

AUTOMATIC RECOGNITION OF NON-ACTED AFFECTIVE POSTURES: A VIDEO GAME SCENARIO

Andrea KLEINSMITH and Nadia BIANCHI-BERTHOUSÉ*

UCL Interaction Centre

University College London, UK

ABSTRACT

How people experience affect and emotion, both in conveyance and recognition, partially determines how they interact with others, how they perform in their jobs, and how they carry out general day-to-day activities. It is hence crucial to endow technology with the ability to recognize its users' affective state to increase the technologies' effectiveness. The study aims at recognising non-basic affective states from non-acted body postures in the context of a body-movement based video game situation. A motion capture system was used to capture the postures of players playing the Nintendo Wii sports games. Automatic recognition models are then built and tested for their ability to generalise to both new observers and new postures using a repeated sub-sampling validation method. As input to the automatic recognition software the joint rotation values captured by the motion capture system are used. This allows for the creation of recognition models that are easily adaptable to different contexts. As output, emotion categories related to a game context were chosen. To set a benchmark against which the recognition system is evaluated, an online posture evaluation survey was conducted using the affective body postures collected with the motion capture system. The postures were reconstructed as faceless humanoid avatars and presented for classification by human observers. The agreement levels and reliability ratings between the observers are based on a repeated sub-sampling validation method. The results showed that human agreement levels in this study are above chance level and the automatic recognition models' performance is comparable to the benchmarks.

***Keywords:** affective posture, automatic emotion recognition, human emotion recognition*

* **Corresponding author:** UCL Interaction Centre, University College London, London WC1E 6BT, UK. Tel.: +44 (0) 20 7679 0690. Email: n.berthouze@ucl.ac.uk

1. INTRODUCTION

Emotions are a very important aspect of our lives [1]. They support our cognitive processes but they can also impair them; they help us work efficiently and appropriately and they mediate our social interactions. Recognising emotion is hence a very important intelligent and social skill. The most common affective cues used to recognize emotions in others are visual (facial expressions and body postures/gestures) and aural ones. As technology is now part of our social lives, the affective computing and the kansei engineering fields have worked in the last decade towards creating technology that has the ability to recognize emotional expressions in its users.

Among the visual affective cues, most of the research in this area has focused on affect that is expressed and perceived from facial expressions in particular, whereas research on body posture has been much less emphasised. Reasons for this difference may be due to the lack of formal models for body posture as there are for the face (e.g., the Facial Action Coding System (FACS) [2]), as well as the complexity of the body. Furthermore, in behavioural science research it was long thought that facial expressions were the main cue for emotion recognition, and gestures and postures were considered to act mainly as a support to emotion recognition [3]. Recent studies in neuroscience and psychology have instead proved that postures and gestures are as valuable as facial expressions [4]. The recent interest in bodily expressions in the affective computing area is also due to the new type of movement-based interaction technology that is becoming available to a larger market. However most of the work done in this area has focused on acted expressions [5,6,7] with a few exceptions [8,9,21] on natural expressions but generally driven by stereotypical rules such as dance. As such they cannot be used as a model for everyday expression recognition software. In real life, real expressions, i.e., non-acted expressions, are subtler and more complex to recognize.

The focus of our study is to create software to recognize non-acted affective postures for applications which could replace a human interaction partner, such as video games. The overall main hypothesis is that affect can be recognised from whole body postures using a low-level description of the body. To investigate this more fully, two sub-hypotheses are made: one, that humans can recognise affect from natural body posture at above chance level; two, that automatic recognition models can achieve accuracy rates similar to benchmarks based on human observers. The benchmark and the system are built upon a dataset collected from players playing Nintendo Wii sports games. The postures are collected between matches and represent the affective expressions that occur after having made or lost a point in the match.

2. BACKGROUND

According to [10,11], changes in a person's affective state are also reflected by changes in body posture. Ekman and Friesen [12] conjecture that postural changes due to affective state aid a person's ability to cope with the experienced affective state. In fact, as seen in behavioural studies [13,14], some affective expressions may be better communicated by the body than by the face. Furthermore, another study by Ekman and Friesen [15] showed that people tend to control their facial expressions more than their bodily expressions when they

are trying to hide their emotions. As counterpart, people trust more expressions of the body than expressions of the face when these two are incongruent [16].

Many of today's affective recognition systems of body posture and movement have focused on extracting emotion information from dance sequences. One of the well known works in this area is by Camurri and colleagues [8]. They examined cues involved in emotion expression in dance for four affective states, anger, fear, grief and pride. Decision trees were chosen to build and test automatic recognition models. The results for the best decision tree model built on testing data ranged between 31% and 46%. For the same set of data, the recognition rate for the human observers was 56% [17].

Turning to non-dance based automatic bodily expression recognition, Berthouze et al. [18,19] proposed a system to recognize basic emotions (happy, sad, fear, anger) and affective dimensions from posture (valence, arousal, action tendency). The system is based on low level features describing the distances between body joints and it was tested on acted postures. The results showed very high levels of recognition rates comparable to human recognition performance. Bernhardt and Robinson [7] have built affect recognition models for non-stylised, acted knocking motions using Pollick et al's motion capture database [20]. They considered three basic emotion categories (angry, happy, sad) and neutral. Their model takes into account individual idiosyncrasies in order to reduce the complexity of the modeling. After training, the classifier was tested on the motion samples from a single actor. The results showed a 50% recognition rate for the motions without removing the personal biases, while recognition significantly increased to 81% using the unbiased motions.

The studies presented in [21,22] focus on non-acted postures. Their system models a more complete description of the body, attempting to recognise three discrete levels of a child's interest [21] and self-reported frustration [22] from postures, facial expressions, and task performance while the child uses a computer to solve a puzzle. Of the three types of input examined, the highest recognition accuracy was obtained for posture activity (55.1%) [21]. Another interesting work for non-acted gestures is presented in [9]. The paper proposes an affective gesture recognition system that recognizes the intensity of the affective state (sad, frustrated, happy and joy) of a child through his/her body gestures while playing computer games. The results showed a 79% agreement between the system and a human observer in the emotion recognition process and a strong agreement for the recognition of the level of intensity of the emotion except in the case of subtle expressions.

To conclude, most of the work has focused on acted bodily expressions or on exaggerated expressions (e.g., dance) that do not occur in everyday situations. Our work proposes a system that recognizes affective states from a low-level description of non-acted body postures. In this paper, we assign to the term *affect* in a broad meaning to include cognitive states. A benchmark created upon an extensive analysis of the agreement between human observers is used to evaluate the system. The next section will present the method used to create the posture data. Section 4 describes the surveys implemented to build the benchmark on human observers. Finally, Section 5 presents the automatic recognition system and evaluates its performance against the human benchmark.

3. CORPORA COLLECTION

The first step in assessing if non-basic emotions can be recognised from non-acted postures was to obtain postural data. A Gypsy 5 electro-mechanical motion capture system (Animazoo Ltd.) was used to numerically record the body motions of the participants during video game play. Eleven players, six females and five males, ranging in age from 20 to 30, were recruited for participation. The players were asked to play sports games with the Nintendo Wii for at least 30 minutes and have their body motions recorded.

After collecting the motion capture data, the apex instants of the motion capture files were manually located. The apex postures were selected during the non-play windows, i.e., during the gaming session in which the player views a replay of the point that was just played. Due to the nature of collecting *non-acted* postural expressions, a player-defined apex instant does not exist nor was there a definitive static posture for many of the motions. Thus, sections of each motion capture file in which affect was displayed needed to be located first. Three university students were recruited as novice coders. The purpose is to examine how untrained lay people interpret affect from posture. They were asked to locate the start and end frames of the replay windows which they felt contained affective bodily expressions. The coders also provided potential affective state labels for these sections to obtain a list of possible affective states to be used in the forced-choice posture judgment survey described in the next section. The labels are grouped according to the affective state that was ultimately chosen for the survey, partially determined from an article by Lazzaro [23] which describes some of the typical affective states associated with general game play: Concentrating (Determined, Focused, Interested); Defeated (Bored, Defeated, Give up, Sad, Tired), Frustrated (Angry, Frustrated) and Triumphant (Confident, Excited, Motivated, Happy, Victory). A total of 103 affective postures were chosen.

4. HUMAN RECOGNITION OF AFFECTIVE POSTURES

The goal of this section is to examine the extent to which human observers can recognise non-basic affective states from static images of non-acted whole body postures. As previously stated, as of yet there are no recognised benchmarks for evaluating human recognition rates, thus chance level is considered the target, as it is the current metric used in affective computing. A benchmark computed on the observers agreements will be used as the benchmark for evaluating the performance of the automatic recognition models discussed.

4.1. Postures judgement survey

To test human performance in recognising affect from posture, the ground truth must be built; labels need to be assigned to the affective postures. The view taken in this research project is that there is no inherent ground truth affective state label that can be attached to the postures. The players are not used to label their own postures because "*self-reported feelings at the end of a task are notoriously unreliable*" [21] and it is not feasible to stop the players during the gaming session to ask them their current affective state. Furthermore, because the complete affective state is expressed through the combination of a variety of modalities in the non-acted scenario in particular, it is difficult for the players to be aware of which modality they used, or if their bodies were expressing their true feelings. Thus, the approach used in this research is to use outside observers' judgments of the postures. An online posture

evaluation survey was conducted using computer avatar stimuli built by using the selected set of postures (see Section 3). Computer avatars were used instead of human photos in order to create a non-gender, non-culturally specific ‘humanoid’ in an attempt to eliminate bias. The goal right now is to understand how posture alone allows for the recognition of affect. As future work, an ideal system would integrate many types of information such as facial expressions, age, gender, etc.

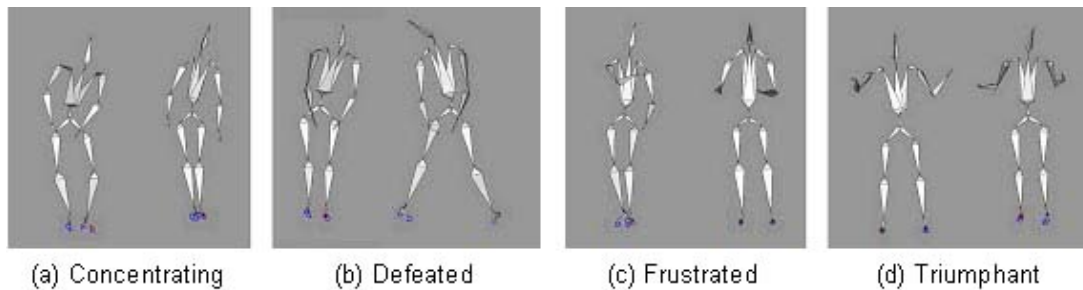


Figure 1: Examples of avatars for each category according to the results from the survey

The posture judgment survey was carried out using two classification methods: an online task as a series of webpages and a card sorting task [25]. These two different methods were used in an attempt to reduce potential boredom experienced by the observers. Both tasks were performed on the entire set of 103 posture stimuli. A forced-choice experimental design was implemented. The survey participants, referred to as *observers* hereafter, were asked to view the postures and associate one of four affective state labels, *concentrating*, *defeated*, *frustrated*, and *triumphant*, to each posture. For both tasks, the posture stimuli were presented in a randomised order. Eight observers participated: five males and three females between the ages of 23 and 31. Each observer made five evaluations on the entire set of postures: one online evaluation and four card sorting evaluations. At least 12 hours elapsed between evaluations.

4.2. Measuring human agreement performance

The results of the survey are presented in Figure 2. Each pie chart represents the frequency of use for each affective state label for an individual posture. The pie charts (i.e., postures) are organized according to the most frequent label assigned to them. Examples of labelled postures (avatars) are provided in Figure 1. The percentage of agreement for each of the 103 postures is above chance level of 25% (considering four affective state categories). By using the most frequent label as the ground truth for each posture, across the 40 evaluations (60 concentrating, 22 defeated, 5 frustrated, 16 triumphant), the total overall agreement and the agreement for each affective state was as follows: Overall agreement 58%, Concentrating 57%, Defeated 64%, Frustrated 39%, Triumphant, 61%. Concentrating is the category containing the most postures. One comment made by a few observers was that they often used this category when they felt that the posture being evaluated did not fit into any of the other categories. Within the Defeated category, disagreement occurs with Concentrating and Frustrated with almost no disagreement for Triumphant as could be expected. Frustrated is the category with the most disagreement in labelling, however, only five postures achieved highest agreement for Frustrated (ranging from 35% to 50%) than for the other labels. These postures are quite different from one another. Most noticeable is that the Frustrated postures

appear more animated than the Concentrating or Defeated postures. Disagreement within the Triumphant category occurs mainly with Frustrated, and Concentrating for a smaller number of postures.

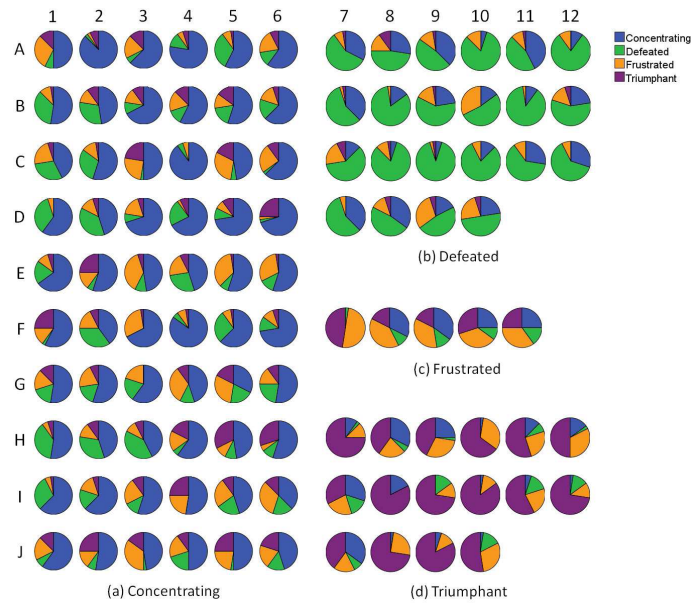


Figure 2: The overall agreement on the set of affective postures for the 40 evaluations made by the eight Wii posture judgment survey observers

Table 1: The overall agreement and their inter-observer reliability (i.e., Cohen’s Kappa) between each pair of observer subsets

Trial	Agreement	Kappa	95% CI	Strength
1	62.14%	0.436	0.313, 0.559	Moderate
2	52.43%	0.295	0.158, 0.432	Fair
3	70.87%	0.542	0.411, 0.673	Moderate
4	69.9%	0.523	0.392, 0.654	Moderate
5	76.7%	0.623	0.498, 0.748	Substantial
6	76.7%	0.600	0.471, 0.729	Moderate
7	63.11%	0.462	0.339, 0.585	Moderate
8	59.22%	0.372	0.233, 0.511	Fair
9	67%	0.470	0.329, 0.611	Moderate
10	68.93%	0.497	0.362, 0.632	Moderate

4.3. Creating the benchmark

To create benchmarks for the human recognition of affective states from posture, the first step is to assess observer reliability. The second step is to create the benchmarks that will be used to evaluate the performance rates of the automatic recognition models. Inter-observer reliability was measured to test the consistency between subsets of observers. The observers were divided into 3 subsets 10 times using random repeated sub-sampling with replacement.

Subsets 1 and 2 (each containing 3 observers) were used to determine the human level of agreement. Subset 3 (containing 2 observers) will be used later to train automatic recognition models. The results are listed in Table 1. Each row constitutes a trial and lists the overall agreement between subsets 1 and 2, Cohen's kappa, the 95% confidence interval and the strength of agreement [24]. The strength of agreement is mainly moderate across the 10 trials, which indicates good agreement beyond chance [25]. The agreement levels for all 4 affective states are above chance level, thus outperforming the target rate. The highest agreement levels are seen for Defeated and Triumphant possibly because they are the two most strongly opposite affective states being studied. The lowest agreement level, as expected is for Frustrated. The average agreement across the 4 categories is 66.7% and it is set as our benchmark for evaluating the system.

Table 2: ANOVA analysis for the low-level features: normalized mean values. The letters “a, b, c,” indicate the pair of states showing significant differences for the corresponding features

Low-level feature	Concentrating	Defeated	Frustrated	Triumphant	P
x rot. torso	.53a	.63abc	.44b	.48c	.003
x rot. L. collar	.45ab	.36acd	.52bc	.49d	.001
x rot. R. collar	.45ab	.36acd	.52bc	.49d	.001
y rot. R. collar	.47ab	.49acd	.59bc	.54d	.039
z rot. L. shoulder	.53ab	.58cd	.48ac	.44bc	.000
y rot. L. shoulder	.55a	.65b	.44	.34ab	.000
y rot. R. shoulder	.50a	.61b	.34	.32ab	.000
z rot. L. elbow	.66a	.54ab	.66	.73b	.003
x rot. L. elbow	.75a	.62ab	.83	.84b	.011
y rot. L. elbow	.64a	.52ab	.62	.72b	.001
z rot. R. elbow	.57a	.50b	.66	.75ab	.000
x rot. R. elbow	.67a	.55b	.77	.86ab	.001
y rot. R. elbow	.13ab	.03ac	.28	.46bc	.000
x rot. L. wrist	.58	.47a	.59	.68a	.004

5. AUTOMATIC RECOGNITION OF AFFECTIVE POSTURES

5.1. Low-level posture description and feature analysis

Each posture stimulus selected according to the procedure set in Section 3 is associated to a vector containing a low-level description of the posture. The description is built upon the 3D joint Euler rotations captured by the motion capture system. Each rotation value is normalised to [0,1] by taking into account the fact that the maximum range of rotation differs for each joint. Some of the joint rotations were removed as they were considered redundant (e.g., baricenter rotations). The remaining joints were: head, neck, collar, shoulders, elbows, wrists, torso, hips and knees.

To evaluate the discriminative power of the low-level posture configuration features for distinguishing between the affective states, each feature was subjected to one-way ANOVAs. The results are summarised in Table 2. Even though a standing scenario was chosen, it is interesting to note that the important low-level posture description features are mainly the arms and upper body. This could indicate that the majority of the movement really is upper body for the type of scenario used. Looking more closely at the results, it can be seen that significant differences occurred for the x rotation of the torso (the degree of forward or backward bending of the body) between the more ‘active’ affective states (Frustrated and Triumphant) and the less ‘active’ states (Concentrating and Defeated). In the case of the x rotation of the collar (the degree of forward slumping or backward straightening of the collar) significant differences occurred between Concentrating and Defeated against Frustrated, and Defeated against Triumphant. Significant differences were also obtained for the z and y rotations of the shoulders. For the z rotation of the shoulder (the lateral and vertical extension of the arm at the shoulder), the significant differences occurred with Concentrating and Defeated against Frustrated and Triumphant. For the y rotation of the shoulder (the rolling of the shoulder causing a lateral to and from frontal movement), the significant differences occurred with Concentrating and Defeated against Triumphant. Similar to the z rotation of the shoulders, for the y rotation, much more ‘movement’ is implied in the Triumphant postures.

Table 3: Automatic recognition performance (average and standard deviation): observer generalization

System total	Concentrating	Defeated	Frustrated	Triumphant
60% (11.61)	66% (18.27)	62% (15.76)	16% (18.82)	64% (14.29)

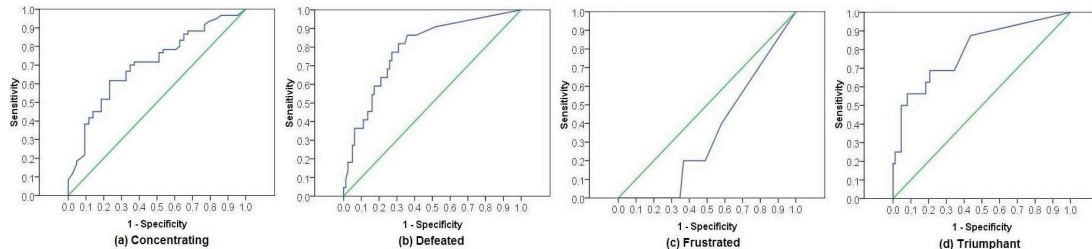


Figure 3: The ROC curves for the 4 affective states for the automatic recognition model built for generalising to novel postures. (a) Concentrating; (b) Defeated; (c) Frustrated; (d) Triumphant

5.2. Automatic recognition

The next task is to build and evaluate automatic recognition models of non-basic affective states from posture. Recognition model testing was conducted for the ability to generalise to: i) new observers and ii) new postures. The input for creating the models is the vector of low-level posture description features as described above, and a non-basic affective state label for each static posture stimulus. To test the automatic recognition models’ ability to generalise to new observers, the models were built with the WEKA back-propagation algorithm and trained using the previously unused subset 3 and then tested with subset 1. The results are summarised in Table 3. The average recognition across all 10 trials was 59.81%, $SD =$

11.61%. The automatic recognition models for seven of the 10 trials achieved recognition rates above the benchmark (66.7%). If we look at each affective state, we notice that the performance on Frustration is very low. This was to be expected given the very small data set available (5 postures) and the low level of agreement among the observers.

Automatic recognition models were also tested for their ability to generalise to novel postures. The 103 postures were associated with the ground truth label most frequently assigned by the 10 observers' 40 evaluations and the vectors of posture description features. An automatic recognition model was built and tested using the WEKA back-propagation algorithm with 10 fold cross-validation. The model achieved a recognition rate of 60.19%. The model's performance on the individual affective state categories are illustrated in the ROC curves in Figure 3. The ROC curves show positive results for all the categories except Frustrated. Again, this is due the low number of postures for this category.

6. CONCLUSION

This paper presented experiments designed to investigate the recognition of non-basic affective states from non-acted body posture information. It was hypothesised that human observers could achieve above chance agreement levels when attributing non-basic affective states to static images of a faceless humanoid avatar. It was also hypothesised that automatic recognition models could be grounded on a set of low-level posture information and achieve recognition rates similar to a benchmark computed from the human observers. Posture data was collected from participants wearing a motion capture suit and playing sports video games with the Nintendo Wii™. Because the players were unaware of the purpose of the study, it is believed that their bodily expressions of affect were non-acted and unsolicited. After locating the affectively expressive sections of the motion capture data, a different set of participants, the observers, was asked to judge static posture images taken from the motion capture data by associating one of four non-basic affective states (concentrating, defeated, frustrated and triumphant) to the postures. To analyse human agreement rates, the level of agreement and reliability was ascertained for the observers' judgments first resulting in an average of 75% agreement (i.e., the benchmark). A repeated sub-sampling method was implemented for testing the models' ability to generalise to new observers. Ten-fold cross-validation was implemented to test the models' ability to generalise to new postures. As hypothesised, the observer agreement levels were above chance level and the performances of the system were similar to the benchmark that was set.

ACKNOWLEDGMENT

This research received funding from the EC under grant agreement MIRG-CT-2006-046434.

REFERENCES

- [1] Damasio A.R., *Descartes' error: Emotion, reason and the human brain*, Avon Books, New York, 1994.
- [2] Ekman P., and Friesen W., *Manual for the facial action coding system*, Consulting Psychology Press, California, 1978.

- [3] Argyle M., *Bodily communication*, Methuen & Co. Ltd, London, 1988.
- [4] Meeren H., van Heijnsbergen C., and de Gelder, B., Rapid perceptual integration of facial expression and emotional body language, *Proceedings of the National Academy of Sciences of the USA*, Vol. 102, No. 45, pp. 16518-16523, 2005.
- [5] Coulson M., Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence, *Journal of Nonverbal Behavior*, Vol. 28, pp. 117-139, 2004.
- [6] Kleinsmith A., De Silva R., and Bianchi-Berthouze N., Cross-cultural differences in recognizing affect from body posture, *Interacting with Computers*, Vol. 18, No. 6, pp. 1371-1389, 2006.
- [7] Bernhardt D., and Robinson P., Detecting affect from non-stylised body motions, *In: Proc. 2nd Intl Conf of ACII, LNCS 4738*, Portugal, pp. 59-70, 2007.
- [8] Camurri A., Lagerlof I., and Volpe G., Recognising emotion from dance movement: Comparison of spectator recognition and automated techniques, *International Journal of Human-Computer Studies*, Vol. 59, No. 1-2, pp. 213-225, 2003.
- [9] De Silva P.R., Osano M., Marasinghe A., and Madurapperuma A.P., Towards recognizing emotion with affective dimensions through body gestures, *Seventh IEEE International Conference on Automatic Face and Gesture Recognition (FG'06)*, pp.269-274, 2006.
- [10] Wallbott H.G. and Scherer K.R., Cues and channels in emotion recognition, *Journal of Personality and Social Psychology*, Vol. 51, No. 4, pp. 690-699, 1986.
- [11] Mehrabian A., and Friar J., Encoding of attitude by a seated communicator via posture and position cues, *Journal of Consulting and Clinical Psychology*, Vol. 33, pp. 330-336, 1969.
- [12] Ekman P., and Friesen W., Head and body cues in the judgment of emotion: A reformulation. *Perceptual and Motor Skills*, Vol. 24, pp. 711-724, 1967.
- [13] Paterson H.M., Pollick F.E., and Jackson E., Movement and faces in the perception of emotion from motion. *Perception, ECVP Glasgow Supplemental*, Vol. 31, No. 118, pp. 232-232, 2002.
- [14] de Gelder B., Towards the neurobiology of emotional body language, *Nature Reviews Neuroscience*, Vol. 7, No. 3, pp. 242-249, 2006.
- [15] Ekman P., and Friesen W., Detecting deception from the body or face, *Journal of Personality and Social Psychology*, Vol. 29, No. 3, pp. 288-298, 1974.
- [16] van Heijnsbergen C.C.R.J., Meeren H.K.M., Grezes J., and de Gelder B., Rapid detection of fear in body expressions, an ERP study. *Brain Research*, Vol. 1186, pp. 233-241, 2007.
- [17] Camurri A., Mazzarino B., Ricchetti M., Timmers R., and Volpe G., Multimodal analysis of expressive gesture in music and dance performances, *Gesture-based Communication in Human-Computer Interaction*, pp. 20-39, 2004.
- [18] Bianchi-Berthouze N., and Kleinsmith A., A categorical approach to affective gesture recognition, *Connection Science*, Vol. 15, No. 4, pp. 259-269, 2003.
- [19] Kleinsmith A., and Bianchi-Berthouze N., Recognizing affective dimensions from body posture, *In: Proc. 2nd Intl Conf of ACII, LNCS 4738*, Portugal, pp. 48-58, 2007.
- [20] Pollick E., Paterson H.M., Bruderlin A., and Sanford A.J., Perceiving affect from arm movement, *Cognition*, Vol. 82, pp. 51-61, 2001.
- [21] A. Kapoor, R.W. Picard, and Y. Ivanov. Probabilistic combination of multiple modalities to detect interest. *Proc 17th Intl Conf on Pattern Recognition*, 3:969-972, August 2004.
- [22] A. Kapoor, W. Bursleson, and R.W. Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724-736, August 2007.

- [23] Lazzaro N., Why we play games: Four keys to more emotion without story. *Technical report, XEODesign, Inc.*, March, 2004.
- [24] Landis J.R., and Koch G.G., The measurement of observer agreement for categorical data, *Biometrics*, Vol. 33, No. 1, pp. 159-174, 1977.
- [25] Banerjee M., Capozzoli M., McSweeney L., and Sinha D., Beyond kappa: A review of interrater agreement measures, *The Canadian Journal of Statistics/ La Revue Canadienne de Statistique*, Vol. 27, No. 1, pp. 3-23, 1999.
- [26] Rugg G., and McGeorge P., The sorting techniques: A tutorial paper on card sorts, picture sorts and item sorts, *Expert Systems*, Vol. 14, No. 2, pp. 80-93, 1997.