A Review of Face Recognition Algorithms and Their Application in Age Estimation

Petra Koruga, Jurica Ševa, Miroslav Bača
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
{petra.koruga, jurica.seva, miroslav.baca}@foi.hr

Abstract. *Age estimation of humans is one of the problems in the field of computer vision which is insufficiently researched. To efficiently estimate the age of an individual based on his/her face image, characteristic points of the face have to be determined. In this paper, the most common face recognition algorithms, like Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA) and Elastic Bunch Graph Matching (EBGM) will be described. These algoritms will then be compared, and the application in age estimation will be given. At the end of the paper, ideas for future research will be stated.*

Keywords. face recognition, age estimation, PCA, ICA, LDA, EBGM

1 Introduction

The problem of face recognition can be stated as follows: Given a set of face images labeled with the person's identity (the learning set) and an unlabeled set of face images from the same group of people (the test set), identify the name of each person in the test images[1]. Many algorithms have been developed in order to solve this problem. This paper will give a review of four most commonly used algorithms and make the connection with applying them in age estimation.

Age estimation by humans is used constantly in everyday life, the problem arises when age estimation needs to be automated. Age estimation can be widely used and has great potential: determining the age of imigrants in situations in which there are no documents that can determine age, for web pages that allow access only for persons above certain age. It can improve face recognition systems (most of face recognition systems are sensitive to changes caused by aging), and can also be used for finding missing people during several years (especially children). It can also improve the human-machine interaction based on age of a person, predict the way a person ages, and to fight pedofilia (removing photos of underaged children from the Internet and personal computers)[2]. Previous research regarded facial age estimation using anthropometric model[2]. In this research, facial landmarks were marked manually, which has been recognized as a problem and that is the reason behind this paper.

2 State of the Art

Field of computer vision, and face recognition algorithms in particular, has been researched from many angles and many new algorithms have been developed. Turk and Pentland [3] developed a system that locates a head of a person and then recognizes the person by comparing characteristics of the face to known individuals. This system projects face images onto a feature space that spans significant variations among known face images. These significant features are called eigenfaces because they are eigenvectors (principal components) of the set of faces. Belhumeur et al.[1] developed an algorithm insensitive to variations in lighting direction and facial expression. They used a pattern classification approach, and considered each pixel in an image as a coordinate in a high-dimensional space. This method is based on Fisher's Linear Discriminant and is called Fisherface method. Etemad and Chellappa [4] proposed a new algorithm called Discriminant eigenfatures for face recognition. Algorithm is based on PCA, but the main difference is that PCA preforms eigenvalues analysis on the covariance matrix, and their algorithm on separation matrix. Wiskott and Fellous[5] presented a system for face recognition where faces are described in the form of image graphs. Facial landmarks are described by sets of wavelet components also called jets. Recognition is based on comparison of image graphs. Moghaddam et al. [6] proposed a technique for face recognition based on Bayesian analysis of image differences. Bartlett et al. [7] developed a method called Independent Component Analysis, a generalization of PCA, which uses high-order statistics to analyse high-order relationships between pixels. Li[8] presented a face recognition system based on combination of Support Vector Machines and Elastic Graph Matching. Chai et al.[9] proposed a novel face recognition method based on dual-tree complex wavelet transform and independent component analysis. Wei et al.[10] developed a technique for face recognition that combines Wavelet transformation, PCA and support vector machine.

3 Face Recognition Algorithms

A number of face recognition algorithms and methods have been developed. This paper will give an overview of four most commonly used algorithms.

3.1 Principal Component Analysis

PCA is an algorithm for face recognition based on principal components, also called eigenvectors. The idea of PCA is to extract relevant information from a face image, encode that information as efficiently as possible, and compare one face encoding with a database of models encoded in the same way[3]. The idea is to find eigenvectors of the covariance matrix of the set of face images, treating an image as a point (or vector) in a high dimensional space. Eigenvectors are actually a set of features

Figure 1: An example of eigenfaces in face recognition[3]

that characterize the variation between face images, and they can be displayed as a ghostly face called eigenface[3].
A face image I(x,y) is a two dimensional m by n array of (8-bit) intensity values. Image is also a vector of dimesion m x n[11].

The process of PCA has the following steps summarized by Mane et al.[12]:
X1, X2, X3...XN is the training set of images. The average face of this set is defined by:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} Xi \tag{1}$$

Calculate the covariance matrix to represent the scatter degree of all feature vectors related to the average vector. The covariance matrix C is defined by

$$C = \frac{1}{N}\sum_{i=1}^{N}(Xi-\overline{X})(Xi-\overline{X})^T \tag{2}$$

The Eigenvectors and corresponding eigenvalues are computed by using

$$CV = \lambda V \tag{3}$$

Where V is the set of eigenvectors associated with its eigenvalue λ. Next step is sorting the eigenvector according to their corresponding eigenvalues from high to low. Each of the mean centered image project into eigenspace using

$$W_i = V_i^T(X_i - \overline{X}) \tag{4}$$

In the testing phase each test image should be mean centered, than projected into the same eigenspace

as defined during the training phase. This projected image is than compared with projected training image in eigenspace. Images are compared with similarity measures. The training image that is closest to the test image will be matched and used to identify.
Given an s-dimensional vector representation of each face in a training set of images, Principal Component Analysis (PCA) tends to find a t-dimensional subspace whose basis vectors correspond to the maximum variance direction in the original image space. This new subspace is normally lower dimensional (t«s). If the image elements are considered as random variables, the PCA basis vectors are defined as eigenvectors of the scatter matrix[13].

3.2 Independentl Component Analysis

The problem with PCA makes one important assumption: the probability distribution of input data must be Gaussian. The mentioned assumption doesn't have to be true. Face images have more general distribution of probability density functions along each dimension, the representation problem has more degrees of freedom. In that case PCA fails because the largest variances do not correspond to meaningful axes of PCA[4].
ICA accounts for higher order statistics and it identifies the independent source components from their linear mixtures (the observables). ICA thus provides a more powerful data representation than PCA as its goal is that of providing an independent rather than uncorrelated image decomposition and representation[14].
ICA of a random vector searches for a linear transformation which minimizes the statistical dependence between its components [15]. In particular, let $X \in R^N$ be a random vector representing an image, where N is the dimensionality of the image space. The vector is formed by concatenating the rows or the columns of the image which may be normalized to have a unit norm and/or an equalized histogram. The covariance matrix of X is defined as

$$\Sigma_X = E\{[X - E(X)][X - E(X)]^t\} \tag{5}$$

where $E(\cdot)$ is the expectation operator, t denotes the transpose operation, and $\Sigma x \in R^{NxN}$. The ICA of X factorizes the covariance matrix Σx into the following form

$$\Sigma_X = F \Delta F^t \tag{6}$$

where Δ is diagonal real positive and F transforms the original data X into Z

$$X = FZ \tag{7}$$

such that the components of the new data Z are independent or the most independent possible [15]. To derive the ICA transformation F, Comon [15] developed an algorithm which consists of three operations: whitening, rotation, and normalization. First, the whitening operation transforms a random vector X into another one U that has a unit covariance matrix.

$$X = \Phi \Lambda^{1/2} U \tag{8}$$

where Φ and Λ are derived by solving the following eigenvalue equation

$$\Sigma_X = \Phi \Lambda \Phi^t \tag{9}$$

where

$$\Phi = [\phi_1, \phi_2, \ldots, \phi_N] \tag{10}$$

is an orthonormal eigenvector matrix and

$$\Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_N\} \tag{11}$$

is a diagonal eigenvalue matrix of Σx. One notes that whitening, an integral ICA component, counteracts the fact that the Mean Square Error (MSE) preferentially weighs low frequencies [16]. The rotation operations, then, perform source separation (to derive independent components) by minimizing the mutual information approximated using higher order cumulants. Finally, the normalization operation derives unique independent components in terms of orientation, unit norm, and order of projections [15].

3.3 Linear Discriminant Analysis

Both PCA and ICA do not use face class information. Linear Discriminant Analysis (LDA) finds an efficient way to represent the face vector space by

exploiting the class information. It differentiates individual faces but recognizes faces of the same individual [17].
LDA searches for vectors in the underlying space that best discriminate among classes. For all the samples of all classes, two measures are defined[18]:
Within-class scatter matrix

$$S_w = \sum_{j=1}^{c}\sum_{i=1}^{N_j}(\mathbf{x}_i^j - \mu_j)(\mathbf{x}_i^j - \mu_j)^T \tag{12}$$

where x_i^j is the i-th sample of class j, μ_j is the mean of class j, c is the number of classes, and N_j the number of samples in class j.
Between class scatter matrix

$$S_b = \sum_{j=1}^{c}(\mu_j - \mu)(\mu_j - \mu)^T \tag{13}$$

Where μ represents the mean of all classes. The goal is to maximize the between-class measure while at the same time minimizing the within-class measure[17].

$$W_{LDA} = \arg\max_{W} \frac{W^T \cdot S_B \cdot W}{W^T \cdot S_W \cdot W} \tag{14}$$

3.4 Elastic Bunch Graph Matching

Elastic Bunch Graph Matching is an algorithm developed by Wiskott et al.[5].All human faces share a similar topological structure. EBGM represents faces as graphs, with the nodes positioned at fiducial points [19].
Labeled graph is a basic object representation form for EBGM algorithm. It is composed of nodes and edges, nodes are labeled with wavelet response coefficients bundled in feature vectors and edges are labeled with distance vectors between adjacent nodes. The node labels describe the image information locally, and the edge labels encode the topological relationship of local facial feature positions. Nodes of the labeled graph are usually defined at the points which are useful for recognition; these points are called feature points. A structure vector is a catenation of all the edge labels of a face, it

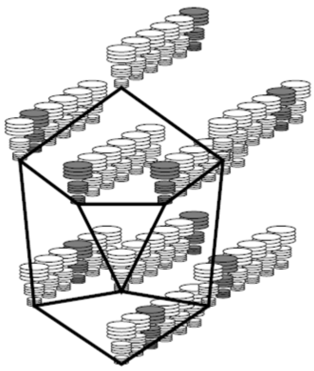

Figure 2: An example of face bunch graph[5]

describes the shape of the face[8].
Stored model graphs can be matched to new images to generate image graphs, which can then be incorporated into a gallery and become model graphs[5].
The assembly of some labeled graphs which are taken from different face images is called face graph bunch and it covers a wide range of possible variations of facial features, expressions, illuminations [8].
Elastic graph matching is performed by match a new face image to the face bunch graph. Suppose a labeled graph representing a face image consists of N nodes and E edges, the face bunch graph has M labeled graphs and the structure bunch has F structure vectors, its nodes are labeled with bunches of feature vectors $J_n^{B_m}$ where n=1,...,N , m =1,...,M, and edges are labeled with distance vectors

$$\Delta\vec{x}_e^{B_f} \tag{15}$$

(e=1,...,E , f=1,...,F) [8]. For a face image represented by a labeled graph I with nodes n and edges e, the matching cost function between the face image and face bunch graph can be defined as:

$$C(I,B) = \sum_n \max(S_a(J_n^I, J_n^{B_m})) - \lambda \sum_e D(\Delta\vec{x}_e^I, \Delta\vec{x}_e^{B_f}) \tag{16}$$

Where J_n^I is the feature vector of node n and $\Delta\overrightarrow{X_e^I}$ is the distance vectors of edges e in labeled graph I.

$$S_a(J_n^I, J_n^{B_m}) \tag{17}$$

is the similarity of two feature vectors.

$$D(\Delta\bar{x}_e^I, \Delta\bar{x}_e^{B_f}) \tag{18}$$

is the Euclidean distance of two distance vectors[8]. The aim of elastic graph matching is to seek the maximal value, which is a process of optimizing the matching of node labels and edge labels simultaneously. To find the feature points in a new image, a two-stage matching process is needed. The first matching stage is to find the best structure vector from the structure bunch, that is to find the shape of a face, by this way the approximate positions of feature points can be obtained. The second matching stage is realized by adjusting the position of every node in a pseudo-random sequence, in this stage the feature points can be located precisely[8]. After having located feature points on a face image, local facial features are extracted for face register or recognition. Recognition is performed by comparing the local facial features of a testing image to all trained images and selecting the one with the highest similarity value. The similarity function used for comparing two images is defined as:

$$S_G(G^I, G^M) = \sum_n S_a(J_n^I, J_n^M) \tag{19}$$

Where G^I is the testing image graph, G^M is the trained image graph.

$$S_a(J_n^I, J_n^M) \tag{20}$$

is the similarity of two feature vectors corresponding to node n in two images[8].

4 Application of algorithms in age estimation

Age estimation is one of the fields of computer vision. Our previous research regarding age estimation was to apply an anthropometric model to facial images and to find similarities in facial ratios of individuals of the same age. The basis of that and other similar research is to locate important facial landmarks. One of identifyed problems in that research was that facial landmarks were defined manually. In order to automate that process, an adequate algorithm needs to be chosen. As was seen in previous chapters, many algorithms have been developed and can be used for landmark localization and ratios calculation. The most adequate of analysed algorithms is Elastic Bunch Graph Matching algorithm because it sees a face as a graph and nodes of that graph are placed at important facial landmarks. In that way, by calculating distances between nodes of a graph, distances between facial landmarks are calculated, and that is a step towards calculation of facial ratios.

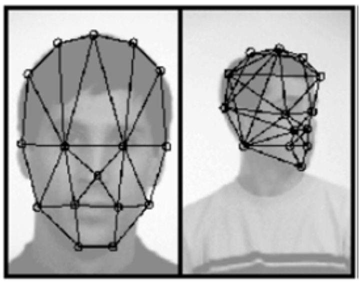

Figure 3: An example of connected facial landmarks[5]

5 Conclusion and future research

Extensive research has been done on the topic of face recognition, especially on the face recognition algorithms and methods. Some of the papers on face recognition algorithms have been described as a part of this paper. Next, the most commonly used algortihms were described and connected with age estimation. Future research will connect this research with previous research on age estimation using anthropometric model and concentrate on applying elastic bunch graph matching algorithm to a set of images and extracting face graphs from those images.

6 Acknowledgments

Shown results come out from the scientific project Methodology of biometrics characteristics evaluation (016-01611992-1721) and technological project Multiple biometric authentication using smart card

(2008-043) financed by the Ministry of Science, Education and Sport, Republic of Croatia.

References

[1] Belhumeur Peter N., Hespanha Joao P., Kriegman David J.:Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, European Conference on Computer Vision, 1996.

[2] Koruga Petra, Baca Miroslav, Schatten Markus: Analysis of Craniofacial Morphology Changes during Aging and their Connection with Facial Age Estimation, ITI conference, 2011

[3] Turk Matthew, Pentland Alex: Eigenfaces for Recognition, Journal of Cognitive Neuroscience, Volume 3, Number 1, 1991

[4] Etemad Kamran, Chellappa Rama: Discriminant Analysis for Recognition of Human Face Images, Optical Society of America, Volume 14, Number 8, 1997

[5] Wiskott Laurenz, Fellous Jean-Marc, Kruger Norbert, Malsburg Christoph: Face Recognition by Elastic Bunch Graph Matching, Intelligent Biometric Techniques in Fingerprint and Face Recognition, Chapter 11, pp.355-396, 1999

[6] Moghaddam Baback, Jebara Tony, Pentland Alex: Bayesian Face Recognition, Pattern Recognition, Vol. 33, No. 11, pps. 1771-1782, 2000

[7] Bartlett Marian Stewart: Face recognition by Independent Component Analysis, IEEE Transactions on Neural Networks, Volume 13, Number 6, 2002

[8] Li Yun-feng: A Face Recognition System Using Support Vector Machines and Elastic Graph Matching, International Conference on Artificial Intelligence and Computational Intelligence, 2009

[9] Chai Zhi, Ma Kai-Kuang, Liu Zhengguang: Complex Wavelet-based Face Recognition Using Independent Component Analysis, Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009

[10] Wei Li Xian, Sheng Yang, Qi Wang, Ming Li: Face Recognition Based on Wavelet Transform and PCA, Pacific-Asia Conference on Knowledge Engineering and Software Engineering, 2009

[11] S.Wei: A Shape Analysis in Copmuter Vision Final project report: Face recognition, Department of Electrical Engineering, McGill University, Canada, 1998

[12] Mane Arjun V., Kazi M.M., Manza Ramesh R., Kale Karbhari V.: Human Face Recognition Using Superior Principal Component Analysis

[13] Dinesh Kumar, C.S. Rai and Shakti Kumar: Face Recognition using Self-Organizing Map and Principal Component Analysis, 0-7803-9422-4/05, 2005 IEEE

[14] Karhunen J., Oja E. ,Wang L., Vigario R. , Joutsensalo J.: A class of neural networks for independent component analysis. IEEE Transactions. on Neural Networks, 1997

[15] P. Comon: Independent component analysis, a new concept? Signal Processing, 1994

[16] Liu Chengjun, Wechsler Harry: Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition, Second International Conference on Audio and Video-based Biometric Person Authentication, 1999

[17] Delac Kresimir, Grgic Mislav, Liatsis Panos: Appearance-based Statistical Methods for Face Recognition, 47th International Symposium ELMAR, 2005

[18] Martinez Aleix M., Kak Avinash C.: PCA versus LDA, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 23, Number 2, 2001

[19] Fazl-Ersi Ehsan, Zelek John S., Tsotsos John K.: Robust Face Recognition through Local Graph Matching, Journal of Multimedia, Volume 2, Number 5, 2007