Stanisław OSOWSKI[1,2], SIKORSKA-ŁUKASIEWICZ[2]

[1]Warsaw University of Technology, Institute of the Theory of Electrical Engineering, Measurement and Information Systems
[2]Military University of Technology, Institute of Electronic Systems

PCA transformation and Support Vector Machine for recognition of the noisy images

***Abstract**. The paper presents the application of principal component analysis (PCA) and Support Vector Machine (SVM) for recognition of face images. PCA is a well known method of optimal reduction of the dimensionality of the vectors, while preserving the most important part of the original information. It maps the N-dimensional original vector **x** into K-dimensional output vector **y**, where K<N through the transformation matrix **W**. The most important PCA components represent the original image by the most important features, hence this representation is well suited for noisy images. The transformed vector **y** contains the entries forming the diagnostic features used as the input signals to the SVM neural network, working as the classifier. The paper will present the results of recognition of human faces taken from the FERET data base.*

***Streszczenie** Praca przedstawia zastosowanie transformacji PCA oraz sieci neuronowej SVM do rozpoznawania obrazów twarzy, w szczególności zaszumionych. PCA jest transformacją liniową umożliwiającą optymalną redukcję wymiaru wektora **x** przy zachowaniu najważniejszej porcji oryginalnej informacji zawartej w tym wektorze. Przy wektorach z dużą zawartością szumu odtworzona informacja jest w dużej mierze odszumiona (szum odpowiada najmniej istotnej części informacji, która podlega obcięciu). Metoda ta dobrze nadaje się do tworzenia cech diagnostycznych, które mogą stanowić sygnały wejściowe dla klasyfikatora SVM dokonującego rozpoznania obrazów. W pracy pokazane zostały wyniki przeprowadzonych eksperymentów rozpoznania twarzy z bazy danych FERET. (**Rozpoznawanie zaszumionych obrazów z użyciem PCA i sieci SVM**)*

Keywords: principal component analysis, Support Vector Machine, diagnostic features, image recognition.
Słowa kluczowe: analiza składników głównych, maszyna wektorów nośnych, cechy diagnostyczne, rozpoznawanie obrazów

Introduction

Recognition of images, especially distorted ones, belongs to difficult numerical problems, especially when the distortions are of high level [1,2,3]. In this paper we will present the Principal Component Analysis (PCA) based technique combined with Support Vector Machine (SVM) for recognition of the images. PCA is a well known linear transformation enhancing the most important elements of the data and suppressing the least important portion of the information. Hence it is well suited for distorted data, where the noise corresponds to the minor principal components, which are rejected in reduction of the dimensionality of data.

The most important principal components represent the diagnostic features, that may be applied as the input signal to the final system performing the recognition and classification task. In this paper we use Support Vector Machine as the final classifier.

The numerical experiments have been performed for recognition of the six human faces taken from FERET data base [3]. The original images have been used in learning the SVM classifier. The testing set was formed from the distorted images, at random noise of 'salt&pepper' and Gaussian distribution. The results of these experiments have shown, that PCA based method of creation of diagnostic features and SVM as a classifier form the robust basis for recognition of the noisy images.

Fundamentals of PCA

Let $\mathbf{x} = [x_1, x_2, ..., x_N]^T$ denotes the random vector of zero mean value and $\mathbf{R}_{xx}$ the auto-covariance matrix of this set, i.e., $\mathbf{R}_{xx}=E[\mathbf{xx}^T]$. This matrix is symmetric and non-negative of the real eigen-values $\lambda_i \geq 0$. Let us settle the eigen-values in decreasing order starting from the highest one. In similar way we order the orthogonal eigen-vectors $\mathbf{v}_i$ associated with them. The eigenvalues decomposition allows to present the correlation matrix in the form

$$\mathbf{R}_{xx} = \sum_{k=1}^{N} \lambda_i \mathbf{v}_i \mathbf{v}_i^T \quad (1)$$

The impact of each eigen-vector $\mathbf{v}_i$ to the reconstruction of the correlation matrix $\mathbf{R}_{xx}$ is measured by the value of the corresponding eigen-value λ_i. In practice it is enough to consider only *K* first highest eigen-values, rejecting the rest. Usually at high values of N we have $K << N$. On the basis of the considered (ordered) eigen-vectors we can define the PCA matrix **W**

$$\mathbf{W} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K]^T \quad (2)$$

This matrix determines the PCA transformation defined by the linear relation

$$\mathbf{y} = \mathbf{Wx} \quad (3)$$

The quality of representation of the vector **x** by the reduced dimension vector **y** might be estimated on the basis of the reconstructed vector **x** (with the hat) defined by the relation

$$\hat{\mathbf{x}} = \mathbf{W}^T \mathbf{y} \quad (4)$$

It was proved that the PCA transformation minimizes the expected value of the reconstruction error $E_r = E\left[\|\mathbf{x} - \hat{\mathbf{x}}\|^2\right]$ at the assumed K, which is described by the relation

$$E_r = \sum_{i=K+1}^{N} \lambda_i \quad (5)$$

Fig. 1 presents the graphical scheme of PCA transformation. The output signals y_1, y_2, y_K represent the principal components, characterizing the original vector **x**. The weights forming $\mathbf{v}_1$, $\mathbf{v}_2$ …, $\mathbf{v}_K$ represent the rows of the matrix **W**.

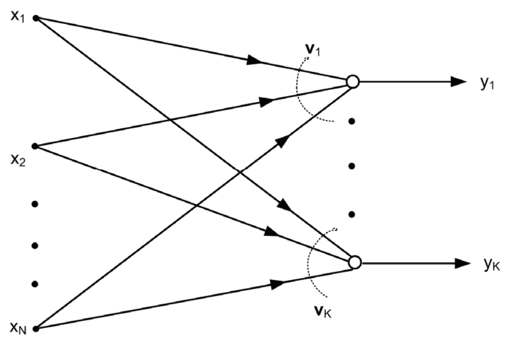


Fig. 1 The graphical scheme of PCA transformation of the original vector **x** into vector **y**

PCA transformation of the images

In our work we will consider the set of images belonging to *M* classes. Each image is represented the *N*-dimensional vector (the vectorial form of the column-wise or row-wise rearranged image matrix). For typical 512x512 images the size of this vector $N=512^2=262144$ is extremely high and it will be very difficult (practically impossible) to form and process the correlation matrix of such size even in modern computer systems. Let us assume that the number of image vectors is p, where usually $p<<N$. They are arranged in the form of the matrix **X** of the size pxN. To avoid the problem of processing the NxN dimension correlation matrix $\mathbf{R}_{xx}$ we form the small dimension matrix $\mathbf{RS}_{xx}$ of the size pxp, i.e., $\mathbf{RS}_{xx}=\frac{1}{p}\mathbf{X}\mathbf{X}^T$. The true PCA of real (very large) dimension is created on the basis of this small dimension matrix using the eigen-decomposition of $\mathbf{RS}_{xx}$ (equation 2). This decomposition generates the set of small size eigen-vectors $\mathbf{vs}_1$, $\mathbf{vs}_2$,…$\mathbf{vs}_p$. In our application the PCA matrix (the small one) is built on the basis of K ($K<p$) eigenvectors $\mathbf{vs}_i$ associated with K largest eigen-values λ_i, forming the matrix **VS.** The return to the normal (high) size of these vectors is achieved by the transformation

(6) $$\mathbf{V}=\mathbf{X}^T*\mathbf{VS}$$

where **V** is the matrix representing the original (high size) eigen-vectors (arranged column-wise). Then the final PCA matrix **W** is determined as follows $\mathbf{W}=\left[\mathbf{V}_1,\mathbf{V}_2,...,\mathbf{V}_K\right]^T$, where $\mathbf{V}_i$ represent the first succeeding columns of the matrix **V** (up to K). Thanks to this approach we avoid the problem of processing very high dimensional matrices in eigen-value decomposition.

The PCA decomposition in our solution will be used to generate the numerical descriptors of the image. In this approach the original image described by the vector **x** is transformed to the less dimensional vector **y** using equation (3). The elements of vector **y** form the input signals to the classifier performing the role of pattern recognition. In our solution we use the Support Vector Machine of the linear kernel as the working horse of the classification system. SVM is known from its high efficiency in classification tasks.

Support Vector Machine classifier

The SVM is a feedforward network of one hidden layer (the kernel function layer). It is known as an excellent classifier of good generalization ability [4,5]. The learning problem of SVM is formulated as the task of separating the learning vectors into two classes of the destination values either $d_i=1$ (one class) or $d_i=-1$ (the opposite class), with the maximal separation margin. The separation margin, formed in the learning stage, provides some immunity of this classifier to the noise, inevitably contained in the real data under testing [5].

The great advantage of SVM is the unique formulation of learning problem leading to the quadratic programming with linear constraints, which is very easy to solve. The SVM of the linear kernel has been used in our application. The hyperparameter (the regularization constant C) has been adjusted by repeating the learning experiments for the set of their predefined values and choosing the best one at the validation data sets. The experiments have shown that this classifier is relatively insensitive to the choice of C value (it was enough to set C>10).

To deal with a problem of many classes we have applied the strategy "one against one" [4]. In this approach the few SVM networks are trained to recognize between all combinations of two classes of data. For M classes we have to train M(M-1)/2 individual SVM networks. In the retrieval mode the input vector belongs to the class of the highest number of winnings in all combinations of classes.

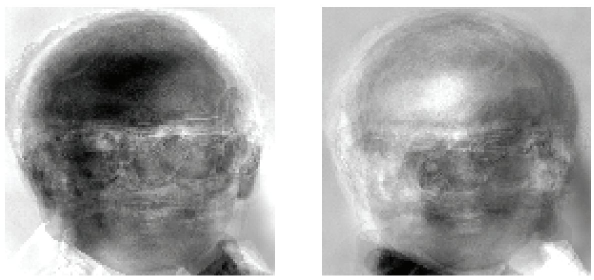
Results of numerical experiments

In the experimental part of this paper we will present the results of recognition of six classes of images. The experiments will be aimed on the recognition of the distorted representatives of these classes at different distortion level and various types of distortion ('salt&paper', and Gaussian). In the experiments we will use 6 classes of human faces taken from FERET data base [3]. All of them correspond to males and represent the faces in different positions. Fig. 2 depicts the few exemplary representatives of one class at different pose (the first row) and 5 images representing the other 5 classes (the second row).

Fig. 2 The representatives of 6 different face images used in the recognition experiments

The first row of this figure represents the image of one chosen person in different poses. In the second row the individual representatives of the other persons taking part in experiments are depicted. Each of these persons has been also photographed in similar poses as the first one. Each class was represented by the images taken in different positions (frontal, quarter and half right, and quarter and half left, etc.). Six persons photographed in different positions created six classes of faces under recognition. The size of each image was 128 x128.

All images have been preprocessed using PCA according to the described above procedure. The introductory experiments have shown that representation of each image by 18 principal components is satisfactory for the recognition task. Fig. 3 presents two most important eigen-faces (the eigen-vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ converted to format of the image). They resemble the ghost-like images containing the transformed features of all 30 images taking part in formulation of PCA. However closer analysis of both images proves that the first eigen-face is a bit more informative; however, both present different look on details of all original images.



a) b)

Fig. 3 Two most important eigen-faces generated by PCA transformation: a) the first eigen-face, b) the second eigen-face

The aim of our analysis is to find how stable is the recognition process, when the recognized images are noisy. In creating the distorted versions of the images we have applied two kinds of random noise. One of them was 'salt&pepper' and the second the noise of Gaussian distribution. Each original image was multiplicated by applying different measures of randomness. In the case of random Gaussian noise the applied variance was either 0.1, 0.05 or 0.02. In the case of 'salt&pepper' noise the original pixel was substituted by either 1 or 0. At noise density equal D adding 'salt&pepper' to the original image affects approximately the number of pixels equal D*N, where N is the total number of pixels in the image. We have applied different values of D, for example D=0.5, D=0.3 and D=0.2. In this way each time we have created 30x4=120 noisy images, that will form the testing sets. Fig. 4 presents the exemplary images corresponding to one person, affected by four types of distortion.

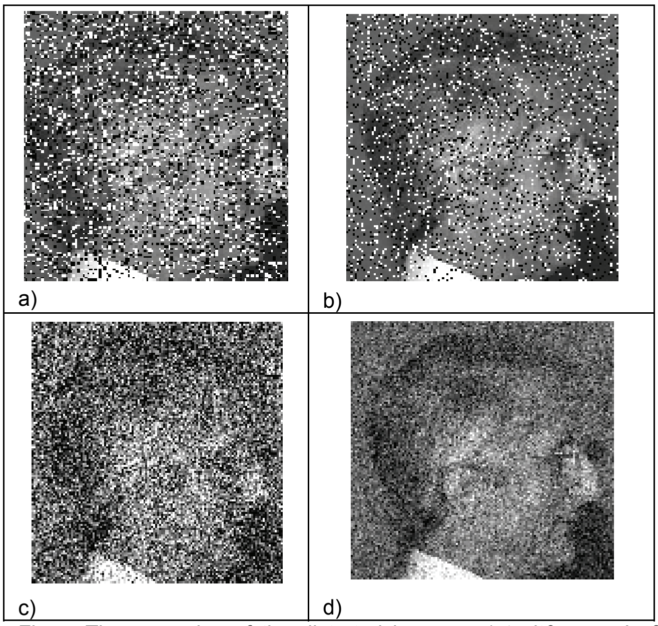


Fig. 4 The examples of the distorted images: a) 'salt&pepper' of D=0.3, a) 'salt&pepper' of D=0.2, c) Gaussian noise of var=0.1, d) Gaussian noise of var=0.02

Each image (the original and distorted) was represented by 18 principal components, that formed the input signal to the linear kernel SVM. The linear kernel was much more efficient than the nonlinear ones, tried also in experiments. To get the most objective results the experiments have been done in cross-validation mode, by repeating all steps of processing and learning 30 times. At each trials the distorted images have been generated randomly. The final result was the mean of all tests (only testing data, not taking part in learning were of interest).

The experiments have been performed in two ways. In the first set of experiments the SVM network was trained on all 30 original (non-distorted images) and the distorted copies of these images formed the testing set. Different scale of distortion have been tried. In the case of D=0.3 ('salt&pepper' noise) and variance of the Gaussian noise var<0.03) the average error of recognition was zero. Increasing the distortion we got the recognition with some error. At D=0.4 and var=0.03 the average error was equal 2.44% at std=1.34%. At half of images distorted by D=0.5 and half by D=0.4 (the images practically fully covered by the noise) and var=0.03 the average error has increased to 6.39% at std=1.43%. The experiments have shown that the recognition rate is less susceptible to the Gaussian noise. At D=0.3 and variance of the Gaussian noise changing from 0.03 to 0.1 the average error of recognition was practically the same and equal 1.22% at std=0.76%.

The next set of experiments was a bit different. This time only half of the original images (15) have been used in learning and the second half applied in testing together with the distorted images at D=0.3, D=0.2, var=0.05 and var=0.02. However, to balance the learning set in these experiments we have added four examples of the distorted images representing each person to the learning set. In this way the number of learning data was equal 39. The testing was done on the rest of distorted set of images (111 examples). We have repeated the experiments 30 times in cross-validation mode, exchanging randomly the learning and testing sets. This time the average misclassification error was equal 6.79% at standard deviation of 4.18%.

It is interesting to analyze the misclassification results of the images. They changed from trial to trial. The typical confusion matrix corresponding to one trial is presented in Table 1. The diagonal elements of this matrix represent the percentage of well recognized classes. The non-zero value of the *ij*th off-diagonal element indicates what percentage of *i*th class was recognized as the *j*th one.

Table 1 The confusion matrix of the recognition of 6 classes of images

Class	1	2	3	4	5	6
1	100%	0	0	0	0	0
2	0	100%	0	0	0	0
3	0	0	100%	0	0	0
4	0	0	0	85%	15%	0
5	0	0	0	0	100%	0
6	0	0	0	0	0	100%

As it is seen most of the images have been recognized perfectly. The off-diagonal entry different from zero (the misclassification rate of the class) has appeared only once (the fourth class was confused with the fifth one three times which corresponded to15% of error).

Conclusions

The experiments have shown that application of PCA in combination with SVM classifier represents efficient approach to the distorted face recognition problem. The main source of this efficiency is the PCA transformation, of high relative insensitivity to the noise, distorting the images.

Acknowledgement

Portions of the research in this paper use the FERET data base of facial images collected under the FERET program.

RFERENCES

[1] Belhumeur P., Hespanha J., Kriegman D., Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Trans. PAMI*, 1997, vol. 19, No 7, pp. 711-720

[2] Phillips P. J., Moon H., Rizvi S., Rauss P., The FERET evaluation methodology for face recognition algorithms, *IEEE Trans. PAMI*, 2000, vol. 22, No 10, pp. 1090-1104

[3] Phillips P. J., Wechsler H., Huang J., Rauss P., The FERET database and evaluation procedure for face recognition algorithms," Image and Vision Computing J, 1998, vol. 16, No. 5, pp 295-306.

[4] Osowski S., Sieci neuronowe SVM w zastosowaniu do klasyfikacji wzorców, Przegląd Elektrotechniczny, 2002, vol. 78, No 2, ss.29-36

[5] Schölkopf B., Smola A., Learning with kernels, Cambridge, *MIT Press, MA.* 2002

***Authors**:*

prof. dr hab. inż. Stanisław Osowski, Warsaw University of Technology, Institute of the Theory of Electrical Engineering, Measurement and Information Systems, Military University of Technology, Institute of Electronic Systems, Email: sto@iem.pw.edu.pl.

mgr inż. Katarzyna Sikorska-Łukasiewicz, Institute of Electronic Systems, Military University of Technology, sikorska@gmail.pl.