[9] On the solution of the image recognition method of principal components and linear discriminant analysis

Mokeyev V.V., Tomilov S.V.
South Ural State University (National Research University)

Abstract

Some aspects of image analysis based on PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) are considered in the paper on the solution of the image recognition. The main idea of this approach is that, firstly, we project the face image from the original vector space to a face subspace via PCA; secondly, we use LDA to obtain a linear classifier. Efficiency of application of PCA and LDA to a problem of recognition of face images without their preliminary normalization is investigated in the paper. When the number of images per class is not large, it is proposed to supplement the training set by the images obtained by rotating, scaling and mirroring. On images of the ORL and Feret databases influence expansion of the training set on performance of unnormalized face images recognition is studied. Also a problem of increasing the high performance of principal component calculation for large image samples is considered. A linear condensation method is used as a new technique to calculate the principal components of a large matrix. The accuracy and high performance of the developed algorithm is evaluated.

Keywords: face recognition, principal component analysis, linear discriminant analysis, linear condensation method.

Citation: Mokeyev V.V. On the solution of the image recognition problem by a principal component method and linear discriminant analysis / Mokeyev V.V., Tomilov S.V. // Computer Optics. – 2014. – Vol. 38(4). – P. 871-880.

Introduction

Most image recognition systems are based on finite-element methods for constructing the lower-dimension feature space. One of the most common methods to reduce image dimensions is PCA (Principal Component Analysis) [1]. Many algorithms using PCA are currently offered to solve the search and face recognition problem.

The image recognition technique using finite-element methods involves two steps. The first step includes construction of a classifier using a training set of images. At the second step the recognition of unknown images using the constructed classifier is performed.

Different methods are used to construct the classifier; among them we may specify LDA (Linear Discriminant Analysis) [2,3]. LDA allows us to convert the original image space into the low-dimensional feature space in which images are grouped in classes around their centers, and class centers are removed from each other insofar as possible.

Papers [4,5] propose an approach for solving image recognition problems, which is based on PCA and LDA joint usage. The training set is formed for principal components calculation consisting of the images grouped in classes. Images of one class describe a face of one person. A class can contain dozens or even hundreds of images of the same face. First, we use PCA to reduce image dimensions and then LDA is applied to separate image classes.

Processing of preliminary images is usually performed during images recognition that leads them to a standard form (scaling, centering, background intercepting, brightness adjustment). It is also undesirable to have glasses, beards, facial expressions, etc. since during recognition the algorithm begins to response more sensitive, for example, to the existence of glasses than to interclass differences. Image normalization requires some extra computations; this is not critical at the step of constructing the classifier, but at the second stage it is not always acceptable in systems that can operate in real time. Besides, image normalization can lead in some cases to the loss of image informative value that may restrict the growth of recognition quality. Therefore, one of the tasks of interest is to develop face recognition systems without preliminary image processing (normalization). The paper considers the unnormalized face image recognition technique using PCA+LDA. One of the factors improving face recognition quality is the increase of a number of images per class. If the number of images per class is small, the training set may be expanded by the images obtained through mirroring, rotation and scaling of original face images. This method has little effect on recognition quality in the case of face image normalization since the normalization procedure aims at reducing such differences between images. However, in the case of unnormalized images such method may be useful when the training set contains a small number of images. The proposed unnormalized image recognition technique provides good recognition quality only when the number of images per class is large that increases time costs for constructing the classifier, but reduces costs for recognition of unknown images, since it does not require the normalization procedure.

Due to the fact that at the step of constructing the classifier the training set can reach large sizes, the computational complexity of principal components may substantially increase. Therefore the paper considers the problem of computational efficiency of principal components for lots of images. Papers [6-8] describe approaches reducing the computational complexity of principal components for large image sets. Algorithms of the two-dimensional principal component analysis are investigated in paper below [6]. Main components of two-dimensional PCA reflect differences between image rows and columns, and do not take into account differences between individual image points which can be meaningful in image recognition. The main advantage of two-dimensional PCA is that it can reduce matrix dimensions when computing the principal components. The principal component synthesis approach is based on partitioning the image set, obtaining particular solutions and synthesizing principal components from particular solutions [7]. The linear condensation method [8] uses matrix deflation in computing the principal components. Matrix deflation ideas for calculating the eigenvalues have shown high efficiency of solution of the generalized eigenvalue problem using frequency-dynamic [9] and frequency condensation [10] methods. The linear condensation method adapts these ideas to the standard eigenvalue problem. This paper offers to use a block-orthogonal condensation algorithm, which is considered to be the development of a multilevel linear condensation algorithm, to calculate the principal components of large sets of high-dimensional images [8].

1. Basic interrelationship between PCA and LDA

The method based on PCA and LDA consists of two steps: first, we project a facial image from the original features space into the facial eigenvalues subspace using PCA; then we use LDA to obtain the linear classifier. Let us suppose that there exists the set of images each of which can be described by the vector x_i ($i = 1,2,3,\ldots, m$), where m – is the number of different images in the training set. The dimension n of the vector x_i equals to the number of pixels in the image. Therefore all images can be represented as a matrix whose rows are the vectors x_i.

The average vector of training images $\nu = \frac{1}{m}\sum_i^m x_i$ is subtracted from each image in the training set. Thus, a new space $\mathbf{X}^0$ is obtained with the dimension of $m \times n$ whose rows are $x_i^0 = x_i - \nu$ vectors.

The PCA+LDA approach can be considered as a phased linear transformation of the original image space to the projection of a smaller-dimensional space. The first step includes the transformation which allows us to reduce the dimension of each image from n to p ($p < n$). The principal component method is a dimension reducing technique based on extracting the desired number of principal components from multi-dimensional data. The first principal component is a linear combination of original features (pixels) which has a maximum variance, and the i-principal component is the linear combination of original features (pixels) with the highest variance among $m-i+1$ principal components and orthogonal to $i-1$ first principal components.

It is known that the matrix $\mathbf{X}^0$ may be represented as the singular value decomposition.

$$\mathbf{X}^0 = \mathbf{U}_{PCA}\, \Lambda\, \mathbf{V}_{PCA}^T \quad (1)$$

where $\mathbf{U}_{PCA}$ ($m \times p$) and $\mathbf{V}_{PCA}$ ($n \times p$) – are matrices of left and right eigenvectors $\mathbf{X}^0$, Λ – is the diagonal matrix ($p \times p$) whose diagonal element s $\eta_1, \eta_2, \ldots, \eta_p$ are positive eigenvalues of the matrix $\mathbf{X}^0$. Here p – is the number of eigenvectors; usually p should not exceed the matrix range $\mathbf{X}^0$.

The key element of PCA is the calculation of the principal component matrix $\mathbf{V}_{PCA}$. The principal component matrix $\mathbf{V}_{PCA}$ is formed by the right eigenvectors $\mathbf{V}_{0i}^{PCA}$, which correspond to the largest eigenvalues.

The eigenvector matrix $\mathbf{V}_{PCA}$ may be determined as follows:

$$\mathbf{V}_{PCA}^T = \Lambda^{-1}\mathbf{U}_{PCA}^T\mathbf{X}^0 \quad (2)$$

The left eigenvector matrix $\mathbf{U}_{PCA}$ is formed from the eigenvectors of the following equation:

$$(\mathbf{A}-\lambda\mathbf{I})\mathbf{u}_0^{PCA}=0 \quad (3)$$

where $\mathbf{I}$ – is an identity matrix with the order m, $\mathbf{u}_0^{PCA}$ – is the eigenvector, and λ – is the eigenvalue, $\mathbf{A}=\frac{1}{m}(\mathbf{X}^0)^T\mathbf{X}^0$ – is a covariance matrix.

The main image components are calculated by the following formula:

$$\mathbf{Z}=\mathbf{X}^0\mathbf{V}_{PCA}$$

where $\mathbf{Z}$ — is the matrix ($m\times p$) of image principal components whose i – row represents the principal component vector of the i -image. We shall further denote the image principal component vector as z_i^k, where k — is a class number, i — is an image number.

Let us denote the mean vector of image principal components belonging to k class as ν_k, and the mean value of the principal components of all images as ν

$$\nu_k=\frac{1}{m_k}\sum_{i=1}^{m_k}z_i^k,\ \nu=\frac{1}{K}\sum_k^K\frac{1}{m_k}\sum_{i=1}^{m_k}z_i^k \quad (4)$$

Here K – is the class number, and m_k – is the number of facial images per class k.

The interclass differences matrix $\mathbf{A}_b$ can be calculated as

$$\mathbf{A}_\omega=\frac{1}{m}\sum_{k=1}^{K}\sum_{i=1}^{m_k}(z_i^k-\nu_k)(z_i^k-\nu_k)^T \quad (5)$$

The intraclass differences matrix is determined as

$$\mathbf{A}_b=\frac{1}{m}\sum_k^K m_k(\nu_k-\nu)(\nu_k-\nu)^T \quad (6)$$

LDA tries to find such transformation which would maximize the relationship between interclass and intraclass differences as shown below:

$$\mathbf{V}_{lda}=\arg\max_{\mathbf{V}\in R^{n\times r}}\frac{\left|\mathbf{V}^T\mathbf{A}_b\mathbf{V}\right|}{\left|\mathbf{V}^T\mathbf{A}_\omega\mathbf{V}\right|} \quad (7)$$

In order to determine $\mathbf{V}_{lda}$ the eigenvalue problem is to be solved

$$(\mathbf{A}_b-\lambda\,\mathbf{A}_\omega)\,v_0^{lda}=0 \quad (8)$$

Solution of the above equation (8) represents the generalized eigenvalue problem. Paper [11] shows that in order to solve the above equation (8) we should efficiently use the general Jacobi eigenvalue method.

Combining PCA and LDA we shall obtain the linear transformation matrix which first projects the image to the subspace of principal components $\mathbf{Z}$ and then to the classification space:

$$\mathbf{W}=\mathbf{V}_{lda}\mathbf{V}_{pca} \quad (9)$$

where $\mathbf{V}_{lda}$ – is the linear discriminant transformation in the principal components space. Image recognition shall be performed after this linear transformation in the discriminant components space using different metrics such as Euclidean distances.

2. Quality research of unnormalized facial recognition via PCA+LDA

The ORL database contains images of 40 persons, each of which is described with 10 different facial images with different facial expressions, perspectives and detailed silhouettes. All images in the database are in grayscale with 256 brightness gradations. The size of each image is 92×112 pixels. Fig. 1 shows examples of images of eight faces taken from the ORL database.

The size of facial images in the Feret (Facial Recognition Technology) database is 512×768 pixels [12]. However, in experiments we have used images of 128×128 pixels in size thus to reduce computational costs. Therefore each image was represented as the vector of 16384 in dimension. Facial images are extracted from the Feret database and describe 121 persons. Facial images are selected using the Viola-Jones' face detection algorithm, but they don't undergo the geometric normalization procedure. There are 10 images of various sizes for each person obtained in different light intensity at different shooting angles with improvised facial expressions. The total number of images in the training set is 1210. Figure 2 shows eight facial images taken from the Feret database.

The PCA+LDA approach is used for image recognition and includes two stages. Principal components of the training set are calculated at the first stage thus reducing the dimension of each image up to p of principal components. In order to form the principal components the eigenvectors of equation (1) shall be used with the largest eigenvalues. Then interclass and intraclass difference matrices ($p\times p$) are formed and $r\le p$ of discriminant components are calculated. The nearest class-center classifier shall be used in image recognition. The images not included in the training set are used as a test set.

The quality of facial image recognition is described with a recognition rate which is numerically equal to a proper percentage of correctly recognized images of the total amount of all presented images. The recognition rates K_{test} shall be determined for the test set.

The first part of experimental researches is carried out with the ORL database images. Experiments are performed to study how the number of images in the training set class can influence on the recognition quality. The first three series of experiments are conducted for training sets containing 2, 3 and 4 images per each class which are randomly selected. When using PCA+LDA methods for image recognition, the nearest neighborhood classifier shall be applied [11].

Fig. 1. Examples of facial images selected from the ORL database

Fig. 2. Examples of facial images selected from the Feret database

In experimental researches a cross-validation procedure is used which averages the recognition rates obtained while classifying the set of images into training and test sets. The training set images are randomly selected from the ORL database. All remaining images have formed the test set. Ten experiments are performed per each series.

In paper [13] experimental researches of normalized images from the ORL database have been performed using the following different methods: PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), DLA (Discriminative Locality Alignment [14]), NDLPP (Null Space Discriminant Locality Preserving Projections [15]), RLPDA (Regularized Locality Preserving Discriminant Analysis [13]).

Table 1 shows the recognition rates obtained using the aforementioned methods [13]. The same table shows the recognition rates obtained by the classifier constructed using PCA+LDA. Unnormalized face images have been used when constructing the classifier. The obtained results are in good agreement with experimental data proposed by other authors.

As Table 1 shows, the recognition rates obtained using PCA+LDA methods without image normalization have rather high values though they are inferior to such methods as NDLPP and RLPDA.

Table 1. Comparing the recognition rates of the ORL database

Method	Number of images per class		
	2	3	4
PCA [13]	69.6	78.6	83.6
LDA [13]	80.1	88.0	91.5
DLA	73.3	87.1	92.6
NDLPP	83.0	91.3	94.6
RLPDA	80.7	90. 4	94.8
PCA+LDA	82.2	90	94.2

The following series of experiments shall be conducted on training sets of the first three test series which are supplemented with the images obtained by rotating, scaling, and mirroring of original images of the training set. The obtained training sets are used to construct classifiers using PCA+ LDA.

Table 2 shows the recognition rates obtained using the constructed classifiers. The number of images per class resulted from expanding sets is given in brackets. The first row contains the results obtained by the training set supplemented with mirrored images. The number of images per class shall be doubled, and the recognition rates can also slightly be increased. The second row contains the results obtained by expanding the training set with the images decreased and increased by 5%. The number of images per class is increased three times, and the quality of facial recognition is significantly improved.

The third row contains the recognition rates obtained by expanding the training set with the images rotated by 4^{o} clockwise and counterclockwise. The number of images per class is increased three times.

The fourth table row gives the results obtained by expanding the training set with the images which are first mirrored and then rotated by 4^{o} clockwise and counterclockwise. The number of images per class in this case is increased 6 times, and the recognition rates are slightly higher than the results obtained by expanding the training set with only mirrored or only rotated images.

The fifth row contains the results obtained by the following expansion of the training set. First, we supplement the images obtained from original images through reducing and increasing by 5%. Then the expanded set shall be supplemented with the images obtained through rotating by 4^{o} clockwise and counterclockwise. The result is that the number of images per class is increased 9 times, and the recognition rates are considered to be the greatest.

Table 2. Comparing the recognition rates when expanding the training set of the ORL database

Training set expansion method	Number of original images per class		
	2	3	4
Mirroring	84.3 (4)	91.1 (6)	94.5 (8)
Scaling	85.6 (6)	96.4 (9)	95.8 (12)
Rotating	85.4 (6)	91.8 (9)	94.6 (12)
Mirroring and rotating	85.7 (12)	92.2 (18)	95.1(24)
Scaling and rotating	85.6(18)	96.4 (27)	96. 5 (36)
Mirroring, rotating and scaling	85.6 (36)	96. 3 (54)	94.1 (72)

The combination of image mirroring, scaling and rotating can lead to the maximum expansion of the training set. However, the classifier constructed on the basis of this set does not show the highest facial recognition quality.

The second part of experimental researches is performed on the image training set from the Feret database. The set includes 1210 facial images of 121 persons. There are 10 images of various sizes for each person obtained in different light intensity at different shooting angles with improvised facial expressions.

Experiments with varying number of facial images per class are performed. The first three series of experiments are conducted for training sets which contain 3, 4 and 5 images in each class of the Feret database. When performing researches, the cross-validation procedure is also used. The specified number of images selected from the training set is randomly selected from the Feret database. All remaining images have formed the test set. Ten experiments are performed in each series.

Paper [16] gives the results of investigating the accuracy of facial recognition performed in several databases including the Feret database. Facial images undergo the normalization procedure including centering, rotating, scaling, brightness alignment, etc. Different methods are used for image recognition, i.e. PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), NPE (Neighborhood Preserving Embedding [17]), MFA (Marginal Fisher Analysis [18]), LSDA (Locality Sensitive Discriminant Analysis [19]), RLPDE (Regularized Locality Preserving Discriminant Embedding [16]). Table 3 shows the recognition rates calculated by the following methods, i.e.: PCA, DLDA, NPE, MFA, LSDA, RLPDE and PCA+LDA. The recognition rates obtained by these methods are compared with the results of PCA+LDA without images normalization.

Table 3. Comparing the recognition rates of the Feret database

Method	Number of images per class		
	3	4	5
PCA	59.7	60	60
LDA	64	63.7	64
NPE	61,3	60	61.3
MFA	66.4	61.3	63.3
LSDA	64	63.6	65.6
RLPDE	75.4	75.4	75.4
PCA+LDA	52.6	59.3	65.6

As shown in the above table, increasing the number of images per class does not result to significant improvement of facial recognition quality if the following methods are used, i.e. PCA, LDA, NPE, MFA, LSDA, RLPDE. This may be due to the fact that the normalization procedure does not only eliminate facial differences, but can also distort facial images. For classifications constructed by means of PCA+LDA using unnormalized images, the facial recognition quality shall be considerably improved with increasing the number of images per class.

The next three series of experiments are performed for training sets of unnormalized images which are increased by supplementing images obtained through rotating, scaling and mirroring the original images. Ten experiments shall be performed per each series; the results thereof shall be averaged. Table 4 shows the recognition rates obtained by LDA+PCA classifiers constructed through expanded training sets. The number of images per class resulted from expanding the set is given in brackets.

The table shows the results which have been obtained when the training set has been expanded by mirrored images (the first row), by the images increased and decreased by 5% (the second row), and by the images rotated by 4^o clockwise and counterclockwise (the third row).

The fourth row gives the results obtained by expanding the training set with the images which are first increased and decreased by 5% and then all image of the expanded training set are rotated by 4^o clockwise and counterclockwise. As can be seen from the table below, the classifications constructed on such training set can provide the best recognition results.

Table 4. Comparing the recognition rates when expanding the training set of the Feret database

Training set expansion method	Number of images per class		
	3	4	5
Mirroring	58.2 (6)	64.2 (8)	70 (10)
Scaling	64.4 (9)	70.5 (12)	75.3 (15)
Rotating	64 (9)	70 (12)	72 (15)
Scaling and rotating	65.7 (27)	73 (36)	76 (45)
Mirroring and rotating	65.5 (18)	69.9(24)	74.5 (30)
Scaling, rotating and mirroring	67.7 (54)	72.8 (72)	75.6 (90)

The combination of image rotating by 4° clockwise and counterclockwise and mirroring of all obtained images allows us to get the results shown in the fifth row of the table.

The combination of all three ways of expansion of the training set results to the maximum increase of the number of images per class, however the classifier constructed therethrough cannot demonstrate the best results.

As the above table shows, the expansion of the training set by means of derivative images obtained through rotating and scaling of original images shows a steady rise of the recognition rate. Images mirroring along with their scaling and rotating does not improve the recognition quality. This may be due to the fact that such extension of the training set may result to the growth of similar images that reduces the information content of the testing set and the quality of classifiers constructed therethrough.

It should be noted that expansion of the image training set obtained by rotating, scaling and mirroring can be reasonable only in case of lack of a large number of original images.

Thus taking the aforesaid into consideration, it may be said that increase of the number of images per class allows us to construct rather high-quality classifications based on PCA+LDA methods with no image normalization procedure applied.

The increasing number of images per classes results to increase of computational costs when constructing the classifier, which is related to the fact that the computational complexity of the principal components shall significantly rise with increasing number of images in the training set.

3. Calculation of the principal components of large image sets

The main computational complexity of the principal components is related to computation of the eigenvectors in equation (3). If there are many images in the training set, the matrix order in (3) becomes larger and considerable computational resources shall be required to compute the principal components.

We can select the following groups of methods among the eigenvalue problem-solving techniques: iterative similarity transformation methods [20]. The group of iterative methods includes the power method, the Lanczos method, etc. The main disadvantage of iterative methods is that the convergence rate of solutions can affect the ratio of the desired eigenvalue to the nearest eigenvalue.

Similarity transformation methods are applied to obtain from the original matrix a new matrix with the same eigenvalues but of a simpler form. The most well-known methods are Jacobi, Givens and Householder methods. The Householder method allows us to obtain the desired result faster than the Givens method since it is related to performing fewer but more complex transformations. Matrix deflation methods (condensation methods) refer to the group of similarity transformations methods since their purpose is to obtain a lower-order matrix which would be similar to the original matrix in the sense that the eigenvalues of these matrices within the given range would coincide with the given accuracy.

Due to the fact that the eigenvalues are close enough to each other, especially at the top of a spectrum, iterative methods are ineffective when calculating eigenvectors of large matrices. In paper [8] it is shown that the Householder method demonstrates a higher rate than the power method.

The effective way to solve a noncomplete problem of the eigenvalues for large matrices is the matrix deflation. Since the number of required eigenvectors is usually much less than the matrix order, this approach is particularly effective for large matrices. The frequency-dynamic condensation method is proposed in paper [9], which reduces the matrix order when solving the generalized eigenvalue problem. The method has been thereafter widely used to solve various engineering analysis problems [10]. The idea of the frequency-dynamic condensation method has provided background for the linear condensation method which, when solving the standard eigenvalue problem, reduces the matrix order while maintaining its eigenvalues within the given interval [8].

Let us introduce the equation (3) in the following form:

$$(\mathbf{I}-\mu\mathbf{A})u_0^{pca}=0 \quad (10)$$

where $\mu=1/\lambda$. Then the problem shall be formulated as follows: to determine the smallest eigenvalues and their corresponding eigenvectors of the equation (10). Paper [8] describes the multilevel linear condensation algorithm for calculating the eigenvalues within the interval from 0 up to μ_2 and their corresponding eigenvectors. The algorithm includes five steps. The first step represents a multilevel matrix deflation procedure which starts with the fact that all elements of the vector u_0^{pca}, which will be further called the features, shall be sorted in descending order of diagonal coefficients of the covariance matrix $\mathbf{A}$. A group of features with minimal diagonal coefficients is to be selected at each level of the matrix deflation procedure. The decision to exclude the selected features shall be made if the following condition has been fulfilled:

$$\mu_{\min} > k_c \mu_2 \qquad (11)$$

Here $\mu_{\min} = 1/\lambda_{\max}$ – is the smallest eigenvalue of a block of variables $\mathbf{A}_{ss}^k$ to be deleted, and k_c is called the intercept parameter. If the condition (11) has been fulfilled, the matrix deflation shall be performed. The condition (11) can be rewritten as $\lambda_{\max} < \lambda_2 / \kappa_c$.
As is known, the sum of matrix diagonal coefficients is equal to the sum of the eigenvalues λ. Therefore the set of the features to be deleted is formed from the features which correspond to the smallest diagonal coefficients of the matrix $\mathbf{A}$. In fact, it is assumed that the smaller the sum of the eigenvalues λ, the smaller the maximum eigenvalue of the matrix being analyzed. In matrix deflation we will consecutively obtain the matrices

$$\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k = \begin{bmatrix} \mathbf{A}_{bb}^k & \mathbf{A}_{bs} \\ \mathbf{A}_{sb}^k & \mathbf{A}_{ss}^k \end{bmatrix},$$

where k – is the deflation level of the matrix $\mathbf{A}$.
The effectiveness of the multilevel linear condensation algorithm depends on how much the matrix order is deflated, however the deflation degree is not always high enough. Paper [8] demonstrates performance of the multilevel linear condensation algorithm via examples of calculation of 67 principal components of sets with different dimensions (from 500 to 5000). However further researches have shown that in calculation of the large number of principal components (several hundred) we fail to considerably deflate the matrix order. This is because the eigenvalues near the upper boundary μ_2 are very tight and we fail to select the features to be excluded without violating the above condition (11).
Therefore, the multilevel linear condensation algorithm ceases to be effective. In order to speed up calculation of the eigenvectors it is suggested to use the block-orthogonal condensation algorithm.

The block-orthogonal condensation algorithm is based on the multilevel matrix deflation procedure. However the block of features (candidates to be deleted) at each level shall be reduced to a diagonal form using the orthogonal transformation.
Let the equation (10) in the k-level of the matrix deflation has the following form:

$$(\mathbf{I}_k - \mu \mathbf{A}_k) u_{k-1}^{pca} = 0 \qquad (12)$$

We represent the matrix $\mathbf{A}_k$ and the vector u_{k-1}^{pca} in the following form:

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{A}_{bb}^k & \mathbf{A}_{bs} \\ \mathbf{A}_{sb}^k & \mathbf{A}_{ss}^k \end{bmatrix}, \quad u_{k-1}^{pca} = \begin{Bmatrix} u_b^{k-1} \\ u_s^{k-1} \end{Bmatrix}.$$

Here the index b refers to the features to be held, and the index s refers to the features to be deleted.
Matrix diagonalization of the features to be deleted shall be performed by using the orthogonal transformation

$$\mathbf{P}_s^T \mathbf{A}_s^k \mathbf{P}_s = \acute{\mathbf{O}}_s \qquad (13)$$

The matrix $\mathbf{P}_s$ is the orthogonal matrix which is composed of the eigenvectors of the matrix $\mathbf{A}_{ss}^k$ obtained when the following equation has been solved:

$$(\mathbf{I}_s - \mu \mathbf{A}_s^k) p_s = 0 \qquad (14)$$

Thus,

$$u_k^{pca} = \begin{bmatrix} \mathbf{I}_b & 0 \\ 0 & \mathbf{P}_s \end{bmatrix} \begin{Bmatrix} u_b^{k-1} \\ u_s^{k-1} \end{Bmatrix} = \mathbf{P} u_{k-1}^{pca} \qquad (15)$$

Let us note that $\mathbf{P}_s$ is the orthogonal matrix, so the following relation is meaningful:

$$\mathbf{P}_s^T \mathbf{I}_s \mathbf{P}_s = \mathbf{P}_s^T \mathbf{P}_s = \mathbf{I}_s \qquad (16)$$

Substituting equation (15) to equation (12) and multiplying on the right by the matrix $\mathbf{P}^T$ with regard to (16), we receive

$$(\mathbf{I}_k - \mu \mathbf{A}_k^*) u_k^{pca} = 0 \qquad (17)$$

where $\mathbf{A}_k^* = \mathbf{P}^T \mathbf{A}_k \mathbf{P}$.
In view of set dividing and the aforementioned equation (13), the matrix $\mathbf{A}^k$ may be represented as

$$\mathbf{A}_k^* = \begin{bmatrix} \mathbf{I}_b & 0 \\ 0 & \mathbf{P}_s \end{bmatrix}^T \begin{bmatrix} \mathbf{A}_{bb}^k & \mathbf{A}_{bs} \\ \mathbf{A}_{sb}^k & \mathbf{A}_{ss}^k \end{bmatrix} \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & \mathbf{P}_s \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{bb}^k & \mathbf{A}_{bs}^k \mathbf{P}_s \\ \mathbf{P}_s^T \mathbf{A}_{sb}^k & \Sigma_{ss} \end{bmatrix} \qquad (18)$$

If diagonal coefficients of the matrix Σ_{ss} are arranged in descending order, the reciprocal value of the first diagonal coefficient of the matrix Σ_{ss} will be equal to the smallest eigenvalue of the block of features to be deleted ($\mu_{\min}$). Testing of the condition (11) allows determine whether it is possible to remove features included in the block of features to be deleted. If it is impossible to delete features, the condition (11) shall be checked for the second diagonal coefficient. In fact, this means that we reduce the set of features being deleted per unit and make a decision to delete a reduced block of features. If the condition (11) has not been fulfilled, we again reduce the block of features per unit, i.e. the condition (11) is to be checked up to the third diagonal coefficient of the matrix Σ_{ss}, etc. If the condition (11) has been fulfilled, it means that we have identified the block features which can be deleted. If the condition (11) has not been fulfilled, it means that the deflation process has been completed.

The remaining steps correspond to the steps of the multilevel linear condensation algorithm [8]. The only difference that occurs at the stage of recovery of the eigenvectors relates to introduction of orthogonal transformations (15) when calculating values of the features to be deleted.

The block-orthogonal condensation algorithm along with the multilevel linear condensation algorithm provides an approximate solution of the eigenvalues problem. The point is that the eigenvectors of the equation (4) form the orthogonal basis $m \times m$, while the eigenvectors computed by the block-orthogonal condensation algorithm constitute the orthogonal basis with the dimension of $m \times p$. Thus, the eigenvectors error calculated by the block-orthogonal condensation algorithm is related to the fact that they are not orthogonal eigenvectors starting with $p+1$ and more. The error of the solutions to be obtained may be reduced by using the intercept parameter. Increasing the intercept parameter we reduce the eigenvectors error. However, the degree of matrix contraction is herewith reduced and the solution time is increased.

It is offered to evaluate the eigenvectors error calculated by the block-orthogonal condensation algorithm using the following formula:

$$e_i = \frac{\sum_{k=1}^{N} (\mathrm{v}_{0ik} - \mathrm{v}_{0ik}^*)^2}{\sum_{k=1}^{N} \mathrm{v}_{0ik}^{\;2}},$$

where v_{0i}^* – is the i -eigenvector calculated by the block-orthogonal condensation, v_{0i} – is the i -eigenvector obtained using the Householder method.

The eigenvectors computational accuracy is evaluated based on the ORL database. The eigenvectors are calculated using the block-orthogonal condensation algorithm and the Householder method. The eigenvectors computational accuracy using the block-orthogonal condensation algorithm, as shown above, depends on the intercept parameter. Table 5 shows how the eigenvectors computational error may vary depending on the intercept parameter.

Table 5. The eigenvectors computational error depending on the intercept parameter

Intercept parameter	Eigenvectors error, %	
	80 PC	200 PC
1.1	75.06	82.5
1.5	20.98	26.7
2	3.12	2.56
2.5	1.02	1.32
3	0.29	0.46

As Table 5 shows, the more is the eigenvectors computational error, the less the intercept parameter. However, for any value of the intercept parameter the eigenvectors calculated by the block-orthogonal condensation algorithm represent the set of orthogonal vectors, therefore they may be used as the principal components.

Efficiency evaluation of the principal components calculated by the block-orthogonal condensation algorithm to solve the facial image recognition problem shall be performed on the ORL database. Images stored in the database are divided into training and test sets. The cross-validation procedure is used to investigate the facial image recognition which averages the recognition rates obtained from different training sets. The training set shall be formed from 8 images of each class of the ORL database which are randomly selected. All remaining images shall draw the test set. The nearest neighbor classifier is used for facial recognition and provides higher values of the recognition rate if compared to the nearest class-center classifier [11].

So, 10 experiments shall be performed with the principal components obtained by the Householder method, and 10 experiments shall be performed with the principal components calculated by the block-orthogonal condensation algorithm at different values of the intercept parameter. The recognition quality for each test is evaluated according to the recognition rate of the test set. Based on the processed experimental results we obtain the mean values K_{mn} and the root-mean-square deviations K_{msd} of recognition rates of the test set.

Table 6 shows the results obtained using the principal components calculated by the block-orthogonal condensation algorithm (the intercept parameter is equal to 1.1, 2 and 3). The results obtained by the Householder method fully coincide with the recognition rates calculated by the block-orthogonal condensation algorithm with the intercept parameter equaled to 3.

As Table 6 shows, despite the fact that the eigenvectors computational error is the greatest when the intercept parameter is 1.1, their usage as principal components allows to obtain even higher recognition rate.

Table 6. Mean values and root-mean-square deviations of recognition rates of the test set

Number of principal components	Intercept parameter					
	1.1		2		3	
	K_{mn}	K_{msd}	K_{mn}	K_{msd}	K_{mn}	K_{msd}
24	95.5	0.87	95.2	0.98	95.2	0.98
26	96.1	0.92	96.1	0.92	96.1	0.92
28	97.2	1.15	97.1	1.02	97.1	1.02
30	97.2	0.79	96.8	0.65	96.7	0.87
32	97.3	0.40	97.2	0.98	97.1	0.84
34	97.2	0.79	97.2	0.79	97.2	0.79
36	97.6	0.92	97.6	0.92	97.6	0.92
38	97.6	0.71	97.5	0.83	97.5	0.83
40	97.6	0.92	97.6	0.92	97.6	0.92
42	98.1	0.66	98.0	0.87	98.0	0.87
44	98.0	0.65	97.8	0.84	98.0	0.64
46	97.6	0.71	97.3	0.92	97.3	0.92
48	97.7	0.79	97.38	0.922	97.3	0.92
50	97.6	0.71	97.75	0.791	97.6	0.92

In some experiments the principal components calculated by the block-orthogonal condensation algorithm with the intercept parameter 1.1 demonstrate a better recognition quality that indicates the prospects of this approach for calculating the principal components.

Performance of the block-orthogonal condensation algorithm is investigated by comparison of the computation time of the principal components of image sets stored in the Feret database. By the computation time we mean the calculating time for the principal components u_0^{pca} by solving the equation (3) using PC.

To demonstrate the performance of the block-orthogonal condensation algorithm we evaluate 290 principal component matrices of various sizes, and the results are compared with the results obtained by the Householder method.

Table 7 shows how the computation time of 290 matrix principal components of various dimensions may vary. The results obtained by the Householder method, the multilevel linear condensation and block-orthogonal condensation algorithms are presented herein for reference.

As seen from the table, the block-orthogonal condensation algorithm exceeds in its performance not only the Householder method but also the multilevel linear condensation algorithm.

Table 7. Relative computational time for principal components depending on matrices' dimensions

Matrix order	Household method	Multilevel linear condensation	Block-orthogonal condensation
1210	61	12	12
4080	4742	776	477
4832	9000	1283	879
5793	19103	2298	1122
6899	35411	3467	1320

Conclusion

Some aspects of the Principal Component Analysis and the Linear Discriminant Analysis have been considered to solve the facial recognition problem. The classifier construction technique has been offered using the training set of unnormalized images that reduces the costs at the stage of recognition of unknown images through eliminating the normalization procedure.

The quality research of unnormalized facial images has been performed depending on the number of samples per class according to the ORL and FERET databases.

The performed researches have shown that the usage of unnormalized facial images in construction of classifiers using PCA+LDA is a promising direction in construction of facial recognition systems, since it allows to reduce recognition costs of new images due to partial or complete cancellation of the facial normalization procedure. A good quality of facial recognition by unnormalized images can be achieved by increasing the number of images per class. If the number of images per class is not large, the training set may be supplemented with images obtained through their rotating, scaling and mirroring.

The problem of increasing the computational efficiency of the principal components of large image sets has also been considered hereby. In this case it is proposed to use the block-orthogonal condensation algorithm for calculating the principal components. It is shown that this algorithm allows significantly reduce the computational complexity of the principal components.

References

1. Kirby, M. Application of the KL procedure for the characterization of human faces / M. Kirby, L. Sirovich // IEEE Trans. Pattern Anal. Mach. Intell. – 1990. – Vol. 12(1). – P. 103-108.

2. Etemad, K. Discriminant Analysis for Recognition of Human Face Images / K. Etemad, R. Chellappa // Journal of the Optical Society of America A. – 1997. – Vol. 14(8). – P. 1724-1733.

3. Lu, J. Face Recognition Using LDA-Based Algorithms / J. Lu, K.N. Plataniotis, A.N. Venelsanopoulos // IEEE Trans, on Neural Networks. 2003. – Vol. 14(1). – P. 195-200.
4. Belhumeur, P.N. Eigenfaces vs Fisherfaces: recognition using class specific linear projection / P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman // IEEE Trans. Pattern Anal. Mach. Intell. – 1997. – Vol. 19. – P. 711-720.
5. Martinez, A.M. PCA versus LDA / A.M. Marline, A.C. Kak // IEEE Trans, on Pattern Analysis and Machine Intelligence. – 2001. – Vol. 23(2). – P. 228-233.
6. Kukharev, G.A. Algorithms of two-dimensional principal component analysis for face recognition / G.A. Kukharev, N.L. Shchegoleva // Computer Optics. – 2010. – Vol. 34(4). – P. 545-551.
7. Mokeev, A.V. On accuracy and performance principal component synthesis method / A.V. Mokeev // Business Informatics. – 2010. – 3(18). – P. 64-68.
8. Mokeyev, V.V. On effectiveness increase of principal components computation in image analysis problem / V.V. Mokeyev // Digital Signal Processing. – 2011. – Vol. 4. – P. 29-36.
9. Grinenko, N.I. Problems of studying vibrations of structures by the finite-element method / N.I. Grinenko, V.V. Mokeev // International Applied Mechanics. – 1985. – Vol. 21(3). – P. 25-30.
10. Mokeev, V.V. On the problem of finding the eigenvalues and eigenvectors of large matrix system, arising in use a finite element method / V.V. Mokeev // Computational Mathematics and Mathematical Physics. – 1992. – Vol. 32(10). P. 1652-1657.
11. Mokeyev, V.V. On solution of small sample size problem with linear discriminant analysis in face recognition / V.V. Mokeyev, S.V. Tomilov // Business Informatics. – 2013. – Vol. 1(18). – P. 37-43.
12. Phillips, P.J. The FERET evaluation methodology for face recognition algorithms / P.J. Phillips, H. Moon, P.J. Rauss, S. Rizvi // IEEE Trans. Pattern Anal. Mach. Intell. – 2000. – Vol. 22(10). – P. 1090-1104
13. Gu, X. Regularized locality preserving discriminant analysis for face recognition / X. Gu, W. Gong, L. Yang // Neurocomputing. – 2011. – Vol. 74. –P. 3036-3042.
14. Zhang, T. Patch alignment for dimensionality reduction / T. Zhang, D. Tao, X. Li, et al. // IEEE Transaction on Knowledge and Data Engineering. – 2009. – 21(9). – P. 1299-1313.
15. Yang, L. Null space discriminative locality preserving projections for space recognition / T. Zhang, Gong X. Gu, et al. // Neurocomputing. – 2008. – Vol. 71. – P. 2045-2054.
16. Pang, Y.H. Regularized locality preserving discriminant embedding for face recognition / Y.H. Pang, Teoh A.B.J., Abas F.S. // Neurocomputing. – 2012. – Vol. 77. –P. 156-166.
17. Stan, Z. Li Non linear mapping from multi-view face patterns to a Gaussian distribution in a low dimensional space / Stan Z. Li, Xiao Rong, Yu L. Z., Zhang HongJiang // in: Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Face sand Gestures in Real-Time Systems. – 2001. – P. 47–54.
18. Yan, S.C. Graph embedding and extensions: a general framework for dimensionality reduction / S.C. Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, S. Lin // IEEE Trans. Pattern Anal. Mach. Intell. – 2007. – Vol. 29(1). – P. 40-51.
19. Cai, D. Locality sensitive discriminant analysis / D. Cai, X. He, K. Zhou, J. Han, H. Bao // in: Proceedings of the International Joint Conference Artificial Intelligence. – 2007. – P. 708-713.
20. Parlett, B.N. The symmetric eigenvalue problem. NJ, Englewood Cliffs: Prentice-Hall, 1980.

COMPUTER OPTICS