Classification approach to determining emotional state of person based on analysis of video information*

1st Ivan A. Gurin
"Computer Sciences and Control"
of Russian Academy of Sciences
Moscow, Russia
covielle12@gmail.com

2nd Alexey B. Petrovski
"Computer Sciences and Control"
of Russian Academy of Sciences
Moscow, Russia
a.b.petrovsky@gmail.com

3rd Alla V. Zaboleeva-Zotova
Volgograd State Technical University
Volgograd, Russia
zabzot@gmail.com

Abstract—This work considers recognition of person's emotional state with body movements and postures as a problem of group classification of multi-attribute objects. The analysis of existing approaches to the recognition of the human emotional state based on video information was made. An approach proposed includes using markerless recognition of body movements based on video information, identification of poses and movements with a model of human motor activity, characteristics description of postures and movements in a limited natural language and the group classification method of poses and movements having verbal description for emotional state identification.

Keywords—human emotional state, analysis of the video information, movement and postures recognition, group classification

I. INTRODUCTION

Modern information technologies are becoming more and more entrenched in all spheres of human activity. One of the actual and demanded themes is recognition of human emotional state based on the automated analysis of video information of his movements and poses, as human emotions in many respects define his actions, acts and behavior. Physiologists have established that emotional excitation of a person is accompanied by coordinated reductions and relaxation of facial and body muscles, changes in heart, breathing rhythm, stomach and intestinal motility, etc. Tracking and analyzing reactions to emotional excitation, it is possible to define human emotions. There are many works devoted to the development of emotion detection and recognition systems (EDRS) [14], [15]. A rather large number of publications refer to the recognition of emotions by facial mimicry [16], [17]. Essentially less works in which the emotional state is determined by human body movements and poses [2], [19]. However, the problem of reliable

automated recognition of human emotions is far from being solved. In the given work the approach to define the human emotional state as solving the problem of group classification of multi-attribute objects which are the poses described in the limited natural language is offered. Verbal characteristics of poses are formed based on the analysis of body movements, which are recognized by video information and identified using a vector model of the human.

II. HUMAN POSTURE IDENTIFICATION

The first stage of automated recognition of a human emotional state is identification of poses and movements based on the analysis of video information. This task is not trivial and can be solved in different ways. In this paper we will limit ourselves to determining the emotional state of the person only by positions. First, we need to recognize the pose on the input video and present the input data in some digitized form. For this purpose, we use certain model, which contains information about the position of the human body. The models can be very different and depend primarily on the purpose of the analysis. For example, if the purpose is to determine the relative position of objects, the model can be quite simple. If the purpose is to read hand gestures, then a model capable of describing the position of fingers is needed. In our case, the model should allow setting any posture expressing some emotional state. The vector model of a human being described in works [18], [19] is well suited for automated recognition of the emotional state of a person. The specified model of a human being consists of 22 nodes and 19 body elements - on each of 22 points of bending in a human body, in which the change of angles between body elements is fixed. The position of each node is set by three angles of rotation relative to the axes X, Y, Z (Figure 1). The

vector model of the person is presented by file format bvh (BioVision Hierarchy) which contains hierarchical structure of a skeleton and parts of a body, the description of position of nodes and the information on dynamics of movement, i.e. changes of position of nodes and duration of these changes. Durations are measured in frames. It is worth noting that in a reality one defines a pose or body movement not by positions of bending points on someone's body, but by positions of particular body parts. Besides, a huge number of poses differ from each other by position of only a small number of nodes. And motions frequently are position changing not of concrete node, but set of the nodes modeling a part of a body. Besides, psychologists tell an emotional state not so much by a posture as a whole, but by noticing specific behavior of separate parts of the body (such as compressed fists, crossed hands, the lowered head, etc.). All this leads to the idea that it makes sense to present a posture as a set of positions of different bodyparts. Each part of the body is defined by a certain number of nodes, the position of which, in turn, set the angles of rotation relative to the axes X, Y, Z. It means that each part of a body can also be set as a vector. It should also be noted that for motions this approach is also suitable, because, to say it rough, motions are just position changes of one or more parts of the body. The difference lies only in the vectors that define the body parts: for motions, the parameters responsible for the duration must be carefully set and accounted in emotional state identification.
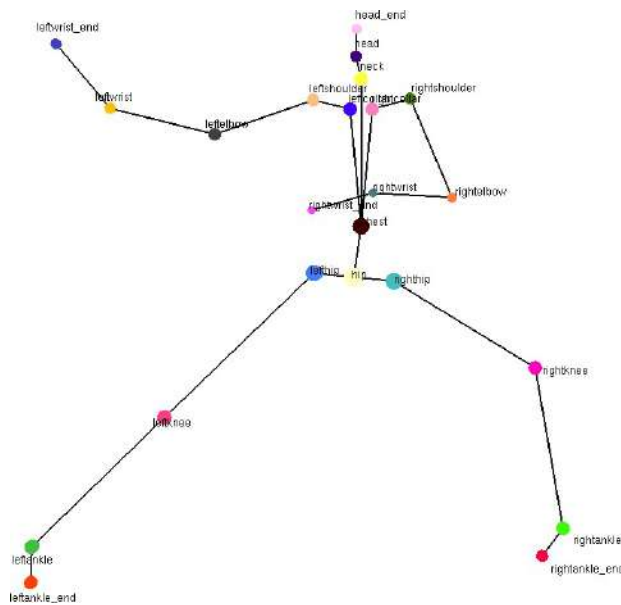

Figure 1. Vector model of the human

Using the specified model, each pose can be represented as a set of vectors describing positions of different parts of a body. For example, the hand position is defined by four nodes <n1, n2, n3, n4>, and each node is defined

by three parameters - rotation angles. Thus, it is possible to divide the whole pose into separate parts of a body, each of which is set by some vector describing position of the nodes included into a certain part of a body.

III. DESCRIPTION OF HUMAN POSTURE

At the following stage it is necessary to transform model representation of each posture which is specified as a set of vectors describing positions of different parts of a body into the description of a posture in the limited natural language habitual for expert psychologists. It can be done, for example, in the following way. Based on the expert's knowledge, the range of values of all elements of the vector representing the posture is divided into intervals, which can be used to match with some verbal characteristics of the posture in the natural language. Obviously partition of the area of values of parameters into intervals cannot be universal as from the physiological point of view people's bodies differ. However, it is possible to divide people into groups (for example, by sex or age - from "children" to "old people") and to build a partition for these groups. Thus, for each image of the pose we get a set of expressions consisting of features that describe this pose.

Figure 2. Example posture

For example, figure 2 shows posture which can be described by the following set of expressions: back is straight; body standing position; head is not bent or overturned; hands are bent in elbows and crossed at chest level, one hand lies on the other; palm of one hand embraces forearm of the other; fingers are not bent, pressed against each other. The description of a pose in natural language, as a matter of fact, represents the position of various parts of a body, thus the certain position of any part of a body is a feature of the given pose. In view of the fact that position of each part of

a body is described by some quantity of nodes which position in turn is set by three corners of rotation, we receive that features of a posture is a set of certain elements from the vector model describing a posture.

IV. Determining human emotional state

The third stage is the definition of the emotional state of the person by posture. As shown in works in the field of psychology, there are 61 characteristic poses of the person and 8 basic emotional conditions. However there is no unequivocal conformity between a pose and emotional conditions. We shall consider definition of human emotional state as the decision of a problem of group classification of multi-attribute objects with teachers where objects are poses described in the limited natural language. Let's use for this purpose method of Multi-Dimensional Harmonized Classification Alternatives [7], [8], [9]. The initial array of multi-signed objects is a set of text descriptions of pose images obtained after analysis and identification of motions by video images. Each image of a pose is evaluated by number of experts (t, teachers) and belongs to one of the specified characteristic poses. Then each pose is a multi-signed object O_i, i = 1,...,m, that is present in several versions (copies, instances) $O_i^{<s>}$, s = 1,...,t, which are the essence of the pose image, differing in the values of the attributes $K_1, \ldots, K_n$. Features $K_1, \ldots, K_n$ describe positions of body parts in a limited natural language, each of which is defined by the scale $X_l = x_l^1, \ldots, x_l^{h_l}$, l = 1,...,n, where the gradation $x_l^{e_l}, e_l = 1, \ldots, h_l$ – is a verbal position characteristic. The basic emotional states are classes $D_1, \ldots, D_g$. The expert estimation of the pose image is given by the set of sorting features $R = r_1, \ldots, r_g$, where each of rf shows membership of class D_f, f = 1,...,g. Let's represent each version of $O_i^{<s>}$, i = 1,...,m, s = 1,...,t as a multiset:

$$A_i^{<s>} = \{k_{A_i}^{<s>}(x_1^1) \circ x_1^1, \ldots, k_{A_i}^{<s>}(x_1^{h1}) \circ x_1^{h1};$$
$$\ldots$$
$$k_{A_i}^{<s>}(x_n^1) \circ x_n^1, \ldots, k_{A_i}^{<s>}(x_n^{h_n}) \circ x_n^{h_n}; \quad (1)$$
$$k_{A_i}^{<s>}(r_1) \circ r_1, \ldots, k_{A_i}^{<s>}(r_g) \circ r_g\}$$

above the set X=$\{x_1^1, \ldots, x_1^{h_1}; \ldots; x_n^1, \ldots, x_n^{h_n}; r_1, \ldots, r_g\}$ the gradation of the feature scale. Here the value of the multiplicity function $k_{A_i^{<s>}}(x_l^{e_l})$ shows how many times the value $x_l^{e_l} \in X_l, e_l = 1, \ldots, h_l$ of the attribute K_l is present in the description of the object version $O_i^{<s>}$ by the expert s. The value of the multiplicity function $k_{A_i}^{<s>}(r_f) = 1$, if the expert s has referred the version of the object $O_i^{<s>}$ to the class D_f, and $k_{A_i^{<s>}}(r_f) = 0$ – in the opposite case. Let's assume that all images of the same pose are equivalent. Let's form an expert estimation of each O_i as a multiset of features, which is a sum of multiplicity $A_i^{<s>}$, representing images $O_i^{<s>}$ of this posture:

$$A_i = A_i^{<1>} + \ldots + A_i^{<t>} =$$
$$= \{k_{A_i}(x_1^1) \circ x_1^1, \ldots, k_{A_i}(x_1^{h_1}) \circ x_1^{h_1};$$
$$\ldots \quad (2)$$
$$k_{A_i}(x_{n_1}) \circ x_{n_1}, \ldots, k_{A_i}(x_n^{h_n}) \circ x_n^{h_n};$$
$$k_{A_i}(r_1) \circ r_1, \ldots, k_{A_i}(r_g) \circ r_g\}.$$

where the multiplicity functions are calculated by the rules $k_{A_i}(x_l^{e_l}) = \sum_s k_{A_i^{<s>}}(x_l^{e_l})$, $k_{A_i}(r_f) = \sum_s k_{A_i^{<s>}}(r_f)$. As a result, an array of text descriptions of 61 poses will be formed with expert assessments of their possible belonging to 8 basic emotional states. By means of the Multi-Dimensional Harmonized Classification Alternatives method there are generalized decisive rules of group classification of human postures by their characteristics, which are best coordinated with individual rules of expert sorting. The obtained classification rules allow to determine a person's emotional state by video image of his body in an automated mode.

Conclusion

At present, determining a human emotional state by body movements and poses is an urgent task in the areas of health care, public safety and others. However, this area is still poorly developed. So far scientists-psychologists have not offered the reasonable theory about connection between human emotional state and postures and movements. Most of the work in the field of automated recognition of emotions is aimed at determining the emotional state by mimicry. The works on estimation of emotional condition by pantomimics are few, they often have only research purpose. In this work, recognition of emotional state by video image of body movements is considered as a task of group classification of postures. In the future, this approach is also intended for group classification of movements.

References

[1] M.I. Abramov, "Poligoniometricheskie issledovania podvizhnosti osnovnikh otdelov oporno-dvigatelnogo apparata vokrug prodol-noi osi," in Teoria I praktika fizkulturi 1971, pp 34-36.

[2] A.V. Zaboleeva-Zotova, V.L. Rozaliev, A.S. Bobkov, O.S. Fedorov, Trudi kongressa po intellektualnim sistemam i informacionnim tekhnologiam, 2010, pp 398-405.

[3] N.A. Belaya, "Lechebnaya fizkultura i massazh,", 2001 p 272.

[4] A.S. Bobkov, "Interpretiruushaya model poiska dvizheni cheloveka po slovesnomy opisaniu na ogranichennom estestvennom yazike," 2012, pp 53-57.

[5] C. Darvin, "Proishozhdenie cheloveka i polovoi otbor. Virazhenie emocii u cheloveka i zhivotnih," 1940, 320 p.

[6] K. Izard, "Psihologia emocii,", 2003, pp. 464.

[7] A.B. Petrovski, "Metodi gruppovoi klassifikacii mnogopriznakovih objectov (chast 2),", 2009, pp. 3-14.

[8] A.B. Petrovski, "Metod "MASKA" dlya gruppovoi expertnoi klassifikacii mnogopriznakovih objectov,", 2010, pp. 27-31.

[9] A.B. Petrovski, "Gruppovoi verbalniy analiz resheni,", 2019, p. 287.

[10] M. Coulson, "Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. Journal of Nonverbal Behavior, 28(2), 117–139.", 2004, pp. 117-139.

[11] Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y., "Realtime Multiperson 2D Pose Estimation Using Part Affinity Fields, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", 2017.

[12] Saini, S., Zakaria, N., Rambli, D. R. A., and Sulaiman, S., "Markerless Human Motion Tracking Using Hierarchical Multi-Swarm Cooperative Particle Swarm Optimization", 2015.

[13] Castillo, S., Legde, K., and Cunningham, D. W., "The semantic space for motion-captured facial expressions. Computer Animation and Virtual Worlds", 2018.

[14] Martin Magdin, Ľubomír Benko, Štefan Koprda, "A Case Study of Facial Emotion Classification Using Affdex", 2019.

[15] D. YangAbeer AlsadoonP.W.C. PrasadA. K. SinghA. Elchouemi, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment", 2017.

[16] R.I. Ahmetshin, A.P. Kirpichnikov, M.P.Shleymovich, "Raspoznanie emocii cheloveka na izobrazheniyah", 2015.

[17] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak, "Emotion recognition using facial expressions", 2017.

[18] Hao Guo1, Zhenjiang Miao1, Feiyue Zhu2, Gang Zhang2, Song Li , "Automatic Labanotation Generation Based on Human Motion Capture Data", 2014, pp. 3-14.

[19] F. Noroozi, C.A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, G. Anbarjafar , "Survey on Emotional Body Gesture Recognition", 2017.

Классификационный подход к определению эмоционального состояния человека на основе анализа видеоинформации

Гурин И.А., Петровский А.Б, Заболеева-Зотова А.В.

В работе распознавание эмоционального состояния человека по телодвижениям и позам рассматривается как проблема групповой классификации многопризнаковых объектов. Проведён анализ существующих подходов к распознаванию эмоционального состояния человека на основе анализа видеоинформации. Предложен подход, включающий безмаркерное распознание телодвижений по видеоинформации, идентификацию поз и движений с использованием модели двигательной активности человека, описание характеристик поз и движений на ограниченном естественном языке, выявление эмоциональных состояний с помощью метода групповой классификации поз и движений, имеющих вербальные описания.