

Independent Component Analysis, Principal Component Analysis and Rough Sets in Face Recognition

Roman W. Świniarski¹ and Andrzej Skowron²

¹ Department of Mathematical and Computer Sciences
San Diego State University
5500 Campanile Drive San Diego, CA 92182, USA

and
Institute of Computer Science, Polish Academy of Sciences
Ordona 21, 01-237 Warsaw, Poland
rswiniar@sciences.sdsu.edu

² Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl

Abstract. The paper contains description of hybrid methods of face recognition which are based on independent component analysis, principal component analysis and rough set theory. The feature extraction and pattern forming from face images have been provided using Independent Component Analysis and Principal Component Analysis. The feature selection/reduction has been realized using the rough set technique. The face recognition system was designed as rough-sets rule based classifier.

1 Introduction

Face recognition is one of the most classical pattern recognition projects. Many techniques have been developed for face recognition based on facial images [4,9,11,13,14,15,17,18,19]. One of the prominent methods is based on principal component analysis [11,17,18]. Despite of the fact that many face recognition systems have been designed, new robust face recognition methods are still expected. Independent Component Analysis (ICA) [2,3,5,8,16] is one of new powerful methods of discovery latent variables of data that are statistically independent.

We have tried to explore a potential of ICA methods combined with rough set methods (hybrid method) in face recognition. For comparison, we have also studied an application of classic Principal Component Analysis (PCA) nn-supervised technique combined with rough set method for extraction and selection of facial features as well as for classification. For feature extraction and reduction techniques from face images we have applied the following sequence of processing operations:

1. ICA for feature extraction and reduction, and pattern forming of facial images.

2. Rough set method for feature selection and data reduction.
3. Rough set based method for rule based classifier design.

Similar processing sequence has been applied for PCA for feature extraction and reduction, and pattern forming of facial images.

ICA based patterns, with feature selected by rough sets have shown significant predisposition for robust face recognition.

The paper is organized as follows. First, we present a brief introduction to rough set theory and its application to feature selection, data reduction and rule based classifier design. Then, we provide a short introduction to ICA and its application to feature extraction and data reduction. The following chapter briefly describes PCA and its application to feature extraction and reduction. The paper concludes with description of numerical experiments in recognition of 40 classes of faces represented by gray scale images [11] using proposed methods.

2 Rough set theory and its application of rough sets to feature selection

The rough sets theory has been developed by Professor Pawlak [10] for knowledge discovery in databases and experimental data sets. This theory has been successfully applied to features selection and rule based classifier design [4,6,7,10,12,13, 14]. We will briefly describe a rough set method.

Let us consider an *information system* given in the form of the decision table

$$DT = \langle U, C \cup D, V, f \rangle, \quad (1)$$

where U is the *universe*, a finite set of N objects $\{x_1, x_2, \dots, x_N\}$, $Q = C \cup D$ is a finite set of *attributes*, C is a set of *condition* attributes, D is a set of *decision* attributes, $V = \bigcup_{q \in C \cup D} V_q$, where V_q is the set of *domain (value)* of attribute $q \in Q$, $f : U \times (C \cup D) \rightarrow V$ - is a total *decision function* (information function, decision rule in DT) such that $f(x, q) \in V_q$ for every $q \in Q$ and $x \in U$.

For a given subset of attributes $A \subseteq Q$ a relation

$$IND(A) = \{(x, y) \in U : \text{for all } a \in A, f(x, a) = f(y, a)\}, \quad (2)$$

is an *equivalence relation* on universe U , called the *indiscernibility relation*. By A^* we denote $U/IND(A)$, i.e., the partition of U defined by $IND(A)$.

For a given information system S a given subset of attributes $A \subseteq Q$ determines the approximation space $AS = (U, IND(A))$ of S . For a given $A \subseteq Q$ and $X \subseteq U$ (a concept X), the *A-lower approximation* $\underline{A}X$ of set X in AS and the *A-upper approximation* $\bar{A}X$ of set X in AS are defined as follows:

$$\underline{A}X = \{x \in U : [x]_A \subseteq X\} = \bigcup \{Y \in A^* : Y \subseteq X\}, \quad (3)$$

$$\bar{A}X = \{x \in U : [x]_A \cap X \neq \emptyset\} = \bigcup \{Y \in A^* : Y \cap X \neq \emptyset\}. \quad (4)$$

A *reduct* is the essential part of an information system (related to a subset of attributes) which can discern all objects discernible by the original information system. We have applied of rough sets reduct in the technique of feature selection-reduction of face images.

From a decision table DT , decision rules can be derived. Let $C^* = \{X_1, X_2, \dots, X_r\}$ be C -definable classification of U and $D^* = \{Y_1, Y_2, \dots, Y_l\}$ be D -definable classification of U . A class Y_i from a classification D^* can be identified with the decision i ($i = 1, 2, \dots, l$). Then *ith decision rule* is defined by

$$Des_C(X_i) \implies Des_D(Y_j) \text{ for } X_i \in A^* \text{ and } Y_j \in D^*. \quad (5)$$

These decision rules are logically described as follows:

if (a set of conditions) *then* (a set of decisions).

The set of decision rules for all classes $Y_j \in D^*$ is denoted by

$$\{\tau_{ij}\} = \{Des_C(X_i) \implies Dec_D(Y_j) : \text{where } X_i \cap Y_j \neq \emptyset \text{ for } X_i \in A^*, Y_j \in D^*\}.$$

The set of decision rules for all classes $X_i \in D^*$, where $i = 1, 2, \dots, r$, generated by the set of decision attributes D (D -definable classes in S) is called the *decision algorithm* resulting from the decision table DT .

In our approach C is a relevant relative reduct [1] of the original set of condition attributes in a given decision table.

2.1 Rough sets for feature reduction/selection

Many feature extraction methods for different types of row data (like images or time-series) do not guarantee that attributes (elements) of extracted feature patterns will be the most relevant for classification tasks. One of possibilities for selecting features from feature patterns is to apply rough sets theory [3,5,7,8]. Specifically, defined in rough sets computation of a reduct can be used for selection of some of extracted features constituting a reduct [6,7,10,12] as reduced pattern attributes. These attributes will describe all concepts in a training data set. We have used rough set method to find of reducts from the discretized feature patterns and to select features forming the reduced pattern based on chosen reduct. Choosing of one specific reduct from the set of reducts is yet another searching problem [4,6,7,12,23,14]. We have selected reduct based on dynamic reduct concept introduced in [1,3].

3 ICA - An introduction

ICA is the unsupervised computational and statistical method for discovering intrinsic hidden factors in the data [2,3,5,8]. ICA exploits higher-order statistical dependencies among data and discovers a generative model for the observed multidimensional data. In the ICA model, observed data variables are assumed

to be linear mixtures of some unknown independent sources (independent components). A mixing system is also assumed to be unknown. Independent components are assumed to be nongaussian and mutually statistically independent. ICA can be applied to feature extraction from data patterns representing time series, images or other media [2,5,8,16].

The ICA model assumes that the observed sensory signals x_i are given as the pattern vectors $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbf{R}^n$. The sample of observed patterns are given as a set of N pattern vectors $T = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, that can be represented as a $n \times N$ data set matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{n \times N}$ which contains patterns as its columns. The ICA model for the element x_i is given as linear mixtures of m source independent variables s_j

$$x_i = \sum_{j=1}^m h_{i,j} s_j, \quad i = 1, 2, \dots, n, \quad (6)$$

where x_i is observed variable, s_j is the independent component (source signals) and $h_{i,j}$ are mixing coefficients. The independent source variables constitute the source vector (source pattern) $\mathbf{s} = [s_1, s_2, \dots, s_m]^T \in \mathbf{R}^m$. Hence, the ICA model can be presented in the matrix form

$$\mathbf{x} = \mathbf{H}\mathbf{s}, \quad (7)$$

where $\mathbf{H} \in \mathbf{R}^{n \times m}$ is $n \times m$ unknown mixing matrix where row vector $\mathbf{h}_i = [h_{i,1}, h_{i,2}, \dots, h_{i,m}]$ represents mixing coefficients for observed signal x_i . Denoting by $\mathbf{h}_{c,i}$ columns of matrix \mathbf{H} we can write

$$\mathbf{x} = \sum_{i=1}^m \mathbf{h}_{c,i} s_i. \quad (8)$$

The purpose of ICA is to estimate both the mixing matrix \mathbf{H} and the sources (independent components) \mathbf{s} using sets of observed vectors \mathbf{x} .

The ICA model for the set of N patterns \mathbf{x} , represented as columns in matrix \mathbf{X} , can be given as

$$\mathbf{X} = \mathbf{H}\mathbf{S}, \quad (9)$$

where $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N]$ is the $m \times N$ matrix which columns correspond to independent component vectors $\mathbf{s}_i = [s_{i,1}, s_{i,2}, \dots, s_{i,m}]^T$ discovered from the observation vector \mathbf{x}_i . Once the mixing matrix \mathbf{H} has been estimated, we can compute its inverse $\mathbf{B} = \mathbf{H}^{-1}$, and then the independent component for the observation vector \mathbf{x} can be computed by

$$\mathbf{s} = \mathbf{B}\mathbf{x}. \quad (10)$$

The extracted independent components s_i are as independent as possible, evaluated by an information-theoretic cost criterion such as minimum Kulback-Leibler divergence kurtosis, negentropy [2,3,8].

3.1 Preprocessing

Usually ICA is preceded by preprocessing, including centering and whitening.

Centering

Centering of \mathbf{x} is the process of subtracting its mean vector $\boldsymbol{\mu} = E\{\mathbf{x}\}$ from \mathbf{x} :

$$\mathbf{x} = \mathbf{x} - E\{x\}. \quad (11)$$

Whitening (sphering)

The second frequent preprocessing step in ICA is decorrelating (and possibly dimensionality reducing), called *whitening* [16]. In whitening the sensor signal vector \mathbf{x} is transformed using formula

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \quad \text{so} \quad E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{I}_l, \quad (12)$$

where $\mathbf{y} \in \mathbf{R}^l$ is the l -dimensional ($l \leq n$) whitened vector, and \mathbf{W} is $l \times n$ whitening matrix. The purpose of whitening is to transform the observed vector \mathbf{x} linearly so that we obtain a new vector \mathbf{y} (which is white) which elements are uncorrelated and their variances are equal to unity. Whitening allows also dimensionality reduction, by projecting of \mathbf{x} onto first l eigenvectors of the covariance matrix of \mathbf{x} .

Whitening is usually realized using the eigen-value decomposition (EVD) of the covariance matrix $E\{\mathbf{x}\mathbf{x}^T\} \in \mathbf{R}^{n \times n}$ of observed vector \mathbf{x}

$$\mathbf{R}_{\mathbf{x}\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{E}_x \boldsymbol{\Lambda}_x^{\frac{1}{2}} \boldsymbol{\Lambda}_x^{\frac{1}{2}} \mathbf{E}_x^T. \quad (13)$$

Here, $\mathbf{E}_x \in \mathbf{R}^{n \times n}$ is the orthogonal matrix of eigenvectors of $\mathbf{R}_{\mathbf{x}\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^T\}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of its eigenvalues

$$\boldsymbol{\Lambda}_x = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \quad (14)$$

with positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$. The whitening matrix can be computed as

$$\mathbf{W} = \boldsymbol{\Lambda}_x^{-1/2} \mathbf{E}_x^T, \quad (15)$$

and consequently the whitening operation can be realized using formula

$$\mathbf{y} = \boldsymbol{\Lambda}_x^{-1/2} \mathbf{E}_x^T \mathbf{x} = \mathbf{W}\mathbf{x}. \quad (16)$$

Recalling that $\mathbf{x} = \mathbf{H}\mathbf{s}$, we can find from the above equation that

$$\mathbf{y} = \boldsymbol{\Lambda}_x^{-1/2} \mathbf{E}_x^T \mathbf{H}\mathbf{s} = \mathbf{H}_w \mathbf{s}. \quad (17)$$

We can see that whitening transforms the original mixing matrix \mathbf{H} into a new one, \mathbf{H}_w

$$\mathbf{H}_w = \boldsymbol{\Lambda}_x^{-1/2} \mathbf{E}_x^T \mathbf{H}. \quad (18)$$

Whitening makes it possible to reduce the dimensionality of the whitened vector, by projecting observed vector into first l ($l \leq n$) eigenvectors corresponding to first l eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_l$ of the covariance matrix \mathbf{E}_x . Then, the resulting dimension of the matrix \mathbf{W} is $l \times n$, and there is reduction of the size of observed transformed vector \mathbf{y} from n to l .

Output vector of whitening process can be considered as an input to ICA algorithm. The whitened observation vector \mathbf{y} is an input to unmixing (separation) operation

$$\mathbf{s} = \mathbf{B}\mathbf{y}, \quad (19)$$

where \mathbf{B} is an original unmixing matrix.

An approximation (reconstruction) of the original observed vector \mathbf{x} can be computed as

$$\tilde{\mathbf{x}} = \mathbf{B}\mathbf{s}, \quad (20)$$

where $\mathbf{B} = \mathbf{W}_w^{-1}$.

For the set of N patterns \mathbf{x} forming as columns the matrix \mathbf{X} we can provide the following ICA model

$$\mathbf{X} = \mathbf{B}\mathbf{S}, \quad (21)$$

where $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N]$ is the $m \times N$ matrix which columns correspond to independent component vectors $\mathbf{s}_i = [s_{i,1}, s_{i,2}, \dots, s_{i,m}]^T$ discovered from the observation vector \mathbf{x}_i . Consequently we can find the set \mathbf{S} of corresponding independent component vectors as

$$\mathbf{S} = \mathbf{B}^{-1}\mathbf{X}. \quad (22)$$

3.2 Fast ICA

The estimation of the mixing matrix and independent components has been realized using Karhunen and Oja FastIca algorithm [16]. In computational efficient Karhunen and Oja ICA algorithm the following maximization criterion has been exploited

$$J(\tilde{\mathbf{s}}) = \sum_{i=1}^m |E\{\tilde{s}_i^4\} - 3[E\{\tilde{s}_i^2\}]^2|. \quad (23)$$

This equation corresponds to 4-th order cumulant kurtosis. ICA assumes that independent components must be nongaussian. The base for estimating the ICA model is assumption about nongaussianity. One can observe that that kurtosis is a normalized version of the fourth moment $kurt(y) = E\{y^4\}$. Kurtosis has been frequently used as a measure of nongaussianity in ICA and related fields. In practical computationally, kurtosis can be estimated by using the fourth moment of the sample data.

A negentropy, as robust measure of nongaussianity can be used for approximations of a kurtosis.

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}), \quad (24)$$

where \mathbf{y}_{gauss} is a Gaussian random variable with the same covariance matrix as \mathbf{y} . Since it requires estimation of the probability density of y , hence the approximations of negentropy can be written as

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \text{kurt}(y)^2 \approx \quad (25)$$

$$\approx \sum_{i=1}^p k_i [E\{G_i(y)\} - E\{G_i(\nu)\}]^2, \quad (26)$$

where k_i are some positive constants, and ν is a standardized Gaussian variable of zero mean and unit variance. The variable $y = \mathbf{w}^T \mathbf{x}$ is assumed to be of zero mean and unit variance, and the functions G_i are some nonquadratic functions. The above measure can be used to test nongaussianity. It is always non-negative, and equal to zero if y has a Gaussian distribution.

If one uses only one nonquadratic function G , the approximation results in

$$J(y) \approx [E\{G(y)\} - E\{G(\nu)\}]^2. \quad (27)$$

Karhunen and Oja have introduced a very efficient method of maximization suited for this task. It is here assumed that the data is preprocessed by centering and whitening. The FastICA learning rule finds a direction, i.e. a unit vector \mathbf{w} such that the projection $\mathbf{w}^T \mathbf{x}$ maximizes nongaussianity. Nongaussianity is here measured by the the criterion being an approximation of negentropy given in [2, 5, 8].

3.3 Feature extraction using ICA

In feature extraction which is based on independent component analysis [2,3,5,8] one can consider an independent component s_i as the i -th feature of the recognized object represented by the observed pattern vector \mathbf{x} . The feature pattern can be formed from m independent components of the observed data pattern. The use of ICA for feature extraction is partly motivated by results in neurosciences, revealing that the similar principle of pattern dimensionality can be found in the early processing of sensory data by the brain.

In order to form the ICA patterns we propose the following procedure:

1. Extraction of n_f element feature patterns \mathbf{x}_f from the recognition objects. Composing the original data set T_f containing N cases $\{\mathbf{x}_{f,i}^T, c_i\}$. The feature patterns are represented by matrix \mathbf{X}_f and corresponding categorical classes are represented by column \mathbf{c} .
2. Heuristic reduction of feature patterns from the matrix \mathbf{X}_f into n_{fr} element reduced feature patterns \mathbf{x}_{fr} (with resulting patterns \mathbf{X}_{fr}). This step could be directly possible for example for features computed as singular values of image matrices.
3. Pattern forming through ICA of reduced feature patterns \mathbf{x}_{fr} from the data set \mathbf{X}_{fr} .

- (a) Whitening of the data set X_{fr} including reduced feature patterns of dimensionality n_{fr} into n_{frw} element whitened patterns \mathbf{x}_{frw} (projected reduced feature patterns into n_{frw} principal directions).
 - (b) Reduction of the whitened patterns \mathbf{x}_{frw} into first n_{frwr} element reduced whitened patterns \mathbf{x}_{frwr} through projection of reduced feature patterns into first principal directions of data.
4. Computing the unmixing matrix \mathbf{W} and computing reduced number n_{icar} of independent components for each pattern \mathbf{x}_{frwr} obtained from whitening using ICA (projection patterns \mathbf{x}_{frwr} into independent component space).
 5. Forming n_{icar} element reduced ICA patterns \mathbf{x}_{icar} from corresponding independent components of whitened patterns, with the resulting data set \mathbf{X}_{icar} . Forming a data set T_{icar} containing pattern matrix \mathbf{X}_{icar} and original class column \mathbf{c}
 6. Providing rough sets based processing of the set T_{icar} containing ICA patterns \mathbf{x}_{icar} . Discretizing pattern elements and finding relative reducts from set T_{icar} . Choosing one relevant relative reduct. Selecting the elements of patterns \mathbf{x}_{icar} corresponding to chosen reduct and forming the final pattern \mathbf{x}_{fin} . Composing the final data set $T_{final,d}$ containing discrete final patterns $\mathbf{x}_{fin,d}$ and class column. Composing the real valued data set T_{fin} from the set T_{icar} choosing elements of real-valued pattern using selected relative reduct.

3.4 ICA and rough sets

The ICA does not guarantee that selected first independent components, as a feature vector, will be the most relevant for classification. As opposed to PCA, ICA does not provide an intrinsic order for the representation features of a recognized object (for example an image). Thus, one cannot reduce a ICA pattern just by removing its trailing elements (which is possible for PCA patterns). Selecting features from independent components is possible through application of rough sets theory [3,5,7,8]. Specifically, defined in rough sets computation of a reduct can be used for selection some of independent components based attributes constituting a reduct. These reduct-based independent component based features will describe all concepts in a data set. The rough set method is used for finding of reducts from the discretized reduced ICA patterns. The final pattern is formed from reduced ICA patterns based on the selected reduct.

The results of discussed method of feature extraction/selection depend on a data set type and designer decisions: a) selection of dimension of the independent component space, b) discretization method applied, and (c) the selection of a reduct, etc.

4 ICA-faces. Rough-set-ICA-faces

ICA tries to find a linear transformation of patterns using basis as statistically independent as possible, and to project feature patterns into independent component space. Figure 1 depicts an example of ICA-face. By applying rough sets



Fig. 1. ICA-face

processing one can find the final pattern composed with independent component patterns corresponding to elements of selected reduct. We can call this pattern a rough-set-ICA-face.

5 Face recognition using ICA and rough sets

We have applied independent component analysis and rough sets for the face recognition. We have considered data set of images with 40 classes of 112×92 pixel gray scale images with 10 instances per class [11]. Figure 2 presents examples of face classes. We have applied the following method of feature extraction and pattern forming from 112×92 pixels gray scale face images [11]. The original feature patterns have been extracted from the lower 92×92 pixels sub-image of an original image. First, we have sampled that subimage taking its every second pixel and we have formed the sampled 46×46 pixels subimage. This subimage has been considered as a recognized object. The feature pattern has been created as a $n_{or} \times n_{or}$ ($n_{or} = 46 \times 46 = 2116$) element vector $\mathbf{x}_{orig} = [\mathbf{x}_{row}^1, \dots, \mathbf{x}_{row}^{2116}]^T$ containing concatenated rows of a sampled subimage. As a result of that phase of the pattern forming it was obtained the $N \times (n_{or} + 1)$ feature data set T_f with N ($N = 400$) cases containing $n_{or} = 2116$ element feature patterns and associated classes $(\{\mathbf{x}_{or}^i\}^T, class)$ ($class = 1, 2, \dots, 40$).

Then we have applied ICA (including the whitening phase) to the pattern part \mathbf{X} of the data set T_f in order to transform feature patterns into independent component space and to reduce pattern dimensionality.



Fig. 2. Example of face classes

Additionally the rough set technique has been applied for feature selection and data set reduction.

The original feature patterns have been projected into reduced 100 element independent component space (including 120 element whitening phase). Then we have found 14 element reduce from the ica patterns using rough set method and we have formed 14-element final patterns. The rough sets rule based classifier has provided 88.75% of classification accuracy for the test set.

6 PCA for feature extraction and reduction from facial images

Reduction of pattern dimensionality may improve the recognition process by considering only the most important data representation, possibly with uncorrelated elements retaining maximum information about the original data and with possible better generalization abilities. We have applied PCA, with the resulting

Karhunen-Loève transformation (KLT) [4], for the orthonormal projection (and reduction) of feature patterns \mathbf{x}_f composed with concatenated rows of sampled face images.

6.1 PCA

We generally assume that our knowledge about a domain is represented as a limited size sample of N random n -dimensional patterns $\mathbf{x} \in \mathbf{R}^n$ representing extracted feature patterns. We assume that the pattern part of data set X of feature patterns can be represented by a $N \times n$ data pattern matrix $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N]$. The training data set can be characterized by the square $n \times n$ dimensional covariance matrix \mathbf{R}_x . Let the eigenvalues of the covariance matrix \mathbf{R}_x are arranged in the decreasing order $\lambda_1 \geq \lambda_2 \geq \dots \lambda_n \geq 0$ (with $\lambda_1 = \lambda_{max}$), with the corresponding orthonormal eigenvectors $\mathbf{e}^1, \mathbf{e}^2, \dots, \mathbf{e}^n$. Then the optimal linear transformation

$$\mathbf{y} = \hat{\mathbf{W}}\mathbf{x}, \quad (28)$$

is provided using the $m \times n$ optimal Karhunen-Loève transformation matrix $\hat{\mathbf{W}}$

$$\hat{\mathbf{W}} = [\mathbf{e}^1, \mathbf{e}^2, \dots, \mathbf{e}^m]^T, \quad (29)$$

composed with m rows being the first m orthonormal eigenvectors of the original data covariance matrix \mathbf{R}_x . The optimal matrix $\hat{\mathbf{W}}$ transforms the original n -dimensional patterns \mathbf{x} into m -dimensional ($m \leq n$) feature patterns \mathbf{y}

$$\mathbf{Y} = (\hat{\mathbf{W}}\mathbf{X}^T)^T = \mathbf{X}\hat{\mathbf{W}}^T, \quad (30)$$

minimizing the mean least square reconstruction error. The PCA can be effectively used for the feature extraction and the dimensionality reduction by forming the m -dimensional ($m \leq n$) feature vector \mathbf{y} containing only the first m most dominant principal components of \mathbf{x} . The open question remains, which principal components to select as the best for a given processing goal [3]. One of possible methods (criteria) for selection of a dimension of a reduced feature vector \mathbf{y} is to choose a minimal number of the first m most dominant principal components y_1, y_2, \dots, y_m of \mathbf{x} for which the mean square reconstruction error is less than heuristically set the error threshold ϵ .

6.2 Numerical experiment

The original face feature patterns have been projected into reduced 100 element principal component space.

Then we have found 19 element reduct from the PCA patterns using rough set method and we have formed 19-element final patterns. The rough sets based rule based classifier has provided 86.25% of classification accuracy for the test set.

6.3 Eigen-faces. Rough-set-eigen-faces

PCA of face feature patterns (obtained by concatenation of image rows) makes it possible to find the principal components of data set and to project feature patterns into principal component space. PCA intends to find the representation of patterns based on uncorrelated basis variables. The optimal Karhunen-Loeve transform matrix is composed with eigenvectors of data covariance matrix. Rows of optimal transform matrix represent orthogonal bases. Figure 3 depicts an example of eigen-face. By applying rough sets processing one can find the final pattern composed with reduced principal component patterns corresponding to elements of selected reduct. We call this pattern a rough-set-eigenface.



Fig. 3. Eigen-face

7 Conclusion

A proposed hybrid processing technique including ICA and rough sets has demonstrated significant predispositions for face recognition. The proposed method outperforms a hybrid method which comprises PCA and rough sets.

Acknowledgements

The research has been supported by the grant 3T11C00226 from Ministry of Scientific Research and Information Technology of the Republic of Poland.

References

1. Bazan, J., Skowron, A., and Synak, P.: Dynamic reducts as a tool for extracting laws from decision tables. In *Proc. of the Symp. on Methodologies for Intelligent Systems*, Charlotte, NC, October 16-19, 1994, Lecture Notes in Artificial Intelligence, 869 Springer-Verlag (1994) 346-355
2. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7 (1995)1129-1159
3. A. Cichocki, R.E. Bogner, L. Moszczynski, Pope, A.: Modified Herault-Jutten algorithms for blind separation of sources. *Digital Signal Processing*, 7, (1997) 80 - 93
4. Cios, K., Pedrycz, W., Swiniarski, R.: Data Mining Methods for Knowledge Discovery. Kluwer Acad. Publ., Boston (1998)
5. Comon, P.: Independent component analysis - a new concept? *Signal Processing*, 36 (1994) 287-314
6. Grzymała-Busse, J.W.: Knowledge acquisition under uncertainty - A rough set approach. *Journal of Intelligent & Robotic Systems*, 1(1) 3-16
7. Grzymała-Busse, J.W.: LERS-a system for learning from examples based on rough sets. In: *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Słowiński, R. (ed.), Kluwer Academic Publishers, (1992) 3-18
8. Hyvriinen, A., Oja, E.: Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, 64(3) (1998) 301-313
9. Jonsson, J. , Kittler, J., Li, J.P. , Matas, J.: Learning Support Vectors for Face Verification and Recognition, Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (2000) 26-30
10. Pawlak, Z.: Rough sets, Theoretical aspects of reasoning about data, Kluwer, Dordrecht (1991)
11. Samaria, F., Harter, A.: Parametrization of stochastic model for human face identification. Proceedings of IEEE Workshop on Application of Computer Vision, 1994. ORL database is available at www.cam-orl.co.uk/facedatabase.html.
12. Skowron, A.: The rough sets theory and evidence theory. *Fundamenta Informaticae*, 13 (1990) 245-262
13. Swiniarski, R., Hargis, L.: Rough Sets as a Front end of Neural Networks Texture Classifiers. *Neuralcomputing Journal*. A special issue on Rough-Neuro Computing, 36 2001) 85-102
14. Swiniarski, R.: An Application of Rough Sets and Haar Wavelets to Face Recognition. In Proceedings of RSCTC'2000 In the Proceedings of The International Workshop on Rough Sets and Current Trends in Computing. Banff Canada, October 16-19 (2000) 523-530
15. Swiniarski, R. Skowron, A.: Rough Sets Methods in Feature Selection and Recognition. *Pattern Recognition Letters*. 24(6). (2003) 833-849
16. Swets, D.L, Weng. J.J.: Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on Pattern Recognition and Machine Intelligence*. 10(9)(1996) 831-836
17. The FastICA MATLAB package.
Available at <http://www.cis.hut.fi/projects/ica/fastica/>
18. Turk, M.A. Pentland, A.P.: Face Recognition Using Eigenspaces. Proc. CVPR'91, June (1991) 586-591
19. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition. (1991) 586-591