CLASSIFICATION POSSIBILITIES OF BLOOD SERUM CRYSTALLOGRAMS IN THE KLT DOMAIN

Hozman Jiri
Department of Radioelectronics, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, Prague 6, 16627 Czech Republic

Pata Petr
Department of Radioelectronics, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, Prague 6, 16627 Czech Republic

ABSTRACT

The crystallization in mixture of physiological solution (0.9% NaCl) with blood serum leads to the creation of characteristic dendritic crystals. The composition of the blood serum influences crystallization results in dependence on simultaneous pathological processes in the organism. Because of this crystallographic method, there is a requirement to find a set of classification methods of blood serum crystallograms. One of them is Karhunen – Loève expansion (PCA). An evaluation of classification possibilities of the blood serum crystallograms based on application of Karhunen – Loève expansion is conducted and discussed. As follows from the above mentioned facts, the main aim is to classify the blood serum crystallograms based on supposed diagnoses. The first most important task is to determine the images of the blood serum crystallograms related to healthy and unhealthy humans.

KEY WORDS
Blood Serum Crystallograms Classification, Karhunen – Loève Transform

1. Introduction

There is a lot of classification methods utilised during a digital image processing and image analysis. From the point of view of blood serum crystallogram classification there are three main textural classification approaches [1]: spectral, structural and statistical. As follows from the previous research [2] and [3] some methods have been tested. Every method makes possible to determine the significant parameters of image texture. A convenient combination of these independent parameters enables to perform the blood serum crystallogram classification with a relatively high sensitivity and specificity [2]. Our method, which belongs to the statistical approach, is based on Karhunen-Loève (KL) transform [4]. The KL transform is known for having certain properties that make it convenient for image compression. But the KL transform is suitable for the image classification as well. It is also called Principal Component Analysis (PCA). There are two main approaches how to use this transform in this field. The first approach is based on application of Karhunen-Loève transform for images as a matrix of image data. The second possibility is to use the so-called Karhunen-Loève expansion for arrangement of symptoms obtained from the image data analysis. The first approach is discussed in detail in this paper. The main aim is to find a convenient selection method of ROI from which Karhunen-Loève spectrum can be computed.

Fig. 1 A typical autocorrelation curve of a blood serum crystallogram related to a health human.

2. Karhunen – Loève Transform

The Karhunen Loève transform is derived under presumption based on statistical properties of an image data. By the distribution operator \( R \) we split image matrix \( \hat{X} \) into a set of \( M \) submatrices – ROI [5]. Each submatrix can be considered as an element of vector space \( V_N^N \) with defined a special form of the scalar product

\[
\hat{X} \rightarrow \{ x_{i,j} \}_{i,j=1,...,N}^N
\]

1 The Dirac and Einstein formalism are used in this paper. It is a powerful and well known tool from quantum mechanics. Upper (lower) index is used as a row (column). The \( x_{i,j} \) is used for a Hilbert vector space element and \( [ x_{i,j} ] \) for dual space elements. Greek sign is assigned for expression of an index realization of a stochastic process [6]. The symbols \( I, m, o, s, p, j, n, p, \) are assigned for expression of row (column) indexes. The \( \{ \xi \} \) is an operator of mean value defined in the vector space \( V_N^N \). \( \Xi \) is a set of covariance matrices. \( | \Phi \rangle \) is an eigenimage and also a base vector of \( V_N^N \). \( \beta \) is a symbol used for description of the eigenvalue of the covariance matrix.
It is important to find the optimal dimension of the vector space. There are two aspects [6]:

1. The typical size of objects in the image. Submatrix has to comprise whole objects and their neighbourhood.

2. Computation requirements are rapidly increased by a higher space dimension. Relations among spectral coefficients are not clear.

The shape of 2D autocorrelation function (see fig. 1) of a training image has been chosen as one of the criteria for the optimal space dimension determination. The optimal size of 64x64 has been found for the submatrix distribution operator. Let every such submatrix is one realisation of $N \times N$ dimensional random process. For the correlation between realisations we can write:

$$
\Xi_{ij}^{mn} = E\{ \rho(x_i^j | \{x_{ij}^m \}) \} \rho^{-1},
$$

$$
i, j, m, n = 1, \ldots, N.
$$

Concerning the mean square error, the best orthonormal basis $\{\Phi^m_i\}$ of the space $V_N^j$ can be composed from the eigenvectors

$$
\sum_{i,j=1}^{N} \Xi_{ij}^{mn} \{\Phi^m_i \} = \beta^m_i \{\Phi^m_i \},
$$

$$
m, n, r, s = 1, \ldots, N,
$$

where $\beta^m_i$ are eigenvalues of the set of covariance matrices $\Xi_{ij}^{mn}$.

Fig. 2 Image of the blood serum crystallogram of a healthy human.

Fig. 3 Image of the blood serum crystallogram of an ill human.

The first nine eigenimages of KL expansion of the crystallogram images related to healthy subjects.

Fig. 4 The relation between the first and second PCA components for the training images. The top figure represents measured data and the bottom image represents a model of log-Gaussian distribution function.

The two types of images have been chosen for formulation of this method. The first image class of images contains images of blood serum crystallogram related to health humans (see fig. 2). The images of crystallograms of ill humans have been assigned to the second image class (fig. 3). Each image category contains a set of forty classified images from each class. There were selected square ROIs in every image. The square ROIs were divided into a set of square submatrices with a dimension of 64 x 64 pixels.
The first versus second PCA component

Fig. 6 The first component versus second PCA component of training ‘healthy’ images. Other ‘healthy’ classified images (square) and ‘ill’ images (diamond) are plotted in the graph.

### Tab. 1 The parameters of the gravity centre and log-Gauss distribution of the relation between the first and second PCA components of the training ‘healthy’ images (see fig. 5). The x-centre is equal to the first PCA component and the y-centre to the second one (fig. 6).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_x$</td>
<td>3.065</td>
<td>-</td>
</tr>
<tr>
<td>$\sigma_y$</td>
<td>0.945</td>
<td>-</td>
</tr>
<tr>
<td>$I_x$</td>
<td>2.520</td>
<td>-</td>
</tr>
<tr>
<td>x-centre</td>
<td>12.95</td>
<td>0.24</td>
</tr>
<tr>
<td>y-centre</td>
<td>-0.3287</td>
<td>0.0774</td>
</tr>
</tbody>
</table>

### Tab. 2 The dispersion of point distribution in the plane of the first and second PCA component (see fig. 6).

<table>
<thead>
<tr>
<th>Image</th>
<th>Mean distance</th>
<th>RMSE</th>
<th>Diagnose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>3.40</td>
<td>0.15</td>
<td>Healthy</td>
</tr>
<tr>
<td>O23B695</td>
<td>3.72</td>
<td>0.38</td>
<td>Healthy</td>
</tr>
<tr>
<td>O49D695</td>
<td>4.88</td>
<td>0.74</td>
<td>Ill</td>
</tr>
<tr>
<td>O98D695</td>
<td>6.90</td>
<td>0.88</td>
<td>Ill</td>
</tr>
</tbody>
</table>

### 3. Method

The above-mentioned partial tasks are included in the complex blood serum crystallogram evaluation procedure. The simplified set of presumptions and steps is the following:

**Presumptions:**

1. The submatrices of ROI are generated as a multiple stochastic process. The eigenimages (see fig. 4) as a base of Karhunen-Loève transform can be found.
2. The distribution of relation between the first and any other PCA component (see fig. 5) is log-Gauss.
3. The images of the blood serum crystallograms related to healthy humans have similar statistical properties. For decorrelation of other ‘healthy’ images the Karhunen-Loève transform base images computed from training ‘healthy’ images with a satisfying result can also be used.

**Inputs:**

1. A set of blood serum crystallograms
   a) $N_A$ – a set of ‘healthy’ images (=40),
   b) $N_B$ – a set of ‘ill’ images,
   c) $N_N$ – a set of images with unknown diagnoses.
2. Analysis assumptions
   a) ‘pseudoregistration’ of images [6]. All images are acquired under same conditions.

**Training on a set of $N_A$ images $\{x_a\}_{k=1}^{N_A}$:**

1. Division of images into $M$ ROIs $\{x_j\}_{i=1,\ldots,M}$ with designed operator $R$.
2. Computing of KLT base eigenimages and eigenvalues of the ROI covariance matrix.
3. Selection of the most significant PCA component and computing of KLT spectrum images of the all ROIs.
4. PCA components analysis and plot of the relation between PCA components (see fig. 6):
   a. calculation of “centre of gravity” of PCA components,
   b. determination of the parameters describing log-Gauss distribution (see fig 5).
5. Analyses of input ROIs and determination of their distances from the center of cluster.
6. Histogram of distances derived in 5. for verification of log-Gauss distribution (fig. 7).

**Analysis of a set of images with an unknown diagnosis:**

1. Separation of each image into $M$ submatrices of ROI.
2. Computation of KLT spectrum for every ROI. The eigenimages calculated from the set of training images have been used for the spectrum derivation.
3. Determination of the distribution of deviations of the PCA components from the center of training images – ROI cluster.
4. From this follows that it is possible to determine a value (i.e. probability) of log-Gauss distribution ($G$) in this distance. The mean value of $G$ of each image can be used as a classification criterion:
   a. $G < 2\sigma$ - it is in cluster. $\sigma$ stands for mean value of the dispersion parameters of log-Gauss distribution,
   b. $2\sigma < G < 4\sigma$ - it is not possible to make decision simply,
   c. $G > 4\sigma$ - it does not belong to the cluster.
4. Discussion and results of the method

Comparison of suitability of selection of individual regions from the point of view of applicability for classification has been a very important task. KLT base images have been evaluated as a base on the previous approach. These eigenimages have been used for computing of spectra of both images belonging to the training set, and images from the classification group with an unknown medical diagnose. The classes or classification groups have been selected regarding to their diagnostic possibilities.

These images are represented by the cluster of points in the KLE domain (see fig. 6). The mean radius of this cluster expresses a measure of suitability for classification of this image. It takes into account an orderliness measure of the individual blood serum crystallograms. The dispersion of mutual deviations of points either in the cluster or in the spectral domain (the PCA component) was selected as the criterion of the suitability for medical classification. These conclusions are summarized in the Tab. 1 and 2. The relations between the first and second PCA components of the ‘healthy’ images and two images with unknown diagnoses are shown in fig. 5. It is clear that the image with an unknown diagnosis can be classified according to its data dispersion and location in the first – second PCA component plane.

5. Conclusions

Based on the principles of the method it is obvious that the method is sensitive to a possible change of scanning conditions. We took into account a value and direction of scanning sample illumination, scale, cut etc. Method of blood serum crystallogram classification, based on KL expansion evaluated in this contribution, is applicable. Two classification criteria have been formulated. Both methods are based on Karhunen – Loève spectrum evaluation. The first method is based on the distribution function fit of a function of the first several spectral components. The log-Gauss stochastic distribution function has been chosen. The second approach is based on derivation of a mean distance of elements in the Hilbert vector space of the ROIs.

6. Acknowledgement

This research work has been supported by the research program No. MSM 21000012 at CTU in Prague.

References