

Evaluation of Human Emotion from Eye Motions

Vidas Raudonis

Department of Control Technologies, KTU
Kaunas, Lithuania

Gintaras Dervinis

Department of Control Technologies, KTU
Kaunas, Lithuania

Andrius Vilkauskas

Faculty of Mechanical Engineering and Mechatronics,
KTUK aunas, Lithuania

Agne Paulauskaite - Taraseviciene

Department of Business Informatics, KTU
Kaunas, Lithuania

Gintare Kersulyte - Raudone

Department of Control Technologies, KTU
Kaunas, Lithuania

Abstract— The object of this paper is to develop an emotion recognition system that analysis the motion trajectory of the eye and gives the response on appraisal emotion. The emotion recognition solution is based on the data gathering using head mounted eye tracking device. The participants of experimental investigation were provided with a visual stimulus (PowerPoint slides) and the emotional feedback was determined by the combination of eye tracking device and emotion recognition software. The stimulus was divided in four groups by the emotion that should be triggered in the human, i.e., neutral, disgust, exhilaration and excited. Some initial experiments and the data on the recognition accuracy of the emotion from eye motion trajectory are provided along with the description of implemented algorithms.

Keywords—Emotional status; eye tracking; human computer interaction; virtual human

I. INTRODUCTION

Recently emotional analysis has become an important line of research in human computer interaction (HCI), virtual human [1] or creating domestic robots [19] emotional agents [11]. Emotion recognition is one type of sentiment analysis that focuses on identifying the emotion in images of facial expression, measurements of heart rates, EEG, etc [3], [20]. Automatic recognition of emotion from the natural eye motions is an open research challenge due to the inherent ambiguity in eye motion related to certain emotion [2]. Various techniques have been proposed for emotion recognition. They include an emotion lexicon, facial expression models, EEG measurements with combination of knowledge-based approaches.

Humans interact with each other mostly by speech. However, regular human speech is often enriched with certain emotions. Emotions are expressed by visual, vocal and other physiological ways. There is evidence that certain emotion skills are part of what is called human intelligence that helps to understand better each other [18]. There exist an old saying “the eyes are the mirror of the human soul”. The facial expressions is only one of the ways to display an emotion, the emotion can be deduced from the eye gaze [4], [5]. If there exist a need for more affective human-computer interaction, recognizing of the emotional state from his or her face could

prove to be an invaluable tool. The accurate emotion recognition is important in the field of the virtual human design, which should be with lively facial expression and behaviours, present motivated responses, to environment and intensifying their interaction with human users.

Eye-tracking can be especially useful investigating the behaviour of the individuals with physical and psychological disorders. Such studies typically focus on the processing of emotional stimuli, suggesting that eye-tracking techniques have the potential to offer insight into the downstream difficulties in everyday social interaction which such individual’s experience [6], [7]. Studies such as [5], [8], [10] describe how eye analysis can be used to understand human behaviour, the relationship between pupil responses and social attitudes and behaviours, and how it might be useful for diagnostic and therapeutic purposes.

The emotional part closely correlates to the eye based HCIs [10], [11], [12]. For example it is known, that increases in the size of the pupil of the eye have been found to accompany the viewing of emotionally toned or interesting visual stimuli. Emotional stimuli may also be used to attract the user’s attention and then divert to a control scheme of a HCI [9], [12]. E.g., eye tracking can be used to record visual fixation in nearly real-time to investigate whether individuals show a positivity effect in their visual attention to emotional information [13], [16], [17]. Different viewing patterns can be detected using strongly stimulant pictures, as in the study [14], [15].

This paper describes a system that uses a machine learning algorithm such as artificial neural network in order to detect four emotions: (a) neutral, (b) disgust, (c) funny and (d) interested from the eye motions. The text bellow presents our work on the development of emotional HCI. The proposed emotional status determination solution is presented based on the data gathering from visual sensors, providing the participants with a visual stimulus (strongly stimulant slides). Some initial experiments and the data on the recognition accuracy of the emotional state based on the gaze tracking are provided along with the description of implemented algorithms.

II. TECHNICAL IMPLEMENTATION OF EMOTION RECOGNITION METHOD

Our gaze tracking device is shown in the figure 1 (left). It consists of one mini video camera that is directed to the user's eye and it records the eye images. Next to the camera the infrared (IR) light source is attached which illuminates the user's eye. The camera is design to capture the user's eye in the IR light. Such illumination does not disturb the user, because it is invisible for the naked human eye and gives mostly stable IR luminosity. The eye image captured in the near infrared light always shows dark (black) pupil that is caused by the absorption of infrared light of the eye tissues. Thus, the task of pupil detection can be limited or simplified to the searching goal of the dark and round region in the image. The camera and IR led is fixed on the ordinary eye glasses. Video camera is connected to the personal computer through USB port. Computer records the eye images, estimates the pupil size and coordinates, and draws the attention maps.

The constant adaptation of the eye pupil to the random variation of the illumination condition and different pupil sizes between user groups are the main challenges for any gaze tracking algorithm. In this work we propose the algorithm which is differ from well known method such as Circular Hough Transform or Viola-Jones. First method is based on the voting procedure that is carried out in a parameter space, from which circle candidate is obtained as local maxima in so called accumulator space. When the pupil size or circle diameter is unknown, the algorithm based on the Hough transform computes the accumulator space for every possible circle size and the circle candidate is obtained as global maxima. Such computation task takes relatively a lot of time. Viola-Jones based object tracking algorithm is capable to tracking the object of any shape and size based on the learned features. Unfortunately, it cannot be used when accurate measurements of the pupil size are required. Proposed pupil detection algorithm is based on adaptive thresholding of grayscale image. It enables the precise detection of pupil in the different layers of gray color, regardless of how the lightening is changing. The diameter measurement of the dark region is compared with the limits of the possible minimal and maximal pupil size.

Our gaze tracking algorithm finds the rough pupil center in the iterative manner and it executes the logical indexing on the gray level image using certain threshold of grayness value, which is variable (adaptive). The algorithm is preceded in two major steps. The rough pupil center is obtained in the first, and the accurate coordinates and the diameter of the pupil is obtained in the next step. The detection of the pupil is executed in the region of interest (sliding window) that is defined by three parameters Length, Width and the center coordinates (Cx, Cy). All logical indexing operations are executed in the region of interest. At the detection beginning, the center of eye image is used as starting position for the region. All other positions are defined by located pupil center in the last frame. The values of all pixels which are higher than threshold are equalized to one, otherwise to zero. The threshold Θ is increased or reduced from the default gray level value Θ_0 according to certain conditions which are defined by the current measured diameter r of the object of interest. The threshold Θ is increased by step

$\Delta\Theta$, when current diameter of the object is smaller than the possible limits of the pupil size [Rmin Rmax] and otherwise Θ is decreased by step $\Delta\Theta$ if these limits are exceed. Where Rmin – is the minimal pupil size and Rmax – is the maximal pupil size. These parameters are measured in image pixels. The variation limits of the pupil size of the human eye are taken according to analytical research [25]. The threshold value does not change if current measurement r is between limits. The rough pupil center (coordinates Cx and Cy) is computed in the next step. Used notations: N, M – the number of columns and rows of the eye image, $d(i,j)$ – Euclidean distance between two candidate points, x, y – the coordinates of the candidate pixels, flag and count – is used for iteration purposes and σ - standard deviation of Euclidean distances. More about eye tracking algorithm is published in [12]. The schematic pseudo code of the proposed eye tracking method is shown in the tables 1 and 2.

TABLE I. THE PROPOSED ALGORITHMS FOR EYE PUPIL TRACKING AND REGISTRATION

Pupil detection based on adaptive gray level threshold	
1	//The pupil center is extracted in the grayscale image of the eye $I(u,v)$, where $u,v \in N,M$
2	while ($flag = 0$) do
3	//Collect candidate points
4	$count = count + 1, k = 0;$
5	for $u, v \in N, M$ do
6	if $I(u, v) \leq \Theta$ then
7	$k = k + 1, x_k = u, y_k = v;$
8	for $i, j \in k$ do
9	$d(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$
10	//Make measurements of the point cloud
11	$r = \max_{i \in k}(\max_{j \in k}(d(i, j)))$
12	$C_x, C_y \leftarrow x(k), y(k)$
13	//Verify the conditions
14	if $R_{min} < r < R_{max}$ then , $flag = 1,$
15	else if $R_{min} > r$ then , $\Theta = \Theta + \Delta\Theta$
16	else if $R_{max} > r$ then , $\Theta = \Theta - \Delta\Theta$

TABLE II. THE PSEUDO CODE FOR ACCURATE DETECTION OF PUPIL CENTER

Accurate pupil center detection	
1	//Accurate pupil center $C' = (C'_x, C'_y)$ of the point cloud x_k and y_k is computed using nonlinear least squares approach
2	//Perform data filtration on the candidate points
3	//Collect candidate points
4	$d_i = \sqrt{(\bar{x} - x_i)^2 + (\bar{y} - y_i)^2}$
5	$\bar{d}, \sigma_d \leftarrow d(i), n = 0$
6	for $i \in k$ do
7	if $\bar{d} - 3\sigma_d \leq d(i) \leq \bar{d} + 3\sigma_d$ then
8	$n = n + 1, x'_n = x_i, y'_n = y_i$
9	//Fit smallest surrounding circle to the filtrated data
10	$\mathbf{u} = (C'_x, C'_y, r'), \mathbf{X} = \begin{pmatrix} x'_1 & \dots & x'_n \\ y'_1 & \dots & y'_n \end{pmatrix}, \mathbf{X} = \mathbf{R}^2$
11	$f_j(C'_x, C'_y, r') = \ \mathbf{C}' - \mathbf{X}_j\ ^2 - r^2$
12	$C'_x, C'_y, r' \leftarrow \sum_{j=1}^m f_j(\mathbf{u})^2 = \min_{\mathbf{u}} \sum_{j=1}^m f_j(\mathbf{u})^2$

The computational actions of the proposed algorithm, taken for pupil detection, are shown in the figure 1. The threshold Θ is iteratively changed and the distance between two extremities of the selected pixels is measured. These pixels usually appeared in opposite direction from each other. The pixels are marked with lighter color in figure 1. The threshold value is decreased from 130 to 120 gray level and the same measurements are applied to the new selected pixels. Gray threshold is reduced until the distance between extremities is in the possible variation range of the pupil size. The grayness value was reduced to 90 of the gray level in this case. If selected pixels do not satisfy the predefined condition then Θ is changed. The threshold is increased when the distance between two extremities is smaller than minimum limit of the pupil size variation and decreased when this distance exceeds predefined maximum limit. The least square method is applied to selected pixels and the new accurate pupil center is obtained. The possibility of the proposed algorithm to change the threshold value adaptively overcomes several detection difficulties, such as, the variation of an ambient luminosity, constantly adapting pupil size and noise.

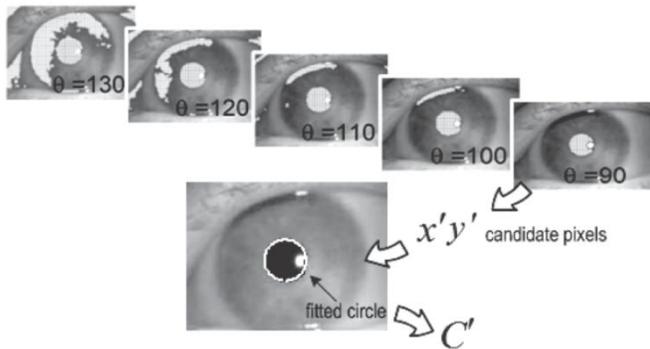


Fig. 1. The process of the adaptively changing threshold

Most eye movement is executed without awareness, i.e., there is no voluntary control. A sufficient amount of studies worldwide prove an interrelation between pupil size, pupil motions and a person's cognitive load or stress. The eye movements and the size of eye pupil strongly depend on the various factors of the environment and mental state of person. To recognize in which mental state is the person, the authors of this article have developed an eye pupil analysis system that is based on application of artificial neural network (ANN). The relation between measured inputs and the emotional human states is not known precisely, i.e., is it linear or nonlinear. Therefore, in the problems when linear decision hyper-planes are no longer feasible, an input space is mapped into a feature using hidden layers of the neural network. The mathematical model based on ANN is selected in order to construct a non-linear classifier.

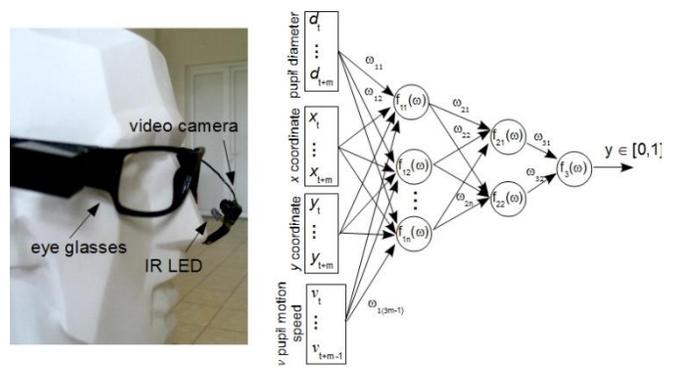


Fig. 2. The ANN model and hardware implementation of experimental investigation

Our experimental emotion detection system is illustrated in figure 2. In addition to the gaze tracking hardware/software the system runs the proprietary real-time emotion analysis toolkit based on an Artificial Neural Networks. We have implemented a 3 layer ANN: consists of 8 neurons, the second of 3 neurons and the output layer of 1 neuron. ANN networks have a variable input number and are trained based on 3 features: the size of the pupil, and the position of the pupil (coordinates x, y) and motion speed v of the eye. For each emotional state we develop different neural network. The artificial neural network can be described using the following formulas:

$$X = ([d_t \dots d_{t+m}][x_t \dots x_{t+m}][y_t \dots y_{t+m}][v_t \dots v_{t+m}]); \quad (1)$$

$$y'_j = f(\sum_{i=1}^{3m} X_i \omega_{ji}); \quad (2)$$

$$y''_k = f(\sum_{j=1}^n y'_j \omega_{kj}); \quad (3)$$

$$y = f(\sum_{k=1}^2 y''_k \omega_{k2}); \quad (4)$$

where, t – is the current sample, X – the input vector of the artificial neural network, y – output, ω – weights of the neural network, d – is the diameter of the recognized pupil, x_t and y_t are the coordinates of the pupil center and v_t is the speed of eye pupil motion. Decision is made by selecting maximal value from four outputs of the artificial neural networks.

III. EXPERIMENTAL SETUP, INVESTIGATION AND RESULTS

At this initial stage of the evaluation we have chosen to analyze 4 very common emotions: neutral (regural, typical state), disgust, funny state and interest state. The emotion analysis system was evaluated on 30 people (20 males, 10 females, age ranging from 24 to 42). All participants were presented with a close-up (field-of-view consisted mostly of the display) PowerPoint slideshow consisted of various photographs sorted on the type of emotion they were supposed to invoke (the playback time limit was 3 minutes for each of the emotional photo collection (same number of photos for each emotion)). The images were selected based on consulting with expert of human psychology. Up to 30 samples (pupil size, and x, y coordinates) are recorded during the one second.

In total there is more than 600 thousands of samples (Number of participants * Number of emotions * Minutes per emotion * Number of seconds * Number of per samples per second) which were divided into training and testing data in the proportions respectively, 40% and 60%.

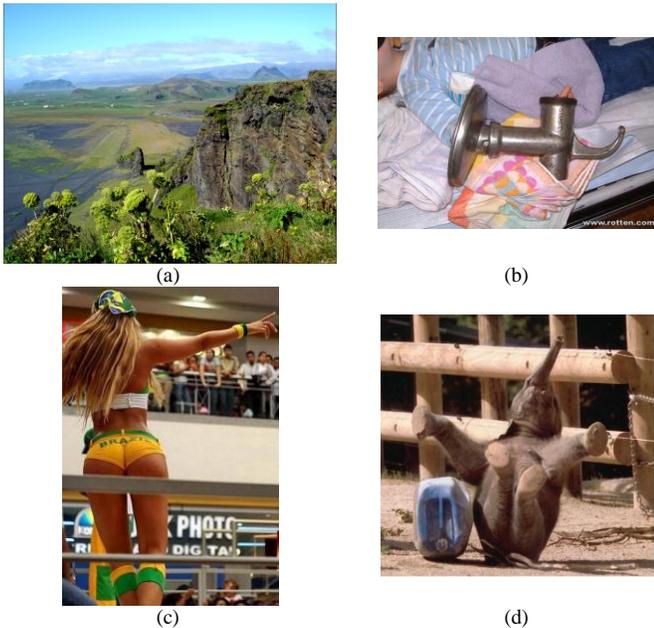


Fig. 3. The examples of visual stimulus which should invoke in the person four different emotions: a) neutral, b) disgust, c) interest and d) funny states.

After each set of emotional pictures there were 30 second pauses in the automated slideshow. During the experiment we have registered the size of the pupil, the coordinates of the center of the pupil in the video frame, as well as movement speed and acceleration. Figure 4 illustrates the fragment of the experimental analysis (6 people) on the size variation of eye pupil based on a current emotion (on the left) and size dispersion of eye pupil based on a current emotion (on the right).

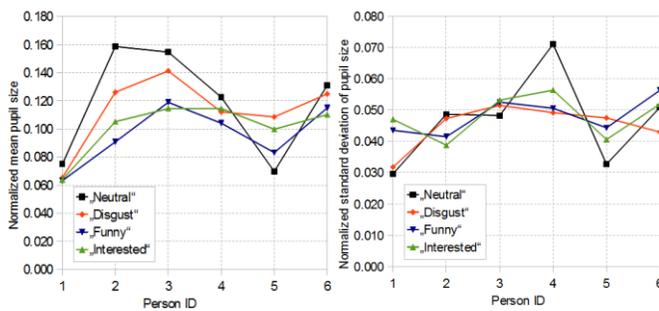


Fig. 4. The bar graph of relationship between average pupil size and the emotional reaction of the person (left), relationship between standard deviation of pupil size and emotional reaction (right).

The figure confirms the fact that the changes of the emotional stage can be recorded by observation of the small eye movements and the variation of the human eye pupil. Different emotional stimuli evoked different pupil sizes to different participant. For example, the pupil size of the first participant who was stimulated by the neutral visual stimulus is

28% bigger than pupil size that was measured when participant was stimulated with interested stimulus (see fig. 4 left). The right figure 4 shows the relation between variation of the pupil size in the time and the emotional stimulus.

The humans emotionally react differently to the different visual emotional stimulus. There can be, that one person feels the same emotions when he observes neutral stimulus and, another person, when he is stimulated with a funny stimulus. Every person interprets differently the emotional stimuli based on the life experience, emotional state at the beginning of the experiment. The correct answer cannot be given concerning on the automatically recognized emotion based only on the pupil size and variations, because pupil size can be affected by the general illumination, stress, physiological human properties and starting emotional status. Proposed emotion sensitive system uses not only information about the eye pupil size, but the classification features are enriched with the coordinates and motion speed.

Figures 5 and 6 illustrates an attention maps, which is computed based on the coordinate variation of the eye pupil center in the two dimensional space (measurement unit – normalized pixels). The attention map consists of green circle which radius depends on the time spent to observe certain part of the visual stimuli and the certain coordinates of the attention point. The attention map of the first participant is shown in the figure 4 and in the figure 6 the attention map of the second participant is shown. Overall registration period of the center of eye pupil is divided into four parts based on the shown emotional stimuli. The attention map shown in the figure 5 is computed from the data when participant were stimulated with a) neutral, b) disgust, c) funny and d) interested visual stimuli.

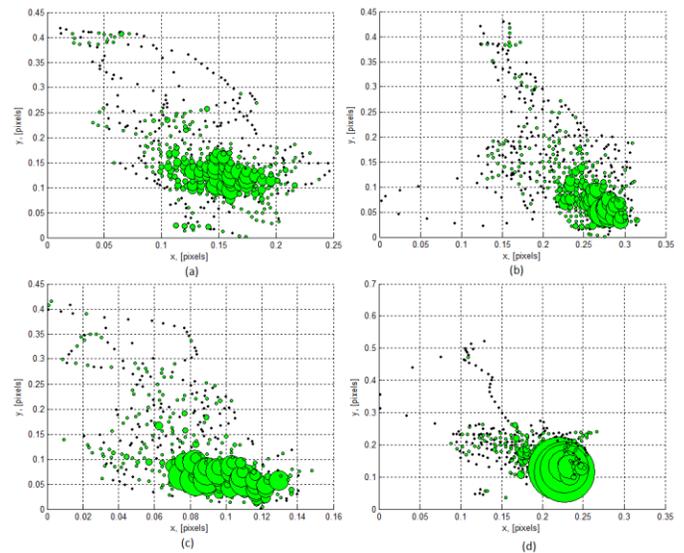


Fig. 5. The average variation of the attention point of the first experiment participant when he was stimulated with a) neutral, b) disgust, c) funny and d) interested stimuli

From the attention maps can be noticed, that the main attention concentration points are spread differently in two dimensional space due the different emotional stimulation. The attention points correlates with felt human emotion. The distribution of the attention points rely on the shown content of

the visual information and the arrangement information on the screen. For example, all participants of the experimental investigation tend to avoid certain part of the disgust images.

The high distribution of the attention points has the data acquired during emotional stimulation using neutral images (see fig. 5a). The natural nature such as mountains and forest is captured in the recent visual stimuli; therefore, it can partly explain such big distribution of the attention points. The attention map of the second participant is slightly different in comparison with the attention maps of the first participant. When second participant was stimulated with the neutral stimulus, he was less concentrated then, when he was stimulated with funny stimulus (see fig. 6 a and c). It depends on the person's individuality, i.e., he finds that the funny stimulus is more interested (that shows the overlapping attention points in the figure 6c).

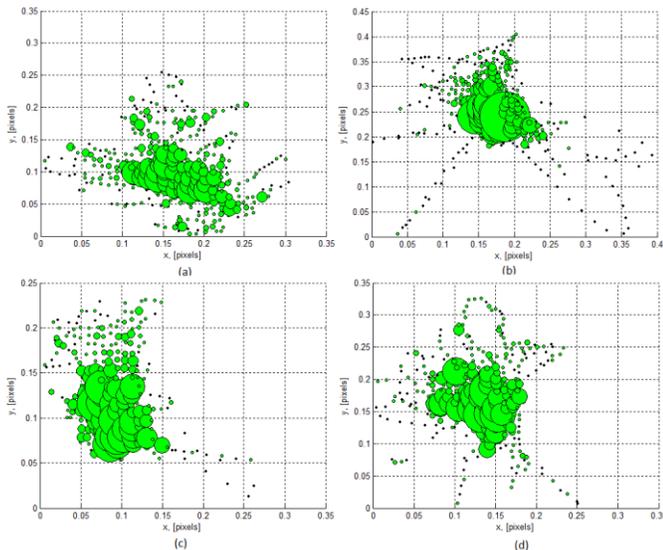


Fig. 6. An illustration of the average variation of the attention point of the second experiment participant when he was stimulated with a) neutral, b) disgust, c) funny and d) interested stimuli

Figure 7 illustrates the variation of the typical movement speed of eye pupil (a fragment of 6 persons) depending on the emotional stimuli. The motion speed of the eye pupil is measured in the pixels per second. The motion speed is presented on the vertical axis and the ID of the experiment participant is presented on the horizontal axis. The different color of the curve represents the different emotional stimuli. As it was with the variation of the pupil size, the motion speed depends on the person, on his starting emotional status, life experience and etc. For example, the motion speeds which were computed for the first participant differs in the relatively small range, i.e., up to 10%. While, the motion speeds of the sixth participant differs more than 46%. Such big difference may appear because of different cognitive capabilities, curiosity or usefulness of the presented information for the person (see fig. 7).

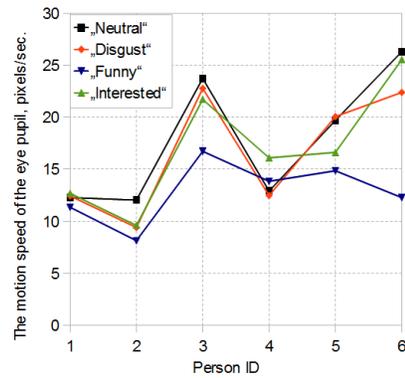


Fig. 7. The relationship between average speed of pupil motion and the emotional stimuli

Figure 8a illustrates the functional relationship between the number of feature samples and the recognition accuracy of different emotions. The overall best recognition accuracy (~90%) was achieved when we used 18 samples per feature. This means that the system can determine the emotion with a 2 second delay with approximately 10 % of deviation. The functional relationship shown in the figure 8a represents the accuracy graph of the one participant. The bar graph shown in the figure 8b represents the average recognition accuracy along the participants. Each bar represents the recognition rate for different emotion. Best recognition accuracy (90.27%) is reached when funny emotion is classified. Up to 16% of recognition fault is generated when system recognize the neutral emotion.

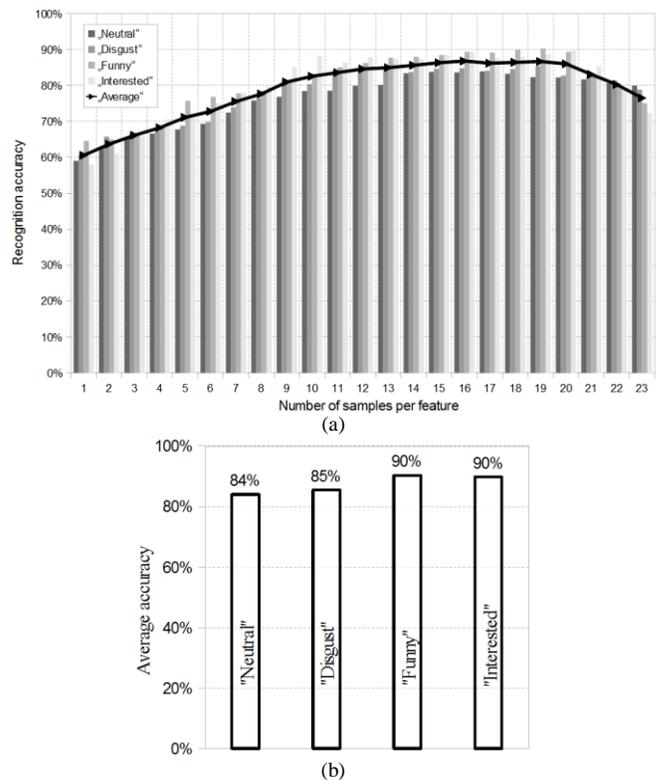


Fig. 8. The functional relationship between recognition accuracy and the number of samples per feature (a) and the bar graph of average accuracy (b)

IV. CONCLUSIONS AND FUTURE WORK

Experimental investigation has approved the fact, that the emotional state is individual and it depends on the persons cognitive perceptions. Although, it is possible to design a system based on the computational intelligence to recognize and detect certain emotional state of the human. The results of the experiments have shown that it is possible to detect the certain emotional state with up to 90% of recognition accuracy. Best recognition accuracy (90.27%) is reached when funny emotion is classified. Up to 16% of recognition fault is generated when system recognize the neutral emotion. The system can determine the emotion with a 2 second delay with approximately 10 % of deviation in average. Therefore, emotion recognition system uses 18 time samples per feature.

Future work will involve the application of remote eye tracking system that should allow recording more natural emotional responses to the visual and audio stimulation. There exist more than four emotional conditions; therefore, the multi-class problem will be solved using support vector machine and decision trees.

ACKNOWLEDGMENT

The works were funded by the Lithuanian Ministry of Education and Science according the project "Micro-sensors, micro-actuators and controllers from mechatronic systems" (GO-SMART, VP1-3.1-SMM-08-K-01-015).

REFERENCES

- [1] Klin A, Schultz R & Cohen D (2000). Theory of mind in action: developmental perspectives on social neuroscience. In *Understanding Other Minds: Perspectives from Developmental Neuroscience*, ed. Baron-CohenS, Tager-FlusbergH & CohenD, 2nd edn, pp. 357–388. Oxford University Press, Oxford .
- [2] M. Betke, J. Gips, and P. Fleming. The cameramouse: Visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 10:1, pages 1–10, March 2002.
- [3] M.A. Miglietta, G. Bochicchio, and T.M. Scalea. Computer-assisted communication for critically ill patients: a pilot study. *The Journal of TRAUMA Injury, Infection, and Critical Care*, Vol. 57, pages 488–493, September 2004.
- [4] Arvid Kappas (Editor), Nicole C. Krämer (Editor) *Face-to-Face Communication over the Internet: Emotions in a Web of Culture, Language, and Technology (Studies in Emotion and Social Interaction)* [Hardcover] 316 pages Publisher: Cambridge University Press; 1 edition (July 25, 2011)
- [5] Hess, Eckhard H. *The tell-tale eye: How your eyes reveal hidden thoughts and emotions*. Oxford, England: Van Nostrand Reinhold. (1975). xi 259 pp.
- [6] Eckhard H. Hess and James M. Polt *Pupil Size as Related to Interest Value of Visual Stimuli Science* 5 August 1960: Vol. 132 no. 3423 pp. 349-350
- [7] Amy D. Lykins, Marta Meana and Gretchen Kambe. *Detection of Differential Viewing Patterns to Erotic and Non-Erotic Stimuli Using Eye-Tracking Methodology*. *Archives of Sexual Behavior*. Volume 35, Number 5, 569-575
- [8] Isaacowitz, Derek M. et al. Selective preference in visual fixation away from negative images in old age? An eye-tracking study. *Psychology and Aging*, Vol 21(1), Mar 2006, 40-48
- [9] Prosevičius, Tomas; Raudonis, Vidas; Kairys, Artūras; Lipnickas, Artūnas; Simutis, Rimvydas. Autoassociative gaze tracking system based on artificial intelligence, *Electronics and Electrical Engineering*. ISSN 1392-1215. 2010, nr. 5(101), p. 67-72.
- [10] Gengtao Zhou; Yongzhao Zhan; Jianming Zhang; , "Facial Expression Recognition Based on Selective Feature Extraction," *Sixth International Conference on Intelligent Systems Design and Applications, 2006. ISDA '06.*, vol.2, no., pp.412-417.
- [11] Mihaela-Alexandra Puică, Adina-Magda Florea, Emotional Belief-Desire-Intention Agent Model: Previous Work and Proposed Architecture, *International Journal of Advanced Research in Artificial Intelligence*, Vol. 2, No. 2, 2013, pp. 1-8
- [12] J. Grath, A domain – independent framework for modeling emotion, *Journal of Cognitive Systems Research*, 2004, vol. 4, No. 5, pp. 269-306
- [13] I.B. Mauss, M.D. Robinson, Measures of emotion: A review, *Cognition and Emotion*, 2009, vol. 23, pp. 209-237
- [14] C.J. Stam, Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field, *Clinical Neurophysiology*, 2005, vol. 116, pp. 2266-2301
- [15] O. Sourina, A Sourin, V. Kulish, EEG data driven animation and its application, In *proceedings of international Conference Mirage*, Springer, 2009, pp. 380-388
- [16] Shylaja S, K N B. Murthy, S N. Nischith, Muthuraj R, Ajay S, Feed Forward Neural Network Based Eye Localization and Recognition Using Hough Transform, *IJACSA*, Vol. 2, No.3, 2011, pp. 104-109
- [17] A. B. Watson, J. I. Yellott, A unified formula for light-adapted pupil size, *Journal of Vision*, 2012, 12(10):12, pp. 1-16
- [18] P.N. Lopes, P. Solovey, R. Straus, Emotional intelligence, personality, and the perceived quality of social relationships, *Personality and Individual Differences*, 2003, Vol. 35., pp. 641-658
- [19] Kohei Arai, Ronny Mardiyanto, Eye-Base Domestic Robot Allowing Patient to be Self-Services and Communications Remotely, *International Journal of Advanced Research in Artificial Intelligence*, Vol. 2, No. 2, 2013, pp. 29-33
- [20] S.Nirmala Devi, Dr. S.P Rajagopalan, A study on Feature Selection Techniques in Bio-Informatics, *International Journal of Advanced Computer Science and Applications*, Vol. 2, No.1, 2011, pp. 138-144