Automatic Affect Classification of Human Motion Capture Sequences in the Valence-Arousal Model

William Li
School of Interactive Arts and Technology
Simon Fraser University
dla135@sfu.ca

Philippe Pasquier
School of Interactive Arts and Technology
Simon Fraser University
pasquier@sfu.ca

ABSTRACT

The problem that we are addressing is that of *affect classification*: analysing emotions given input data. There are two parts to this study. In the first part, to achieve better recognition and classification of human movement, we investigate that the labels on existing Motion Capture (MoCap) data are consistent with human perception within a reasonable extent. Specifically, we examine movement in terms of valence and arousal (emotion and energy). In part two, we present machine learning techniques for affect classification of human motion capture sequences in both categorical and continuous approaches. For the categorical approach, we evaluate the performance of Hidden Markov Models (HMM). For the continuous approach, we use stepwise linear regression models with the responses of participants from the first part as the ground truth labels for each movement.

Author Keywords

Movement; motion capture; affect classification; valence; arousal.

ACM Classification Keywords

I.2.6 Artificial Intelligence: Learning

INTRODUCTION

In the recent growing interest of developing technology to recognize people's affective states [11], more and more studies have shown that body expressions are effective in conveying emotion [4, 27]. As such, there is an increasing demand for development of affect recognition systems which in turn has potential impacts in clinical and entertainment contexts. Thus in this paper, we developed an affect classification system using the valence-arousal (VA) model of human emotion, and is using full body motion capture data as input, which does not contain any information regarding facial expressions or voice. When considering the three aspects of movement, functional (the task of the movement, such as picking up a cup), executional (the pattern of movement, such as using the left or right hand to pick up the cup), and expressive dimensions (the emotions behind the movement) [2], we are essentially measuring the expressive dimension of full body movements.

There are two parts to this study. This first part was conducted to determine whether human participants would classify movement to the same labels given the same model of affect and to establish the ground truths to be used in the continuous affect classification. In the second part of this study, we create categorical models using HMM, as well as a continuous model using stepwise linear regression for affect classification.

This system has both off-line and on-line applications. The main off-line applications involve database labelling, which is especially useful for development of movement databases [22]. A valid and reliable classifier for movement expressivity would allow us to automatically label existing motion capture data according to the valence-arousal model. In on-line scenarios, such a classifier could be used in interactive arts or therapy contexts. Such a system can also be used in generating movement with user-specified valence and arousal [2].

Our goal is therefore to estimate affect expressed by movement using the VA model. The system output is 1 of 9 classes of VA combinations as shown in Figure 1, with each of valence and arousal taking a label of low, neutral, or high for the categorical approach, and a number between -1 and 1 for each of valence and arousal for the continuous approach.

For the rest of the paper, we start by outlines of the related work in affect classification. After that, we describes the data and the processing used in our study. Next is experimental methods, materials, and participants. Then, we present the experiment results in the categorical and continuous models. Lastly, we end with concluding remarks and future work.

BACKGROUND WORK

In affect classification, considerations that come into play include the mood the mover is expressing, the intended mood of the mover, and the perceived mood of the mover. [20, 17]. Malandrakis et al. [19] have shown that there can be a difference even in award-winning movies in the intended and experienced emotions. The difference between good and bad movies is essentially how well the intended emotions are portrayed. With their experiment, they used award-winning films and expert annotaters to narrow the gap between the intended

MOCO'16, July 5–6, 2016, Thessaloniki, Greece.

DOI: http://dx.doi.org/10.1145/2948910.2948936

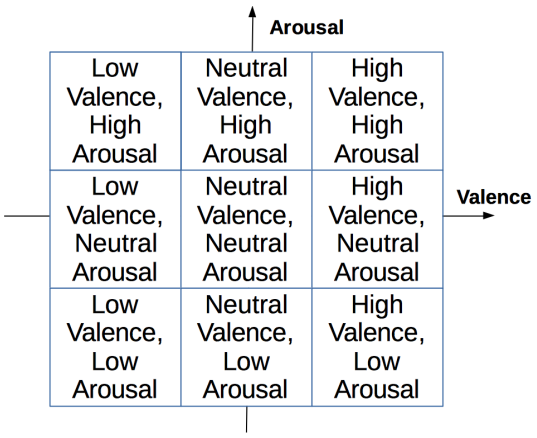


Figure 1. Valence-Arousal Combinations

and expressed emotions. In our experiment, we are able to direct the movers, so we assume the intended affect is identical to the expressed affect. Therefore, it is important to note that the categorical approach in our study did not contain a user survey. As such, the categorical system is more of a predictor of the intended affect, whereas the continuous model is more of a predictor of the perceived affect.

In the field of affective computing, facial expressions are often examined in the determination of affective states [9, 13]. However, Inderbitzin et al. [14] have shown that it is possible to perceive VA states from movement even on faceless generated characters, regardless of viewing angle. They have even identified some canonical parameters that control the expression of emotions in locomotive behavior, such as upright upper body postures being perceived as more emotionally positive and vice-versa for forward leaning postures. Other documented sources also suggest that humans convey emotions through body movement and postures [7, 8]. Analysis of head pose and movement is able to achieve 71.2% accuracy in recognizing depression [3]. Furthermore, studies in movement have shown certain features in expressive movement, such as portrayal of strength, can be linked to specific emotions, such as fear or anger [8, 28].

In affect estimation based on body movement, there have been many studies in using dance with mixed results [5, 15, 23]. Kapur et al. developed classifiers that achieved comparable regconition rates as observers using dance movements [16]. However, as Kleinsmith points out, dance is often exaggerated to convey affect [17].

Looking at non-dance-based systems, Castellano et al. have attempted to infer emotional states using video analysis on movement qualities such as amplitude, speed, and fluidity. Their system was able to discriminate between "high" and "low" arousal emotions and "positive" and "negative" [6]. Pollick et al. conducted a study to compare the performance of their automatic system with human recognition. In their study, they used 3D positioning measurements of the arm in knocking, lifting, and waving motions with two affective states, neutral and angry. They concluded that the automatic system was able to discriminate between the two states more consistently than humans [24]. Samadani et al. developed a system for both full body as well as hand-arm improvisiation movements to discriminate between 4 affective states using HMMs with good results [26].

Nicolau et al. developed a system for estimation of affect modalities in the Valence-Arousal space using multi-modal input (based on facial expression, shoulder gesture, and audio cues). Their approach claims to be unique in that it performs *continuous* affect prediction according to the valence-arousal model. In their paper, they compare both Support Vector Machines (SVM) and bi-directional Long-Short-Term Memory Neural Networks (BLSTM-NN), concluding that BLSTM-NN perform better [21]. However, we have decided not to use BLSTM-NN due to the fact that they were using different sets of input data (extracting data from video and audio as well as mainly focusing on facial expressions); in our case we are using motion capture data with no facial expressions. Furthermore, the lack of a benchmark and standard skeleton markers due to the use of different datasets and body markers in the aforementioned studies makes it difficult to compare and evaluate different systems.

More generally, classification of motion capture data into categories has been explored in different contexts. For example, Adam et al. outlined an approach using clustering and Hidden Markov Models for identification of humans by gait [1].

For our study, we will be using Russell's model of affect [25]. A drawback of Russell's model is that some researchers such as Fontaine et al. [10] are starting to believe more dimensions are needed to describe the emotional space. To our knowledge, automatic systems in affect classification using motion capture data and such a broad coverage of VA states has not yet been attempted.

DATASET

Motion Capture Data

As part of the efforts of the MovingStories project, an open source MoCap database (`http://moda.movingstories.ca/`) [22] has been created. For this study, we are using some of the recordings in this database that have been labelled according to the circumplex model of affect [25]. The data are in the form of MoCap bvh files and accessible in the MoCap database (`http://moda.movingstories.ca/projects/29-affective-motion-graph`). Two professional actors, one male and one female, performed in the videos, carrying out 9 different types of movements: walking in a figure eight pattern, hugging, static improvisation, free improvisation, sitting down, pointing while sitting, walking with sharp turns, improvisation, and lying down. There are 9 takes for each movement, corresponding to the 9 different possible VA combinations shown in Figure 1, covering more emotional states than similar existing datasets (e.g. 4 emotions in the library presented by Ma et al. [18]). Existing labels were created by dividing the Russell's model [25] into low, neutral, and high along both the valence and arousal axis. Using this model, anger would be classified as low on the valence axis but high on the arousal axis.

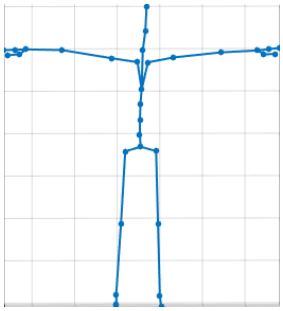


Figure 2. MoCap skeleton

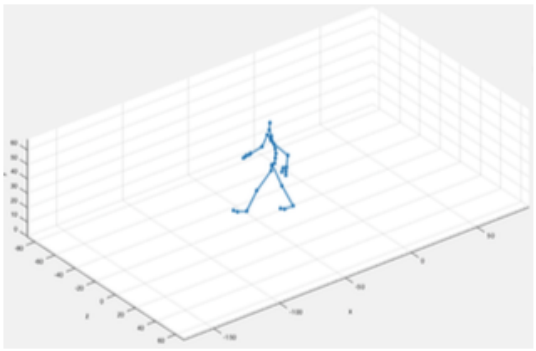


Figure 3. Motion Capture Frame

The data were recorded with a Vicon motion capture system with 53 markers, mapped to a skeleton representation with 30 joints as shown in Figure 2, at 120 frames per second. Sequences varied in length from 2500 frames to 10000 frames. A sample frame rendered in Matlab is shown in Figure 3. Each frame contains the Euler angles for each of the joints, as well as the spatial location and orientation of the skeleton root.

Pre-Processing

Rotational positions for joints are expressed in Euler angles ($-180\,\text{deg}$ to $+180\,\text{deg}$), which are used as features in both the categorical and the continuous approach. We normalized these values to -1 to 1. We also translated the values for the center of mass position to delta values (i.e. $\forall n > 0$, $(dx_n, dy_n, dz_n) = (x_n - x_{n-1}, y_n - y_{n-1}, z_n - z_{n-1})$), thereby eliminating bias due to geometrical translation. Furthermore, we computed rotational velocity and acceleration values for all joints and for the centroid position. We have chosen to use the low level positional information as features as a first approach. We notice this is also the approach taken by related works [24, 21]. It is not clear yet which specific high level features would be effective for our data, ratings, and affective states. However, this is an area to be explored in our future directions.

Ground Truth

Given that this is a supervised machine learning task, it is necessary to choose ground truth values for the labels assigned to samples in the training set. In our case, the database was curated and therefore labeled when the recordings were made. For the categorical approach, we are accepting this annotation as ground truth. For the continuous approach, we used labels provided by external observers as ground truth. We did not use these labels in the categorical approach as ground truth because the performers were only given the categories and not a continuous spectrum. Therefore, the categorical approach can be considered as a classification of the intended emotions rather than the perceived.

In order to ensure the inter-rater reliablility, and in turn the validity of our experiment, we will also examine the Intraclass Correlation Coefficient (ICC) as a measure of the inter-rater reliability of the ratings. To evaluate the validity of our system, we test the categorical approach using MoCap sequences that were labelled the same way as the training data but never seen before by our system. For the continuous approach we examine the coefficient of determination for our regression models.

METHODS

We conducted an online survey to obtain valence and arousal ratings from observers in order to establish ground truth on the continuous scale from which to construct our linear regression models.

Participants

The participants of the validation process were 33 undergraduate second and third year students enrolled in a computer animation course. Other background information about the participants was not collected. All responses were anonymous and could not be traced back to any particular student. All students are willing participants with informed consent in this study and were able to drop out of the study at any time during the survey without consequences. Participants were not offered compensation for this study.

Materials

The video clips used range in length from about 10 to 25 seconds of two different professional actors. Each clip shows either an actor walking in a figure eight pattern or walking in a straight line with several sharp turns. Each clip has been labelled in terms of valence and arousal levels as either high, neutral, or low by the region on the affect grid corresponding to the valence and arousal prompt given to the actors when they performed the movements. We used 3 takes of movements for the survey, 2 for the figure eight pattern, and 1 for the sharp turns. Each take covers all 9 possible combinations of valence and arousal, with an additional neutral valence, neutral arousal clip for each take, resulting in a total of 30 videos, the order of which are randomized for each participant. The responses for the extra videos shown in the survey were used for another experiment and were not included in this analysis. We did not include the other movements such as hugging and free improvisation in order to keep the survey at a reasonable length. The other movements were included in the categorical approach due to the fact that it did not involve participants.

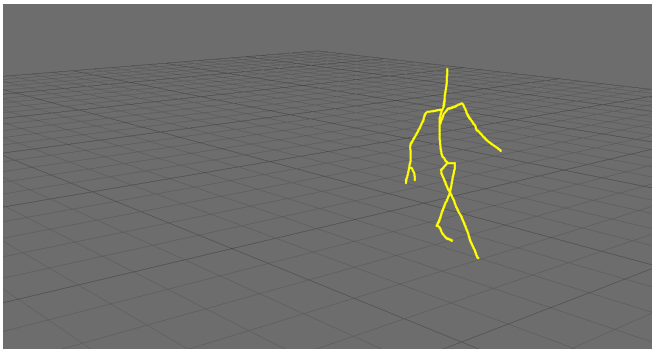


Figure 4. Screenshot of video in survey

The motion capture data are converted to video mp4 format from bvh files, the widely accepted format for motion capture data. The bvh motion capture files only show a skeletal format without facial expressions. As such, participants must rely on movement cues to discern levels of valence and arousal. Figure 4 is an example screenshot from a video in the survey.

Procedure

In the first part of the study, the model of affect was defined and explained for the participants. The series of 30 videos was then presented as an online survey where the participant recorded on a two-dimensional scale interface what he or she perceived to be the valence and arousal for that video. Each response is then recorded in a database. Figure 5 is an example of the interface the participants see after making a response. The specific mood descriptors shown in the figure were chosen because they were the same descriptors given to the actors in their instruction during recording of the movement data used in this study. Participants were free to ask for clarification regarding the model or the interface at any point during the survey. No post-experiment questions were asked.

After the survey, we eliminate biased or outlier results before conducting further analysis. A response was considered biased or an outlier if it was obvious that the participant did not watch the video closely, ie. giving a response after 2 seconds or less. In the second part of the study, the procedure for each different approach is described below.

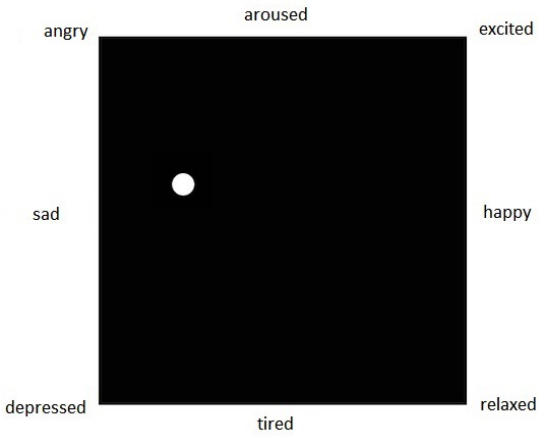


Figure 5. Sample response on the interface. This particular case illustrates a low valence and a relatively neutral arousal.

EXPERIMENT AND RESULTS

For our experiments, we built both categorical and continuous systems in affect estimation. In application, categorical labels are simpler construct. Continuous systems, however, obviously offers a more detailed comparison between movements. In the context of animation, users can specify the exact amount of valence or arousal for their characters. Furthermore, for a sequence of movement, categorical labelling gives discrete classification at each time step whereas continuous labelling can show the trajectory of change in affect over time.

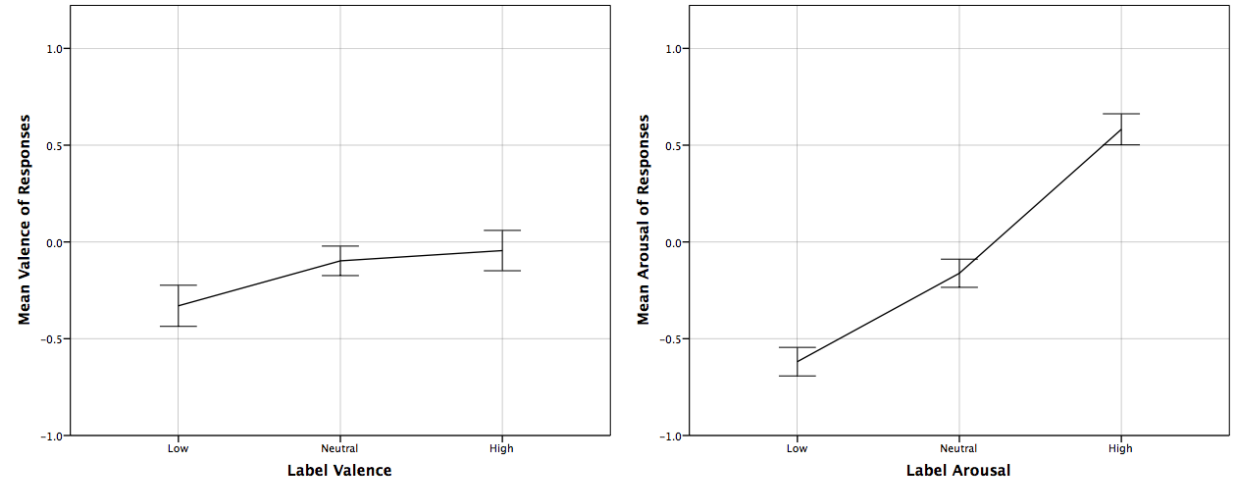
Categorical Model

In our categorical approach, we construct models using the existing VA labels for each movement as the ground truth. The labels represent the VA that were given as instructions to the actors when they performed the movements. They are essentially the VA that the actors were attempting to portray. We also examined the survey results here to see how the observer ratings compare with the VA labels. Observers agree with these categorical labels mostly in the low to neutral range. Furthermore, the categories are more distinguished for observers in arousal than valence. Figures 6 and 7 present a summary of the valence and arousal ratings given by participants from the survey.

The survey shows that participants are consistent in almost all cases in distinguishing between low and high for both valence and arousal, an exception being that responses for neutral and high valence of the sharp and figure eight walk were not significantly different from each other. However, although the relative perception is similar to the existing labels, the participants view almost all neutral and high valence and arousal movements to be lower than the amount the actors attempted to express. In contrast, most low valence and arousal movements were perceived to be closer to neutral. This further suggests there is a difference between the intended and perceived affect when only considering body movement.

We tested Hidden Markov Models (HMM), Support Vector Machines (SVM), and k-Nearest Neighbour (k-NN), with HMM producing the best results. SVM were implemented using *libsvm* (`https://www.csie.ntu.edu.tw/~cjlin/libsvm/`). We achieved the best SVM results of 43% accuracy using a linear kernel, nu-SVC classification with nu-value of 0.8. The k-NN was implemented using the Matlab package (`http://www.mathworks.com/help/stats/knnsearch.html`). We experimented with a variety of distance functions, with the best result coming from the *City-block*, achieving 54% accuracy. Both SVM and k-NN results were the best achieved through cross-validation.

Using Kevin Murphy's Hidden Markov Model toolbox (`http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html`), we trained nine HMM models, one for each combination (NVNA, HVLA, etc.). Each HMM has five states and is built using a Gaussian distribution on the training data,



(a) Valence responses

(b) Arousal responses

Figure 6. Mean participant responses to sharp walk videos. Error bars indicate 95% confidence interval

chosen because these parameters yielded the best results. For any given test input, the model that gave the highest log probability was considered to be the prediction made by the system.

In the first run, all motion-capture sequences of one actor were used as the training and the sequences of the other actor was used as test. The log-likelihood for each test sequence was calculated for each of the nine models. The label of the model that returned the highest log-likelihood was taken to be the prediction. In this case, the HMM achieved 16.87% test accuracy. This low accuracy is likely due to the models overfitting to each particular actor.

In the second run, the first half of all motion-capture sequences was used as the training and the second half of all sequences was used as test. The same parameters and procedures were used otherwise. In this case, the HMM achieved 72.56% test accuracy. This was the best accuracy achieved through cross-validation. The majority of errors came from sequences of movements that differ in gesture, such as improvisation and hugging. It could also be the case that these movements have more varied movement signatures. For example, perhaps the actors use an occasional arm raise in the second half of the MoCap, resulting in a movement that is still logical in the context of a free improvisation movement but is a gesture that the model has never seen before. In contrast, using movements with a smaller number of movement signatures such as the figure eight walk and the sharp walk results in a test accuracy of 89.29%.

Continuous Model

An important issue of using a continuous approach is that inter-rater reliability of VA ratings amongst the participants is challenging [12]. To address this, we first examine the ICC. Table 1 shows the ICC of ratings in both valence and arousal using Cronbach's α as an index. An α of 0.7 or higher is generally considered to be good reliability. In this case, both

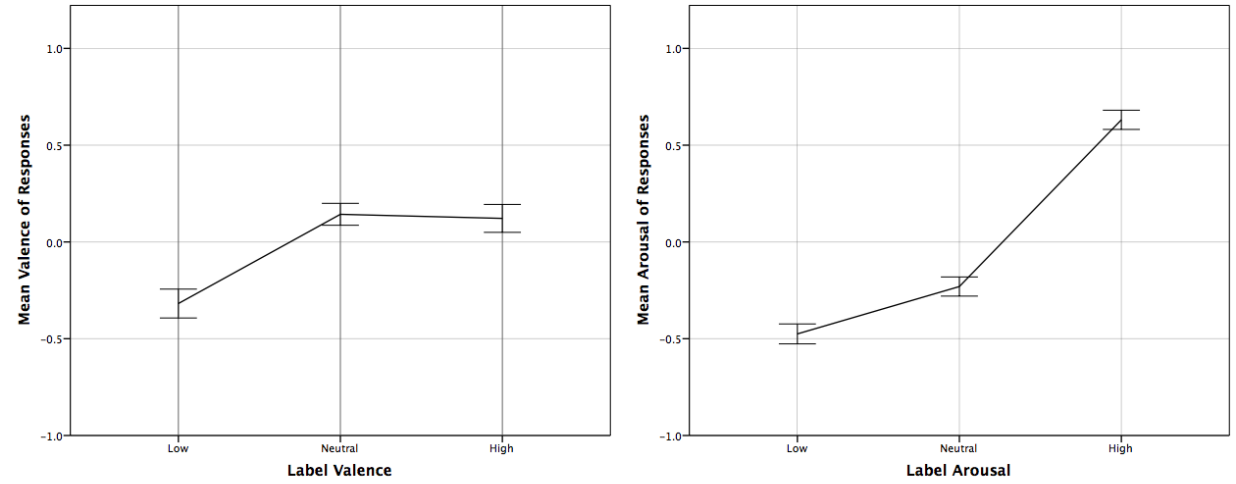
	Measure	Intraclass Correlation
Valence	Cronbach's α	0.89
Arousal	Cronbach's α	0.98

Table 1. Intraclass Correlation Coefficient of Ratings

the valence index, with a 95% confidence interval of 0.814 to 0.945 and the arousal index, with a 95% confidence interval of 0.97 to 0.991, suggest that the ratings provided are reliable enough to use for building models. However, the higher index for arousal suggest that it is easier for observers to agree on the level of arousal than the level of valence, most likely because the energy is more easily differentiable than the emotion without a facial expression.

For the continuous approach, we used a stepwise linear regression model to fit the features extracted from the bvh file to the predictor value for valence and arousal given by the participants. We chose a stepwise linear regression because models generated from different users are expected to be quite different and the stepwise regression can easily and automatically explore different choices of predictive variables for different models. This was implemented using the linear regression package in Matlab (`http://www.mathworks.com/help/stats/stepwiselm.html`). We constructed models based on the figure eight walk and the sharp turn as those were the only movements in the online survey. We built valence and arousal models for each individual user as well as for the entire group. Each model was built from 25 frames of every movement for which the participant provided a response. We downsampled from 120 to 5 frames per second, resulting in a total of 5 seconds of movement. This ensures that the frames used in the construction of models are frames that the participants have viewed.

To evaluate our linear regression models, we examine the coefficient of determination (R^2), which indicates how well the data fit a linear regression. An R^2 is defined in (1), with f_i



(a) Valence responses

(b) Arousal responses

Figure 7. Mean participant resopnses to 8 walk videos. Error bars indicate 95% confidence interval

being the prediction made by the model and $\bar{y}$ being the mean of the ratings. Therefore, a high R^2 indicates that the regression fits the data well. Figure 8 shows the R^2 of valence and arousal models built using the responses of each participant. The individual user valence models have a mean R^2 of 0.86 with a standard deviation of 0.068. The arousal models have a mean R^2 of 0.93 with a standard deviation of 0.046. Using the average rating for each movement, the all-users model results in a R^2 of 0.925 for valence and 0.985 for arousal with $p < 0.001$. Testing these models on 25 frames covering another 5 second span, the individual models produce an average mean squared error (MSE) of 0.183 for valence and 0.145 for arousal. The combined model results in an MSE of 0.0748 for valence and 0.0583 for arousal. We define MSE in (2), where $\hat{y}_i$ is the prediction made by the system and y_i is the rating given by the user.

$$\begin{aligned} R^2 &= 1 - \frac{SS_{res}}{SS_{tot}} \\ SS_{res} &= \sum_i (y_i - f_i)^2 \\ SS_{tot} &= \sum_i (y_i - \bar{y})^2 \end{aligned} \tag{1}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y_i)^2 \tag{2}$$

Lastly, we conducted a Pearson correlation test between valence and arousal on all the participant responses of 1163 data points, resulting in an insignificant ($p > 0.05$) negative correlation. This suggests viewers can perceive valence independently from arousal.

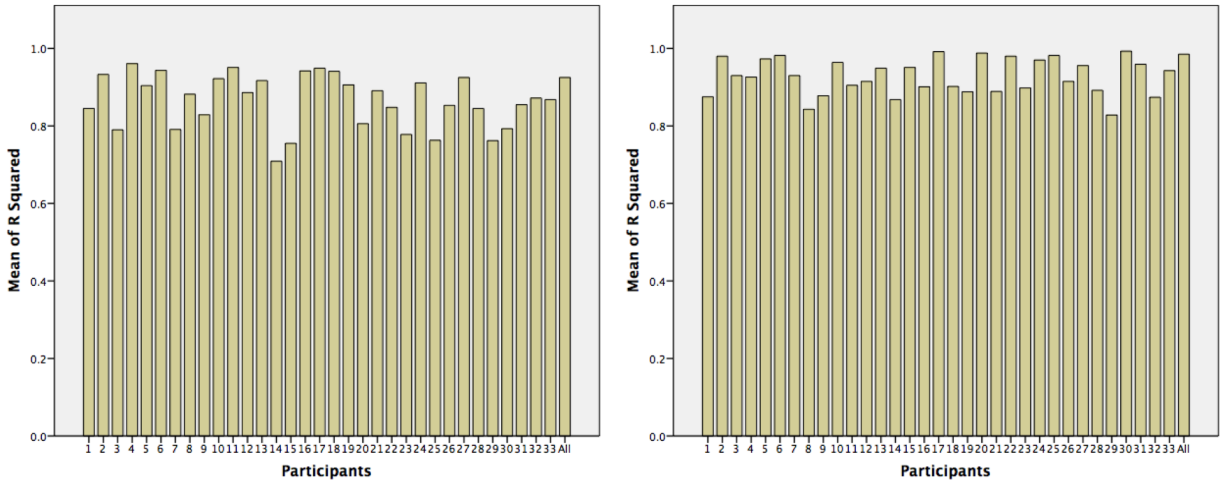
DISCUSSION

In the categorical approach, Hidden Markov Models achieved much better results than SVM and k-nearest neighbour. Actor bias clearly plays a role, and we can improve our results by training on a dataset including a greater number of actors. It also appears that the majority of errors arose from movements with a less defined functional and executional dimensions of movement, such as improvisations, suggesting that certain movements might require a different model of classification or different feature extraction. In comparison with the survey results, HMM using features that are potentially unnoticeable to observers suggests that a significant difference exists between the VA expressed by the mover and the VA perceived by the observer. In that case, the categorical labelling of the HMM is only accurate in the sense that it classifies the intended VA, but not the perceived VA.

In the continuous approach, stepwise linear regression has shown it can produce individual models with high coefficient of determination. However, the individual models generally have a high MSE. A model built from the average responses of all participants had a higher R^2 and a lower MSE than the average individual model, suggesting that using more participant responses will result in a better model. In almost all cases, arousal models show less variability and less error, together with the difference in the ICC between valence and arousal ratings suggest that it is easier for observers to agree on arousal than valence. In contrast to the categorical system, the continuous system makes predictions on the perceived VA rather than the expressed VA. Potential future work includes constructing models using a subset of participants or using ratings of expert movers as ground truth to achieve better models as well as extending to other movements than walking.

CONCLUSION AND FUTURE WORKS

We built systems for automatic affect classification in the valence-arousal space using both categorical and continuous



(a) Valence models

(b) Arousal models

Figure 8. Average R^2 of models

approaches. For the categorical approach, we used the pre-existing labels used to prompt the actors as the ground truth. For the continuous approach, we conducted an online survey to obtain continuous ratings from observers to establish a ground truth. Our experiments determined that people can perceive affect in MoCap without facial expression, agree on the affect they perceive, and train machines to classify affect.

For both categorical and continuous approaches, a drawback of our experiment was that even though there were many movement types in our dataset, it was all from two actors. This can lead to the classification systems overfitting to these two actors. Therefore, it is among our future directions to expand the database to include more actors. Furthermore, we are also looking into exploring dimensionality reduction techniques and higher level feature selection. Our goal is to eventually be able to generalize to a more varied dataset with more actors.

REFERENCES

1. Adam, A. Identifying humans by their walk and generating new motions using hidden markov models. *The University of British Colombia, Topics in AI: Graphical Models and Computer Animation, Technical Report* (2004).
2. Alemi, O., Li, W., and Pasquier, P. Affect-expressive movement generation with factored conditional restricted boltzmann machines. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*, IEEE (2015), 442–448.
3. Alghowinem, S., Goecke, R., Wagner, M., Parkerx, G., and Breakspear, M. Head pose and movement analysis as an indicator of depression. In *Proceedings of the 2013 Conference on Affective Computing and Intelligent Interaction*, IEEE Computer Society (2013), 283–288.
4. Argyle, M. *Bodily Communications*. Methuen & Co. Ltd, 1988.
5. Camurri, A., Lagerlf, I., and Volpe, G. Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies 59*, 1-2 (2003), 213–225.
6. Castellano, G., Villalba, S. D., and Camurri, A. Recognising human emotions from body movement and gesture dynamics. In *Affective Computing and Intelligent Interaction*, vol. 4738 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2007, 71–82.
7. de Gelder, B. Towards the neurobiology of emotional body language. *Nature Reviews Neurosciences 7*, 3 (2006), 242–249.
8. de Meijer, M. The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior 13*, 4 (1989), 247–268.
9. Fasel, B., and Luettin, J. Automatic facial expression analysis: a survey. *Pattern recognition 36*, 1 (2003), 259–275.
10. Fontaine, J. R., Scherer, K. R., Roesch, E. B., and Ellsworth, P. C. The world of emotions is not two-dimensional. *Psychological science 18*, 12 (2007), 1050–1057.
11. Fragopanagos, N., and Taylor, J. Emotion recognition in human-computer interaction. *Neural Networks 18*, 4 (2005), 389–405.
12. Gunes, H., and M., P. Automatic dimensional and continuous emotion recognition. *Int. Journal of Synthetic Emotions 1* (2010), 68–99.
13. He, S., Wang, S., Lan, W., Fu, H., and Ji, Q. Facial expression recognition using deep boltzmann machine from thermal infrared images. In *Proceedings of the 2013 Conference on Affective Computing and Intelligent Interaction*, IEEE Computer Society (2013), 239–244.

14. Inderbitzin, M., Väljamäe, A., Calvo, J. M. B., Verschure, P. F., and Bernardet, U. Expression of emotional states during locomotion based on canonical parameters. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, IEEE (2011), 809–814.

15. Kamisato, S., Odo, S., Ishikawa, Y., and Hoshino, K. Extraction of motion characteristics corresponding to sensitivity information using dance movement. *Journal of Advanced Computational Intelligence and Intelligent Informatics 8*, 2 (2004), 168–180.

16. Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G., and Driessen, P. Gesture-based affective computing on motion capture data. In *Affective Computing and Intelligent Interaction*, vol. 3784 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2005, 1–7.

17. Kleinsmith, A., and Bianchi-Berthouze, N. Affective body expression perception and recognition: A survey. *IEEETransactions on Affective Computing 4*, 1 (2013), 15–33.

18. Ma, Y., Paterson, H. M., and Pollick, F. E. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods 38*, 1 (2006), 134–141.

19. Malandrakis, N., Potamianos, A., Evangelopoulos, G., and Zlatintsi, A. A supervised approach to movie emotion tracking. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2011), 2376–2379.

20. Mehrabian, A., and Russell, J. A. *An approach to environmental psychology*. M.I.T. Press, 1974.

21. Nicolaou, M. A., Gunes, H., and Pantic, M. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *Affective Computing, IEEE Transactions on 2*, 2 (2011), 92–105.

22. Nixon, M., Bernardet, U., Alaoui, S., Alemi, O., Gupta, A., Schiphorst, T., DiPaola, S., and Pasquier, P. Moda: an open source movement database. In *Proceedings of the 2nd International Workshop on Movement and Computing*, ACM (2015).

23. Park, H., Park, J., Kim, U., and Woo, W. Emotion recognition from dance image sequences using contour approximation. In *Structural, Syntactic, and Statistical Pattern Recognition*, vol. 3138 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2004, 547–555.

24. Pollick, F. E., Lestou, V., Ryu, J., and Cho, S. Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research 42*, 20 (2002), 2345 – 2355.

25. Russell, J. A. A circumplex model of affect. *Journal of personality and social psychology 39*, 6 (1980), 1161–1178.

26. Samadani, A.-A., Gorbet, R., and Kulic, D. Affective movement recognition based on generative and discriminative stochastic dynamic models. *Human-Machine Systems, IEEE Transactions on 44*, 4 (2014), 454–467.

27. Van den Stock, J., Righart, R., and De Gelder, B. Body expressions influence recognition of emotions in the face and voice. *Emotion 7*, 3 (2007), 487–494.

28. Wallbott, H. G. Bodily expression of emotion. *European journal of social psychology 28*, 6 (1998), 879–896.