Computer Science Journal of Moldova, vol.23, no.1(67), 2015

Measuring human emotions with modular neural networks and computer vision based applications*

Veaceslav Albu   Svetlana Cojocaru

Abstract

This paper describes a neural network architecture for emotion recognition for human-computer interfaces and applied systems. In the current research, we propose a combination of the most recent biometric techniques with the neural networks (NN) approach for real-time emotion and behavioral analysis. The system will be tested in real-time applications of customers' behavior for distributed on-land systems, such as kiosks and ATMs.

Keywords: Neural networks, emotion recognition, RBFN modular neural networks, Kinect.

1 Introduction

One of the most prominent innovations in the research world in the past decade is the introduction of neuroscience and computer vision based applications to measuring human emotions. Facial imaging, as captured by machine-learning software, is essentially a biometric method whose biological manifestations are strongly linked to the limbic system in the brain, which is notoriously hard to measure.

The aim of this research is to provide statistical observations and measurements of human emotional states in a chosen environment. Using computer vision and machine learning algorithms, emotional states

* This work was supported by project NATO.NUKR.SFPP 984877 "Modeling and Mitigation of Social Disasters Caused by Catastrophes and Terrorism"

of multiple targets can be inferred from facial expressions recorded visually by a camera. In this research, emotions are recorded, recognized, and analyzed to give statistical feedback of the overall emotions of a number of targets within a certain time frame. This feedback can provide important measures for user response to a chosen system. An application example of this research is a camera system embedded in a machine that is used frequently, such as an ATM. The camera will record the emotional state of customers (happy, sad, neutral, etc.) and build a database of users and recorded emotions to be analyzed later.

Recognizing human emotions is most efficiently achieved by visually detecting facial expressions. Facial imaging vision applications have gone through great advancement in the last decade for many practical purposes, such as face recognition and modeling.

Many researchers introduced algorithms for facial expression detection such as in [1], [2], [3], [4], [5], [6]. Recently, Microsoft released the new Kinect sensor along with an SDK that provides an emotion detection API.

Most research is focused on detecting facial expressions in an isolated framework, where each target is analyzed separately. Here we present a collective framework to analyze the "group" emotions and general human behavior.

The algorithm described in [6] and the Kinect API is implemented to record a database of multiple targets. A ground truth database containing manually labelled emotions will be also created for analysis and evaluation purposes.

We propose a hybrid architecture for complex event analysis. The real-time analysis of emotions (facial expression and pulse analysis) is performed with the help of state-of-the art biometric techniques, such as Kinect data analysis. The resulting measurements are compared with the statistical data for distributed on-land systems (e.g. kiosks, ATMs etc.).

Visual information in the proposed system is presented in several steps (Fig.1):

1 First, we use cameras and 3D sensors such as the Microsoft Kinect

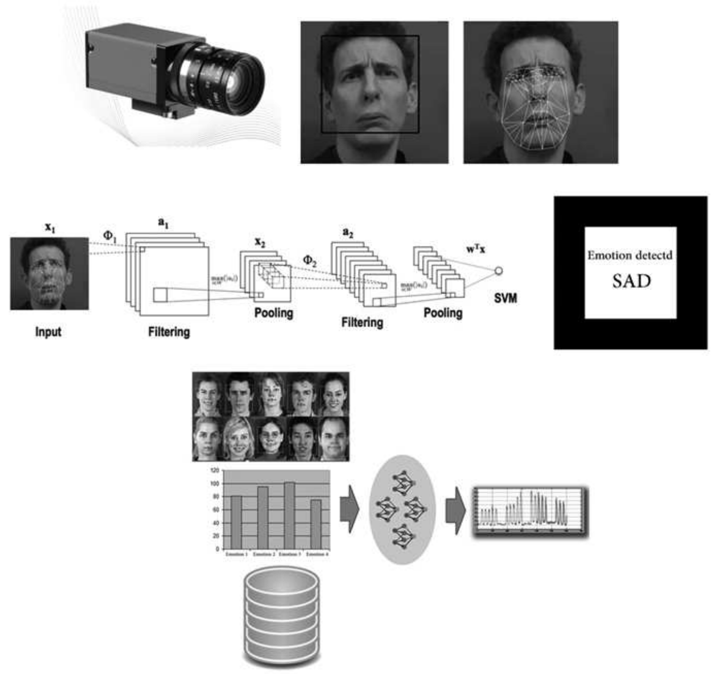

Figure 1. System work-flow overview

to detect facial features in order to recognize and classify emotions.

2 Second, we apply computer vision techniques for feature extraction and pattern recognition.

3 We apply machine learning (neural networks) for emotion detection and classification.

4 We use recorded statistical data from the machine transactions or logs. We train a modular neural network together with emotion records to provide analysis of events. We can use the trained networks for further event analysis and forecasting.

We suggest using modular neural network architecture for time series analysis and predictions, based on the statistical data combined with the biometric measurements.

Modular neural networks has been used in the last decade in a variety of applications, including object recognition [7], forecasting [8], [9] and financial analysis [10]. Here, we suggest modular network analysis for system learning and forecasting. Alternatively, we are planning to implement deep learning techniques (convolutional NNs) and compare the results of modular NNs performance for statistical and biometric data analysis [11]. In the current research, we propose a combination of most recent biometric techniques with the NN approach for real-time emotion and behavioral analysis. The system will be tested in real-time applications of customers' behavior for distributed on-land systems, such as kiosks and ATMs.

The rest of this paper will be organized as follows. First, we will present the overview of important studies on the related topics and provide the psychological basis for object recognition models. Next, we will present the neural network model for emotion recognition. The architecture of basic module of the network is the self-organized map (SOM) of functional radial-basis function (RBF) modules. We will give a detailed mathematical description of the applied approach, formalize the mathematics of the model and provide an explanation on the choice and implementation of the model's learning algorithm. Also, we will demonstrate implementation of the algorithm as a neural network model. Finally, we will describe the usage of the Kinect API and the algorithm we utilize for emotion detection with the Kinect camera.

2 Background

The aim of this research is to provide statistical observations and measurements of human emotional states in a chosen environment. Using computer vision and machine learning algorithms, emotional states of multiple targets can be inferred from facial expressions recorded visually by a camera.

In the proposed system, human emotions are recorded, recognized, and analyzed to give statistical feedback of the overall emotions of a number of targets within a certain time frame. This feedback can provide important measures for user response to a chosen system. An application example of this research is a camera system embedded in a machine that is used frequently, such as an ATM. The camera will record the emotional state of customers (happy, sad, neutral, etc.) and build a database of users and recorded emotions to be analyzed later.

One of the most prominent innovations in the research world in the past decade is the introduction of neuroscience and computer vision based applications to measuring human emotions. Facial imaging, as captured by machine-learning software, is essentially a biometric method whose biological manifestations are strongly linked to the limbic system in the brain, which is notoriously hard to measure.

Recognizing human emotions is most efficiently achieved by visually detecting facial expressions. Facial imaging vision applications have gone through great advancement in the last decade for many practical purposes, such as face recognition and modelling. Many researchers introduced algorithms for facial expression detection (we will describe the most relevant algorithms later in this section).

Before talking about the ways how to model emotions or how to measure emotions, we need to define the concept of emotion.

Nowadays, a number of opinions on this subject exists. Some researchers describe emotions as states of mind, others say that they are reactions. In this paper, we will regard Damasio's definition of emotions [12], i.e. we will follow Damasio in treating emotions as guides or biases to behaviors and decision making, action plans in response to internal or external stimuli and integral part of cognition and developmental processes.

It is important to distinguish between emotions and moods. While they affect each other to a large extent, they are fundamentally different. Emotions are fast reactions to stimuli. They last seconds (hardly more than minute) and they can be affected by one's current mood. Moods last longer: hours-days, and can be traced to one's personal inclinations, illness (depression) or other factors.

Finally, it is also important to distinguish between primary emotions and secondary emotions [13]. Primary emotions are fast immediate responses on stimuli, like "fight-or-flight" behaviors. Secondary emotions are more related to cognition and represent "blends" of different emotions with cognitive assessment. For example, "hope" can be considered as such emotion. Moods can sometimes be seen similar or interchangeable concept with secondary emotions, but it really depends on interpretations.

There is a number of models of emotions developed for different purposes like formalization, computation or understanding. Since this text is not a thorough review of models of emotions, we will only discuss a couple of them. But, the thing that must be mentioned first of all, is that all models of emotions can be classified into discrete and continuous. Discrete models work with limited sets of emotion. There might be from two (like "anger" and "happiness") to many. Continuous models rather represent the full spectrum of human emotions in some space (usually 3D or 2D). There are combinations of those, when you bring some uncertainty into discrete models in the form of probabilities of certain emotions happening at time t and thus implement some "quasi continuous" model.

The best known and the most widely used discrete model of emotions was developed by Paul Ekman. He developed his model over years and ended up with six basic emotions: anger, disgust, fear, happiness, sadness, and surprise [14]. Important feature of Ekman's model is that those six basic emotions are semantically distinct [15].

Dimensional models can be traced back to Wundt [16]. He proposed to use 3D space with axes "pleasure", "arousal", and "dominance" to describe emotions [17]. Any emotion or some blend of emotions can be represented as a point in this PAD cube. There are variation of this model, where different axes are used, but most of them end up with 3D space. Another 3D model of emotions which should be mentioned, is developed by Lövheim [18]. He sticks levels of serotonin, dopamine and noradrenaline directly to dimensions of emotions.

Other model of emotions, which need to be mentioned is OCC [19]. The theory of Ortony, Clore and Collins assumes that emotions develop as a consequence of certain cognitions and interpretations. Therefore it exclusively concentrates on the cognitive elicitors of emotions. The authors postulate that three aspects determine these cognitions: events, agents, and objects. According to the authors' assumptions, emotions represent valenced reactions to these perceptions of the world.

Experiencing emotions is not the only component of humans functionality related to emotions. There are other aspects like humans skills or abilities to control emotions, to understand self and other emotional states. This can be generally called "emotional intelligence". It was found some years ago that manager's emotional quotient (EQ) can be more important than intelligence quotient (IQ) in efficient decision making especially when in the lack of information.

A number of ways to measure emotions has been proposed. One of the most famous ways is the Self-Assessment Manikin (SAM) [20]. In SAM, users have to rate three simple parameters to express their emotional state. This method can be quite precise when working with clear and intense emotions. Modern approaches to measure emotions utilize various kinds of audio and video analysis. Most of methods are based or originate from facial action coding system [21] which Ekman and Friesen by turn originated from Carl-Herman Hjortsjö. This approach was later implemented by [22]. Another interesting paper might be [23]. Finally, we need to mention the approaches, based on the physiological measurements. Now, with current technology, we can measure emotional states in an unobtrusive way: heart rate measurements with blind ICA on face skin color, for example, show really high reliability. Based on those abilities, previously hard-to-use techniques can be used [5], [24].

Since this paper focuses on neural networks for emotions recognition, we introduce the most relevant to this study research in this field. The system was proposed by Ekman and colleagues in 1978 [21]. It measures facial expressions in terms of activity and underlying facial muscles. This approach was first automated by Bartlett et al. in 1996 [22]: they classified facial actions using image processing and machine learning techniques. From the mathematical point of view, this system utilizes 3 approaches: principal component analysis, local image features and optical flow based template matching. It is used now in several branches of behavioural science.

Next approach, probabilistic feature analysis, was proposed by Muelders et al. in 2005 [3]. It is based on the hypothesis that there are special relationships between the parts of the face function as a source of information in the facial perception of emotions. Here, the probabilistic feature model was introduced instead of linear models, which empirically extracted relevant facial features. It also formalizes a mechanism for the way in which information about separate facial features is combined in processing the face. The next approach, which was called "Graphical representation of the proportion of perceived particular emotion" was first proposed by Maja Pantic and colleagues in 2012 [4]. They presented a fully automatic recognition of facial muscle actions that compound expressions and explicit modelling of temporal characteristics.

The architecture, proposed in this paper, is close to system, proposed by Ekman and Freisen, but the type of equipment we use and the architecture of the neural network system are different.

3 Model Architecture

The choice of neural network model was performed as follows: we tested several existing neural network architectures and selected the one with better performance results (computational costs and recognition rate). The modular architecture of our model allows using the Kinect output (point cloud).

The architecture of our model is based on the notion of the self-organized map (SOM), proposed by Kohonen [25]. This kind of neural network is trained using unsupervised learning to produce a two-dimensional map of the input space of the training samples. The quality of SOM to use a neighbourhood function for preserving the topological properties of the input space is used in our simulations to create the similarity map of the IT cortex. First, we provide the justification of the applied approach and the description of the related studies.

To provide better understanding of the applied approach, we need to describe two main notions: the notion of self-organised map and the RBF-network. The RBF-networks were briefly described in the previous chapter. This part provides a detailed description.

The self-organized maps were introduced by Kohonen in 1990. The prototype for this network was the self-organization characteristics of the human cerebral cortex. Studies of the cerebral cortex showed that the motor cortex, somatosensory cortex, visual cortex and auditory cortex are represented by topologically ordered maps. These topological maps form to represent the structures sensed in the sensory input signals [26].

SOM detects regularities and correlations in its input and adapt their future responses to that input. The neurons of competitive networks learn to recognize groups of similar input vectors in such a way that neurons, which are located physically near each other in the neuron layer, respond to similar input vectors.

The main idea in the SOM learning process is that for each input vector the winner unit is selected. The winner unit (which is called best matching unit, BMU) and the nodes in its neighbourhood are changed closer to the input data. If the number of available inputs is restricted, they are presented re-iteratively to the SOM algorithm. The Kohonen' network is trained with the method of successive approximations. The sample SOM map is presented in Fig.2.

A radial basis function neural network (RBFN) is an artificial neural network that uses radial basis functions as activation functions. The typical RBFN architecture consists of three layers: an input layer, a hidden layer of J basis functions, and an output layer of linear output units. The activation values of the hidden units are calculated as the closeness of the input vector x_i to an I-dimensional parameter vector f_i^j associated with hidden unit u_j (Fig.3).

In a RBF network there are three types of parameters that need to be chosen to adapt the network for a particular task: the center vectors u, the output weights w_j, and the RBF width parameters σ.

The variety of the training algorithms for RBFNs exists. One of the possible training algorithms is gradient descent. In gradient descent training, the weights are adjusted at each time step by moving them in a direction opposite from the gradient of the objective function (thus

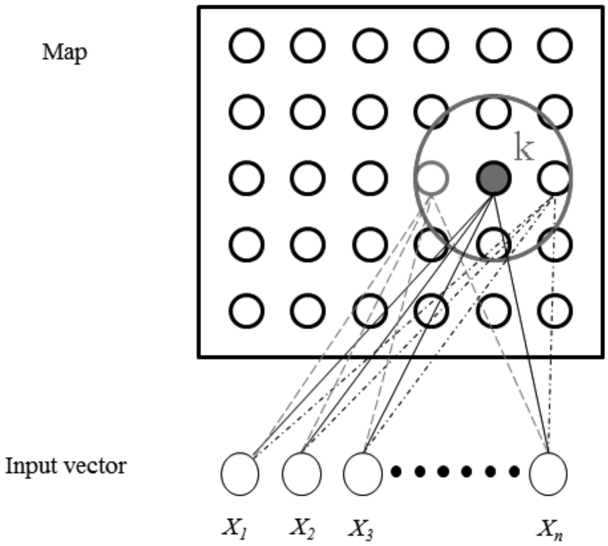

Figure 2. The self-organised map. x_i denotes the units of the input vector, *Map* represents the resulting SOM, k is the BMU, grey circle defines the neighbourhood of BMU

allowing the minimum of the objective function to be found). The conventional SOM algorithm has a number of restrictions, and the main one is its ability to deal only with the vectorized data. To solve this problem, a number of modifications of the conventional SOM have been proposed. We used one of these modifications as a basis for constructing our model.

Tokunaga and Furukawa have proposed a significant variation of the conventional SOM, called the modular network SOM (mnSOM) [8]. In their model, each vector unit of the conventional SOM is replaced by a functional module. These modules are arrayed on a lattice that represents the coordinates of the map. Authors regard the case of a multi-layer perceptron (MLP) module as the most commonly used type of neural network. This architecture was designed to keep the backbone algorithm of the SOM untouched. The algorithm of the mnSOM is a

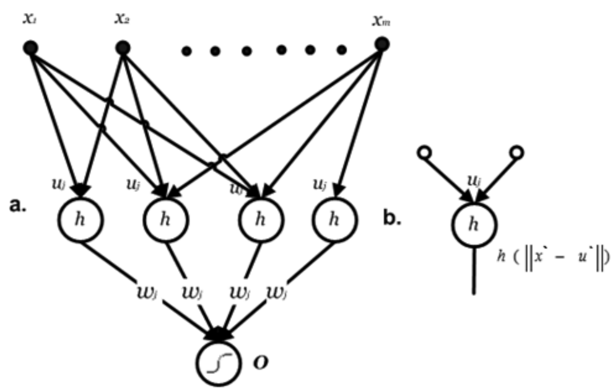

Figure 3. The structure of an RBF-network: a. x_i denotes the units of the input vector, o denotes the output of each units, u_j are the centres, σ defines variance, w_j define weights, and n is the number of hidden units, where j defines the j-th hidden unit. b. The components of the input vector x are compared in the each center u via the RBF h

generalization of a conventional SOM that inherits many properties from a conventional version of the algorithm and also adds several new original properties.

This architecture has the number of advantages. First, every module in the mnSOM has the capability of information processing and can form a dynamic map that consists of an assembly of functional modules. Second, the mnSOM combines supervised and unsupervised learning algorithms: at the MLP-level, the network is trained by a supervised learning algorithm, i.e., the back propagation at the MLP module level, while the upper SOM level is described in an unsupervised manner.

For the purposes of this study we used RBF network modules. The usage of RBFs instead of the MLPs adds the following properties to such a network while preserving the ability to form a dynamic map:

- there is no need for an algorithm for avoiding local minima;
- the network can recognize the object and store its representation

in its inner centre.

The generalized algorithm for processing the SOM of functional models can also be applied in this case.

The model of the main module of the proposed network represents a modification of the conventional SOM, where each vector unit of the conventional SOM is replaced by a functional RBF module. These modules are arrayed in a lattice that represents the coordinates of the feature map.

The architecture of the SOM of RBFs module has a hierarchical structure: it consists of two levels, which we will call the RBF-level and the SOM-level of the network. At the first level, the architecture of our network represents k RBF-networks, which are the modifications of the Poggio and Edelman network. Since each module represents a certain "functional feature" determined by the model architecture, the SOM-level the SOM of RBFs represents a map of those features.

The proposed network solves an approximation problem in a high-dimensional space. Recognizing an object is equivalent to finding a hyper-plane in this space that provides the best fitting to a set of training data. The training data represents a vector with coordinates of 2D projections of 3D objects, taken at each degree of rotation.

Let x_i denote the units of the input vector, o define the output of each RBF-module, u_j^k define the RBF centres, σ_j^k define the variance, w_j^k define weights, and n to be the number of hidden units, where j defines the j^{th} hidden unit and k defines the k^{th} RBF-module. Then the conventional SOM algorithm can be rewritten as follows.

In the first step, the weights w_j^k are defined randomly in the interval [0 0.5]. In the evaluative process, we calculate all outputs for all of the inputs in single RBF-unit according to the following rule:

$$o(x) = \sum_{j=1}^{n} w_j exp\{\frac{-(x-u_j)^2}{{\sigma_j}^2}\}. \qquad (1)$$

This calculation process is repeated for all of the RBF-units using the same input x. After evaluation of all of the outputs for all inputs, the errors for all the datasets are calculated:

$$E_i^k = \frac{1}{2}(y - \frac{1}{1+e^{-o(x)}})^2, \tag{2}$$

where y defines the desired output. E_i^k defines the error for the i^{th} dataset for the k^{th} module. The desired output equals 1 for all the RBF modules.

In the competitive process, the module that minimizes the error is determined as the winner module.

In the cooperative process, the learning weights are calculated using the neighbourhood function α, which decreases with the calculation time:

$$\alpha(r_i) = \frac{e^{\frac{-(r_i - r_v)}{2\xi_j^2}}}{\sum_{i-1}^{n} e^{\frac{-(r_i - r_v)}{2\xi_j^2}}}, \tag{3}$$

where r_i denotes the position of the i^{th} RBF-unit in the map space, r_v expresses the position of the module with the minimal error and ξ is the parameter of the neighbourhood function. The neighbourhood function area is decreased monotonically each epoch of learning.

In the adaptive process, all of the modules are updated by the back-propagation learning algorithm:

$$\Delta w_j^k = \eta \partial E_i^k / \partial w_j^k (t-1) \tag{4}$$

and

$$w_j^k(t) = w_j^k(t-1) + \Delta w_j^k \alpha(r_i). \tag{5}$$

So (4) can be rewritten as follows:

$$\Delta w_j^k = \eta(y - \frac{1}{1+e^{-o(x)}})(y - \frac{1}{(1+e^{-o(x)})^2})e^{\frac{-(x-u_j)^2}{2\sigma_j^2}}. \tag{6}$$

The centres of the RBF-units are updated according to the following rules

$$\Delta u_j^k = \eta \partial E_i^k / \partial u_j^k (t-1) \tag{7}$$

and

$$u_j^k(t) = u_j^k(t-1) + \Delta u_j^k \alpha(r_i) \tag{8}$$

i.e.,

$$\Delta u_j^k = \eta(y - \frac{1}{1+e^{-o(x)}})(y - \frac{1}{(1+e^{-o(x)})^2})w_j \frac{x-u_j}{2\sigma_j^2} e^{\frac{-(x-u_j)^2}{2\sigma_j^2}}. \tag{9}$$

The learning is repeated until all of the modules are updated. Training continues until the network reaches a steady state.

In order to investigate the ability to classify complex 3D objects, such as faces, we extend our SOM of RBFs model by adding a hierarchical pre-processing module. The first level of the model consists of local orientation detectors. These detectors are Gabor-like filters [27].

The model contains orientation detectors for four preferred orientations. The next level contains position-invariant bar detectors. The combination of the features, extracted at earlier stages in the proposed architecture are then processed with the RBFxSOM.

The input image is divided into small overlapping patches, which are then processed with the neurons of the network, resembling simple cells of the cortex [28].

Therefore, on the first layer of the model, patterns on the input image (250×250 pixels) are first filtered through a layer (S1) of simple cell-like receptive fields. We use Gabor filters with four orientations (0, 45, 90 and 135 degrees) with diameter of 11 pixels, in steps of 2 pixels. S1 filter responses are dot products with the image patch falling into their receptive field. Receptive field (RF) centres densely sampled the input retina.

A Gabor filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be appropriate for pattern recognition. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The Gabor filters are self-similar: all filters can be generated from one mother wavelet by dilation and rotation.

J. G. Daugman discovered that simple cells in the visual cortex of mammalian brains can be modelled by Gabor functions. Thus, image analysis by the Gabor functions is somewhat similar to visual perception in the human visual system [29].

The impulse response of the Gabor filter is defined by a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property, the Fourier transform of a Gabor filter′s impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function. The filter has a real and an imaginary component representing orthogonal directions [27]. The two components may be formed into a complex number or used individually.

A set of Gabor filters with different frequencies and orientations, called filter banks, are usually used for extracting feature information from the image. The filters are convolved with the signal, resulting in a so-called Gabor space. This process is closely related to processes in the primary visual cortex [27]. Jones and Palmer showed that the real part of the complex Gabor function is a good fit to the receptive field weight functions found in simple cells in a cat's striate cortex [28]. Relations between activations for a specific spatial location are very distinctive between objects in an image. Furthermore, important activations can be extracted from the Gabor space in order to create a sparse object representation.

Responses of the simple cells, extracted on the first layer of the network, are then pooled into the next layer of the complex-like cells, which can be compared with the complex cells in areas V2-V4.

As we use several successive layers of simple and complex cells, let s_i^1 denote the first layer of the network, i.e. the first layer of simple cells. Cells in the next layer pool S1 cells with the non-linear maximum-like operation and thus can be referred to as complex cells, or the first layer of complex cells.

Therefore, c_i^1 for i-th neuron during the presentation of image patch v is defined as follows:

$$c_i^1 = max_{j \in A_i}\{v_\alpha(j) \circ \xi_j\}, \tag{10}$$

where A_i – the set of the afferent i, $\alpha(j)$ – the centre of the receptive field of the afferent j, $v_\alpha(j)$ – a square normalized image patch with the centre in $\alpha(j)$, corresponding to the receptive field ξ_j (square normalized) input signal j, and $\circ$ is the dot product.

The responses of C1 cells are then combined in higher layers. We define two ways of processing them:

- C1 cells tuned to different features can be combined to yield S2 cells that responded to co-activation of C1 cells tuned to different orientations.
- C1 cells tuned to same features can be combined to yield C2 cells responding to the same feature as the C1 cells, but with bigger receptive fields.

The second layer of simple cells S2 contains a set of features, all pairs of orientations of C1 cells looking at the same part of space.

The second layer of complex cells C2 represents the pooling stage with the parameter p that defines the strength of pooling. It can be defined as follows:

$$c_j^2 = \sum_j \frac{\exp(p \cdot |s_j|)}{\sum_k \exp(p \cdot |s_k|)} s_j, \tag{11}$$

which performs a linear summation (scaled by the number of afferents) for $p = 0$ and the maximum operation for $p \to \infty$.

The responses of C2 units feed into the RBFxSOM module.

4 Results and Discussion

The network output represents the activation map, the activation of each module shows the belonging of the detected expression to one of five basic emotions. For the purposes of this study, we selected five basic emotions, which are located on a square plane, divided into 25 parts. The winning module represents the most plausible emotion. This approach allows defining the most plausible emotion or emotions (since the most active module can be defined between two emotions). In this paper, we used only five emotions, but the usage of a larger number of emotion labels is also possible (Fig.4).

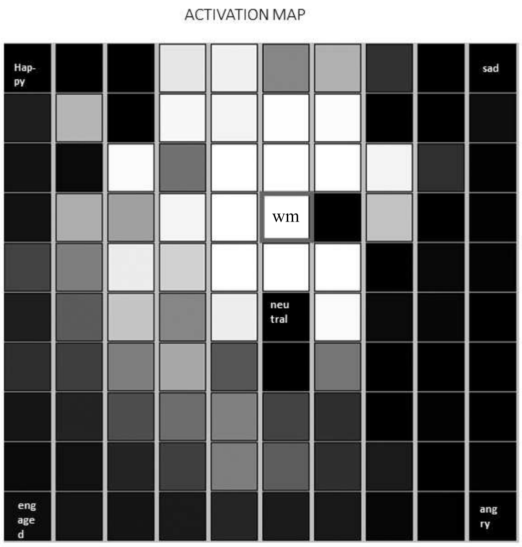

Figure 4. The output of an RBF-network: the activation map describes the performance of the neural network architecture: the networks activated one of five basic emotions (black color refers to absence of activation and the white color refers to active module). In this case, the most active or winning module (the square marked by "wm") is closer to "neutral" expression

We described the properties of the proposed neural architecture for hierarchical visual perceptual processing, composed of modules resembling human visual system. By introducing this architecture, our model appeared to be capable of performing recognition and classification of simple emotions and creating a similarity map of these emotions.

From the point of view of neural network architecture, the performance of the proposed model can be compared with the HMAX model [30], [31], since it inherits the general hierarchical architecture of HMAX. These models would perform similarly within the image recognition task, because they share the same type of the image processing module. That type of processing mechanism is an extension of classical models of complex cells built from simple cells, consisting of a hierarchy of layers with linear (S-units, performing template matching) and non-linear operations (C-units, performing a maximum-like pooling operation) [32]. However, the proposed model differs from HMAX in two major points. First, the HMAX model utilises the different type of classification mechanism (which is built with the separate module), while our model can perform classification at the IT level. This essential difference is achieved by using of the SOM of functional modules for the IT-level recognition, which allows creating of the similarity map of the objects. Second, we use abstract 3D objects, presented in projections during the rotation of the object, which adds the rotation invariance property to the model, while preserving the invariance to scale and translation.

From the point of view of emotion recognition system, current approach is close to the one proposed by Paul Ekman [14], but differs in the type of machine learning techniques, equipment and the number of emotions (he utilises six basic emotions: anger, disgust, fear, happiness, sadness, and surprise, while we use only five expressions: sad, angry, neutral, happy, engaged).

The research, described in this work, constitutes the tiny part of the spacious area of visual object recognition. It includes the description of main research in the domain of emotion recognition with the means of computer vision system with the emphasis on the neural network architectures. It could be further extended in order to add the number of emotions and to implement different machine learning techniques.

References

[1] J. Cohn. *Foundations of human computing: facial expression and emotion*, ICMI 06: Proceedings of the 8th international conference

on Multimodal interfaces, ACM, New York, NY, USA, pp. 233–238, 2006.

[2] Ying-Li Tian, Takeo Kanade, Jeffrey F Cohn. *Facial expression analysis.* In: Handbook of face recognition, pp. 247–275, Springer, 2005.

[3] M. Meulders, P. De Boeck, I. Van Mechelen, A. Gelman. *Probabilistic feature analysis of facial perception of emotions.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 54, 2005, pp. 781–793.

[4] B. Jiang, M. F. Valstar, M. Pantic. *Facial Action Detection using Block-based Pyramid Appearance Descriptors.* Proceedings of the ASE/IEEE International Conference on Social Computing (SocialCom 2012). Amsterdam, The Netherlands, September 2012.

[5] Y. Hsieh and C.-L. Chin. *The emotion recognition system with heart rate variability and facial image features,* in IEEE International Conference on Fuzzy Systems (FUZZ), 2011, pp. 1933–1940.

[6] Marian Stewart Bartlett, Gwen Littlewort, Mark G. Frank, et al. *Fully automatic facial action recognition in spontaneous behavior,* 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 223–230.

[7] N. Efremova, N. Asakura, T. Inui, N. Abdikeev. *Inferotemporal network model for 3d object recognition*, The proceedings of the International Conference on Complex Medical Engineering IEEE/ICME, 2011, pp. 555–560.

[8] K. Tokunaga, T. Furukawa. *Modular network SOM*, Neural Networks, vol.22, Nr.1, Pergamon, 2009, pp. 82–90.

[9] T. Furukawa. *Som of SOMs*, Neural Networks, 22(4), 2009, pp. 463–478.

[10] James M. Hutchinson, Andrew W. Lo, Tomaso Poggio. *A nonparametric approach to pricing and hedging derivative securities*

via learning networks, The Journal of Finance, Vol. 49, No. 3, Papers and Proceedings Fifty-Fourth Annual Meeting of the American Finance Association, Boston, Massachusetts, January 3-5, 1994. (Jul., 1994), pp. 851–889.

[11] Juergen Schmidhuber. *Deep learning in neural networks: an overview*, http://arxiv.org/abs/1404.7828, 2014.

[12] A. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam Publishing, 1994, 336 p.

[13] C. Becker-Asano, I. Wachsmuth. *Affective computing with primary and secondary emotions in a virtual human. Autonomous Agents and Multi-Agent Systems*, vol. 20, no. 1, pp. 32–49, 2009.

[14] Tim Dalgleish, Mick Power (Eds.) *Handbook of Cognition and Emotion*. New York, NY: John Wiley and Sons Ltd., 1999, ch. Basic Emotions, pp. 45–60.

[15] E.Y. Bann, J.J. BRYSON. *The conceptualisation of emotion qualia: Semantic clustering of emotional tweets*, Proceedings of the 13th Neural Computation and Psychology Workshop, 2013.

[16] W. Wundt. *Classics in the history of psychology.* ttp://psychclassics.yorku.ca/Wundt/Physio/

[17] Albert Mehrabian. *Framework for a comprehensive description and measurement of emotional states.* Genet Soc Gen Psychol Monogr. 1995 Aug;121(3):339–361.

[18] Hugo Lövheim. *A new three-dimensional model for emotions and monoamine neurotransmitters.* Medical Hypotheses, Vol. 78, Issue 2, 2012, pp. 341–348.

[19] A. Ortony, G.Clore, A.Collins. *The cognitive structure of emotions.* Cambridge: Cambridge University Press, 1988, 224 p.

[20] M. Bradley, P. Lang. *Measuring emotion: The self-assessment manikin and the semantic differential,* Journal of Behavior Therapy and Experimental Psychiatry, vol. 25, no. 1, pp. 49–59, 1994.

[21] P. Ekman, W. Friesen. *Facial action coding system: A technique for the measurement of facial movement*, Consulting Psychologists Press, Palo Alto, 1978.

[22] M.S. Bartlett, J.S. Hager, P. Ekman, T.J. Sejnowski. *Measuring facial expressions by computer image analysis*, Psychophysiology, vol. 36, no. 2, pp. 253–263, 1999.

[23] Yongmian Zhang, Qiang Ji. *Active and dynamic information fusion for facial expression understanding from image sequences*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 699–714, 2005.

[24] D.S. Quintana, A.J. Guastella, T. Outhred, I. B. Hickie, A.H. Kemp. *Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition*, International journal of psychophysiology: official journal of the International Organization of Psychophysiology, 86(2), 2012; DOI: 10.1016/j.ijpsycho.2012.08.012.

[25] Teuvo Kohonen. *Self-organizing maps. Berlin: Springer-Verlag*, 2001, 490 p.

[26] Andries P. Engelbrecht. *Computational intelligence: an introduction books*. Wiley, 2007, 628 p.

[27] J. Daugman. *Complete discrete 2-d gabor transforms by neural networks for image analysis and compression*, IEEE Trans on Acoustics, Speech, and Signal Processing, vol. 36, issue 7, 1988, pp. 1169–1179.

[28] J. Jones, L. Palmer. *An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex*, Journal of Neurophysiology, 58 (6), 1987, pp. 1233–1258.

[29] Hans G. Feichtinger, Thomas Strohmer (Eds.). *Gabor analysis and algorithms: theory and applications*. Springer, 1998, 496 p.

[30] M. Riesenhuber, T. Poggio. *Hierarchical models of object recognition in cortex*, Nat Neuroscience, 2, 1999, pp. 1019–1025.

[31] M. Riesenhuber, T. Poggio. *Models of object recognition*, Nat Neuroscience, Suppl. 3, 2000, pp. 1199–1204.

[32] K. Fukushima. *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*, Biological Cybernetics, 36(4], 1980, pp. 193–202.

Veaceslav Albu, Svetlana Cojocaru,

Received March 18, 2015

Institute of Mathematics and Computer Science
Academy of Sciences of Moldova
5 Academiei str., Chişinău, MD-2028, Moldova

E–mails
Veaceslav Albu: vaalbu@gmail.com
Svetlana Cojocaru: Svetlana.Cojocaru@math.md