Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Image and Signal Processing (CIISP 2007)

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 follwing 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 characteristics of SOM is that it can simultaneously perform the feature extraction and it performs the classification as well.

This work was supported in part by the Romanian Space Agency (Romanian Ministry of Education and Research) under Grants nr. 16 R /11.11.2005 (project code S-183) and nr. 27/05.10.2006 (project code 4667).

V. E. Neagoe, A. C. Mugioiu, and C. T. Tudoran are with the Department of Electronics, Telecommunications and Information Technology, Polytechnic University of Bucharest, Romania, P. O. Box 16-37, Bucharest 16, Romania 062510, E-mail: victoremil@gmail.com .

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).

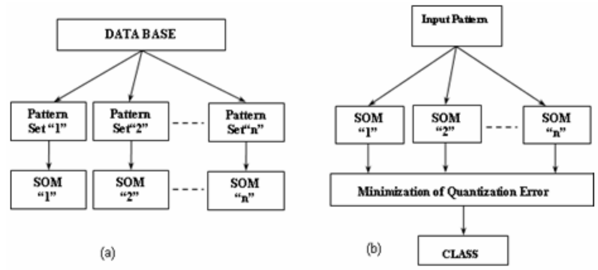

Fig. 1. The CSOM model.
(a) Training phase. (b) Classification phase.

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

$$P(x, y) = [R(x,y)\ G(x,y)\ B(x,y)]^t, \quad (1)$$

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

$$K = \begin{bmatrix} A^t \\ B^t \end{bmatrix}, \quad (2)$$

$$A=\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \qquad \text{and } B=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \qquad (3)$$

(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)

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

given by the equation

$$C(x, y) = = K \bullet P(x, y). \qquad (5)$$

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 :

λ_1 *= 10861.736*

λ_2 *= 342.231*

λ_3 *= 10.218.*

The corresponding eigenvectors defining the color KLT are

$A^t = (0.5540 \quad 0.5831 \quad 0.5942)$

$B^t = (0.8133 \quad -0.2268 \quad -0.5358)$

$C^t = (-0.1777 \quad 0.7801 \quad -0.5999)$

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.

(a) (b)

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.

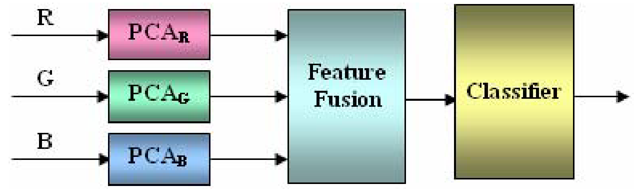

Fig. 3. Color face recognition using feature fusion of the R, G, B channels.

The system in Fig. 3 uses the fusion of the eigen-features (PCA with “m” elements on each channel) corresponding to the R, G, and B color components, followed by the classifier.

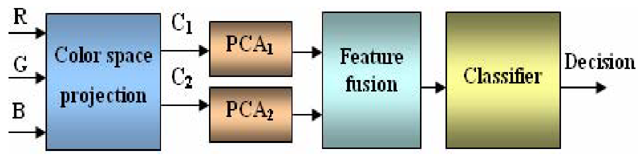

Fig. 4. Color face recognition with 2D color space projection and feature fusion.

Fig. 4 shows the new system of color face recognition (proposed by Neagoe in [12], [13]), using the previous presented color projection model; it contains the following processing stages :

a) Color conversion of the R, G, and B components into the two optimized new components C1 and C2, according to KLT;

b) Principal Component Analysis (PCA) for each of the two color channels (C1 and C2) ;

c) Feature fusion (amalgamation of the “m” eigen-components of each of the two channels) ;

d) Classification (CSOM/SOM/Nearest Prototype).

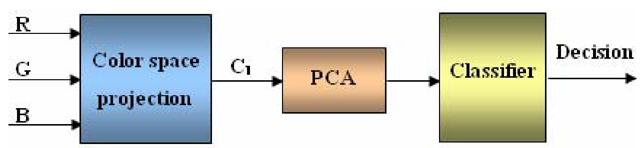

Fig. 5. Color face recognition using a single color projection component.

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, Depart. 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.

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

	Total no. of neurons	Feature fusion of the R, G, B channels	2D Color space projection and feature fusion	1D Color space projection
CSOM	1x151	97.02	99.21	96.62
	2x151	97.48	99.34	97.02
	3x151	99.87	100	99.87
	4x151	99.93	100	99.93
	5x151	99.93	100	99.93
SOM	151	64.44	61.66	69.40
	302	96.16	95.89	91.85
	453	98.28	99.74	98.08
	604	98.61	99.93	98.94
	755	98.94	99.67	99.40
Nearest Prototype		98.74	99.87	98.81

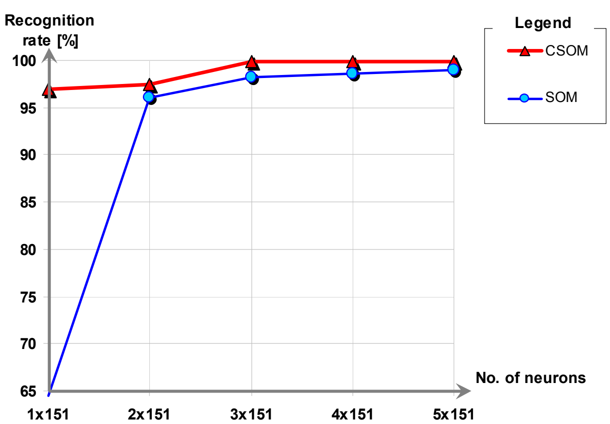

Fig. 7. Recognition rate for the feature fusion of the R, G, B channels.

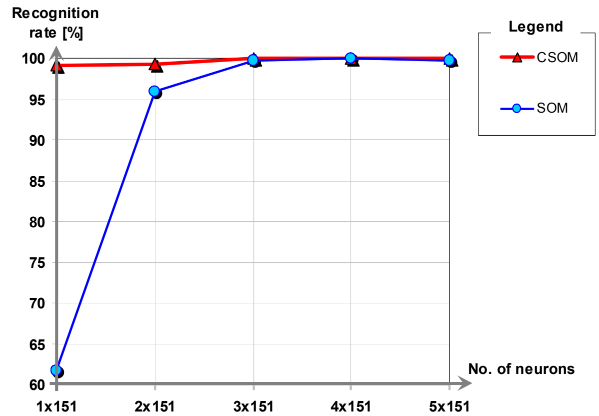

Fig. 8. Recognition rate for the 2D color space projection and feature fusion

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, Depart. 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).

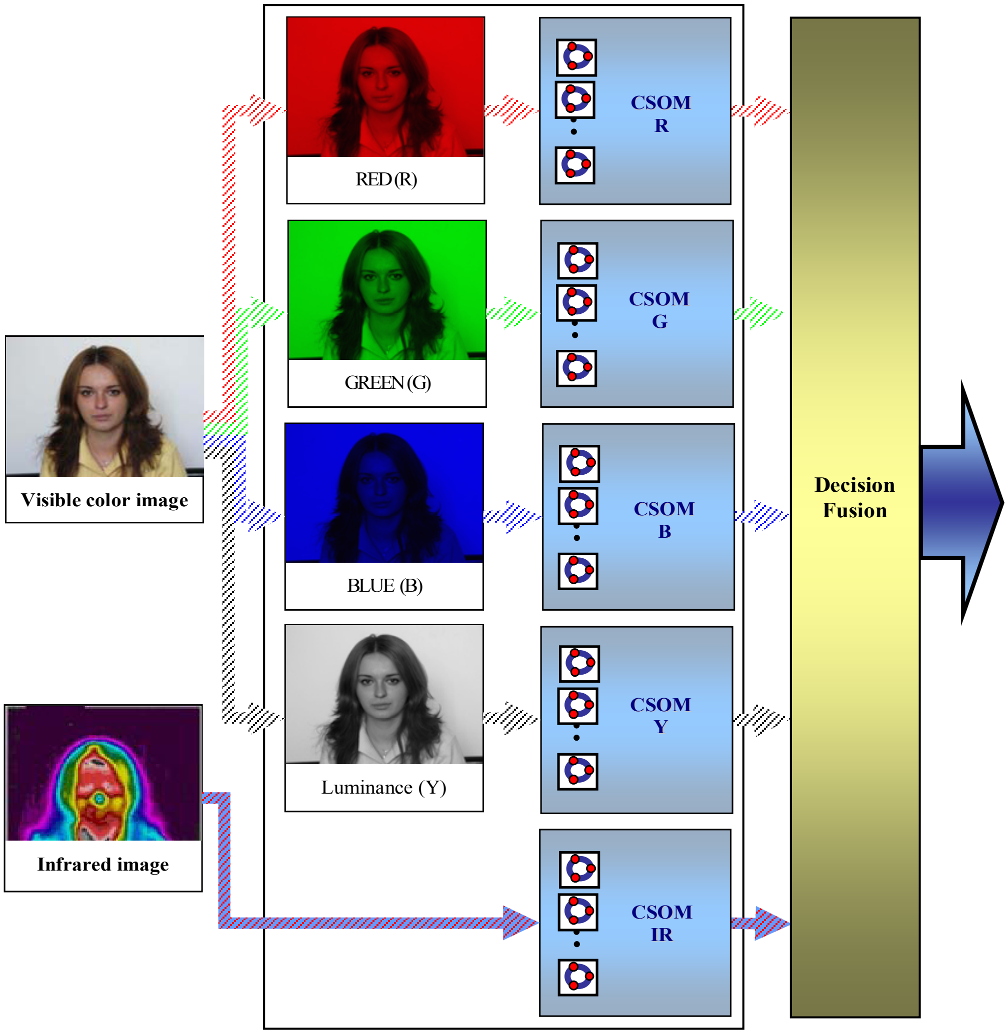

Fig. 9. Real time CSOM face recognition using visible and thermal infrared imagery.

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.

Fig. 10. Visual and infrared images corresponding to five subjects of VICFACE database.

Fig. 11. Facial expressions of the same subject : (a) normal, (b) happiness, (c) sadness.

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)

Illumination conditions	Recognition score [%]								
	R (Red)	G (Green)	B (Blue)	Y (Luminance)	IR (Infrared)	Decision fusion (R, G, B, IR)		Decision fusion (Y, IR)	
						(a)	(b)	(a)	(b)
Frontal uniform illumination	100	100	100	100	100	100	100	100	100
Nonuniform illumination (light comes from top)	86.84	77.19	71.05	79.82	100	98.24	100	99.12	100

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 longwawe infrared imagery in the decision fusion implies the nearly invariance of the system to changes in ambient illumination.

REFERENCES

[1] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction", *IEEE Trans. on Patt. Analysis and Mach. Intell.*, vol. 19, pp. 721–732, July 1997.

[2] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis", *IEEE Trans. Patt. Analysis and Mach. Intell.*, vol. 24, pp. 603-619, May 2002.

[3] P. Gonzales and R. Woods, *Digital image processing,* Addison Wesley, 1992.

[4] J. Heo, S. Kong, B. Abidi, and M. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition", *IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum* in conjunction with *CVPR 2004*, Washington, D.C., July 2004, pp. 94-99.

[5] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems", *IEEE Trans. Patt. Analysis and Mach. Intell.*, vol. 16, pp. 66-75, January 1994.

[6] C. F. Jones III and A. L. Abbott, "Optimization of color conversion for face recognition", *EURASIP Journal on Applied Signal Processing*, nr. 4, 2004, pp. 522-529.

[7] T. Kohonen, *Self-Organizing Maps*, Springer- Verlag, Berlin, 1995.

[8] V. E. Neagoe, "A neural approach to compression of hyperspectral remote sensing imagery", in: Reusch, B. (ed.): *Computational Intelligence. Theory and Applications*, *Proc. of International Conference on Computational Intelligence – "7th Fuzzy Days in Dortmund",* October 2001, Dortmund, Germany, Springer, Berlin, 2001, pp. 436-449.

[9] V. E. Neagoe, "Concurrent self-organizing maps for automatic face recognition", *Proc. 29th International Conference of the Romanian Technical Military Academy*, November 15-16 2001, Bucharest, published by Technical Military Academy, Romania, Section 9 (Communications), ISBN: 973-8290-27-9, (2001), pp. 35-40.

[10] V. E. Neagoe and A. Ropot, "Concurrent self-organizing maps for pattern classification", *Proc. First IEEE International Conference on Cognitive Informatics, ICCI 2002*, August 19-20, 2002, Calgary, Canada, pp. 304-312.

[11] V. E. Neagoe and A. Ropot, "Concurrent self-organizing maps - a powerful artificial neural tool for biometric technology", *Proceedings of World Automation Congress WAC'04*, Seville, vol. 3, 2004, ISBN 1-889335-20-7, IEEE Catalog 04EX832C.

[12] V. E. Neagoe, "An optimum 2D color space for pattern recognition", *Proc. of the 2006 International Conference on Image Processing, Computer Vision&Pattern Recognition*, Las Vegas, Nevada, June 26-29, 2006, vol. 2, pp. 526-532.

[13] V. E. Neagoe, "An optimum color feature space and its applications for for pattern recognition", *WSEAS Transactions on Signal Processing,* issue 12, vol. 2, pp. 1537-1543, December 2006.

[14] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, "Face recognition with visible and thermal infrared imagery", in *Computer Vision and Image Understanding*, Vol. 91, pp.72-114, 2003.

[15] L. Torres, J. Y. Reutter, L. Lorente, "The importance of the color information in face recognition", *Proc. of the IEEE Conference on Image Processing, ICIP'99*, October 25-28, 1999, Kobe, Japan, Vol. 3, pp. 627-631.

[16] L. Xu, A. Krzyzak, and C. Y. Suen, " Methods of combining multiple classifiers and their applications to handwriting recognition", *IEEE Trans. Systems, Man, and Cybernetics,* vol. 22, No. 3, May/June 1992, pp. 418-435.