3D Face Recognition Using Concurrent Neural Modules

VICTOR-EMIL NEAGOE , IONUT MITRACHE, AND DANIEL CARAUSU
Depart. Electronics, Telecommunications & Information Technology
Polytechnic University of Bucharest
Splaiul Independentei No. 313, Sector 6, Bucharest
ROMANIA
Email: victoremil@gmail.com

Abstract: - We investigate 3D face recognition by proposing an algorithm with the following processing stages: (a) thresholding of depth maps of 3D range images; (b) normalization and alignment; c) feature extraction by Gabor Wavelet Filtering (GWF); d) Principal Component Analysis (PCA); e) classification using the concurrent neural model previously proposed by the first author called Concurrent Self-Organizing Maps (CSOM), representing a winner-takes-all collection of self-organizing neural network modules. For comparison to CSOM, we also evaluate the performances of several statistical classifiers (1-NN and K-Means). The implemented neural versus statistical classifiers are evaluated using GavabDB database containing 3D face images of 61 subjects. The best experimental result of CSOM leads to the recognition rate of 95.08 %, by comparison to the rate of 83.60 % obtained using k-Means and to that of 88.52 % given by NN.

Key-Words: - 3D face recognition, range images, Gabor Wavelet Filters, neural classifier, Concurrent Self-Organizing Maps

1 Introduction

Automated face recognition can be defined as a system that looks through a stored set of signatures in the gallery and picks the one that best matches the features of the unknown individual.

The 3D face recognition growths to be a further evolution of 2D recognition problem, because a more accurate representation of the facial features leads to a potentially higher discriminating power. A recent work dedicated on 3D face modelling compared intensity images against depth images with respect to the discriminating power of recognizing people [8]. From their experiments, the authors concluded that depth maps give a more robust face representation, because intensity images are heavily affected by changes in illumination. The main advantage of the 3D based approaches is that the 3D model retains all the information about the face geometry. The 3D facial representation seems to be a promising tool coping many of the human face variations, extra-personal as well as intrapersonal.

There has been increasing interest in using artificial neural networks (ANN) for pattern recognition. A classifier is considered to be good or not according to its ability to generalize. The investigation of sample size problem for neural network classifiers leads the conclusion that the generalization error decreases as the training sample size increases. However, in contrast to statistical pattern recognition, neural networks have a good behaviour regarding small size problem.

Self-Organizing Map (SOM) (also called Kohonen network) is an artificial unsupervised neural network characterized by the fact that 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.

Starting from the idea to consider the SOM as a cell characterizing a specific class only, Neagoe [6] proposed and evaluated a new neural recognition supervised model called Concurrent Self-Organizing Maps (CSOM), representing a collection of small SOM modules, 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 neural network modules 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.

2 Proposed 3D Face Recognition Algorithm

The implemented method has the following processing stages:

a. Thresholding of the depth maps of 3D range images;
b. Normalization and cropping of depth maps;

c. Feature extraction by Gabor Wavelet Filtering (GWF);
d. Dimensionality reduction based on Principal Component Analysis (PCA);
e. Neural classification using Concurrent Self-Organizing Maps (CSOM)

2.1 Depth Map Thresholding

We have chosen Otsu method as a thresholding algorithm. Otsu [4] suggested a criterion by which the best threshold for images with bimodal histogram can be determined. The threshold is chosen in such a way that minimizes the weighted sum of within group variances for the two groups that result from separating the gray levels at the threshold value.

Fig. 1 (a) shows an original image from GavabDB database and Fig. 1 (b) shows the results of the thresholding operation based on Otsu's method.

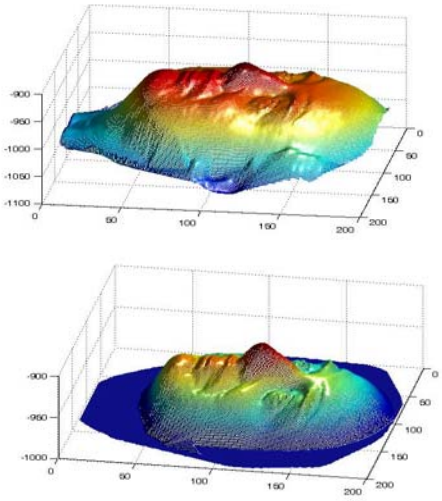


Fig. 1. (a) An original image of the GavabDB database. (b) The result of the thresholding operation.

2.2 Face cropping and normalization

The main object of 3D face normalization is to crop the images so that only the front of the face is used in the model and finally to bring the depth values in the range [0-255].

The nose is a very useful feature since it can be used as a reference point to align all the images. Furthermore, it is highly rigid in the sense that its shape does not change with facial expression.

The 3D normalization algorithm steps are the following:

a. Normalization of the depth values on Z-axis- Depth values are translated into the range [0-255]. Fig. 2 shows an image having the depth values translated from the range [-1020,-959] to [0, 255].

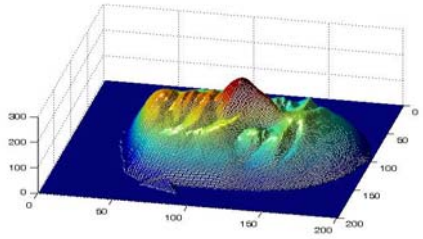


Fig. 2. The depth values on Z-axis are translated from the range [-1020,-959] to [0-255].

b. Nose localization - The nose is the closest part of the face to the 3D scanner; it has the highest depth value among all the facial points. By using a 3 x 3 window that calculates the sum of the depth values of its corresponding pixels, the nose is detected as the coordinates of the central pixel of the window with the maximum value. Fig. 3 shows a detected nose image.

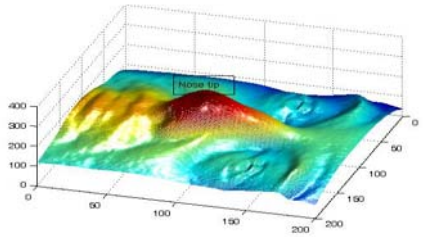


Fig. 3. A nose detection result using a 3 x 3 window that calculates the sum of the depth values of its corresponding pixels.

c. Image cropping - After detecting the nose, all images in the database are cropped using a standard 200 x 200 pixels in size considering the nose lies exactly in the centre of each image at the (100,100) x-y coordinate. Fig. 4 shows an example of a cropped image.

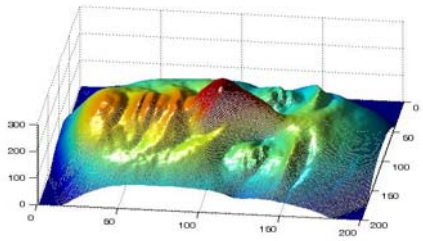
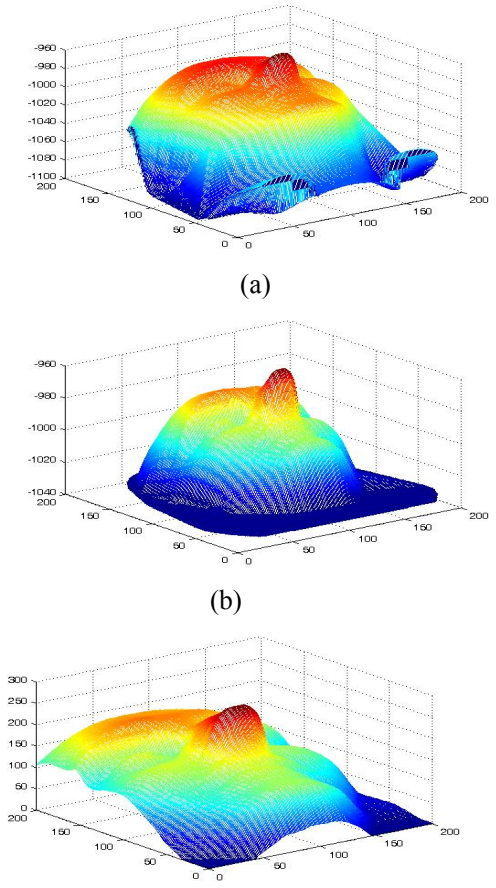


Fig. 4. The cropped and normalized image using a 200 x 200 window centred on the nose tip.

Fig. 5 shows another example of a cropped and normalized image of the GavabDB database.

After the 3D images have been cropped and normalized as described at the previous point, we have obtained 200 x 200 **range images,** centred on

the nose tip and having all z-axis values belonging to the range [0-255].



(a) (b) (c)

Fig. 5. A face image from the GavabDB before any processing; (b) its thresholded version; (c) the cropped and normalized equivalent.

Fig. 6 shows a couple of generated range images for the same subject having different facial expressions.

Fig. 6. Range images generated from GavabDB database after cropping and normalization operations.

2.3 Feature representation based on Gabor Wavelet Filters

Gabor responses are robust against variations caused by changes in facial expression, head pose and lighting conditions. A modified version of the Gabor filters adapted to the 3D face recognition is called 3D Spherical Gabor Filters (3D SGF). The 3D SGF is designed to cope with extensive view variations. To solve the missing point problem, caused by self-occlusion under large rotation angles, a 2D Gabor histogram, rather than the widely used integral operation is used in the computation of the distance between images.

A 3D SGF is defined as

$$g(x,F) = \hat{g}(x,y,z)\exp\left(j2\pi F\sqrt{x^2+y^2+z^2}\right) \quad (1)$$

where F is the centre frequency of the 3D SGF and

$$\hat{g}(x,y,z) = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\left[-\frac{\left(x^2+y^2+z^2\right)}{(2\sigma^2)}\right] \quad (2)$$

The 3D SGF is spherical symmetric. This rotation invariant characteristic makes the 3D SGF responses feasible for face recognition despite of different viewpoints. Given the image I(x, y, z), the Gabor responses of the 3D SGF at point **x**=(x, y, z) are defined as

$$J_n(x) = \int I(x')g(x-x',F_n)dx'. \quad (3)$$

A Gabor wavelet is determined by the following parameters: the central frequency *f*, the orientation θ and the ratio between frequency and the sharpness of Gaussian axis γ, η. When the values of γ and η are normally fixed, a set of Gabor wavelets with different frequencies and orientations should be designed to extract Gabor features

$$\gamma = \eta = \sqrt{2},\ f_u = {F_{\max}}\Big/{\left(\sqrt{2}\right)^v},\ \theta_v = {(v\pi)}\Big/{8}, \quad (4)$$

where u = 0,...,4, v = 0,...7. Once the ratio is fixed, the size of the Gaussian envelope monotonically decreases with the value of the central frequency. The higher the central frequency of the Gabor sinusoidal carrier, the smaller the area the Gaussian envelop will cover in spatial domain. This is reasonable since the high frequency signal changes faster.

A Gabor wavelet with parameters $f_\mu, \theta_v, \gamma, \eta$ can be defined as

$$\varphi_{u,v}(x,y) = \frac{f_u^2}{\pi\gamma\eta}\exp\left(-\left(\left(\frac{f_v}{\gamma}\right)^2 x_x^{\,2} + \left(\frac{f_v}{\eta}\right)^2 y_r^{\,2}\right)\right)\exp(j2\pi f_u x_r) \quad (5)$$

$$x_r = x\cos\theta_v + y\sin\theta_v$$
$$y_r = -x\sin\theta_v + y\cos\theta v$$

Given a bank of 40 Gabor wavelets, $\{\varphi_{\mu,v}(x,y), u = 0,...,4,\ v = 0,...7\}$, the image features at different locations, frequencies and orientations can be extracted by convolving the image $I(x, y)$ with all of the 40 Gabor wavelets. Thus, the image feature set has been expressed as

$$S = \left\{O_{u,v}^I(x,y) : u \in \{0,...,4\}\right\}, v \in \{0,...,7\}. \quad (6)$$

2.4 Dimensionality reduction using Principal Component Analysis (PCA)

Karhunen-Loève transform (KLT) called also Principal Component Analysis (PCA) is a statistical technique for optimal lossy compression of data under least square sense that provides an orthogonal basis vector-space to represent original data.

For each n-dimensional feature vector, denoted by the subscript i (i=1..N), the K-L transform defined by the relation

$$Y_i = T^t X_i \quad (7)$$

assigns an output vector of K-L coefficients (Yi) to an input vector (Xi).

We have truncated the vectors of K-L coefficients to a number of m terms considering different levels for energy preservation, see Table 1.

2.5. Neural classifier with concurrent self-organizing modules vs. statistical classifier

Concurrent Self-Organizing Maps (CSOM) [6], [7] 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 "M" of classes.

The CSOM training technique is a supervised one, but for any individual net the SOM specific unsupervised training algorithm is used. We built "M" training patterns sets and we used the SOM training algorithm independently for each of the "M" 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 Figs. 7 and 8.

2.5.1 Training of each $SOM^{(k)}$ module (k=1,…,M)

Assume that the module $SOM^{(k)}$ has $J^{(k)}$ neurons; particularly, one can choose

$$J^{(1)} = \ldots = J^{(M)} = \frac{J}{M} \quad (8)$$

where J is the number of CSOM neurons and M is the number of classes.

For each $SOM^{(k)}$ module, a specific training data subset is prepared containing all the training vectors having the label "k", as shown in Fig. 7.

Assume also that the number of vectors having the class label "k" is $N^{(k)}$, so that

$$\sum_{k=1}^{M} N^{(k)} = N \quad (9)$$

where N is the total number of training vectors. Usually, $J^{(k)} >> N^{(k)}$, to use the interpolation capacity of CSOM.

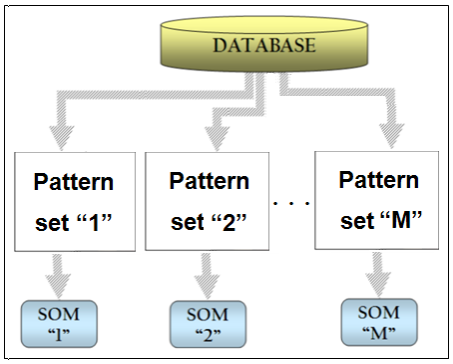


Fig. 7. The training phase of the CSOM model.

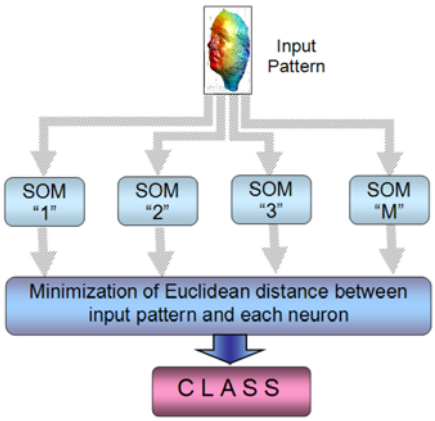


Fig. 8. The classification phase of the CSOM model.

2.5.2 Recognition Phase

For the ***recognition,*** the test pattern has been applied in parallel to every previously trained SOM module. 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. 8).

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 [6].

For comparison, we have considered the classical statistical classifiers of nearest neighbour (NN) and K-Means (the nearest mean)

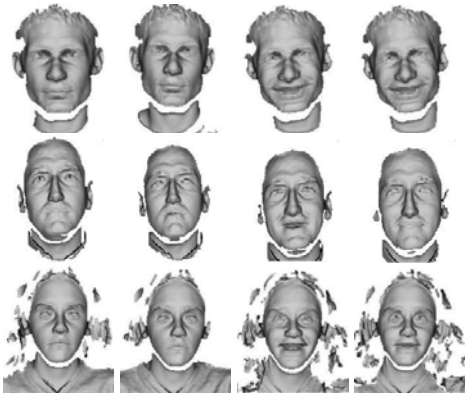
3 Experimental Results

3.1 3D faces database

In our work, we adopt the GavabDB database [7], created by the GAVAB research group of computer science department at the University of King Juan Carlos in Madrid.

GavabDB is a database designed to simulate the noisy face meshes obtained by typical commercial 3D scanners. It accommodates various errors and deteriorations that can occur in face meshes in practice.

This database contains face meshes of 61 individuals, with 9 meshes numbered from 1 to 9 for each person, captured under different settings. For incomplete meshes, the occluded patches typically can be reconstructed relying on the symmetric nature of human faces.



(a) (b) (c) (d)

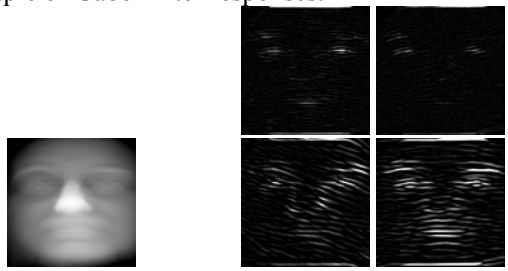
Fig. 9. Samples of faces from GavabDB database. (a) and (b): neutral gesture; (c): laugh gesture; (d): smile gesture.

We have chosen 4 meshes for each individual from the total of 9 meshes: faces no. 4 and no. 5 (frontal - neutral gesture), as well as no. 8 (frontal -laugh gesture) and 9 (frontal - smile gesture) for our experiments. As shown in Fig. 9, we have selected faces from columns (a), (c), and (d) for training (183 vectors) and faces from column (b) for testing (61 vectors).

3.2 Gabor Wavelet representation of faces

Convolving the image with complex Gabor filters with 5 spatial frequency (ν = 0,…,4) and 8 orientation (μ = 0,…,7), totally 5 x 8 = 40 filters, one captures the whole frequency spectrum, both amplitude and phase.

Fig. 10 shows an original face image and a couple of Gabor filter responses.



(a) (b)

Fig. 10. (a) Example of a range image ; (b) Gabor filter responses

Feature vectors are generated convolving the depth images with the 40 Gabor filters on the grid nodes. We have been experimented with different sizes for the grid starting with 4 x 4 and ending with 15 x 15.

The feature vector size varies between 40 x 16 = 640 components corresponding to the 4 x 4 grid to 40 x 225 = 9000 components for 15 x 15 grid.

3.2 Principal Component Analysis (PCA)

We have used Karhunen-Loève transform to reduce the dimensionality of the features data consisting of an ensemble of 183 vectors, each of them with the size m $\in$ [640, 9000] (dimensionality of the features varies with the size of the grid).

One can notice that to preserve a ~100 % of the energy 121 components are required for the 15 x 15 grid, but only at least 71 components for the 5 x 5 grid.

Table 1. Energy preservation factor as a function of the number of features for different grid sizes.

Number of features / grid size		10	20	30	39	49	58	71	94	121
Energy preservation [%]	5x5	87	95.2	98	99.2	99.6	99.8	99.98	100	100
	7x7	75	87	92.2	94.8	96.5	97.8	98.8	99.65	100
	10x10	69.3	82.3	88.4	91.9	94.5	96.1	97.7	99.3	100
	12x12	67	80.5	87.2	90.8	93.7	95.5	97.3	99.2	100
	15x15	65	79	86.2	90	93	95	97	99	100

Table 1 shows the energy preservation as a function of the number of components corresponding to the different grid sizes used to extract the features.

3.5. Recognition performances obtained by neural vs. statistical classifiers

We further present the results of our experimental results of 3D facial image recognition using the above mentioned selection from GavabDB database. We have evaluated the influence of CSOM architecture and size on the recognition score (see Tables 2 and 3 as well as Fig. 11).

The presented results correspond to the grid size of 10 x 10 for 3D imagery and to retain 181 PCA components; this assures an energy preservation factor of 100 % for any choice of test images.

Table 2. Recognition score for 1D CSOM architectures.

Total no. of neurons	61x2	61x4	61x6	61x8	61x10	61x12	61x14	61x16	61x18	61x20
CSOM with linear modules	88.52	93.44	90.16	91.80	90.16	91.80	90.16	90.16	90.16	91.80
CSOM with circular modules	88.52	91.80	93.44	91.80	91.80	93.44	90.16	91.80	93.44	91.80
NN	88.52									
k-Means	83.60									

Table 3. Recognition score for 2D CSOM architectures.

Total no. of neurons	61x4	61x16	61x36	61x64	61x100	61x144	61x196	61x256	61x324	61x400
CSOM with square modules	90.16	93.44	93.44	93.44	93.44	93.44	93.44	93.44	93.44	91.80
CSOM with cylindrical modules	90.16	93.44	93.44	93.44	91.80	93.44	93.44	95.08	93.44	93.44
CSOM with toroidal modules	90.16	93.44	93.44	93.44	93.44	93.44	95.08	93.44	93.44	93.44
NN	88.52									
k-Means	83.60									

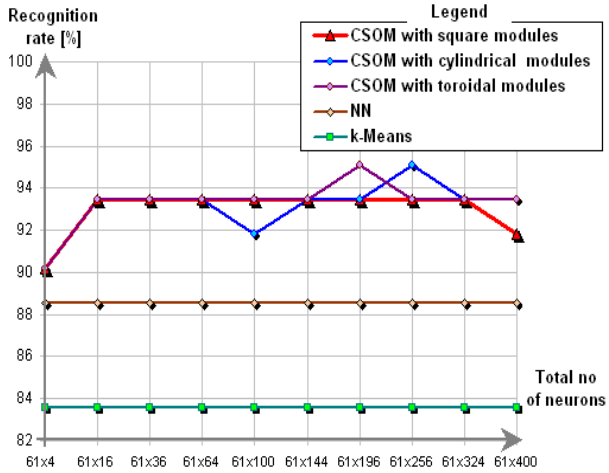


Fig. 11. Recognition score for 2D CSOM architecture as a function of the total number of neurons (M=61 classes).

4 Concluding Remarks

1. The paper proposes an algorithm for 3D face recognition with the following processing stages: (a) thresholding of depth maps of 3D range images; (b) normalization and alignment; c) feature extraction by Gabor Wavelet Filtering (GWF); d) Principal Component Analysis (PCA); e) classification using the neural model previously proposed by the first author called Concurrent Self-Organizing Maps (CSOM).
2. An optimum bimodal histogram algorithm for the depth maps thresholding has been implemented. The threshold is chosen in such a way that minimizes the weighted sum of within group variances for the two groups that result from separating the gray levels at the threshold value.
3. An automated cropping and normalization of the 3D meshes has been presented. After the nose is detected based on the fact that it has the highest depth value among all the facial points, all images are cropped considering the nose lies exactly in the centre of each image.
4. We have experimented the proposed algorithm for a set of 244 3D facial images selected from GavabDB database containing 61 subjects; the results correspond to the grid size of 10 x 10 and to retaining of 181 PCA components. The implemented neural CSOM classifier is evaluated for various architectures and sizes. Namely, we have taken into account both a 1-D architecture (linear and circular) and also 2D one (square, cylindrical, toroidal). The number of neurons/module are changed from 2 to 20 for the 1D module and from 4 to 400 for 2D modules.
5. The best recognition score of 95.08 % is obtained using either a toroidal CSOM with 14 x 14 neurons/module or a toroidal CSOM with 16 x 16 neurons/module. By comparison, k-Means method leads to the score of 83.60 %, while using NN (nearest neighbour) one obtains a score of 88.52 %.

References:

[1] B. Achermann, X. Jiang, H. Bunke, Face Recognition Using Range Images, *Proc. of the International Conference on the Virtual Systems and Multimedia*, September 1997, pp. 129-136.

[2] C. Beumier, M. Acheroy, Automatic 3D face authentication, in *Proc. Image and Vision Computing*, 2000, Vol. 18(4), pp. 315–321.

[3] X. Li and A. Jain (Eds.), *Handbook of Face Recognition*, Springer, 2005.

[4] K. Lin, On improvement of the computation speed of Otsu's image thresholding, *Journal of Electronic Imaging*, Vol. 14, No. 2, 2005.

[5] A.B. Moreno and A. Sanchez. GavabDB, A 3D Face Database. Advances and Applications, C. Garcia et al (eds), *Proc. 2nd COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications*, Ed. Univ. Vigo, pp. 77-82, 2004.

[6] V.-E. Neagoe and A. Ropot, Concurrent Self Organizing Maps for Pattern Classification, *Proc. of the 1st International Conference on Cognitive Informatics*, August 2002, Calgary, Canada, pp. 304-312.

[7] V.-E. Neagoe, I. Mitrache, S. Preotesoiu, A Feature-Based Face Recognition Approach Using Gabor Wavelet Filters Cascaded with Concurrent Neural Modules, *Proc. of World Automation Congress, World Automation Congress (WAC) 2006*, July 24-26, 2006, Budapest, published by TSI Press, San Antonio, Texas, USA.

[8] C. Xu, Y. Wang, T. Tan, L. Quan, Depth vs. intensity: Which is more important for face recognition?, in *Proc. 17th Internat. Conf. on Pattern Recognition*, August 2004, vol. 4 pp. 342–345.