The Third International Conference "Problems of Cybernetics and Informatics"
September 6-8, 2010, Baku, Azerbaijan. Section #3 "Modeling and Identification"
www.pci2010.science.az/3/36.pdf

METHODS AND ALGORITHMS TO CREATE THE INTELLECTUAL FACE RECOGNITION SYSTEM

Rauf Sadykhov[1] and Sergei Tulyakov[2]

Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus
[1]*rsadykhov@bsuir.by*, [2]*sergei.tulyakov@gmail.com*

Abstract

In this paper we present our work on upright frontal face recognition system. The aim of our system is to recognize faces on machine readable travel documents (MRTD). Our goal is to create a system which will be able to handle large image databases. In order to increase detection speed and hit rate we introduce some heuristics. For recognition purposes eigenface approach is used. We also introduce a measure to estimate how well eigenface basis is constructed.

Introduction

During two last decades the technologies of face detection and recognition have attracted considerable research interest. These technologies are used in biometric systems, border control systems, video surveillance, human-computer interface, access control systems, face expression recognition, content based image retrieval.

The task of face recognition consists of two separate problems: face detection and face identification. Methods to detect faces can be divided into four main groups: (i) knowledge-based methods, (ii) feature invariant approaches, (iii) template matching methods and (iv) appearance-based methods [1]. Knowledge-based methods use expert knowledge to encode typical face. Feature invariant approaches are aimed to find facial features even when imaging conditions (pose, viewpoint or illumination) vary. Template matching methods store several standard patterns to describe a whole face or facial features separately.

Nowadays the appearance-based methods gained most popularity due to their hit rate and speed [1]. These methods are based on scanning the input image on different scales with fixed window size in order to find faces. Each window then is given as an input to some classifier, trained to separate two classes of objects (face/non-face).

Our approach is based on cascade of weak classifiers proposed by Viola et al.[2]. In order to increase detection speed and hit rate we introduce some heuristics that will be discussed in the next sections. Identification stage is done using eigenface approach[3].

Stages of face recognition

In our work we apply already developed methods along with suggestions to improve speed, hit rate and false hit rejection on Machine Readable Travel Documents (MRTD). MRTD portraits must conform the quality requirements contained in Doc 9303 (e.g. the portrait shall be colour neutral showing the applicant with the eyes open and clearly visible; the portrait shall show the eyes clearly with no light reflection off the glasses and no tinted lenses).

Face recognition procedure consists of several stages. After input image is loaded, system starts features detection stage. This stage consists of face search and eye search. If system failed to find face or eyes on image it prompts the user to mark them manually. If the user didn't manage to find features, then the image either does not conform quality requirements or is not a portrait at all. When this stage is successfully completed, system moves to the next stage, where the image processing starts. This stage ends up with storing image prepared for identification in the database. Identification stage uses known images to determine the most similar image.

Features detection

Face and eyes are detected using cascade classifier proposed by Viola et al. [2] along with extended set of Haar-like features proposed by Lienhart et. al. [5]. Since appearance based methods scan image on different scales, a region that represents a face is found several times. If we know exactly that there is only one face in the image, we should choose region with maximum number of neighbors (see figure 2, B).

Also, if we use "find the biggest face" strategy, we would also be able to reject false hit. We have run experiments to find out is it better to search the biggest face or to find the candidate which has most neighbors in terms of speed. We used database that consists of 10 000 faces. "The biggest face" works 8 times faster rather than most neighbours approach.

Next, we need to find eye points. In order to increase speed we narrow the region of interest. We have computed regions on image that most likely will contain eyes (see figure 2, C). This idea helps significantly increase detection speed. If eyes were found inside these regions system proceeds to next step. Here we can also apply "the biggest eye" approach. But our experiments showed that in terms of hit rate it is better to search for eye regions that have most neighbors.

If there were no eyes found in predicted region this can mean following: (i) eyes are closed or incorrectly covered with glasses, (ii) image has low quality, (iii) image is rotated. The correction of image slope will be discussed in the next section.

A) B) C) D)

Figure 2. Feature detection stage

Face extraction

The purpose of this stage is extraction and processing of the face region in order to achieve the following properties: (i) imaginary line drawn between eyes is parallel to the top edge of the image, (ii) distance between eyes is set to a certain value and (iii) image is cropped and masked.

After the slope correction image must be upright. New coordinates of the features that where found previously can be adjusted using following system of equations:

$$\begin{cases} x' = x_0 + (x - x_0)\cdot\cos(\varphi) - (y - y_0)\cdot\sin(\varphi) \\ y' = y_0 + (x - x_0)\cdot\sin(\varphi) + (y - y_0)\cdot\cos(\varphi) \end{cases},$$

where x' and y' are the new coordinates of the point, (x_0, y_0) are coordinates of the rotation center, φ is rotation angle.

Then, every image is resized to set distance between eyes to a certain value. This is done to ensure that all images have the same size in pixels. Image is warped only using eye regions. If we consider to resize image using both *x* and *y* axis we change the proportions of the face and that will probably decrease recognition abilities of the system. In order to extract face each image is cropped. All these steps are shown on figure 3.

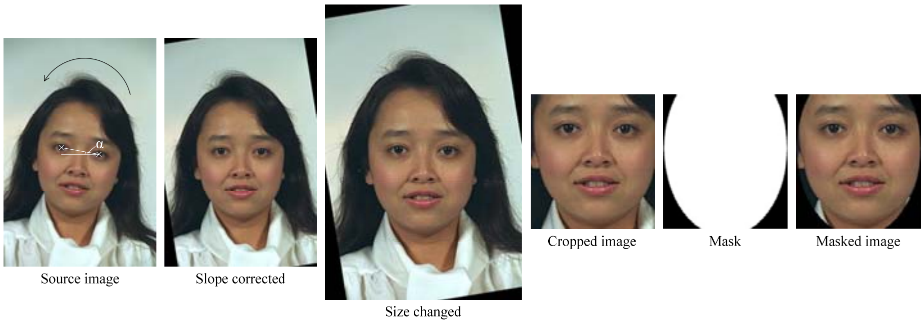

Figure 3. Feature detection stage

Identification

The goal of identification step is to find individuals among known which are the most similar to the input individual. Here the eigenface approach [3] is applied. The reason is that this approach is able to handle large amounts of data. Face image is considered to be a two-dimensional *N* by *M* array. An image can also be treated as a vector of dimension $N \cdot M \times 1$. So each image is considered to be a point in the $N \cdot M$ - dimensional subspace. Face images, being similar in overall configuration, will not be randomly distributed in this huge image space. In order to decrease dimensionality principal component analysis (PCA) can be used. This approach consists of the following steps: (i) compute the average face vector Ψ, (ii) obtain a set of face vectors Φ centered at expectation Ψ, (iii) compute the covariance matrix and find its *K* eigenvectors with the biggest eigenvalues and (iv) project each image onto new subspace to obtain coordinates of each image in this subspace.

Average face can be computed as follows:

$$\Psi = \frac{1}{n}\sum_{i=1}^{n} F_i ,$$

where F is training set of face vectors, n – number of faces in training set. Then, centered at expectation set of face vector is $\Phi_i = F_i - \Psi$.

Average face Ψ is computed using faces of people of different race, gender, age, color, with and without glasses, beards and mustache. Our experiments showed that if we increase n average face remains almost the same. The Euclidean distance between mean faces based on 100 and 200 images is 4.01; 200 and 300 images – 3.3; 300 and 400 images – 0.66; 400 and 500 images – 0.58.

Covariance matrix is determined as follows:

$$C = AA^T ,$$

where $A = [\Phi_1 \Phi_2 ... \Phi_n]$ is $N \cdot M \times n$ matrix, and C is $(N \cdot M)^2$ matrix. Matrix C is too large to compute its eigenvectors ν_i. Turk et al. [3] propose to compute ν_i as follows:

$$\upsilon_i = Au_i ,$$

where u_i are eigenvectors of $n \times n$ matrix $A^T A$.

Now each face can be represented as a linear combination of first *K* eigenvectors as follows:

$$\Phi_i = \sum_{j=1}^{K} w_j \upsilon_j ,$$

where $w_j = \upsilon_j^T \Phi_i$. To classify an image the Euclidean distance is applied:

$$\varepsilon_k^2 = \left\| \Omega - \Omega_k \right\|^2 ,$$

where Ω_k is a vector describing *k*th face class which minimizes ε_k^2. A face is classified as belonging to class *k* when the minimum ε_k^2 is below some threshold Θ. Otherwise, the image is unknown to the system and can be added to the database.

Recognition quality depends on how well eigenface basis is constructed. Let us introduce an idea to estimate how well basis is constructed. Let A be a set of faces which are used to construct basis. Let B be a set of faces which are used to test basis. $|A|,|B|$ are the numbers of images in defined sets respectively. Let $L(\upsilon)_i$ be a loss vector:

$$L(\upsilon)_i = \left(B_i - B'(\upsilon)_i\right)^2 ,$$

where $B'(\upsilon)_i$ – face images restored from basis, υ – basis:

$$B'(\upsilon)_i = \left(\upsilon^T W_i\right) + \Psi ,$$

where W_i – representation of *i*th image in the basis υ, Ψ – mean face.

The target function can be written down as follows:

$$e(\upsilon) = \sum_{k=1}^{|B|} L(\upsilon)_k$$

The task to minimize $e(\upsilon)$ is practically impossible. However, this target function can be used to compare existing eigenface bases. To illustrate this idea let us build two different bases: υ_{A1} consists of people of all races, gender, with or without mustaches (image set $A1$); υ_{A2} – white male faces without mustache (image set $A2$). Each set of images consists of 100 images.

The experiments yielded the following results: $e(\nu_{A1}) = 54.33$ and $e(\nu_{A2}) = 74.5$. Since $e(\nu_{A1}) < e(\nu_{A2})$ we can expect basis ν_{A1} to restore faces more precisely than basis ν_{A2}.

Conclusion

The proposed "the biggest face" approach along with regions that most likely contain eyes significantly increase the speed of feature detection stage and make the system able to handle large amounts of data. The method of basis ranking allows us to choose the best one which is useful for more successful recognition process.

One of the future research ideas is to apply Bayesian face similarity [6]. Another promising research direction is to divide people into several groups using such features as gender, race, age. Each new image then firstly will be compared to individuals in groups it most likely might belong to. This approach will increase identification speed and accuracy.

References

1. M. Yang, "Recent Advances in Face Detection", Tutorial, *IEEE International Conference on Pattern Recognition*, Cambridge, 2004.
2. P. Viola and M. Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision*, 2004, pp. 137-154.
3. M. Turk, and A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, vol. 3, no.2, pp. 71-86.
4. International Civil Aviation Organization. *Machine Readable Travel Documents. Doc 9303*, 2006.
5. R. Lienhart and J. Maydt, "An Extended Set of Haar-Like Features for Rapid Object Detection", *In Proceedings of The IEEE International Conference on Image Processing*, pp. 900-903.
6. B. Moghaddam, T. Jebara and A. Pentland. "Bayesian Face Recognition", *Pattern Recognition*, vol. 33, no.11, 2000, pp. 1771-1782.