

Using Body Movement and Posture for Emotion Detection in Non-Acted Scenarios

Michael Garber-Barron
Cognitive Science Department
Rensselaer Polytechnic Institute
Email: Barron2@rpi.edu

Mei Si
Cognitive Science Department
Rensselaer Polytechnic Institute
Email: Sim@rpi.edu

Abstract—In this paper, we explored the use of features that represent body posture and movement for automatically detecting people’s emotions in non-acted standing scenarios. We focused on four emotions that are often observed when people are playing video games: triumph, frustration, defeat, and concentration. The dataset consists of recordings of the rotation angles of the player’s joints while playing Wii sports games. We applied various machine learning techniques and bagged them for prediction. When body pose and movement features are used we can reach an overall accuracy of 66.5% for differentiating between these four emotions. In contrast, when using the raw joint rotations, limb rotation movement, or posture features alone, we were only able to achieve accuracy rates of 59%, 61%, and 62% respectively. Our results suggest that features representing changes in body posture can yield improved classification rates over using static postures or joint information alone.

I. INTRODUCTION

Emotion detection has wide applications in games, simulations, and computer based training systems [1], [2], [3]. Interactive pedagogical and entertainment systems can benefit greatly from the knowledge of the user’s emotional experiences during their interaction. This will enable the systems to provide more adaptive experiences to those individuals.

A range of approaches have been applied for automatically detecting emotion, including recording and analyzing the person’s physiological responses, electroencephalography (EEG) signals, facial expressions, and body movements. In particular, approaches that utilize facial expressions, physiological and EEG signals have been extensively studied and have shown promising results [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15].

In recent years, along with the advancement of computer graphics hardware, and software, the use of camera devices have become increasingly popular in interactive environments and have been used for detecting the user’s body posture and movement. By exploiting these advances, various research has applied a range of features related to body posture and movement for emotion detection [16], [17], [18], [19], [20], [21], [22], [23], [24].

In this work, we are also exploring the use of body movements and postures for emotion detection. In particular, we examine the degree to which basic and meta-features can contribute to emotion detection. We are interested in developing our approach for detecting emotions in non-acted

scenarios. In contrast, typically, emotion detection studies use acted displays of emotion or are done in performance-based scenarios [19], [18], [21], [16], and the subjects/actors may express their emotions in an exaggerated fashion.

II. RELATED WORK

The use of body cues for detecting emotion has been examined in a variety of scenarios using a range of features. In this section we discuss the related work pertaining to the features, emotions and scenarios that have been previously studied.

Kapur et al. [16] classified the emotions of participants by using the combined information of the velocity and position of their body joints. While standing, participants were directed to act out 4 specific emotions (sadness, joy, anger, and fear), and the movement of their body was recorded. Using this information, they achieved recognition rates of up to 92%. In a similar acted scenario, Kleinsmith and Bianchi-Berthouze [20] classified nine emotions: anger, confusion, fear, happiness, interest, relaxed, sadness, startled, and surprise. Participants were directed to express specific emotions while their full body was recorded. Rather than using movement and position, they used basic posture cues related to the distance of adjacent joints (wrist to elbow, knee to hip, etc.) as features for their algorithm. Their system was able to accurately classify these emotions 70% of the time.

Meta-features have also been applied for emotion detection. Glowinski et al. [17] created an expansive set of expressive features consisting of various metrics such as posture cues (body symmetry), movement (energy, jerk), along with basic features such as the position and velocity of the head. Acted portrayals of emotion were clustered into four clusters of emotions. After applying their features, the portrayals were then automatically assigned to the four categories and out of 120 portrayals only one was incorrectly classified.

Gunes et al. [22] achieved similarly high results in an acted seated scenario. Body posture was represented through changes in body extension, and arm and upper body position (arms crossed, hand on face, shoulder shrug, etc.) By applying this set of features, they were able to accurately classify emotions 96% of the time among eight categories (anger, disgust, fear, happiness, sadness, surprise, happy_surprise, uncertainty.)

Bernhardt and Robinson [24] observed lower classification rates in more natural, acted tasks. Participants were required to perform actions with a specified emotional disposition, like angrily knocking on a door. Primarily using temporal information of the arm (velocity, acceleration, and jerk), their model discriminated between happy, sad, angry, and neutral portrayals with success rates ranging from 50% to 80%.

Emotion detection has also been explored in non-acted settings. Kapoor et al. [25] identified the participants' interest level while they were performing computer tasks in a seated pose. The body lean of participants was used and they achieved classification results of 55%.

Similarly D'Mello and Graesser [26] examined the application of basic and meta-features in a seated environment. Features related to body lean and body pressure were used to classify the emotions (boredom, confusion, delight, flow, frustration) of the participants while they interacted with a computer based tutor. They achieved classification rates of 40%.

Savanghi et al. [27] applied a set of meta-features to differentiate people's emotions in a seated, non-acted task. Extending Castellano et al. [28]'s work, they evaluated the participants' engagement level while playing chess and interacting with a small robot. By applying meta-features that incorporated movement (quantity of movement) and posture (back curvature, body openness) information, they were able to determine the participants' level of engagement correctly 82% of the time.

Less commonly, emotion detection has been applied to non-acted full body tasks. Kleinsmith and Bianchi-Berthouze [29] recorded the joint rotations of participants playing Wii sports video games. By using the changes in the rotation of the joints over sequences of poses they were able to classify emotions correctly 57% between the categories of defeated, triumphant and neutral. Using the joint rotation information statically, Kleinsmith et al. [30] achieved results of approximately 60% for recognition between four categories of emotion (frustration, triumph, concentration, defeat).

Our work used the data set from Kleinsmith and Bianchi-Berthouze's work [29]. Like Kleinsmith and Bianchi-Berthouze, our goal is to detect the subjects' emotions using their full body while they were playing games naturally. The features used by Kleinsmith and Bianchi-Berthouze [30], [29] primarily concentrated on the use of raw joint information. Sanghvi et al. [27] achieved promising results when meta-features related to the movement and posture cues were integrated for detecting emotions in a non-acted scenario. However, they only applied their algorithms to the upper body. On the other hand, high success rates have been achieved using meta-features in standing settings [17], [22] for detecting emotions, but the studies were done in acted situations. Based on these previous works, we hypothesize that meta-features that represent the subjects' posture and movement can aid in the recognition of emotions when the subjects are standing and in natural non-acted scenarios. In this paper, we evaluate this hypothesis by comparing the accuracies of emotion detection

when using basic features and meta-features.

III. EMOTION DETECTION FEATURES

A. Feature Definition

The categories of features we applied consist of seven types of cues: symmetry and limb alignment, head alignment and offset, body openness, average rate of change, relative movement, smooth-jerk, and location of activity. Next we will discuss each of these types of features in turn.

Symmetry and Limb Alignment:

Roether et al. [31] found that asymmetries in the elbow, hip, knees, and shoulders can differentiate between the emotions of sad, happy, and angry. Likewise, Wallbott [32] observed that the position of the arms relative to the body and face provided salient information for emotion recognition. Gunes et al. [22] observed a similar effect when applying information about the relative position and shape of the arms to each other. Glowinski et al. [17] also demonstrated that the symmetry of the arms provided an effective cue for automatically clustering emotions.

In our work the symmetry of the subject's posture is represented by a sequence of features. Our algorithm for evaluating symmetry is similar to that used by Glowinski et al. [17]. We examined how each side of the body is oriented while accounting for the rotation and position of the person's body center. We estimated the asymmetries resulting from the joint rotation angles slightly differently. Algorithm 2 (**Directed_Symmetry**) is used for estimating the asymmetries of joint locations. The result returned for this algorithm reflects the direction of the asymmetries. Algorithm 1 (**Pose_Symmetry**) is used for estimating the asymmetries caused by the misalignment of joints and does not calculate the direction of the asymmetries. These metrics are applied to both the joints in the upper body and in the lower body for generating separate features in the Posture Group in Table 1.

Algorithm 3 (**Pose_Difference**) addresses the asymmetries between the two sides of the body in another way. It returns the mean misalignment of the rotation angles of corresponding joints on the left and the right side of the body. We applied this algorithm to both the joints in the upper body, and in the lower body for generating features in the Posture Group in Table 1.

Head Alignment and Offset:

The relationship between the head and the body is a crucial piece of information to identify emotion, and has been recognized and applied in numerous emotion detection studies [33], [34], [35], [21]. Walbott [32], [36] identified the position of the head as important for classifying emotions such as joy, anger, and disgust. Ekman and Friesen [37] concluded that the position and tilt of the head provided more information than the rest of the body for distinguishing emotions.

We represent the head's relationship to the body in several ways. The head's rotation is compared to the rotation of the hips (HeadAlignment) and chest (HeadChestRatio). In addition, we also measure the location of the head relative to

the hips (HeadOffset). The details for each metric are provided in Algorithm 4 (**Head_Offset_Alignment**).

Body Openness:

Openness relates to the degree to which a person’s body and limbs are extended or closed. In Laban and Ullman [38]’s annotation scheme, they used this feature for differentiating different movements of the body. Monteparre et al. [39] observed that expanded body poses were associated with anger and happiness, and contracted poses with sad and neutral states. Several affect detection studies have [22], [19], [17] applied variations of this feature in acted settings, and Cammurri et al. [18] used this feature for emotion detection in performance and dance settings.

While openness is typically calculated using features from the person’s upper body, in this work we consider the status of the lower body and is shown in Algorithm 5 (**Leg_Hip_Openness**). The function is calculated by examining the distance between the locations of each leg, and the distance of the ankles to the hips. The distance between the legs represent the openness of the lower body. The distance from the hips to the knees represents the degree to which the torso is bent towards the floor.

Average Rate of Change:

One of the basic features for describing movement is speed. Both Wallbott [32] and Montepare et al. [39] observed that the speed of body movements is used by people to evaluate other individuals’ emotions. Montepare et al. [39] demonstrated that subjects typically labeled faster body movements as happy or angry, and slower movements with sad and neutrality. Wallbott [32] observed a similar pattern - greater movement was associated with emotions such as joy and anger, and less motion was associated with sadness. Castellano et al. [40] observed that using head speed as a feature enhanced their model’s ability to predict the emotions of pianists during performances.

In this work we calculate speed as the average rate of change of a feature over a specified interval of time. This is illustrated in Algorithm 6. We apply the average rate of change to the rotation angle vectors of each joint and to the proposed meta-features. The relevant features are shown within the limb rotation movement, and posture movement groups in Table 1.

Relative Movement:

Meta-features related to the level of movement have been applied to emotion detection. Inspired by Laban’s movement research [38], Castellano et al. [40] and Camurri et al. [18] proposed a feature known as quantity of movement. This feature represents the amount of movement detected in the body. Glowinski et al. [17] proposed a similar concept that combines the force and velocity of a participant’s movements.

To create a similar effect we record the amount of change of a specified feature over a window of time, and compare it to the average movement of that same feature over the entire set of recordings. This algorithm is intended to represent the level of change of a feature at a particular moment relative to the average level of change (see Algorithm 7 for details). Relative movement is applied to the features displayed in the

limb rotation movement, and the posture movement group in Table 1.

Smooth-Jerk:

The frequency of change in body position is a relevant movement cue. Montepare et al. [39] found that fast changing and jerky movements are strongly associated with anger, whereas sad and neutral states are associated with smoother body movements. Boone and Cunningham [41] found that children used changes in the direction of the torso and faces to help distinguish other individuals’ anger intensity. Bernhardt and Robinson [24], and Glowinski et al. [17] have also used this feature for differentiating and predicting emotions in their research.

We defined this algorithm as the relative variance of a feature over time as shown in Algorithm 8. Using this algorithm, less rapid changes will generate smaller values, and sudden changes will generate bigger values. Smooth-Jerk is applied to features displayed in the limb rotation movements, and posture movement groups in Table 1.

Location of Activity:

The activity level of different limbs can signify different emotions [42]. This algorithm encodes the degree that each body part is active by using the whole body as a baseline. This feature is computationally defined by normalizing the average change in the amount of rotation of a specified limb relative to the rest of the body (see Algorithm 9 for details.) In the current scenario, the location of activity is only applied to the head as using the other limbs did not measurably impact our predictions (see the posture group in Table 1 for details.)

Algorithm 1 Pose_Symmetry(left, right, center)

#left, right, center: Each parameter is a three dimensional vector corresponding to a joint on the left side, right side, and the center of the body

$$\text{HorizontalSymmetry} = \left| \frac{|center_x - left_x| - |center_x - right_x|}{|right_x - left_x|} \right|$$

$$\text{VerticalSymmetry} = \left| \frac{|center_y - left_y| - |center_y - right_y|}{|right_y - left_y|} \right|$$

Return $\frac{\text{HorizontalSymmetry}}{\text{VerticalRotationSymmetry}}$

Algorithm 2 Directed_Symmetry(left, right, center)

SymmetryDirected = **Pose_Symmetry**(left, right, center)

LeftSideDistance = **Distance**(left, center)

RightSideDistance = **Distance**(right, center)

#Change the sign of the variable depending on the direction of any asymmetry

if *LeftSideDistance* ≥ *RightSideDistance* **then**

Return *SymmetryDirected* * -1

Return *SymmetryDirected*

Algorithm 3 Pose_Difference($L_Joint1, R_Joint1, L_Joint2, R_Joint2, L_Joint3, R_Joint3$)

$L_Joint1, R_Joint1, L_Joint2, R_Joint2, L_Joint3, R_Joint3$: Three pairs of vectors that represent the rotation angles for corresponding joints on the left and right side of the body

$Distance1 = \text{Distance}(L_Joint1, R_Joint1)$
 $Distance2 = \text{Distance}(L_Joint2, R_Joint2)$
 $Distance3 = \text{Distance}(L_Joint3, R_Joint3)$

Return $\text{Average}(Distance1, Distance2, Distance3)$

$\text{Distance}(vector1, vector2)$: Returns the Euclidean Distance between two vectors

Algorithm 4 Head_Offset_Alignment($Head_Loc, Head_Rot, Hips_Loc, Hips_Rot, Chest_Rot$)

Avg_Hips_Loc : A vector representing the midpoint of the left and right hips
Avg_Hips_Rot : A vector representing the mean rotational angle of the left and right hips
$Head_Rot, Chest_Rot$: Vectors representing the rotation angles of the head and the chest
$Head_Loc, Chest_Loc$: Vectors representing the locations of the head and chest

$HeadOffset = \text{Distance}(Head_Loc, Avg_Hips_Loc)$
 $HeadAlignment = \text{Distance}(Head_Rot, Avg_Hips_Rot)$

$Ratio_x = \left(\frac{Head_Rot_x}{Chest_Rot_x} \right)$
 $Ratio_y = \left(\frac{Head_Rot_y}{Chest_Rot_y} \right)$
 $Ratio_z = \left(\frac{Head_Rot_z}{Chest_Rot_z} \right)$

$HeadChestRatio = \text{Average}(Ratio_x, Ratio_y, Ratio_z)$

Return $\{HeadOffset, HeadAlignment, HeadChestRatio\}$

Algorithm 5 Leg_Hip_Openness($L_Ankle, R_Ankle, L_Knee, R_Knee, L_Hip, R_Hip$)

$L_Ankle, R_Ankle, L_Knee, R_Knee, L_Hip, R_Hip$: Vectors representing the locations of the joints on the left and the right side of the body

$Avg_Ankles = \text{Average}(L_Ankle, R_Ankle)$
 $Avg_Hips = \text{Average}(L_Hip, R_Hip)$

$Hip_Ankle_Dist = \text{Distance}(Avg_Hips, Avg_Ankles)$

$Knee_Dis = \text{Distance}(L_Knee, R_Knee)$

Return $\frac{Hip_Ankle_Dist}{Knee_Dis}$

Algorithm 6 Average_Rate_Of_Change($Window, Feature$)

$Feature_t$: The value of the feature at time t can be either a vector or single value
$Window$: The duration of time, set to a default of 100ms

Return $\sum_t \frac{\text{Distance}(feature_t, feature_{t-1})}{Window}$

Algorithm 7 Relative_Movement($Feature, Window, History$)

$History$: Entire length of recording

Return $\frac{\text{Average_Rate_of_Change}(Window, Feature)}{\text{Average_Rate_of_Change}(History, Feature)}$

Algorithm 8 Smooth_Jerk($Feature, Window, History$)

Return $\frac{\text{Variance}(Window, Feature)}{\text{Variance}(History, Feature)}$

$\text{Variance}(Window, Feature)$: Variance of the feature over a window of time

Algorithm 9 Location_of_Activity($Limb, Arms, Legs, Chest, Hips, Window, History$)

$Limb$: The rotation angles of a limb
$Arms$: A vector representing the sum of the rotation angles of the wrist, shoulders, elbows
$Legs$: A vector representing the sum of the rotation angles of the ankles, hips, and knees
$Chest$: A vector representing the rotation angles of the chest
$Head$: A vector representing the rotation angles of the head

$Activity = 0$

for each Vector in $\{Head, Arms, Legs, Chest\}$ **do**
 $Activity += \text{Relative_Movement}(\text{Vector}, \text{Window}, \text{History})$

Return $\frac{\text{Relative_Movement}(Limb, Window, History)}{Activity}$

B. Dataset

We used the UCLIC Affective Body Posture and Motion Database [29] for training and testing our model. The data was collected in a non-acted standing setting and contains information about the participants' full body. The database contains motion capture data of eleven participants playing various Nintendo Wii Sports games for a minimum of thirty minutes. The data in the recordings are the Euler angles (rotation angles) of the participants' joints along the X, Y, and Z axis. The recorded joints include the head, chest, neck, shoulders, elbows, wrists, collars, hips, ankles, and knees. We calculated the locations of each joint using the provided rotation angles.

Kleinsmith et al. [29] picked 103 frames on the basis of their emotional expressiveness out of the entire set of recordings,

and labeled them with one of the four possible emotions: concentrated, frustrated, triumphant, and defeated, which are typically observed emotions during game play. Each category can be described with a set of emotional terms as listed below [29]:

- Concentrated: determined, focused, interested.
- Defeated: defeated, surrendered, sad.
- Frustrated: angry, frustrated.
- Triumphant: confident, excited, motivated, happy, victory.

We removed two frames due to mismatches between the location of the labels and the associated frames in the data set. Therefore, 101 labeled frames were used in our study.

C. Apply Machine Learning Algorithms For Emotion Detection

We explored machine learning techniques that have often been applied for emotion detection in this work, including a Neural Network(NN), Support Vector Machine(SVM), Random Forest, a Logistic Model Tree(LMT), and K-Nearest-Neighbor(KNN). We used the Weka [43] software package as the basis for implementing these algorithms. With the exception of the NN, and KNN, the default parameters for each of the algorithms in Weka were used. For the NN the learning rate was set to .2. For the KNN, the number of instances to compare was set to 5 because we sought to have selection be based on an average as opposed to a single closest instance which is the default setting. Finally, bagging was applied over all of the above algorithms [44] using a majority vote for label prediction.

The data was comprised of both the rotation angles of the joints and the generated features described in section III-A. The input was normalized. Afterwards, ten-fold cross validation was performed.

In order to evaluate the effectiveness of the different types of features, we divided the features into four groups as shown in Table 1. The features in Table 1 include both the raw features and the meta-features generated from the functions discussed above. If not specified, the default parameters of the functions were used. The features in the posture group are given abbreviated names so that they can be referenced when describing the meta-features in other groups.

The posture group is comprised of features related to the generated body posture cues (see Algorithms 1- 5). The limb rotation movement group is comprised of the features that describe the relative movement of different body parts, the smoothness of the movement, and the rate of change of each of the primary limbs' orientation (head, right arm, left arm, left leg, and right leg). It should be noted that the arm limb is represented as the average of values associated with the wrist, elbow, and shoulder joints. Likewise, the leg limb is the average of values related to the ankle, knee, and hip joints.

The posture movement group was created as an attempt to integrate the temporal cues with the posture group. This group consisted of the application of relative movement, smooth-jerk, and rate of change to the posture features described in Algorithms 1 to 5. Finally, the raw joint rotation group consisted

TABLE I
FEATURE GROUPS

<p>Group 1. Posture Group (10 Features):</p> <p>P0: Pose_Difference(Left Arm, Right Arm rotations) P1: Pose_Difference(Left Leg, Right Leg, Head rotations) P2: Pose_Symmetry(Left Arm, Right Arm, Head rotations) P3: Directed_Symmetry(Left Arm, Right Arm, Head location) P4: Pose_Symmetry(Left Leg, Right Leg, Hip location) P5: Directed_Symmetry(Left Leg, Right Leg, Hip location) H1: HeadOffset - generated in Head_Offset_Alignment H2: HeadAlignment - generated in Head_Offset_Alignment H3: HeadChestRatio - generated in Head_Offset_Alignment L1: Leg_Hip_Openness</p>
<p>Group 2. Limb Rotation Movement (18 Features):</p> <p>Average_Rate_of_Change, Smooth_Jerk, and Relative_Movement functions applied to each of the following features:</p> <p>Left Arm (Average of left shoulder, wrist, elbow rotations), Right Arm (Average of right arm rotations), Head, Chest, Left leg (Average of left hip, knee, ankle rotations), Right leg (Average of right leg rotations)</p>
<p>Group 3. Posture Movement Group (26 Features):</p> <p>Average_Rate_of_Change, Smooth_Jerk, and Relative_Movement functions applied to each of the following features:</p> <p>P1, P3, P4, P5, D2, H1, H2, H3</p> <p>Average_Rate_of_Change function applied to L1 Location_of_Activity function applied to the Head</p>
<p>Group 4. Joint Rotation Group (45 Features):</p> <p>The rotation angles along the X,Y,Z axis for each of the following joints:</p> <p>Neck Joint Head Joint Chest Joint Left Shoulder Joint Right Shoulder Joint Left Elbow Joint Right Elbow Joint Left Wrist Joint Right Wrist Joint Left Collar Joint Right Collar Joint Left Knee Joint Right Knee Joint Left Hip Joint Right Hip Joint</p>

of the raw data - the rotation angles of each joint (head, neck, elbows, shoulders, collars, knees, ankles, hip, and wrists.)

IV. RESULTS AND DISCUSSION

Our overall accuracy for differentiating the four emotions are comparable to that of human raters [29]. Overall, we found that using meta-features can help in the detection of these four emotions in standing, non-acted scenarios. By integrating the movement and posture cues (posture movement group) we achieved a classification rate of 66.5% for differentiating the four emotions. When movement features were only generated from joint angles (the limb rotation movement group) or used alone (the raw joint rotation group) without considering the posture cues, classification accuracy decreased by 4% and 7%.

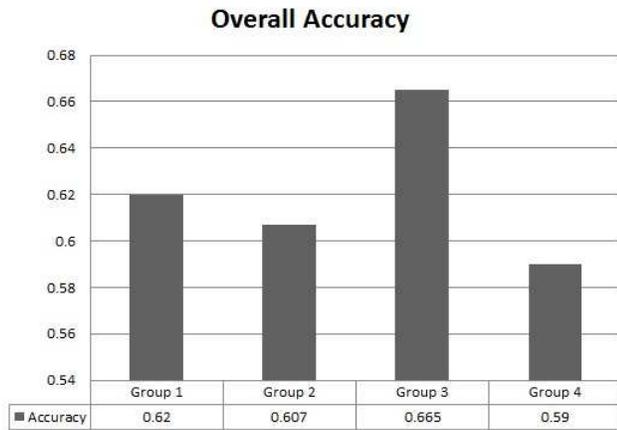


Fig. 1. Overall Accuracy for Differentiating the Four Emotions

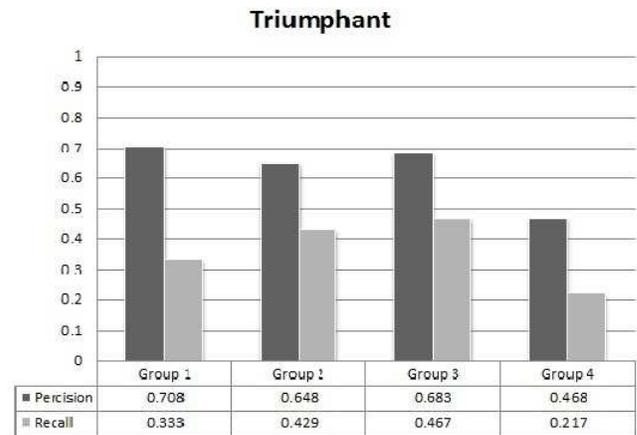


Fig. 3. Results for Identifying Triumphant

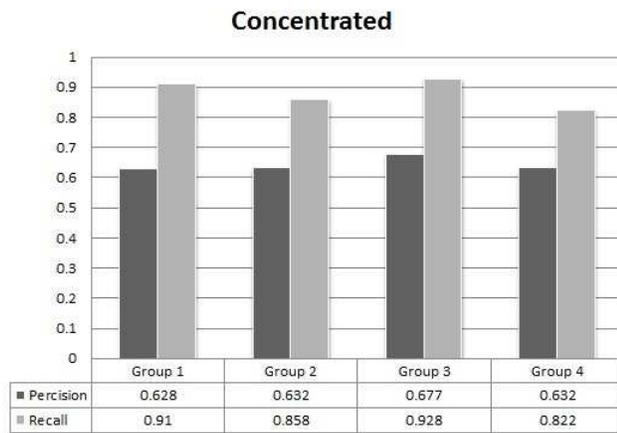


Fig. 2. Results for Identifying Concentrated

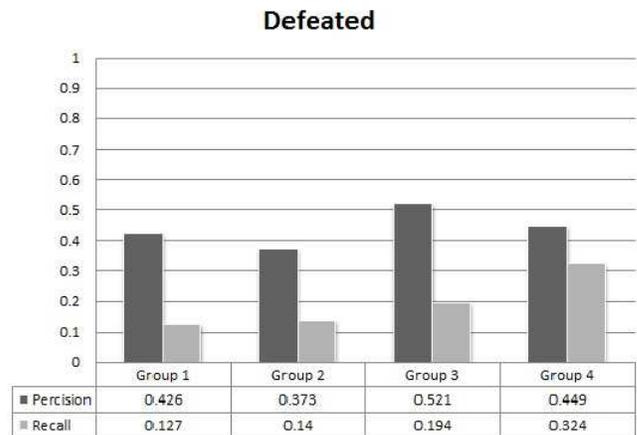


Fig. 4. Results for Identifying Defeated

Finally, using the raw joint rotations alone, we only obtained a success rate of 59%.

We have also observed that the effectiveness of the feature groups vary for identifying different emotions. These results are shown in Figs. 2, 3, 4 and discussed below.

For identifying triumphant, precision rates were higher when using the posture features and the posture movement features (70% and 68.3%). We achieved slightly lower precision rates (64%) using the limb rotation movement features. The rates dropped below 50% when the raw joint feature rotations were used alone (46.8%).

In terms of recall, results were higher using movement, i.e. features from the posture movement group and the limb rotation movement group (47% and 43%). When using static cues we achieved much lower recall rates (33% for the posture group, and 23% for the raw joint rotation group.) We think this difference is likely due to the fact that faster body movements are correlated with positive emotions related to triumphant [39]. Therefore, movement features are more effective for identifying triumphant than the static features.

For identifying concentrated, we achieved the highest recall rate of 92.8% and a precision rate of 67.7% using the posture movement features. We obtained higher recall rates (91%) using the posture group as well. In contrast the precision rates were approximately the same for all groups except the posture movement group (62.8%, 63.2%, 63.2% - as shown in Fig. 2).

Using any of the four feature groups, the recall rate for triumphant is higher than 80%. However, we suspect this may be an artifact because concentrated accounts for 60% of our training data and the bagged machine learning algorithms may be tuned to favor this emotion.

For identifying defeated we achieved the highest precision results using the posture movement features (52%). However, for recall, the raw joint rotation group (32%) outperformed the three other groups (12%, 14%, 19%). We suspect that the generally poorer recall performance is due to the fact that emotions like sad and defeated are associated with subtle static postures [36] that are similar to the postures labeled as concentrated.

Our classifier was not able to identify any frustrated case

using all four groups of features. This is probably because there are few data points for this category - only five data points were labeled as frustrated in the original data set.

V. CONCLUSION

Effectively detecting emotions in real time can facilitate a range of applications, e.g. help game designers dynamically adjust the content of the game and create a personalized experience for players. Previous work along this line primarily used joint angles and static cues for predicting players emotions. In this work, we explored the use of meta-features that incorporate the players body movements and posture for detecting emotions in a non-acted standing setting. In particular, we focused on four emotions users often experience when playing video games: frustration, defeat, triumph, and concentration. Our results suggest that body posture and movement features should be used in addition to basic cues (joint angles) for emotion detection.

VI. FUTURE WORK

Future work has been planned along several directions. Firstly and most importantly, we intend to examine whether the features we applied in this research are useful for identifying a wider range of emotions that a player can experience. In particular, we are interested in collecting new data and testing our models in nontraditional game environments. These include interacting with a pedagogical system or having a standing conversation with an avatar.

Secondly, we plan to improve the accuracy of our emotion detection algorithms by incorporating additional features that represent the status of the game interaction. This way, we won't be completely relying on a bottom-up approach for detecting emotions, and we will be able to consider the cognitive factors of the person as well as their experiences. For example, if the player/user has just lost in a game, it is more likely for the user to experience negative emotions than positive emotions.

Finally, emotion is a complex psychological and physiological phenomenon. People express emotions in many different channels. Moreover, information conveyed through these channels are not always consistent with each other, e.g., a person may be able to control their facial expressions but still reveal their true emotion through their body languages. We plan to improve our algorithms for emotion detection by trying to detect and represent such discrepancies.

ACKNOWLEDGMENT

We thank the researchers of the University College London Interactive Centre for granting us access to the UCLIC Affective Body Posture and Motion Database.

REFERENCES

- [1] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," in *the proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, ser. MindTrek '08. ACM, 2008, pp. 13–17.
- [2] S. D'Mello, A. Graesser, and R. Picard, "Toward an affect-sensitive autotutor," *Intelligent Systems, IEEE*, vol. 22, no. 4, pp. 53–61, july-aug. 2007.
- [3] M. Moussa and N. Magnenat-Thalmann, "Applying affect recognition in serious games: The playmancer project," in *the proceedings of Motion in Games*, ser. Lecture Notes in Computer Science, A. Egges, R. Geraerts, and M. Overmars, Eds. Springer Berlin / Heidelberg, 2009, vol. 5884, pp. 53–62.
- [4] R. L. Mandryk and M. S. Atkins, "A fuzzy physiological approach for continuously modeling emotional experience during interaction with computer games," Tech. Rep. 4, 2007.
- [5] C. Becker, H. Prendinger, M. Ishizuka, and I. Wachsmuth, "Empathy for Max (Preliminary Project Report)," in *the proceedings of the International Conference on Active Media Technology (AMT-05)*, 2005.
- [6] R. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.
- [7] J. Scheirer, R. Fernandez, J. Klein, and R. Picard, "Frustrating the User on Purpose: A Step Toward Building an Affective Computer," *Interacting with Computers*, vol. 14, pp. 93–118, 2002.
- [8] T. Saari, N. Ravaja, M. Turpeinen, and K. Kallinen, "Emotional Regulation System for Emotionally Adapted Games," in *the proceedings of the FuturePlay 2005 conference*. Michigan State University, USA., 2005, pp. 13–15.10.
- [9] S. McQuiggan, S. Lee, and J. Lester, "Predicting User Physiological Response for Interactive Environments: An Inductive Approach," in *the proceedings of the Second Conference on Artificial Intelligence and Interactive Entertainment*, Marina del Rey, California, 2006, pp. 60–65.
- [10] H. Prendinger and M. Ishizuka, "The Empathic Companion: A Character-based Interface that addresses users' affective states," *International Journal of Applied Artificial Intelligence*, vol. 19, pp. 267–285, 2005.
- [11] J. Healey and R. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, June 2005.
- [12] P. Rani, N. Sarkar, and C. Liu, "Maintaining Optimal Challenge in Computer Games Through Real-Time Physiological Feedback," in *the proceedings of the HCI International Conference*, Las Vegas, USA., July 2005.
- [13] B. Herbelin, P. Benzaki, F. Riquier, O. Renault, and D. Thalmann, "Using physiological measures for emotional assessment: a computer-aided tool for cognitive and behavioural therapy," in *the proceedings of the 5th International Conference on Disability, Virtual Reality and Associated Technology*, Oxford, UK, 2004, pp. 307–314.
- [14] G. Chanel, J. Kronegg, D. Gr, and T. Pun, "Emotion assessment: arousal evaluation using eegs and peripheral physiological signals," in *the proceedings of the International Workshop on Multimedia Content Representation, Classification and Security*, 2006, pp. 530–537.
- [15] K. Schaaff and T. Schultz, "Towards an eeg-based emotion recognizer for humanoid robots," in *the proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan, 2009.
- [16] A. Kapur, N. Virji-Babul, G. Tzanetakis, and F. Driessen, "Gesture-based affective computing on motion capture data," in *the proceedings of Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, J. Tao, T. Tan, and R. Picard, Eds. Springer Berlin / Heidelberg, 2005, vol. 3784, pp. 1–7.
- [17] N. Glowinski, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Towards a minimal representation of affective gestures," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 106–118, 2011.
- [18] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe, "Multimodal analysis of expressive gesture in music and dance performances," in *the proceedings of Gesture-Based Communication in Human-Computer Interaction*, ser. Lecture Notes in Computer Science, A. Camurri and G. Volpe, Eds. Springer Berlin / Heidelberg, 2004, vol. 2915, pp. 357–358.
- [19] G. Castellano, S. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *the proceedings of Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, A. Paiva, R. Prada, and R. Picard, Eds. Springer Berlin / Heidelberg, 2007, vol. 4738, pp. 71–82.
- [20] A. Kleinsmith and N. Bianchi-Berthouze, "Recognizing affective dimensions from body posture," in *the proceedings of Affective Computing and*

- Intelligent Interaction*, ser. Lecture Notes in Computer Science, A. Paiva, R. Prada, and R. Picard, Eds. Springer Berlin / Heidelberg, 2007, vol. 4738, pp. 48–58.
- [21] H. Gunes and M. Piccardi, “Automatic temporal segment detection and affect recognition from face and body display,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 64–84, 2009.
- [22] H. Gunes and M. Piccardi, “Affect recognition from face and body: Early fusion versus late fusion,” in *the proceedings of the IEEE International Conference of Systems, Man, and Cybernetics*, 2005, pp. 3437–3443.
- [23] A. Camurri, S. Hashimoto, K. Suzuki, and R. Trocca, “Kansei analysis of dance performance,” in *the proceedings of the International Conference on Systems, Man and Cybernetics*, 1999, pp. 327–332. [Online]. Available: <http://www.iit.tsukuba.ac.jp/kenji/pdf/SMC99.pdf>
- [24] D. Bernhardt and P. Robinson, “Detecting affect from non-stylised body motions,” in *the proceedings of Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, A. Paiva, R. Prada, and R. Picard, Eds. Springer Berlin / Heidelberg, 2007, vol. 4738, pp. 59–70.
- [25] A. Kapoor, R. Picard, and Y. Ivanov, “Probabilistic combination of multiple modalities to detect interest,” in *the proceedings of the 17th International Conference on Pattern Recognition*, vol. 3, Aug. 2004, pp. 969 – 972 Vol.3.
- [26] S. D’Mello and A. Graesser, “Automatic detection of learner’s affect from gross body language,” *Appl. Artif. Intell.*, vol. 23, pp. 123–150, Feb 2009.
- [27] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. McOwan, and A. Paiva, “Automatic analysis of affective postures and body motion to detect engagement with a game companion,” in *the proceedings of the 6th international conference on Human-robot interaction*, ser. HRI ’11. New York, NY, USA: ACM, 2011, pp. 305–312. [Online]. Available: <http://doi.acm.org/10.1145/1957656.1957781>
- [28] G. Castellano, A. Pereira, I. Leite, A. Paiva, and P. McOwan, “Detecting user engagement with a robot companion using task and social interaction-based features,” in *the proceedings of the 2009 international conference on Multimodal interfaces*, ser. ICMI-MLMI ’09. New York, NY, USA: ACM, 2009, pp. 119–126. [Online]. Available: <http://doi.acm.org/10.1145/1647314.1647336>
- [29] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, “Automatic recognition of non-acted affective postures,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 4, pp. 1027–1038, Aug. 2011.
- [30] A. Kleinsmith and N. Bianchi-Berthouze, “Form as a cue in the automatic recognition of non-acted affective body expressions,” in *the proceedings of Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, S. D’Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds. Springer Berlin / Heidelberg, 2011, vol. 6974, pp. 155–164.
- [31] C. Roether, L. Omlor, and M. Giese, “Lateral asymmetry of bodily emotion expression,” *Current Biology*, vol. 18, no. 8, pp. R329–R330, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0960982208002303>
- [32] H. Wallbott, “Bodily expression of emotion,” *European Journal of Social Psychology*, vol. 28, no. 6, pp. 879–896, 1998.
- [33] A. Kapoor, W. Bursleson, and R. Picard, “Automatic prediction of frustration,” *International Journal of Human-Computer Studies*, vol. 65, no. 8, pp. 724 – 736, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1071581907000377>
- [34] S. Petridis, H. Gunes, S. Kaltwang, and M. Pantic, “Static vs. dynamic modeling of human nonverbal behavior from multiple cues and modalities,” in *the proceedings of the ACM International Conference on Multimodal Interfaces*, 2009, pp. 23–30.
- [35] M. Coulson, “Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence,” *Journal of Nonverbal Behavior*, vol. 28, no. 2, pp. 117–139, 2004.
- [36] H. G. Wallbott and K. R. Scherer, “Cues and channels in emotion recognition,” *Journal of Personality and Social Psychology*, vol. 51, pp. 690–699, 1986.
- [37] P. Ekman and W. Friesen, “Head and body cues in the judgement of emotion,” *Perceptual and Motor Skills*, vol. 24, no. 3, pp. 711–724, 1967.
- [38] R. Laban and L. Ullman, *The Mastery of Movement*, 3rd ed. Plays, 1971.
- [39] J. Montepare, E. Koff, D. Zaitchik, and M. Albert, “The use of body movements and gestures as cues to emotions in younger and older adults,” *Journal of Nonverbal Behavior*, vol. 23, pp. 133–152, 2008.
- [40] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, “Automated analysis of body movement in emotionally expressive piano performances,” *Music Perception*, vol. 26, pp. 103–120, 2008.
- [41] R. Boone and J. Cunningham, “Children’s decoding of emotion in expressive body movement: the development of cue attunement,” *Developmental Psychology*, vol. 34, no. 5, pp. 1007–1016, September 1998.
- [42] M. Meijer, “The contribution of general features of body movement to the attribution of emotions,” *Journal of Nonverbal Behavior*, vol. 13, pp. 247–268, 1989, 10.1007/BF00990296.
- [43] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and H. Ian, “The weka data mining software: An update,” *SIGKDD Explorations*, vol. 11, 2009.
- [44] L. Breiman, “Bagging predictors,” in *Machine Learning*, 1996, pp. 123–140.