THE TECHNIQUES FOR CREATION OF FACE IDENTIFICATION SYSTEMS BASED ON SVM AND HMM

Rauf Sadykhov[1], Igor Frolov[2], and Natallia Pavlyonok[3]

Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus
[1]*rsadykhov@bsuir.by*, [2]*frolovigor@yandex.ru*, [3]*n.pavlionok@gmail.com*

1. Introduction

Nowadays personal identification is a very important issue. There is a wide range of applications in different spheres, such as video surveillance security systems, control of documents, forensics systems and etc. Systems for person identification from video sequences require fast and robust recognition algorithms. It is very important to avoid problems that are typical for video. These main problems are the following:

– low image quality;
– high noise level;
– face orientation and size;
– facial expressions;
– real-time requirements etc.

Our goal is to develop face recognition techniques and create the system for face identification. In this paper we propose two different techniques. Each of them has its advantages and contains two main stages: data reducing and classification. Below we discuss different points of training methods and data reduction.

2. Support Vector Machines and Neural Network Principal Component Analysis

The Support Vector Machines (SVMs) [1] present one of kernel-based techniques. A special property of SVMs is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence they are also known as maximum margin classifiers. SVMs are used for classification of both linearly separable and inseparable data. SVMs based classifiers can be successfully apply for text categorization, face identification.

Basic idea of SVMs is building the optimal hyperplane for linearly separable patterns. This approach can be extended to patterns that are not linearly separable by transformations of original data to map into new space due to using kernel trick.

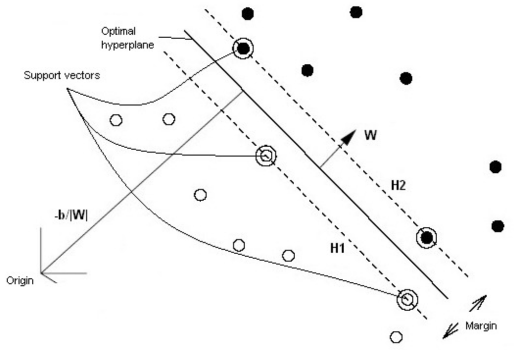


Figure 1. Linear separating hyperplanes for the separable case

In the context of the Figure 1 above, illustrated for 2-class linearly separable data, the design of the conventional classifier would be just to identify the decision boundary *w* between the two classes. However, SVMs identify support vectors (SVs) *H1* and *H2* that will create a margin between the two classes, thus ensuring that the data is "more separable" than in the case of the conventional classifier.

Formally put, if we let $\{x_i, y_i\}$ be the training data *x* with class label *y*, $y_i \in \{-1, +1\}$, then any point lying on the hyperplane separating the two classes (i.e., the decision boundary) will satisfy $\langle w \cdot x \rangle + b = 0$, with *w* being the normal to the hyperplane and $|b|/\|w\|$ being the perpendicular distance of the hyperplane from the origin.

The final formulation definition can be defined as:

$$\begin{cases} \frac{1}{2}\|w\| \to \min, \\ y_i(w \cdot x_i + b) \geq 1, \qquad i = 1,..., l. \end{cases} \tag{1}$$

This quadratic programming problem is the learning task that can be reduced to minimization of *the Wolfe dual lagrangian*[2]:

$$\begin{cases} W(\lambda) = \sum_{i=1}^{l} \lambda_i - \frac{1}{2}\sum_{i=1}^{l} \lambda_i \cdot \lambda_j \cdot y_i \cdot y_j \cdot (x_i \cdot x_j), \\ \lambda_i \geq 0, \quad \sum_{i=1}^{l} y_i \cdot \lambda_i \qquad i = 1,..., l. \end{cases} \tag{2}$$

We can build the classifier with good generalization performance using SVMs. But we would like use this classifier for the face identification and we deal with image in this case and therefore the source data dimension forms about 10.000 features (image with size 112×92 pixels can be considered as a vector of dimension 10.304, or equivalently a point in a 10.304 dimensional space).

We have chosen the Principal Component Analysis method for the data reduction. There are some various techniques to calculate principal components.

The neural network PCA (NNPCA) is used in our work. *The Generalized Hebbian Algorithm* by Sanger[3] is one among the best known learning algorithms that allow a neural network (Figure 2) to extract a selected number of principal components from a multivariate random process. It applies to a single-layered feedforward neural network that may be described by equation (3) with the rule for updating weights (4).

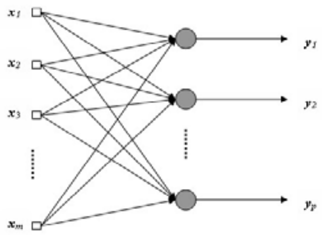


Figure 2. A single-layered feedforward neural network for extracting p principal components

$$y_j(n) = \sum_{i=1}^{p} w_{ij}(n)\, x_i(n), \qquad j = 1, 2, ..., m. \tag{3}$$

$$\Delta w_{ij}(n) = \eta \left[y_j(n) x_i(n) - y_j(n) \sum_{k=1}^{j} w_{ki}(n) y_k(n) \right]. \tag{4}$$

The weight of first eigenvector has been estimated and its value lies within the range from 75 to 84%. Therefore, to decrease the computation expenses we used only one eigenvector for calculation one principal component.

3. Hidden Markov Models and Discrete Wavelet Transformation

Since 1960s Hidden Markov Models (HMM) were successfully applied for processes modeling and in machine recognition of speech [4]. For reducing computations linear HMM were used.

First in static face recognition HMM were used by Samaria [5]. Later linear [6] and pseudo 2D HMM, which are the combination of 2 linear models, were proposed. For minimizing calculations in [6] discrete cosine transform (DCT) was applied.

Simple structure of HMM for face recognition (FR) is shown in Figure 3. It is linear model which considers special properties of face images.

States of the model correspond to main features of human face such as hair, forehead, eyes, nose and mouth. Transitions are probable only between neighbor states in view of natural order.

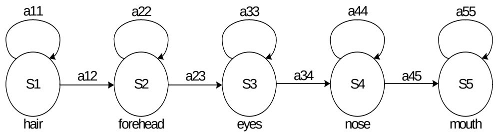
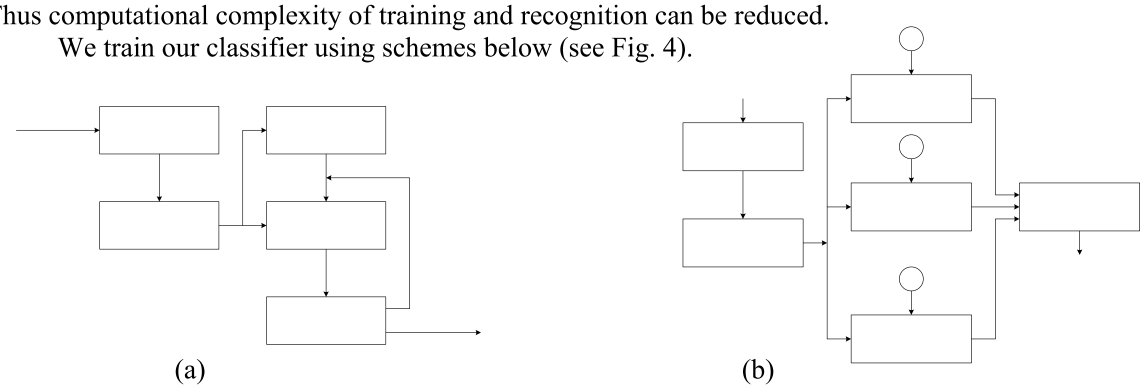


Figure 3. HMM-structure for FR

For describing such model only sparse state transition probability matrix is needed. Thus computational complexity of training and recognition can be reduced.

We train our classifier using schemes below (see Fig. 4).



(a) (b)

Figure 4. Schemes for (a) training HMM-classifier; (b) recognition

Blocks and features extraction are assigned for reducing computational complexity. Model reestimation is literally the solution of *Explanation Problem* [4]. Model convergence is by means of *Baum-Welch algorithm* [4] established.

While recognizing input image is divided and reduced just as while training. For each model probability of image generation is by means of *forward-backward algorithm* [4] calculated.

Picture is divided according to a states representation as shown in Figure 5.

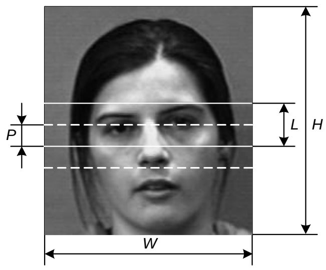


Figure 5. Block extraction

Each image of size *WxH* is divided into *T* overlapping blocks of size *WxL*. The amount of overlapping is *P* [6]. Thus the number of blocks is:

$$T = \frac{H-L}{L-P} + 1 \qquad (5)$$

The choice of parameters *L* and *P* can affect algorithm characteristics e.g. high amount of overlapping increases the recognition rate because allows capture the features more precisely but also it increases complexity and calculation time.

In the Figures 6 visualizations of 2D Discrete Cosine Transformation (DCT) [6] and Discrete Wavelet Transformation (DWT) [7] of single block are shown.

(a)

(b)

Blocks
Extraction

(c)

Figure 6. Transformations visualizations:
(a) initial block; (b) block DCT; (c) block DWT

Features
Extraction

Evidently in Figure 6b, significant part of block is in the first half concentrated. That's why the second part can be neglected. Using DWT the input sequence can be 4 times more reduced.

Thus for our face identification system we apply co-called DWT+HMM technique.

4. Results

We have tested our system for face identification on The Database of Faces, AT&T Laboratories, Cambridge University, UK [8]. This data base contains sequences of 10 images which represent 40 people of various racial origins and age. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). The size of each image is 112x92 with 256 grey levels per pixel . Examples of test images are shown in Figure 7.

Figure 7. The Database of Faces

Classifiers were trained using different number of images (1-5). For DWT+HMM classifier we have estimated size of blocks as *90x20* (*L=20; P=10*).

In table below were following system characteristics analyzed:

– recognition rate (R_r, %) – rate of correct recognition;
– training time (T_t, s);
– average recognition time (T_r, s) – average time of an image (frame) processing.

Table 1. Integrated results

Number of images	DWT+HMM			NNPCA+SVM		
	R_r, %	T_t,s	T_r, s	R_r, %	T_t,s	T_r, s
1	93.75	1.2	0.46	66.39	3.9	0.47
2	96.25	4.2	0.49	73.75	4.1	0.47
3	98.50	5.2	0.43	83.57	4.7	0.42
4	99.33	7.2	0.43	90	5.1	0.39
5	100.00	8.7	0.43	96.5	5.4	0.37

5. Conclusion

In this paper two different systems intended to decide the problem of face identification is described. The DWT+HMM system shows higher accuracy but the NNPCA+SVM system use smaller amount of features. The results show a strong correlation between accuracy in classification and time when the classification is made. The developed systems achieve high accuracy results, are robust with respect to noise, low image quality, facial expressions and orientation.

Literature

[1] Burges C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition, Boston, 1998.
[2] Vapnik V. Universal Learning Technology: Support Vector Machines, NEC Journal of Advanced Technology №2, 2005.
[3] Sanger T.D. Optimal Unsupervised Learning in A Single-Layer Linear Feedforward Neural Network.
[4] Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition / Proceedings of the IEEE. − February, 1989. − Vol. 77, № 2.
[5] Samaria F. HMM based architecture for face identification / Image and Computer Vision. − October, 1994. − Vol. 12.
[6] Nefian A.V. Hidden Markov models for face recognition / Proc. IEEE ICASSP'98. − May, 1998.
[7] Mallat S.G. A theory for multiresolution signal decomposition: The wavelet representation / IEEE Trans. on Pattern Analysis and Machine Intelligence, 1989. – Vol.11, №7.
[8] The Database of Faces. – www.cl.cam.ac.uk/research/facedatabase.html