Concurrent Self-Organizing Maps for Multispectral Facial Image Recognition

Victor-Emil Neagoe, Senior Member, IEEE, Alexandru-Cristian Mugioiu, and Cristian-Tudor Tudoran

Abstract – This paper is dedicated to multispectral facial recognition, based on the model of Concurrent Self-Organizing Maps (CSOM), previously proposed by first author. The first approach of this paper is to apply CSOM classifier for color face recognition. Main variant of this approach has the following processing stages: (a) color conversion from the 3D RGB space into an optimum 2D selected color feature space; (b) Principal Component Analysis (PCA) for each resulted color component; (c) feature fusion; (d) CSOM/SOM classification. The proposed system is experimented using the ESSEX database of color facial images; it contains 151 subjects, where each is represented by 20 pictures of 200 x 180 pixels. The obvious advantage of CSOM over SOM is proved. The second approach of this paper is the implementation of a real time CSOM face recognition system using the decision fusion that combines the recognition scores generated from visual channels (R, G, B, and Y classifiers) with the thermal infrared classifier. As a source of color and infrared images, we used our VICFACE database of 38 subjects. Any picture has 160 x 120 pixels; for each subject there are pictures corresponding to various face expressions and illuminations, in the visual and infrared spectrum. The spectral sensitivity of infrared images corresponds to the longwave range of 7.5 – 13 μm. The very good experimental results are given, proving nearly invariance to illumination conditions.

I. INTRODUCTION

The Self-Organizing Map (SOM) (also called Kohonen network) [7] is an artificial unsupervised network characterized by the fact that its neighbouring neurons develop adaptively into specific detectors of different vector patterns. The neurons become specifically tuned to various classes of patterns through a competitive, unsupervised or self-organizing learning. The spatial location of a neuron in the network (given by its co-ordinates) corresponds to a particular input vector pattern. Similar input vectors correspond to the same neuron or to neighbour neurons. One important characteristic of SOM is that it can simultaneously perform the feature extraction and it performs the classification as well.

Starting from the idea to consider the SOM as a cell characterizing a specific class only, Neagoe proposed and evaluated in [8], [9], [10] a new neural recognition model called Concurrent Self-Organizing Maps (CSOM), representing a collection of small SOM units, which use a global winner-takes-all strategy. Each SOM is trained to correctly classify the patterns of one class only and the number of networks equals the number of classes. The CSOM model proved to have better performances than SOM, both for the recognition rate and also for reduction of the training time.

All about the world, governments and private companies are putting biometric technology at the heart of ambitious projects, ranging from access control and company security to high-tech passports, ID cards, driving licenses, and company security. One of most important areas of biometric technology is face recognition. A common feature found in almost all technical approaches proposed for face recognition is the use of only the luminance associated to the face image. Although the majority of images are recorded in the color format nowadays, most face recognition systems convert the color information to luminance component data and do no use color information. We further investigate the contribution of color cue for image recognition as well as the conversion from the RGB color space into an optimum plane [12], [13]. The selected color features will provide the input information for CSOM classifier.

On the other hand, multisensor data fusion is an emerging technology drawn from artificial intelligence, pattern recognition, statistical estimation, and other areas. Fusion multisensor data has significant advantages over simple source data, obtaining a more accurate estimate of a physical phenomenon. Data fusion provides new modeling opportunities. Now, we shall apply data fusion for biometric technology.

One variant of data fusion investigated here is to combine between the R, G, B channel data of color imagery, or by the features extracted from each color channel and to evaluate the recognition performance of feature fusion. A similar variant is to fuse the features extracted from each of the two color components of the plane into which the 3D color space has been projected.

Another special technique of pattern recognition is decision fusion, by combining the classification powers of several classifiers [5], [16]. Particularly, the problem becomes that of
combining classifiers based on visible (color) spectrum information with a classifier using thermal infrared spectrum. Face recognition in the thermal domain has received less attention in the literature in comparison with recognition in visible spectrum imagery, especially due to much higher cost of thermal sensors versus visible video equipment as well as due to its lower image resolution. However, recently, as a consequence of infrared image technology advances, made attractive to consider thermal sensors in the context of face recognition [4], [14]. We focused our attention on longwave infrared (LWIR) imagery, in the spectral range of 7.5-13 μm. Thermal infrared imagery of faces is nearly invariant to changes in ambient illumination. Consequently, for thermal facial imagery no illumination compensation is necessary. Here, we investigate decision fusion by combining matching scores generated by the visible and thermal infrared channel classifiers for face recognition.

The paper is structured as follows.

Second section presents the essentials of Concurrent Self-Organizing Maps (CSOM) model.

Third section has as aim the application of CSOM for color face recognition. We firstly present an approach to improve the color-based pattern recognition performance by optimizing the color conversion. We use the model proposed by Neagoe in [12] and [13], based on Karhunen-Loève transformation (KLT), to project the 3D RGB space into a 2D optimized space. We simulate and evaluate performances of three systems for color face recognition, having as classifiers the CSOM/SOM. First scheme uses feature fusion of the R, G, B channels; second scheme is based on 2D color space projection and feature fusion. Third variant uses a single color projection component. The experimental results are given.

Fourth section presents a real time face recognition system using the decision fusion based on Dempster-Shaffer theory that combines the recognition scores generated from visual (R, G, B) and longwave infrared (IR) CSOM classifiers.

II. CONCURRENT SELF-ORGANIZING MAPS (CSOM) FOR PATTERN CLASSIFICATION

Concurrent Self-Organizing Maps (CSOM) [8], [9], [10] is a collection of small SOM modules, which use a global winner-takes-all strategy. Each network is trained to correctly classify the patterns of one class only and the number of networks equals the number of classes. The CSOM training technique is a supervised one, but for any individual net the SOM specific training algorithm is used. We built “n” training patterns sets and we used the SOM training algorithm independently for each of the “n” neural units. Namely, each SOM module is trained with the patterns characterized by the corresponding class label. The CSOM models for training and classification are shown in Fig. 1.

For the recognition, the test pattern has been applied in parallel to every previously trained SOM. The neural module providing the minimum distance neuron is decided to be the winner and its index becomes the class index that the pattern belongs to (see Fig. 1).

In fact, CSOM is a system of systems having improved performances over a single big SOM with the same number of neurons, both from the point of view of recognition accuracy and for reducing the training time as well [10], [11].

III. CSOM FOR COLOR FACIAL RECOGNITION USING COLOR SPACE PROJECTION AND FEATURE FUSION

A. Optimum Color Projection Model

We shall further present the color space analysis model proposed by first author in [12] and [13]. Consider the color pixels in a given image as 3D vectors

$$P(x, y) = [R(x, y) G(x, y) B(x, y)]',$$  \hspace{1cm} (1)

where \(R(x, y)\), \(G(x, y)\) and \(B(x, y)\) are the red, green and blue components of the pixel of co-ordinates \((x, y)\).

We assume that color images exhibit features that can be useful in the conversion from a 3D full color space representation to the 2D space. For color conversion, we have chosen the Karhunen-Loève transformation (KLT), also known as Principal Component Analysis (PCA), by exploiting the correlation of the R, G, and B color channels. It is an optimum projection solution, by minimizing the mean square error for vector dimensionality reduction, when one projects the 3D RGB space into the 2D KLT color space with uncorrelated axes.

To deduce the KLT matrix, one firstly computes the covariance matrix of the color pixels (represented as 3D vectors). Then, one computes the eigenvalues of the covariance matrix. Finally, we deduce the two eigenvectors, corresponding to the largest two eigenvalues. Thus, one obtains the KLT matrix \(K\)

$$K= \begin{bmatrix} A' \\ B' \end{bmatrix},$$  \hspace{1cm} (2)
A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \quad (3)

(A and B are the eigenvectors of the covariance matrix, corresponding to the two largest eigenvalues and t denotes transposition).

Then, the projection of the 3D color vector P(x, y) in the 2D space is the vector C(x, y)

C(x, y) = \begin{bmatrix} C_1(x, y) \\ C_2(x, y) \end{bmatrix}, \quad (4)

given by the equation

C(x, y) = K \cdot P(x, y). \quad (5)

Example 1

Using a training set of 114 color pictures of 160x120 pixels selected from the face database called VICFACE (described in Section IV), we obtained the following results.

The eigenvalues of the color pixel covariance matrix for the considered training set are:

\[ \lambda_1 = 10861.736 \]
\[ \lambda_2 = 342.231 \]
\[ \lambda_3 = 10.218. \]

The corresponding eigenvectors defining the color KLT are

\[ A^t = (0.5540 \ 0.5831 \ 0.5942) \]
\[ B^t = (0.8133 \ -0.2268 \ -0.5358) \]
\[ C^t = (-0.1777 \ 0.7801 \ -0.5999) \]

One deduces that the mean square projection error (corresponding to least eigenvalue) for the whole training lot is of 0.09 % only.

The original RGB image given in Fig. 2(a) is reconstructed from its 2D KLT projection (see Fig. 2(b)). One can remark that the reconstructed picture is similar to the original. However, one can remark a slight change of yellow into pink.

\[ \text{Fig. 2. (a) Original. (b) Reconstruction of (a) from 2D KLT color space.} \]

B. Feature Fusion for Color Face Recognition

We further consider three feature fusion models for color face recognition shown in Figs. 3, 4, and 5.

The system shown in Fig. 5 retains a single color component after KLT projection, selects “m” eigen-features (PCA components) of the corresponding channel and finally performs the classification.

C. Experimental Results

We have used the color face database provided by Dr. Libor Spacek, Dept. of Computer Science, University of Essex, U.K. We considered 3020 images from this database, corresponding to 151 subjects, where each subject is represented by 20 pictures (10 images being chosen for training and the other 10 for test). Any picture has 200 x 180 pixels, in RGB format (with 24 bits/pel). The face database contains images of people of...
various racial origins, most of them being of 18-20 year old, but some older individuals are also present (Fig. 6).

For all the three systems (Figs. 3, 4, and 5), we have applied Principal Component Analysis (PCA), for each color channel, by retaining $m = 100$ features/color component. Then, for the first two systems (3D and 2D), we performed the feature fusion.

The experimental results are given in Table I and Figs. 7 and 8.

![Fig. 6. Several images belonging to the Essex database.](image)

TABLE I
COLOR FACE RECOGNITION SCORE FOR THE SYSTEMS SHOWN IN FIGS. 3, 4, AND 5.

<table>
<thead>
<tr>
<th>Total no. of neurons</th>
<th>Feature fusion of the R, G, B channels</th>
<th>2D Color space projection and feature fusion</th>
<th>1D Color space projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSOM</td>
<td>1x151</td>
<td>97.02</td>
<td>99.62</td>
</tr>
<tr>
<td></td>
<td>2x151</td>
<td>97.48</td>
<td>99.84</td>
</tr>
<tr>
<td></td>
<td>3x151</td>
<td>99.87</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>4x151</td>
<td>99.93</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>5x151</td>
<td>99.94</td>
<td>99.93</td>
</tr>
<tr>
<td>SOM</td>
<td>302</td>
<td>96.16</td>
<td>95.59</td>
</tr>
<tr>
<td></td>
<td>453</td>
<td>98.28</td>
<td>98.08</td>
</tr>
<tr>
<td></td>
<td>604</td>
<td>98.51</td>
<td>98.94</td>
</tr>
<tr>
<td></td>
<td>755</td>
<td>98.94</td>
<td>99.40</td>
</tr>
<tr>
<td>Nearest Prototype</td>
<td>98.74</td>
<td>99.87</td>
<td>98.81</td>
</tr>
</tbody>
</table>

![Fig. 7. Recognition rate for the feature fusion of the R, G, B channels.](image)

IV. CSOM FOR REAL-TIME MULTISPECTRAL FACIAL IMAGE RECOGNITION USING DECISION FUSION

In this section, the fusion of visual and thermal IR images is presented for enhancing robustness of face recognition. Thermal infrared imagery of faces is nearly invariant to changes in ambient illumination. Fusion exploits synergistic integration of information obtained from multiple sources.

We further investigate decision fusion by combining matching scores generated by the visible and thermal infrared channels for face recognition. In Fig. 9, our implemented real-time CSOM face recognition system is shown. The system uses a decision fusion based on Dempster-Shaffer theory of evidence [16]. The input information is provided by the visible and infrared channel classifiers. The two considered recognition systems with decision fusion have either four or two input channels: (1) the color components (R, G, B) and the infrared channel (IR); (2) the luminance (Y) extracted from the input RGB color picture as well as the infrared channel (IR). Consequently, we have five CSOM classifiers.

Each CSOM contains a number of SOM modules equal to the number “n” of classes; each module has a circular architecture with “p” neurons.

For each of the considered decision fusion systems \{(R, G, B, IR) and (Y, IR)\}, we used two variants (“a” and “b”) for choosing the rejection threshold.

For experimental evaluation, we have used the face database called VICFACE made by the team led by Prof. Victor Neagoe, Dept. of Electronics, Telecomm. and Information Technology, Polytechnic University of Bucharest, Romania. The face database has 228 images taken under frontal uniform illumination, and other 228 pictures taken using a nonuniform (top and lateral) illumination; the pictures correspond to 38 subjects. The color pictures are represented in RGB format (24 bits/pixel) and have a spatial resolution of 160x120 pixels. Most of the subjects are students of 23-25 year old (Fig. 10). For frontal illumination, each subject is represented by 6 pictures, two for each of the three expressions: normal, happiness and sadness (Fig. 11).
The infrared section of VICFACE database is composed by 456 thermal infrared images of 160 x 120 pixels; they are obtained using the FLIR ThermaCAM B2.

The spectral sensitivity of infrared images is in the longwave range of 7.5 – 13 μm. In Fig. 10 there are given a few examples of color and infrared images for five subjects.

The experimental results are given in Table II.
TABLE II
FACE RECOGNITION SCORE FOR THE SYSTEM SHOWN IN FIG. 9
(38 CLASSES; CIRCULAR NEURAL ARCHITECTURE; 5 NEURONS/MODULE; 30 EPOCHS FOR TRAINING)

<table>
<thead>
<tr>
<th>Illumination conditions</th>
<th>Recognition score [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R (Red)</td>
</tr>
<tr>
<td>Frontal uniform illumination</td>
<td>100</td>
</tr>
<tr>
<td>Nonuniform illumination (light comes from top)</td>
<td>86.84</td>
</tr>
</tbody>
</table>

V. CONCLUDING REMARKS

1) The paper presents an approach to multispectral facial recognition, based on the model of Concurrent Self-Organizing Maps (CSOM), previously proposed by first author. CSOM is a collection of small SOM modules; it uses a global winner-takes-all strategy. Each neural unit is trained to correctly classify the patterns of one class only.

2) For the same number of neurons, CSOM has better recognition performances than SOM.

3) From the point of view of training time, the advantage of CSOM over SOM is obvious. For n classes, the training time of CSOM is n times less than that of the corresponding SOM with the same number of neurons.

4) We have considered three CSOM classifier systems for color facial image recognition: (a) feature fusion of the R, G, B channels; (b) 2D optimum color space projection and feature fusion; (c) 1D optimum color space projection. All the variants led to significantly better recognition score for CSOM versus SOM. However, by increasing the number of neurons, the performance difference between CSOM and SOM decreases.

5) We performed an implementation of a real time CSOM face recognition system using the decision fusion based on Dempster-Shaffer theory, that combines the recognition scores generated from visual channels [(R, G, B) or Y classifiers] with the thermal infrared (IR) classifier. Inclusion of the longwave infrared imagery in the decision fusion implies the nearly invariance of the system to changes in ambient illumination.

REFERENCES


