Real Time Face Recognition Using Decision Fusion of Neural Classifiers in the Visible and Thermal Infrared Spectrum

V. E. Neagoe, *Senior Member IEEE*, A. D. Ropot, and A. C. Mugioiu

Department of Electronics, Telecommunications and Information Technology
Polytechnic University of Bucharest, Romania, email: victoremil@gmail.com

Abstract

This paper is dedicated to multispectral facial image recognition, using decision fusion of neural classifiers. The novelty of this paper is that any classifier is based on the model of Concurrent Self-Organizing Maps (CSOM), previously proposed by first author of this paper. Our main achievement 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, and B) or Y} with a 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 long wave range of 7.5 – 13 μm. The very good experimental results are given regarding recognition score.

1. Introduction

The Self-Organizing Map (SOM) (also called Kohonen network) [1] is an artificial unsupervised neural network characterized by the fact that its 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 neighbor neurons. One important characteristics 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 [2], [3], [4] 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 over 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; this is still a highly challenging task in pattern recognition and computer vision [5], [6]. Face recognition based only on the visual spectrum has shown difficulties in performing consistently under uncontrolled operating conditions. Face recognition accuracy degrades quickly when the lighting is dim or when it does not uniformly illuminate the face [7], [8]. Light reflected from human faces also varies depending on the skin color of people from different ethnic groups. The use of thermal infrared (IR) images can improve the performance of face recognition under uncontrolled illumination conditions [9], [10], [11]. Thermal IR spectrum comprising mid-wave IR (3-5μm) and long-wave IR (8-12μm) bands has been suggested as an alternative source of information for detection and recognition of faces. Thermal IR sensors measure heat energy emitted, not reflected, from the objects. Hence thermal imaging has great advantages in face recognition in low illumination conditions or even in total darkness, where visual face recognition techniques fail.

Recently, it has been observed that classifiers of different types complement one another in classification performance [12], [13], [14]. This has led to a belief that by using classifiers of different types simultaneously, classification accuracy could be improved. The corresponding special technique of pattern recognition is **decision fusion**, by combining the classification powers of several classifiers. Ideally, the combination function should take advantage of the strengths of the individual classifiers, avoid their weaknesses, and improve classification accuracy. Classical methods for classifier combination [13] include intersection of decision regions, voting methods, prediction by top choice combinations, and use of Dempster-Shafer theory. In this paper, we apply Dempster-Shafer theory of evidence presented in [12], [15], for decision fusion face recognition.

Particularly, the problem becomes that of combining CSOM classifiers receiving visible spectrum information (color components or luminance) with a CSOM classifier using thermal infrared spectrum. One variant of decision fusion investigated here is to combine between the R, G, B channel data of color imagery and infrared (IR) channel; a second decision variant is the combination between the luminance channel Y and IR channel. In a previous paper [16], one uses two optimized color components for pattern recognition instead of the R, G, and B ones. Other approach uses a neural technique for feature extraction from (R, G, B) images [17]. Regarding IR channel, we focused our attention on long wave 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.

The paper is structured as follows.

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

Third section presents an algorithm of decision fusion of N CSOM classifiers using the application of Dempster-Shafer theory of evidence.

In the fourth section one proposes and evaluates a real time face recognition system, using the decision fusion based on Dempster-Shafer theory. This system combines the recognition scores generated from visual classifiers {(R, G, B) or Y} and long wave infrared (IR) CSOM classifier. The experimental results are given.

2. A Neural Pattern Classifier Composed by Concurrent Self-Organizing Maps (CSOM)

Concurrent Self-Organizing Maps (CSOM) [2], [3], [4] is a collection of small SOM modules, which use a global winner-takes-all strategy. Each module 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 "K" training patterns sets and we used the SOM training algorithm independently for each of the "K" 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
Figure 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.

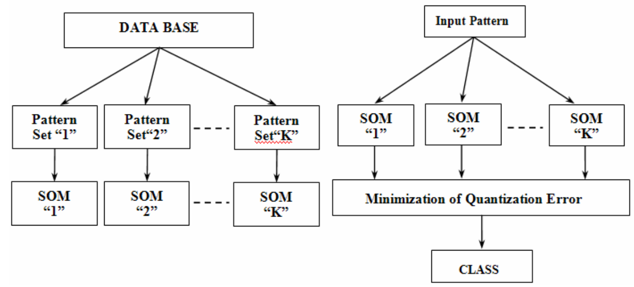

Figure 1: *The CSOM model.*
(a) Training phase. (b) Classification phase.

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].

3. Decision Fusion of Neural Classifiers Using Dempster-Shafer Theory

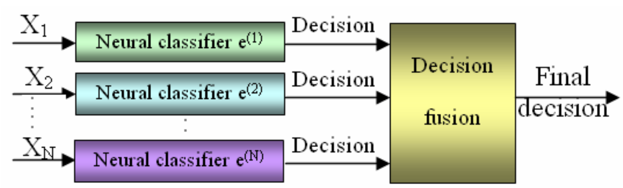

Figure 2: *Decision fusion of multiple concurrent neural modules.*

We further present an algorithm for decision fusion of multiple CSOM classifiers based on *Dempster-Shafer theory of evidence* presented in [12] and [15]. The novelty consists of adapting the theory for the case of CSOM classifiers. Consider the case of N concurrent modular neural classifiers (CSOM) denoted by $e^{(n)}$, where $n = 1, 2, \ldots, N$. Let $\boldsymbol{X}_k$ be the training data matrix for each class (neural module), $k = 1, 2, \ldots, K$, where K is the total number of classes. We will assume here equal amount of training for each of the classes. Also let θ_k be the label for each class k. Now, the feature extraction module of each classifier extracts a feature matrix $\mathbf{X}_k^{(n)}$. We define a modelling function *Ω(.)* which models each class so that

$$\Omega\left(\mathbf{X}_k^{(n)}\right) = U_k^{(n)};\ U_k^{(n)} = \left\{W_{k,i}^{(n)}\right\}$$

$$k = 1, 2, \ldots, K$$

$$n = 1, 2, \ldots, N$$

$$i = 1, 2, \ldots, m$$

Denote by $\left\{W_{k,i}^{(n)}\right\}$ the set of weight vectors corresponding to the neural module labeled with class "k" (for the classifier of index "n"). We denote by "m" the number of neurons of each module. Let **z** be an input test pattern which is modeled in a similar way

$\Omega(\mathbf{z}) = \mathbf{Z}$

For the case of a single classifier, the classification task is to assign class i to pattern **z** if

$D(\mathbf{U}_i, \mathbf{Z}) < D(\mathbf{U}_k, \mathbf{Z})\ \forall k = 1, 2, \ldots, K\ (k \neq i)$,

where $\mathbf{U}_k$ is the model for each class k, and $\mathbf{U}_i$ being the nearest neighbor to **Z**. $D(.)$ is a distance measure between the test pattern model (**Z**) and the training pattern models for each class ($\mathbf{U}_k$, $k = 1, 2, \ldots, K$).

Assume now that we have N classifiers, so that each classifier operates on the test model independently to reach an independent decision.

Since for each classifier, the function $\Omega(.)$ models the patterns in the same manner, we propose the nearest neighbor distance $\overset{(n)}{\min_k}\left\{D\left(\mathbf{U}_k^{(n)}, \mathbf{Z}\right)\right\}$ as the evidence of our belief in the decision made by classifier n. Thus, the belief becomes a decreasing function (say $\psi(.)$) of this distance

$$m^{(n)}(i) = \Psi\left(-\left(\overset{(n)}{\min_k}\left\{D\left(\mathbf{U}_k, \mathbf{Z}\right)\right\}\right)\right)$$

where $m^{(n)}(i)$ is our belief in classifier n for classifying test pattern **z** as class i.

One candidate for the function $\psi(.)$ could be the exponential function:

$$m^{(n)}(i) = \exp\left(-\lambda\left(\overset{(n)}{\min_k}\left\{D\left(\mathbf{U}_k, \mathbf{Z}\right)\right\}\right)\right)$$

Hence the smaller the nearest neighbor distance measure, the greater is our belief in the decision of the classifier. In summary our algorithm works as follows:

1. Each class is modelled using the training data matrix $\mathbf{X}_k$, $k = 1, 2, \ldots, K$ and the function $\Omega\left(\mathbf{X}_k^{(n)}\right) = \mathbf{U}_k^{(n)}$.
2. Input test pattern **z** is also modelled using the same modelling function $\Omega(.)$, i.e. $\Omega(\mathbf{z}) = \mathbf{Z}$.
3. A distance measure, $D(.)$ is then used to evaluate the distance between **Z** and each of the models $\mathbf{U}_k^{(n)}$, $k = 1, 2, \ldots, K$.
4. For each classifier, a label is given to the test pattern **z** which corresponds to minimum distance measure
   $d_k^{(n)} = D\left(\mathbf{U}_k^{(n)}, \mathbf{Z}^{(n)}\right)$
   $n = 1, 2, \ldots, N$
   $k = 1, 2, \ldots, K$
5. We estimate our confidence in each classifier's decision as:
   $m_k^{(n)}(\mathbf{z}) = \exp(-\lambda d_k^{(n)})$

We then combine all evidences $m_k^{(n)}$ $n = 1, 2, \ldots, N$, $k = 1, 2, \ldots, N$ using Dempster-Shafer theory of evidence as follows

$$m(k) = \frac{\prod_{n=1}^{N} m_k^{(n)}(\mathbf{z})}{\sum_{k=1}^{K}\left(\prod_{n=1}^{N} m_k^{(n)}(\mathbf{z})\right)}$$

$k = 1, 2, \ldots, K$

6. Class label j is assigned to test pattern if
   $j = \max_{k=1}^{K}\{m(k)\}$; $k = 1, 2, \ldots K$.

Some special cases to be considered are:

a) if all classifiers reject a pattern, the consensus decision will then be rejection and thus our belief will be given to the frame of discernment $m(\Theta) = 1$.
b) if a subset of classifiers says M rejects a test pattern, then these classifiers will be excluded and the decision will be made on basis of remaining ($N - M$) classifiers.

4. Real Time Face Recognition using Decision Fusion of CSOM Classifiers for Visible and Infrared Thermal Imagery

We further investigate *decision fusion* by combining matching scores generated by the visible and thermal infrared channels for face recognition. In Figure 3, the architecture of our implemented real-time multiple CSOM face recognition system is shown. The system uses a decision fusion based on Dempster-Shafer theory of evidence presented in section 3. The input information is provided by the visible and infrared channel classifiers. The two considered recognition system variants 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 "K" of classes; each module has a circular architecture with "m" 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 (Figure 4). For frontal illumination, each subject is represented by 6 pictures, two for each of the three expressions: normal, happiness and sadness (Figure 5).

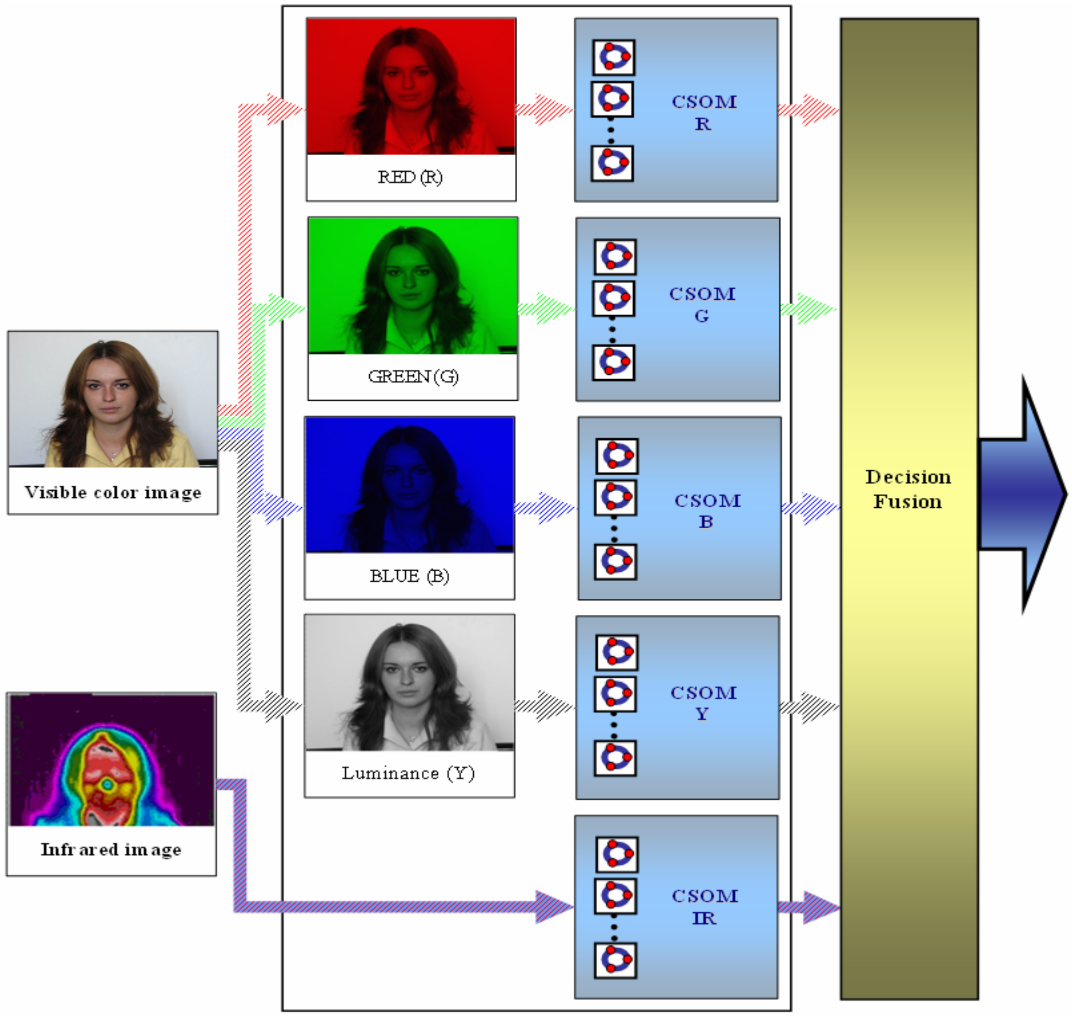

Figure 3: *Architecture of real time multiple CSOM face recognition system using visible and thermal infrared imagery.*

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.

Figure 4: *Visual and infrared images corresponding to five subjects of VICFACE database.*

The spectral sensitivity of infrared images is in the long wave range of 7.5 – 13 µm. In Figure 4 there are given a few examples of color and infrared images for five subjects.

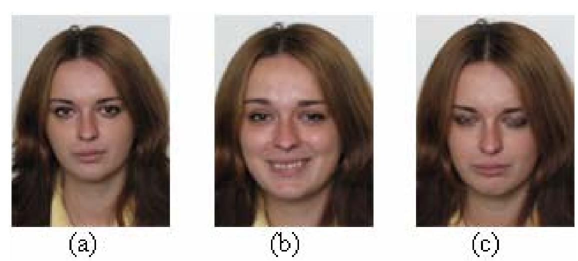

Figure 5: *Facial expressions of the same subject: (a) normal, (b) happiness, (c) sadness.*

The experimental results given in Tables 1 and 2, as well as in Figs. 6 and 7are self-explanatory; we compare CSOM, SOM and NP (Nearest Prototype) classifiers for IR facial image recognition.

Table 1: *Recognition score of CSOM versus SOM for thermal facial image recognition (without feature extraction).*

	1x38	2x38	3x38	4x38	5x38
CSOM	99.12	99.12	100	100	100
SOM	62.28	90.35	93.86	97.37	98.25
Nearest Prototype	98.25	98.25	98.25	98.25	98.25

Table 2: *Recognition score of CSOM versus SOM for thermal facial image recognition (PCA with p=100 retained features/picture).*

	1x38	2x38	3x38	4x38	5x38
CSOM	99.12	99.12	99.12	100	100
SOM	60.52	86.84	91.22	94.73	94.73
Nearest Prototype	97.37	97.37	97.37	97.37	97.37

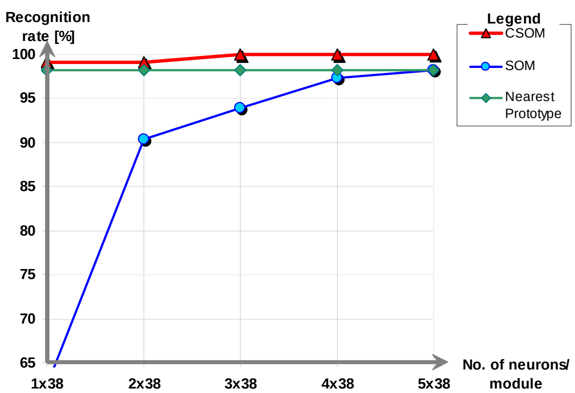

Figure 6 : *Recognition score of CSOM/SOM for thermal facial image recognition (without feature extraction).*

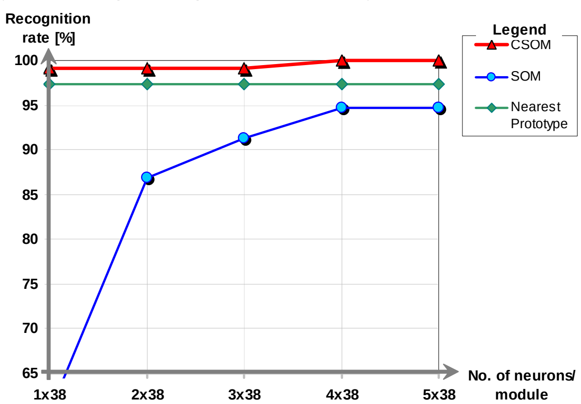

Figure 7: *Recognition score of CSOM/SOM for thermal facial image recognition (PCA with p=100 retained features/picture).*

The face recognition scores using fusion of multispectral CSOM classifiers are given in Table 3.

Table 3: *Recognition score for decision fusion of visual and infrared thermal CSOM classifiers (K=38 modules; m=7 neurons/module; PCA with p=100 features/picture).*

Lighting	R	G	B	Y	IR	Fusion (R,G,B,IR)	Fusion (Y,IR)
Linear decreasing from frontal centre	99.12	99.12	99.12	98.25	97.37	100	100
Linear decreasing from right centre	98.25	98.25	99.12	98.25	97.37	100	100
Low level uniform frontal	99.12	99.12	99.12	99.12	99.12	100	100

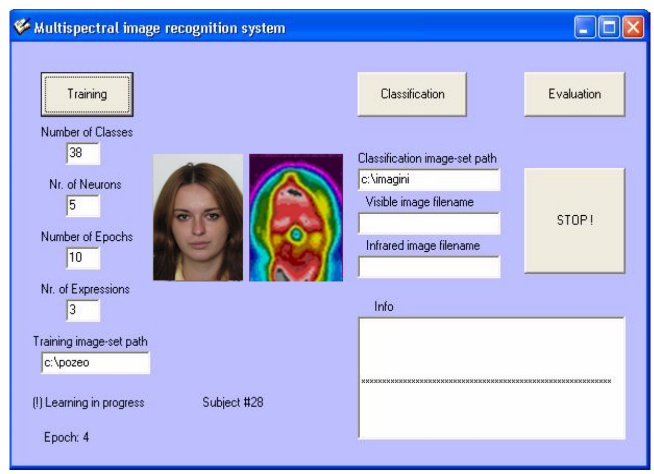

Figure 8: *Display of the experimental software recognition system.*

5. Concluding Remarks

1) This paper presents an approach to facial image recognition, using Dempster-Shafer theory for decision fusion of a special type of neural classifiers. Such a classifier is a set of neural modules based on the model of *Concurrent Self-Organizing Maps (CSOM)*, previously proposed by first author of this paper. Each neural classifier corresponds to a visual or thermal infrared channel. *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) We evaluate the performances of CSOM versus SOM and NP (Nearest Prototype), for face recognition in the IR thermal spectrum. For the same number of neurons, CSOM has better *recognition performances* than SOM and NP.

3) From the point of view of *training time*, the advantage of CSOM over SOM is obvious. For “K” classes, the training time of CSOM is “K**”** times less than that of the corresponding SOM with the same number of neurons.
4) We performed an implementation of a real time CSOM face recognition system using the decision fusion. The novelty consists of adapting the Dempster-Shafer theory for the case of CSOM classifiers. The system combines the recognition scores generated from visual channels {(R, G, B) or Y classifiers} with the thermal infrared (IR) classifier. Inclusion of the long wave infrared imagery in the decision fusion implies the nearly invariance of recognition performances of the system to changes in ambient illumination.
5) One obtains that, for many experimental cases, the recognition score for decision fusion is higher than the best score of the combination classifiers (Table 3).
6) Even the only fusion between luminance (Y) and infrared (IR) information is already very good; since then, the contribution of color seems to be small.
7) The decision fusion performance gains seem rather small since the IR performance is already very good, especially for a small number of neurons; someone can ask if the added complexity is worthwhile for a small addition of performance. However, we consider that the results promise to open an interesting window for applying our neural CSOM model in decision fusion for multispectral facial image recognition.
8) By increasing the number of subjects belonging to the facial image database, as well as by considering images taken from the outdoor environment, one expects to obtain a better evaluation of the proposed fusion model performances for face recognition.

References

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

[2]. V. E. Neagoe, ”Concurrent self-organizing maps for automatic face recognition”, *Proc. 29th International Conference of the Romanian Technical Military Academy*, Nov. 2001, Bucharest, Section 9, ISBN: 973-8290-27-9, pp. 35-40.

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

[4]. 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.

[5]. I. Craw, N. Costen, T. Kato, and S. Akamatsu, “How should we represent faces for automatic recognition?”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 8, 1999, pp. 725-736.

[6]. Zhao, W., C.R.P.P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys 35 (2003) 399–458.

[7]. Y. Adini, Y. Moses, and S. Ullman, “Face recognition: The problem of compensating for changes in illumination direction”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, July 1997, pp. 721-732.

[8]. J. Heo, S. G. Kong, B. R. Abidi, and M. A. 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*, pp. 94-99, Washington DC, July 2004.

[9]. 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.

[10]. D. A. Socolinsky and A. Selinger, “A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery,” *Proc. 16th Int'l Conf. Pattern Recognition*, vol. 4, pp. 217-222, 2002.

[11]. P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, M. Bazakos, Physiology-Based Face Recognition in the Thermal Infrared Spectrum, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 4, April 2007, pp. 613-626.

[12]. 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.

[13]. 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, January 1994, pp. 66-75.

[14]. J. Kittler, M. Hatef, R. Duin, and J. Matas, “On combining Classifiers”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, March 1998, pp. 226-239.

[15]. I. Naseem and M. Deriche, A New Algorithm for Speaker identification Using the Dempster Shafer Theory of Evidence, *Proc. of the 2006 International Conference on Image Processing, Computer Vision & Pattern Recognition*, Las Vegas, Nevada, June 26-29, 2006, vol. 2.

[16]. 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.

[17]. 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.

[18]. V. E. Neagoe, A. Mugioiu and C. Tudoran, “Concurrent self-organizing maps for multispectral facial image recognition”, *Proc. of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP 2007)*, April 1-5, 2007, Honolulu, Hawaii, USA, pp. 330-335.