Body Motion Analysis for Emotion Recognition in Serious Games

Kyriaki Kaza, Athanasios Psaltis, Kiriakos Stefanidis, Konstantinos C. Apostolakis, Spyridon Thermos, Kosmas Dimitropoulos(✉), and Petros Daras

Information Technologies Institute,
Centre for Research and Technology Hellas, Thessaloniki, Greece
{kikikaza,at.psaltis,kystefan,kapostol,
spthermo,dimitrop,daras}@iti.gr
http://www.iti.gr

Abstract. In this paper, we present an emotion recognition methodology that utilizes information extracted from body motion analysis to assess affective state during gameplay scenarios. A set of kinematic and geometrical features are extracted from joint-oriented skeleton tracking and are fed to a deep learning network classifier. In order to evaluate the performance of our methodology, we created a dataset with Microsoft Kinect recordings of body motions expressing the five basic emotions (anger, happiness, fear, sadness and surprise) which are likely to appear in a gameplay scenario. In this five emotions recognition problem, our methodology outperformed all other classifiers, achieving an overall recognition rate of 93 %. Furthermore, we conducted a second series of experiments to perform a qualitative analysis of the features and assess the descriptive power of different groups of features.

Keywords: Body motion analysis · 3D body movement features · Emotion recognition · RBM · Serious games

1 Introduction

One crucial component of games design and development is the accurate measurement of user's experience and undoubtedly, the most important aspect of user's experience is their affective state. Particularly in serious games, the inference of player's affective state could provide not only valuable information for player's engagement and entertainment level, but also indications of whether or not the desirable educational objectives are reached. The majority of state of the art emotion recognition frameworks capitalize mainly on facial expression or voice analysis; however, research in the field of experimental and developmental psychology has shown that body movements, body postures, or the quantity or quality of movement behavior in general, can also help us differentiate between emotions [1, 2].

In particular, specific qualities and characteristics of body movements, such as velocity, direction, turning away/forwards, body expansion/contraction has been examined in the inference of different emotions and combinations of these qualities are

M. Antona and C. Stephanidis (Eds.): UAHCI 2016, Part II, LNCS 9738, pp. 33–42, 2016.
DOI: 10.1007/978-3-319-40244-4_4

suggestive of specific emotions [3, 4]. For example, expressions of joy are characterized by fast, upward directed movements with the arms raised. Similarly, movements indicative of grief tend to be slow, light, downward directed, with the arms closed around the body. Body turning away and body contraction, as an attempt to appear as small as possible is shown to be a strong indicator of fear, while body turning towards is typical of happiness and anger.

As more and more game companies and research centers move towards low cost RGB-depth sensors, a growing interest emerges considering the role of body movement in games, not only in terms of natural user interaction, but of emotion recognition during gameplay as well. Bodily expression provides a means for emotion recognition from a distance [5], and therefore motion analysis is crucial in generating multi-modal data in gameplay environments where players' facial analysis data are either too remote or partially obstructed (e.g. children wearing glasses, hats or other headwear). Additionally, the inclusion of bodily expression as an additional channel for affect communication can help resolve ambiguity observed in the identification of certain basic mental states, such as anger and fear [6].

Over the last years, different approaches have been proposed for emotion recognition based on body movements, gestures and postures [7–11]. These studies obtained quite interesting results, highlighting the importance and feasibility of using body expressions for affect recognition. Specifically regarding affect recognition in games, Piana et al. [12] presented a method that uses features derived from 3D skeleton data and a multi-class SVM classifier for the recognition of six emotions, which was integrated in a platform of serious games for children with Autism Spectrum Condition. This method achieved a 61.3 % recognition rate when evaluated at a dataset of recorded body movements of actors who were asked to express freely the six basic emotions. Savva et al. [13] proposed an automatic recognition method of affective body movement in the context of a Nintendo Wii tennis game which feeds dynamic movement features to a Recurrent Neural Network (RNN) algorithm. This method was tested at a dataset of non-acted movements captured with Animazoo IGS-190 during gameplay and reached a recognition rate of 57.46 %, comparable with the 61.49 % accuracy of human observers' recognition.

Our focus at this paper is to present a method for emotion recognition based on body motion analysis that will be incorporated in serious games that aim at helping children aged 7–10 acquire prosocial skills. Using a set of 3D features, we decided to test the recognition performance of deep learning architectures, such as neural networks (NNs) and Restricted Boltzmann Machines (RBMs), in an emotion recognition task, as well as the descriptive power of different groups of features. This study led to the design of a deep learning network classifier with stacked RBMs. In the following sections, we will briefly analyze the groups of features used (Sect. 2); then, we will present the proposed classifier (Sect. 3); finally, we will evaluate and compare the recognition accuracy of different classifiers and different groups of features, using a dataset of acted movements associated with five emotions (Sect. 4) and will conclude with a discussion about results and future work (Sect. 5).

2 Body Motion Analysis

In this section, we will present a set of movement features which we utilize in our method and which are proven crucial in the process of emotion recognition [14]. The 3D body movement features are extracted from joint-oriented skeleton tracking using the depth information provided by Kinect sensor [15]. We divide the set of features into the following groups: kinematic related, spatial extent related, smoothness related, symmetry related, leaning related and distances related.

2.1 Kinematic Related Features (G1)

Velocity and Acceleration. The velocity and acceleration features per frame are calculated as the mean velocity and acceleration of all frame joints. The velocity and acceleration can be approximated in our case by considering finite differences of position divided by the sampling time interval ΔT [16].

Kinetic Energy. Kinetic energy provides an estimate of the overall energy spent by the user during movement. The amount of movement activity has been shown to be relevantly important for differentiating emotions [17]. The kinetic energy is proportional to the square of velocity. We ignore the mass term in kinetic energy as it is not relevant [16]. So the proportional amount of the kinetic energy of each joint K_i is calculated as:

$$K_i = \frac{1}{2} v_i^2 \tag{1}$$

Then, the kinetic energy of the entire body is calculated as the sum of all joints' kinetic energies.

2.2 Spatial Extent Related Features (G2)

The following features provide an estimate of how the body occupies the 3D space surrounding it. According to research in experimental psychology, the contraction index can be used to infer specific emotional states; people are considered to usually spread out when they are happy, angry or surprised, and similarly reduce their size when in fear [3].

Bounding Box Volume. A bounding box is the minimum cuboid containing the body. Given the 3D positions of the user's limbs' end effectors, we can approximate this volume as the minimum parallelepiped surrounding the user's body.

Contraction Index. Contraction index in 3D is defined as the normalized bounding volume containing the user's body and is related to the definition of ones' "personal space" [14]. The 3D contraction index is then calculated by comparing this bounding volume and an approximation of the volume of the density (DI) of the 3D coordinates calculated as follows:

$$DI = \frac{3}{4}\pi \cdot DI_x \cdot DI_y \cdot DI_z \tag{2}$$

where DI_x, DI_y, DI_z are the approximated density indices calculated respectively on x, y and z axes as described in the following Equations:

$$DI_x = \frac{1}{n}\sum_{i=1}^{n} dx_i \tag{3}$$

$$DI_y = \frac{1}{n}\sum_{i=1}^{n} dy_i \tag{4}$$

$$DI_z = \frac{1}{n}\sum_{i=1}^{n} dz_i \tag{5}$$

in which dx_i, dy_i and dz_i are the distances between the center of mass and the i^{th} joint. The 3D Contraction Index is then calculated as the normalized ratio between DI and the Bounding Volume.

Density. A different measurement of body spatial extent is represented by the density index. Given the center of mass of the user's tracked skeleton C, the density index is calculated as the average sum of Euclidean distances of all tracked joints from C:

$$DEI = \frac{1}{n}\sum_{i=1}^{n} d_{Ci} \tag{6}$$

2.3 Smoothness Related Features (G3)

Curvature (k) measures the rate at which a tangent vector turns as a trajectory bends and provides an indication of joints' trajectories' smoothness. According to Wallbott [2], "a smooth movement seems to be large in terms of space and exhibit a high but even velocity", so it will have low curvature value; by contrast, a sharp trajectory movement will have a high curvature.

The smoothness index for three dimensional curvatures is computed as follows:

$$k_i = \frac{\sqrt{(\dot{x}_i \cdot \ddot{y}_i - \dot{y}_i \cdot \ddot{x}_i)^2 + (\dot{z}_i \cdot \ddot{x}_i - \dot{x}_i \cdot \ddot{z}_i)^2 + (\dot{y}_i \cdot \ddot{z}_i - \dot{z}_i \cdot \ddot{y}_i)^2}}{(\dot{x}_l^2 + \dot{y}_i^2 + \dot{z}_i^2)^{\frac{3}{2}}} \tag{7}$$

In our features set, we include right and left wrist curvature, head curvature and torso curvature.

2.4 Symmetry Related Features (G4)

It has been shown that asymmetry of movements can be related to emotion expression [18]. Each symmetry (SI_x, SI_y, SI_z) is computed from the position of the barycenter and the left and right joints (e.g., wrists, shoulders, feet, knees) as described below:

$$SI_{Xi} = \frac{\left(x_B - x_{Li}\right) - \left(x_B - x_{Ri}\right)}{x_{Ri} - x_{Li}} \tag{8}$$

where x_B is the coordinate of the center of mass, x_{Li} is the coordinate of a left joint i (e.g., left hand, left shoulder, left foot, etc.) and, x_{Ri} is the coordinate of a right joint. In the same way, we compute SI_y, SI_z. The three partial indices are then combined in a normalized index that expresses the overall estimated symmetry:

$$SI = \frac{SI_{Xi} + SI_{Yi} + SI_{Zi}}{3} \tag{9}$$

In our features set, we include wrists, elbows, knees and feet symmetry.

2.5 Leaning Related Features (G5)

Head and body movement and positions are relied on as an important feature for distinguishing between various emotional expressions [13]. The amount of forward and backward leaning of a torso and head joint is measured by the velocity of the joint's displacement along its z component (depth) respective to the body position and orientation, while the amount of right and left leaning is measured by the velocity of joint's displacement along its x component.

2.6 Distance Related Features (G6)

The distances between hands can be indicative of expansion or contraction of gestures, while the distance between hand and head as well as hand and torso could provide estimation for the existence of specific movements (e.g. touching head with one hand in case of grief).

The aforementioned features are extracted per frame and then, their mean value is calculated for the sequence of frames. The total of the mean values for all features constructs the feature vector used as a movement descriptor. Distances and coordinates used in calculations are normalized with respect to height.

3 Deep Learning Network Classifier

In this section, we propose a deep learning network classifier consisting of stacked RBMs, which proved to outperform other classic classifiers at the emotion recognition task, as we will see in Sect. 4.

A Restricted Boltzmann Machine [19] is a parameterized generative model representing a probability distribution. Given some observations, the training data, learning a Boltzmann Machine (BM) means adjusting the BM parameters such that the probability distribution represented by the BM fits the training data as well as possible. Boltzmann machines consist of two types of units, so called visible and hidden neurons, which can be thought of as being arranged in two layers. The visible units constitute the first layer and correspond to the components of an observation. The hidden units model dependencies between the components of observations (Fig. 1).

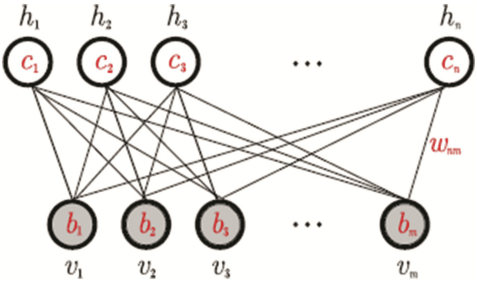

Fig. 1. The undirected graph of an RBM with n hidden and m visible variables

For our emotion recognition task, aiming at a network architecture that will handle independently each group of features, we designed the architecture depicted at Fig. 2.

At our two-layer network, we stacked seven RBMs, six at the first layer and one at the second layer, and train them layerwise, starting at the base layer and move up to the second, with no feedback from the higher layer to the lower layer. Each RBM of the

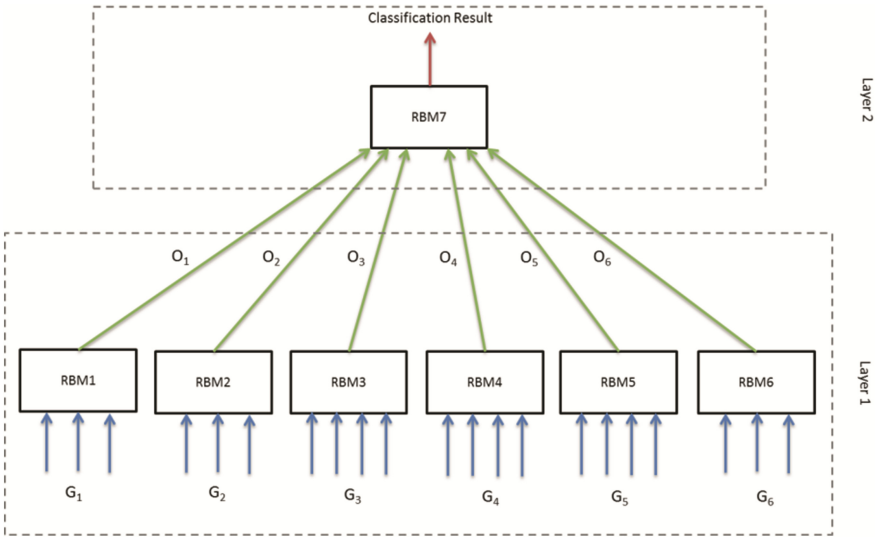

Fig. 2. Deep learning network classifier with 7 stacked RBMs

first layer receives as input the features of a different group of features and it is trained in an unsupervised (since unsupervised pre-training guides the learning towards basins of attraction of minima that support better generalization from the training data set [20]) and afterwards in a supervised way. Then, the output probabilities of the first layer are fed as input to the second layer and the seventh RBM is trained. The output probabilities of the second layer constitute the classification result.

4 Experimental Results

4.1 Dataset

In order to evaluate the performance of our methodology, we created a dataset with Kinect recordings of body movements expressing the 5 basic emotions (anger, happiness, fear, sadness and surprise) which are likely to appear in a gameplay scenario. The predefined set of movements (Fig. 3) associated with these emotions was selected based on social psychology research that identified body movements and postures which, to some degree, are specific for certain emotions [1, 2, 8]. Each emotion was represented with two different types of movements and each recording had duration of 4 s. 14 subjects (5 women and 9 men) participated in the recording session. They were shown a short video with the aforementioned movements and afterwards, they were asked to perform each movement, according to their personal style, 5 times, in front of a Kinect sensor.

Fig. 3. Dataset movements expressing five emotions

4.2 Classification Results

The purpose of the experiments conducted was to evaluate the recognition rate of deep learning network classifiers at the problem of emotion recognition from body motion analysis. For this purpose, we compared the recognition rates of classic classifiers (Naïve Bayes, Linear MultiClass SVM, Non Linear SVM) with the recognition rates of deep learning network classifiers (multilayer perceptron MLP, RBM, our proposed architecture with stacked classifiers) at a Leave-One-Subject-Out cross validation (LOSO cv) training the classifiers over 13 subjects and testing them with the data of the 14th left out subject. The results of these experiments are shown at Table 1. As we can see, the three DL network classifiers achieved higher classification accuracy than classic classifiers, with the proposed classifier outperforming all others with a recognition rate of 93 %.

Table 1. Classification results

Algorithm	Method	Recognition Rate
Naïve Bayes		77.21%
Linear MultiClass SVM	Crammer and Singer (CS)	77.78%
	Weston and Watkins (WW)	78.35%
	one-versus-all (OVA)	79.78%
Non Linear SVM	Crammer and Singer (CS)	84.5%
	Weston and Watkins (WW)	84.71%
	one-versus-all (OVA)	84.78%
MLP		85.53%
RBM		88.9%
Stacked RBMs Classifier		93%

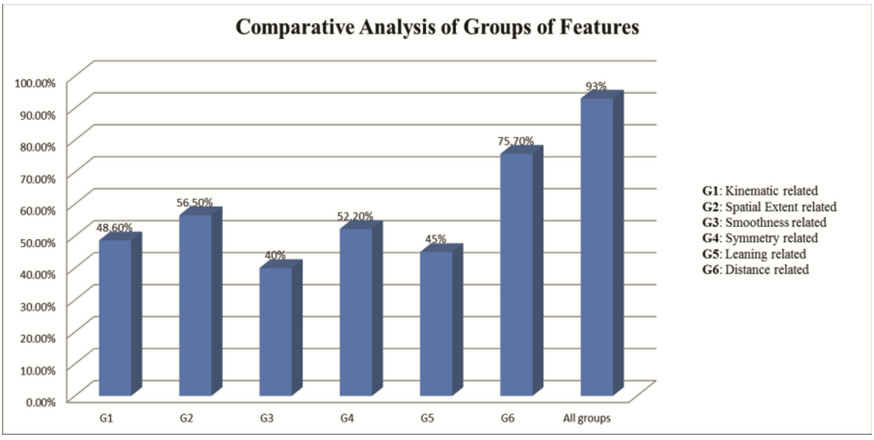

Fig. 4. Recognition rates with different groups of features

4.3 Comparative Analysis of Different Groups of Features

Furthermore, we conducted a second experiment to perform a qualitative analysis of the features and assess the descriptive power of the different groups of features. Our purpose was to examine the contribution of each group of features to the classification process. For the experiments, we trained six different RBMs, each with a different group of features. As it is shown at Fig. 4, only the distances related features, although being quite simplistic, can provide a decent recognition accuracy of 75.7 %, constituting the only group that could be practically used in a recognition task. This means that all groups of features have a significant contribution and are necessary in order to achieve high classification accuracy.

5 Conclusions and Future Work

In this work, we have presented a complete method for affect recognition from body movements that are likely to appear at a gameplay scenario. We have confirmed that a set of geometric and kinetic features can act as adequate descriptors for movements related to emotions and that deep learning classifiers can provide better recognition rates than classic classifiers at this recognition task. Additionally, we have proposed a classifier with stacked RBMs that outperformed all other classifiers in recognition accuracy. Our next goal is to test our method at non-acted emotion expressions recordings during gameplay of prosocial game Path of Trust [21] and as a following step; the affective information derived from body expression analysis will be fused with facial and audio analysis information, in order to further increase the robustness of the algorithm. The final recognition module will be incorporated in Path of Trust game and other serious games for children.

Acknowledgement. The research leading to this work has received funding from the EU Horizon 2020 Framework Programme under grant agreement no. 644204 (ProsocialLearn project).

References

1. Ekman, P.: Differential communication of affect by head and body cues. J. Pers. Soc. Psychol. **2**(5), 726 (1965)
2. Wallbott, H.G.: Bodily expression of emotion. Eur. J. Soc. Psychol. **28**, 879–896 (1998)
3. Boone, R.T., Cunningham, J.G.: Children's decoding of emotion in expressive body movement: The development of cue attunement. Dev. Psychol. **34**, 1007–1016 (1998)
4. de Meijer, M.: The contribution of general features of body movement to the attribution of emotions. J. Nonverbal Behav. **13**(4), 247–268 (1989)
5. de Gelder, B.: Why bodies? twelve reasons for including bodily expressions in affective neuroscience. Philos. Trans. R. Soc. B Biol. Sci. **364**(1535), 3475–3484 (2009)
6. Gunes, H., Shan, C., Chen, S., Tian, Y.: Bodily expression for automatic affect recognition. In: Emotion Recognition A Pattern Analysis Approach, pp. 343–377 (2015)

7. Castellano, G., Villalba, S.D., Camurri, A.: Recognising human emotions from body movement and gesture dynamics. In: Paiva, A.C., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 71–82. Springer, Heidelberg (2007)
8. Coulson, M.: Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. J. Nonverbal Behav. **28**, 117–139 (2004)
9. Kleinsmith, A., De Silva, R., Bianchi-Berthouze, N.: Cross-cultural differences in recognizing affect from body posture. Interact. Comput. **18**(6), 1371–1389 (2006)
10. Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., Volpe, G.: Multimodal analysis of expressive gesture in music and dance performances. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 20–39. Springer, Heidelberg (2004)
11. Kleinsmith, A., Fushimi, T., Bianchi-Berthouze, N.: An incremental and interactive affective posture recognition system. In: Carberry, S., De Rosis, F. (eds.) International Workshop on Adapting the Interaction Style to Affective Factors, in Conjunction with the International Conference on User Modeling (2005)
12. Piana, S., Stagliano, A., Odone, F., Verri, A., Camurri, A.: Real-time automatic emotion recognition from body gestures. In: Proceedings of IDGEI (2014)
13. Savva, N., Bianchi-Berthouze, N.: Automatic recognition of affective body movement in a video game scenario. In: Camurri, A., Costa, C. (eds.) INTETAIN 2011. LNICST, vol. 78, pp. 149–159. Springer, Heidelberg (2012)
14. Piana, S., Stagliano, A., Camurri, A., Odone, F.: A set of full-body movement features for emotion recognition to help children affected by autism spectrum condition. In: IDGEI International Workshop (2013)
15. Kinect for windows (2013). http://www.microsoft.com/en-us/kinectforwindows/
16. Shan, J., Akella, S.: 3D human action segmentation and recognition using pose kinetic energy. In: 2014 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), pp. 69–75 (2014)
17. Camurri, A., Lagerlöf, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. Int. J. Hum Comput Stud. **59**(1), 213–225 (2003)
18. Roether, C.L., Omlor, L., Giese, M.A.: Lateral asymmetry of bodily emotion expression. Curr. Biol. **18**(8), R329–R330 (2008)
19. Smolensky, P.: Information processing in dynamical systems: foundations of harmony theory. In: Rumelhart, D.E., McClelland, J.L. (eds.) Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, pp. 194–281. MIT Press, Cambridge (1986)
20. Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., Bengio, S.: Why does unsupervised pre-training help deep learning? J. Mach. Learn. Res. **11**, 625–660 (2010)
21. Apostolakis, K., Kaza, K., Psaltis, A., Stefanidis, K., Thermos, S., Dimitropoulos, K., Dimaraki, E., Daras, P.: Path of trust: a prosocial co-op game for building up trustworthiness and teamwork. In: B Games and Learning Alliance: Fourth International Conference, GALA (2015)