

Methods and Models for Identifying Human Emotions by Recognition Gestures and Motion

V. L. Rozaliev

CAD Department
Volgograd State Technical University,
Prospect Lenina, 28, Volgograd 400005, Russia
vladimir.rozaliev@gmail.com

A. V. Zaboleeva-Zotova

Russian Foundation for Basic Research
Leninsky prospect, 32A, B-334, GSP-1, Moscow
119991, Russia
zabzot@gmail.com

Abstract—The paper describes a methodology, models and automated system for an identification of human emotions based on analysis of body movements, a recognition of characteristic gestures and poses. In the model of person emotions, the typical body movements are formalized with linguistic variables and fuzzy hypergraphs for temporal events. Emotional states of the real person recognized with the automated system are presented in a limited natural language. Shown the allocation of granules and related postures. Compliance granules poses and basic emotional states by K. Izard is presented. The article shows how to apply system for recognition and translation in real time gestures of the Russian deaf language in the text and the text in gestures. The system is intended for training of children with hearing disabilities and adults who need to learn sign language.

Keywords—human emotions, gestures and movements, fuzzy temporal event, Kinect

I. INTRODUCTION

The main purpose of information technologies, such as Internet, robotics, games, video monitoring, and so on, is to improve a human-computer interaction. But, for instance, a replacement of real persons by automated systems is impossible without overcoming the barrier of man-machine relationship [1]. The inability of machines to recognize and show emotions is one of the impeding factors. The development of telecommunication technologies changes the interpersonal communication. Very soon people will use virtual communications, which will be more effective and easy to learn but could not express emotions. At the same time emotions play a vital role in the human life. Emotions influence on cognitive processes [2] and decision making. So, it is important to recognize and identify the human emotions and use them.

In this paper, we present the methodology, models and the automated system, which are realized a fuzzy approach to identification of human emotions. The suggested approach is based on description and analysis of typical body movements, recognition of gestures and postures specific for the emotional states. Identifying human emotional response realizes the idea of how the human manifests his/her emotions.

II. IDENTIFICATION OF HUMAN BODY MOVEMENTS

The process of identifying human emotional response is based on the idea of how the human manifests his/her emotions [3; 4]. Now various companies are actively developing automated systems for recognition, identification and transmission of emotional reactions. There are different ways for determining emotional states such as by voice, facial expression, body movements, physiological parameters, and so on. [2; 5; 6; 7]

The proposed approach to emotion identification are based on recognition and analysis of human gestures and poses. [8] First of all, we recognize a person on video images using a technology for markerless motion capture with the digital video camera Microsoft Kinect. Video pictures are presented in the special animation format – the BVH-file, which describes poses of body skeleton and contains motion data. Such technology allows visualizing and analyzing different movements of person, determining areas of static or dynamic postures of micro and macro movements.

To detect the borders of movements, the motor activity of person is analyzed. For the separation of postures, we suggest a special notion of activity, which depends on what part of body performs the movement. We describe the typical body movements with linguistic variables and fuzzy hypergraphs for temporal events, and transform these descriptions into the expressions in a limited natural language, which characterize the person emotions. The identification of human emotional reactions such as joy, sadness, anger, etc is provided by the detailed analysis of postures, gestures and motions.

The architecture of computer system developed for identification of human emotional reactions consists of five subsystem. [8; 9] The input of the system are moving images, sound samples and handwriting texts. The output of the system is information about the emotional state of the real person, which is expressed in a limited natural language.

In order to define human emotional reactions by body movements, we use the vector model of skeleton, which is obtained from video information captured with the digital video camera Microsoft Kinect.

Kinect camera allows obtaining three-dimensional image in all lighting conditions and without any requirement to the actor, who is in the frame. Data from Kinect represented as a hierarchy of nodes of the human skeleton. Rotation of one

joint with the other, is presented in the form of quaternions (the role of the rotating vectors perform the bones of the skeleton) and the offset is represented as a 3 dimensional vectors in local to each node coordinate system. To obtain BVH-file, we have developed a new method consists of the following steps: 1. Getting data from the camera Kinect. 2. Determine the displacement of nodes relative to the parent node. 3. Record the hierarchy of key units in accordance with the specifics of the BVH-format. 4. Conversion of quaternions to the Euler angles. With the vector model of body movements, presented as a BVH-file may work most of the currently existing animation packages.

Vector model of the human body is the formalized representation of the movement of the person, where as vectors are presented bones of the human skeleton, and the angles between them correspond to the rotation angles of the main nodes of the human body in relation to each other. The vector model of skeleton consists of 22 nodes, which correspond to different anatomical joints with one, two or three axes of rotation.

We formalize the concept of motor activity of person, which is expressed in gestures, as follows:

$$A(\Delta t) = \sum_{n=1}^m T_n(\Delta t) \cdot k_n \quad (1)$$

Here m is a number of time series describing movement of the body parts, $T_n(\Delta t)$ is a variation of the n -th time series for the time Δt , k_n is a coefficient that characterizes influence of the body parts on the body motion for the n -th time series.

The influence coefficient can be calculated as the sum

$$k_n = \sum_{i=1}^{j_n} (p_i \cdot q_{ni}) \quad (2)$$

where i is a index of the body part, j_n is a number of the moving body parts, q_{ni} is a ratio of the body part in the total body mass, p_i is a gender coefficient of proportionality. According to biomechanical studies, the averaged values of ratio q_{ni} for adults are equal to 6,9% for head, 15,9% for the upper section of trunk, 2,1% for shoulder, 16,3% for the middle section of trunk, 1,6% fore forearm, 11,2% for the lower section of trunk, 0,6% for brush, 14,2% for thigh, 4,3% for lower leg, 1,4% for foot. The gender coefficient p_i is equal approximately to 1 for all parts of man body, and differs for various parts of woman body.

Another important characteristic of body movement is a mobility of the joint M_{joint} , which is measured in morphology by values of the angles of flexion-extension, abduction-reduction, internal-external rotation as follows:

$$M_{\text{joint}} = \text{angle} (\text{Fold+Straightening, Bringing+Abduction, In+Out}). \quad (3)$$

The maximum spine mobility is a sum of the angles of the left and right rotation around the longitudinal axis of the body.

For automatic separation video districts of the individual poses and gestures, we introduce additional parameters, defined by the user: the minimum duration of the movement, the level of activity for poses, the level of activity for the movements. Next, we construct a graph of activity and find areas of the postures and movements.

Poses discussed in detail in the works B. Birkenbil, G. Wilson, D. Morrison, A. Pease, were merged into granules, based on a similar interpretation. As it is impossible to unequivocally define the current posture emotional state of a person, we define the granules, which belong to the posture. This allowed us to increase the reliability of a particular emotional state. Compliance granules poses and basic emotional states by K. Izard is shown on Fig. 1.

III. FORMALIZATION OF BODY MOVEMENTS

In the vector model of skeleton, the typical movements of human body are described with the linguistic variables, which characterize duration of event, and variation of rotation angle. These variables allow to determinate the direction and speed of movement of every joint in the most accurate manner.

The duration of event is measured in the frames of video image. The fuzzy temporal variable ‘‘Duration of event’’ includes the following set of terms: D_0 ‘Zero’, D_1 ‘Very short’, D_2 ‘Short’, D_3 ‘Moderate’, D_4 ‘Long’, D_5 ‘Very long’.

Each group of joints with the similar values of maximal mobility is presented with the linguistic variable ‘‘Variation of rotation angle’’. The variable consists of the following set of terms: B_0 ‘Stabilization’, B_{+1} ‘Very slow increasing’, B_{+2} ‘Slow increasing’, B_{+3} ‘Moderate increasing’, B_{+4} ‘Fast increasing’, B_{+5} ‘Very fast increasing’, B_{-1} ‘Very slow decreasing’, B_{-2} ‘Slow decreasing’, B_{-3} ‘Moderate decreasing’, B_{-4} ‘Fast decreasing’, B_{-5} ‘Very fast decreasing’.

This linguistic variable can be adjusted on various types of the body movements and allow to describe, for instance, the small periodic fluctuations, such as tapping on the table, shaking hands or fingers, wiggle from foot to foot, and so on.

By specifying the name of the analyzed part of body, and the range of movements in the vector model of skeleton, one can obtain the values of rotation angles of the node relative to one of the axes X , Y or Z , which are stored in a separate data array. From this array there is selected a subarray, which contains the values of angles ρ_i falling in the range analyzed. The angles, belonging to different frames for the same node, form a triangular matrix, which elements is determined by the following rule: $\rho_{ij} = \rho_j - \rho_i$ for $j > i$, $\rho_{ij} = 0$ for $j \leq i$. This triangular matrix is used to calculate the values of the membership function of linguistic variable ‘‘Variation of rotation angle’’.

The movement of the joint around an axis has been described in the form of fuzzy temporal events. Since the events are located one after another on the time axis, the motion can be represented as a fuzzy sequential temporal sentence [10]. For example, the variation of the angle of rotation around the axis X for the joint ‘‘right foot’’ in the

interval $[t_4; t_{12}]$ can be described as the following series of fuzzy temporal statements: “For the right foot there is a very slow decreasing the angle of very short duration. This is followed by stabilization of the angle of zero duration. This is followed by a very slow increasing the angle of very short duration”.

The above fuzzy sequential temporal sentence can be rewritten formally as follows:

$$W = (B_{-1} \text{ rtf } D_1) \text{ rtsn } (B_0 \text{ rtf } D_0) \text{ rtsn } (B_{+1} \text{ rtf } D_1). \quad (4)$$

Here rtf is a fuzzy temporal relationship; rtsn is a temporal relationship of the direct sequence; B_0 is the term ‘Stabilization’, B_{-1} is the term ‘Very slow decreasing’, B_{+1} is the term ‘Very slow increasing’ of the linguistic variable “Variation of rotation angle”; D_0 is the term ‘Zero’, D_1 is the term ‘Very short’ of the fuzzy temporal variable “Duration of event”.

IV. EVALUATION OF SIMILARITY BETWEEN THE IDENTIFIED AND ETALON MOVEMENTS

In the model of fuzzy sequential temporal sentence, an adequacy of the analyzed fragment δ_q of a dynamic process and the corresponding attribute q are determined by the validity criterion $J(q/\delta_q)$, which is represented as follows:

$$J(q/\delta_q) = F_q(\delta_q) \& \mu_{L_q}(\delta_q). \quad (5)$$

Here $F_q(\delta_q)$ is the characteristic function that establishes a semantic relationship between fuzzy values of the secondary attributes of dynamic process and values of the primary attributes determining them; $\mu_{L_q}(\delta_q)$ is the membership function of the term L_q of the linguistic variable L .

The validity criterion $J(W/S)$ of fuzzy sequential temporal sentence W with respect to any dynamic process S is written as

$$J(W/S) = \max_{I \in V} (J(W/S)_I), \quad (6)$$

where V is the set of all possible interpretations I . For instance, the validity criterion of fuzzy sequential temporal

sentence W with respect to any dynamic process S for a set of fuzzy temporal events, which is expressed through successive attributes a, b, c , is described by the formula

$$J(W/S)_I = J(a/\delta_a) \& J(b/\delta_b) \& J(c/\delta_c) = (F_a(\delta_a) \& \mu_{L_a}(\delta_a)) \& (F_b(\delta_b) \& \mu_{L_b}(\delta_b)) \& (F_c(\delta_c) \& \mu_{L_c}(\delta_c)). \quad (7)$$

In our case, the analyzed dynamic process is a sequence of frames in the skeleton vector model, which characterizes the rotation of one of the skeleton nodes around the axis X, Y or Z at a certain angle. And the validity criterion is the criterion of similarity between the identified and etalon movements. So, the identified movements are considered as the well recognized with respect to the etalon movements if the value of similarity criterion exceeds the given threshold.

For example, calculate the criterion of similarity between the identified and etalon movements describing a rotation of the node “right ankle”. The etalon movements are presented by a fuzzy temporal event, which is written as follows: “For the right ankle there is a very slow decreasing the angle of zero duration”.

Let the initial data are the following time series: at the time moment t_0 the rotation angle $\rho_0=10.00$ degrees; at the time moment t_2 the rotation angle $\rho_2=6.13$ degrees. So, the duration of event is equal to 2 frames. Then by the graph of membership function $\mu_{D_0}(\delta_t)$ of the term D_0 ‘Zero’ of the fuzzy temporal variable “Duration of event” presented in Fig. 2, we find the value $\mu_{D_0}(\delta_t)=0.70$ for $\delta_t=t_2-t_0=2$ frames. By the graph of membership function $\mu_{B_{-1}}(\delta_\rho)$ of the term B_{-1} ‘Very slow decreasing’ of the linguistic variable “Variation of rotation angle” presented in Fig. 3, we find the value $\mu_{B_{-1}}(\delta_\rho)=0.92$ for $\delta_\rho=\rho_2-\rho_0=6.13-10.00=-3.87$ degrees. Thus, the criterion of similarity $J(W/S)=0.92$. If the threshold is equal to 0.80, then the identified rotation of the node “right ankle” is similar to the etalon movements.

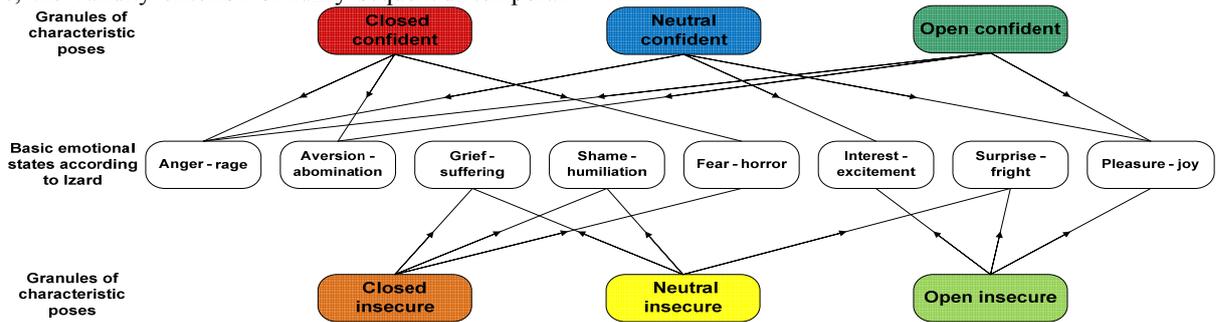


Figure 1. Compliance granules poses and basic emotional states.

V. USE FOR TEACHING CHILDREN WITH HEARING DISABILITIES

We use information about emotional reactions to control the education of children with hearing disabilities. The system is aimed for recognition and translation in real time gestures of the Russian language of the deaf in the text and the text in gestures.

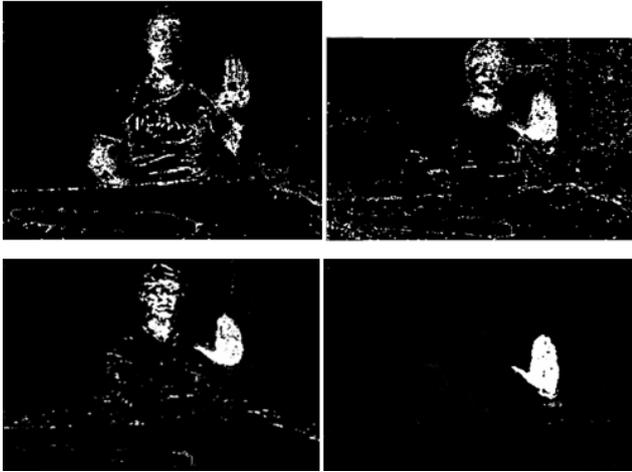


Figure 2. Finding hands on the image.

Hand gestures described via hand position, direction of movement, shape and direction of hands. The first stage of recognition is a segmentation of the image received from Kinect to find the hand or both hands. There are several signs that can be used to detect the object on the image: the appearance, shape, color, distance to the subject and context. When detecting faces in the image, a good sign is the appearance, as the eyes, nose and mouth are always about the same proportions. Therefore, to find hands, we first find the face of a man, define its color and highlight the closest object. Next we apply the developed method for finding the hands and define user gesture. An example of the recognition the user's hand is shown in Fig. 2.

The system works as follows. The user enters text. The system displays an animated image of the gesture. A sample output of the animated gesture is shown in Fig. 3.



Figure 3. Animated demonstration gestures.

User-child repeats movement. The movement is recognized and checked for correctness. If not correct, the movement is shown again. If correct, then enter the new text. If the user starts to receive a closed posture characteristic of anger, resentment, it is informing the administrator and learning process can be stopped.

VI. CONCLUSION

The identification of the human emotional states is closed to the problem of understanding what is the normal behavior. The variety of "normal" behavior is great. So it is difficult to draw the line between acceptable and unacceptable behavior. The automation of the human emotion recognition can help to solve many problems of relationships between people and avoid possible misunderstanding.

Automated systems for the human emotion identification by gestures and movements can be useful and necessary in various areas such as communication of deaf people; education/learning; emergency services; monitoring unstable emotional state of pilots, drivers, operators of complex technical system, etc; monitoring public places to prevent illegal and extremist actions, and so on.

In the future, we intend to use the developed approach to determine the emotional response by the handwritten text, and to animate human gestures and motions described in a limited natural language.

ACKNOWLEDGEMENT

This work is supported in part by the Russian Foundation for Basic Research 12-07-00266-a, 12-07-00270-a, 13-07-00351-a, 13-07-00459-a, 13-07-00461-a.

REFERENCES

- [1] Y. Orlova and V. Rozaliev, "Analysis and evaluation of the emotional reactions of the user in the articulation of interaction with the automated system". *Open education*, 2, pp. 83-85, 2011
- [2] D. Bernhardt, "Emotion inference from human body motion". *University of Cambridge Computer Laboratory, Cambridge*, pp: 227, 2010
- [3] E. Ilyin, "Emotions and feelings". *Piter, St. Petersburg*, pp: 784, 2008
- [4] V. Rozaliev and A. Zaboлева-Zotova. "Modeling the emotional state of a person based on hybrid methods". *International Scientific-Practical Journal of Software and Systems*, 90 (2), pp. 141-146, 2010
- [5] M. Coulson, "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence". *Nonverbal Behavior*, 28, pp. 117-139, 2004.
- [6] N. Hadjikhani and B. Gelder. "Seeing fearful body expressions activates the fusiform cortex and amygdale". *Current Biology*, 13, pp. 2201-2205, 2003.
- [7] R. Laban and L. Ullmann, "The mastery of movement". *Princeton Book Company, London*, pp: 210, 1988
- [8] A. Zaboлева-Zotova, V. Rozaliev and A. Bobkov, "Automated identification of emotional states". In the *Proceedings of the 24th International Conference of Systems Research, Informatics and Cybernetics*, pp: 21-25, 2012
- [9] A. Zaboлева-Zotova, Y. Orlova, V. Rozaliev and A. Bobkov, "Emotional state recognition based on the motion and posture". In the *Proceedings of the Workshop Operations Research and Data Mining*, pp: 161-169, 2012.
- [10] L. Bernshtein, "Models of representation of fuzzy temporal knowledge in databases of temporal series". *Journal of Computer and Systems Sciences International*, 48 (4), pp: 625-636, 2009.