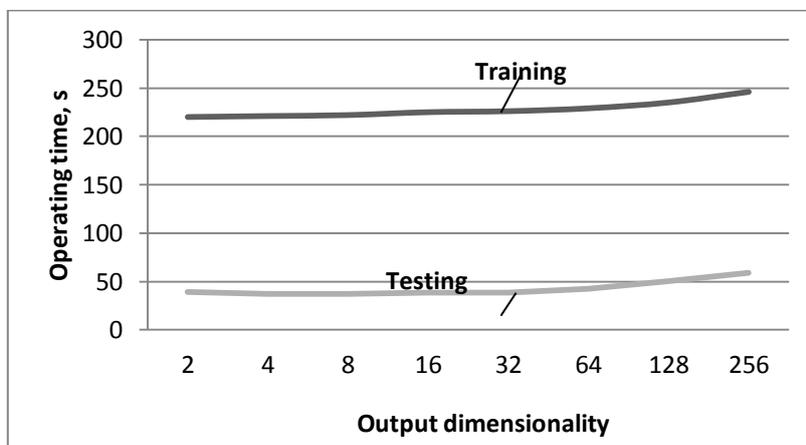**Figure 6.** Dependence of the classification accuracy on the dimensionality.

Figure 7. Dependence of training and testing time on the dimensionality.

In the next 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, and the dimensionality of feature vectors was equal to 64. The results of the experiment are shown in the figures 8 and 9.
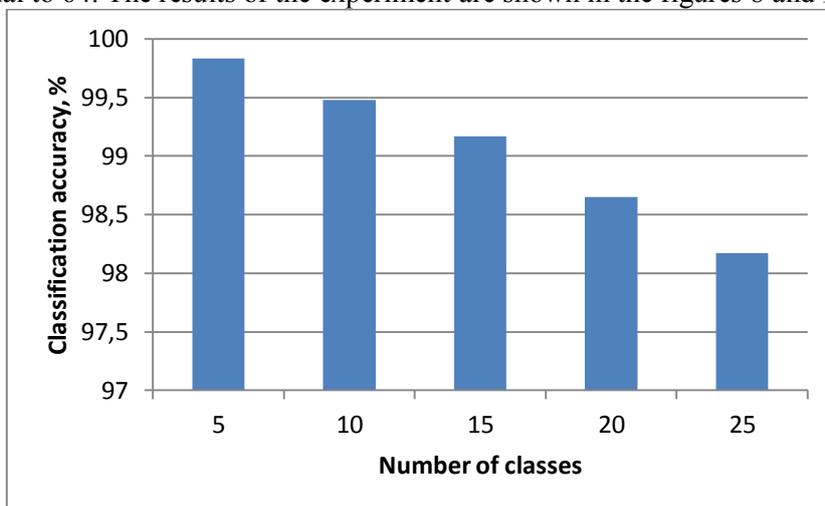

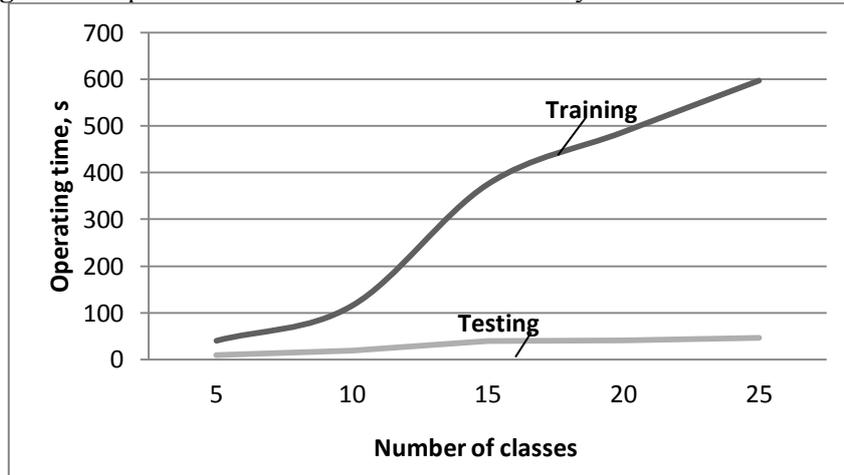**Figure 8.** Dependence of the classification accuracy on the number of classes.

Figure 9. Dependence of training and testing time on the number of classes.

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.

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 the figure 9, 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.

4. Conclusion
In this paper we proposed the method for human identification by gait. The proposed method consists in 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.

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. Another possible direction of further research is the recognition of human actions and the detection of abnormal behavior (see, for example, [17, 18]).

5. References

[1] Josiński H, Świtoński A, Michalczuk A, Kostrzewa D and Wojciechowski K 2013 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* **481** 233-245

[2] Suutala J, Fujinami K and Röning J 2008 Gaussian Process Person Identifier Based on Simple Floor Sensors *European Conference on Smart Sensing and Context* 55-68

[3] Dingbo D, Guangyu G, Chi H and Jian Ma 2014 Automatic Person Identification in Camera Video by Motion Correlation *Journal of Sensors* **2014** 838751

[4] Murukesh C, Thanushkodi K, Padmanabhan P and Feroze Naina M D 2014 Secured Authentication through Integration of Gait and Footprint for Human Identification *J. of Electrical Engineering and Technology* **9(6)** 2118-2125

[5] Wang L, Tan T, Hu W and Ning H 2003 Automatic Gait Recognition Based on Statistical Shape Analysis *IEEE Transactions on Image Processing* **12(9)** 1120-1131

[6] Shelke P B and Deshmukh P R 2014 Person Identification Using Gait: SVM Classifier Approach *International Journal of Emerging Technologies and Engineering (IJETE)* **1(10)**

[7] Kadew T K P and Bowden R 2002 An improved adaptive background mixture model for real-time tracking with shadow detection *Video-Based Surveillance Systems* 135-144

[8] Zivkovic Z 2004 Improved adaptive Gausian mixture model for background subtraction *Proc.of the 17th Int. Conf. on Pattern Recognition* **2** 28-31

[9] Godbehere A B, Matsukawa A and Goldberg K 2012 Visual Tracking of Human Visitors under Variable-Lighting Conditions for a Responsive Audio Art Installation *American Control Conference*

[10] *Background Subtraction* (Access mode: https://docs.opencv.org/3.3.0/db/d5c/tutorial_py_bg_subtraction.html) (13.11.2017).

[11] Fukunaga K 2003 *Introduction to Statistical Pattern Recognition* (London: Academic Press)

[12] Lee J A and Verleysen M 2007 *Nonlinear Dimensionality Reduction* (New York: Springer-Verlag)

[13] Myasnikov E V 2017 Fast Techniques for Nonlinear Mapping of Hyperspectral Data *Proc. SPIE* **10341** 103411D

[14] Myasnikov E V 2016 The use of Interpolation Methods for Nonlinear Mapping *Lecture Notes in Computer Science* **9972** 649-655

[15] Cortes C and Vapnik V 1995 Support-vector networks *Machine Learning* **20(3)** 273-297

[16] *Database CASIA GAIT* (Access maode: http://www.cbsr.ia.ac.cn/english/Databases.asp) (13.11.2017)

[17] Morozov A A and Sushkova O S 2016 Real-time analysis of video by means of the Actor Prolog language *Computer Optics* **40(6)** 947-957 DOI: 10.18287/2412-6179-2016-40-6-947-957

[18] Shatalin R A, Fidelman V R and Ovchinnikov P E 2017 Abnormal behavior detection method for video surveillance applications *Computer Optics* **41(1)** 37-45 DOI: 10.18287/2412-6179-2017-41-1-37-45