Brain Inspired Cognitive Systems
August 29 – September 1, 2004
University of Stirling, Scotland, UK

TOWARDS A BIOLOGICALLY-INSPIRED REPRESENTATION OF HUMAN AFFECT

YingLiang Ma, Psychology Department, University of Glasgow, Glasgow, G12 8QB, UK

Helena Paterson, Psychology Department, University of Glasgow, Glasgow, G12 8QB, UK
Alexander N. Dolia, Department 504, National Aerospace University, 17 Chkalova Street, 61070 Kharkov, Ukraine
School of Electronics and Computer Science, University of Southampton,
Southampton, SO17 1BJ, UK
Sung-Bae Cho, Department of Computer Science, Yonsei University, Seoul 120-749, South Korea
Ales Ude, ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan
Frank Pollick, Psychology Department, University of Glasgow, Glasgow, G12 8QB, UK

ABSTRACT

We propose a method to reveal the features used by humans in the classification of human movement and apply it to the case of classifying arm movements as angry or happy. The method begins with psychophysical experiments investigating the human classification of point-light movements. Then by comparing the results from these perception experiments with the results of principal component decomposition, we can find a particular feature component that has the highest correlation with human perception. In order to verify the component, we reconstruct movements by using either the first two PCA components alone as well as those two components and the feature that correlates highly with human perception. Finally, we used a Parzen window to test the recognition effectiveness of these reconstructed movements.

1. INTRODUCTION

Humans are adept at gleaning a variety of social signals from the visual analysis of body movement. Obtaining an explanation of how this can be achieved presents a challenge to researchers in both visual cognition and computer vision. There has been substantial progress in related problems associated with the processing of emotion from faces, but there is much less known about understanding affect from body movement. In this research we investigate techniques for combining tools of human psychophysics and pattern recognition to find features that can account for human movement classification. In particular, we examine the human classification task of labelling arm movements as happy or angry. This is an important area of investigation in the recognition of affect from human movement since the stimulus features underlying this classification are poorly understood [1].

The recognition of affect from body movement can be thought of as a specific instance of movement style recognition [2]. Recently Davis & Gao [3] proposed an expressive three-mode principal components model to recognize the styles of human actions. They used principal component analysis (PCA) to reduce data dimensionality and trained a set of expressive weights to recognize various styles (e.g. carrying load, gender, walking pace) of human walking. Although they did not work on the classification of human emotions, related work in computer animation has addressed this issue. For example, a Fourier-based approach with basic and additional factors (walk; brisk) has been employed [4] to generate human motion with different emotional properties (e.g. happy walk). Amaya et al. [5] applied digital signal processing techniques to generate an emotional motion from a neutral motion. Bruderlin and Williams [6] used multi-target interpolation with dynamic time-warping to blend between motions. But these methods were restricted to transforming a neutral movement to an emotional one.

Pollick et al. [1] looked at the human classification of emotion from motion using the affective human arm movement data of Amaya et al [5]. By using multi-dimensional scaling they found a two-dimensional psychological space of the human representation of affect from motion. They found that the first dimension of this space could be thought of as activation and correlated highly with movement kinematics. Results suggested that the second dimension corresponded to valence (eg happy versus angry) but a stimulus property of the movements that reliably co-varied with dimension 2 could not be found.

In this paper, we present a computational solution to finding the features that drives the human perception of emotions and apply it to this problem of categorising valence from arm movements. Our approach is to identify a diagnostic principal component (PC) through comparison of PCA with human psychophysics.

The rest of the paper is organized as follows. In section 2, we report the results of human psychophysical experiments in categorizing movements as angry or happy. In section 3, we present a computational analysis of the movements by using PCA to identify expressive features. In addition we use a Parzen window to examine how the categorization of movements reconstructed using the expressive component correlate with human classification. In section 4 we present a conclusion.

2. HUMAN MOVEMENT PSYCHOPHYSICS

2.1 Introduction

Studies by Dittrich et al [7] and Pollick et al [1] have shown that from highly impoverished stimuli such as point-light displays [8] people can recognize emotion from movement. As discussed above Pollick et al [1] proposed that the human representation of emotion from motion is along the two dimensions of arousal and valence [2]. While the arousal dimension is fairly well understood as being depicted in a formless motion cue associated with the speed at which a movement is performed [1, 9], the stimulus cues associated with the other dimension are less well understood. A pair of emotions that span this second dimension of valence, are anger and happiness, since they have been found to be close together on the first dimension [1], but far apart on the second. Hence these are the emotions that we have used in the current research.

In the following experiment we investigated how accurate humans were at classifying angry and happy movements. In addition to accuracy we were also interested in individual differences among actors in producing affective movement as well as observers in classifying the movements. For example, in the recognition of gender from point-light displays there are examples both of actors which are consistently misclassified [10] as well as observers who could not perform the task [11]. This concern with individual differences is based upon extensive pilot studies in trying to find features to match human judgements of affect.

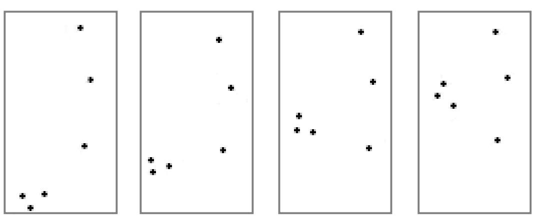


Figure 1. Four frames taken from a lifting movement

2.2 Methods

Participants

Observers were sixteen paid volunteers recruited from the undergraduate population of the university they were naïve as to the purpose of the experiment.

Stimuli

Stimuli were point-light animations (figure 1) of lifting actions, which were recorded as actors depicted angry and happy emotions. The movement database contained movements from 25 actors who were recruited to perform the actions with emotion while their movements were recorded with an Optotrak. With the Optotrak we recorded the 3D position of six infrared light emitting diodes attached to the head, shoulder, elbow, wrist and hand joints. Actors were given a written script that set the situation for each emotion [5] and then performed 10 examples of each emotion, yielding 20 movements per actor.

Movements were processed to find start and end-points and to replace small amounts of missing data. Missing data was replaced by B-Spline interpolation.

For each actor only the most typical angry and happy lifting action was selected for further research (each actor performs 10 repetitions of lifting action). The most typical movement exemplars were defined automatically on an actor-by-actor basis for each emotion independently. In the automatic procedure we first used Functional Data Analysis to construct an average curve from which each exemplar was subtracted. These difference signals were then subjected to Principal Components Analysis (PCA) and the movement with the smallest PCA scores was selected as the most typical. The original position data for the most typical movements was then used to render animations that depicted the six recorded joints as black points on a white background. These procedures yielded 25 angry and 25 happy movements from 25 actors. Because these actors come from student population in university and they are not professional actors, it is expected that not all their performances are consistence with each other. From statistical point of view, actors whose performances are far away from the majority are outliers. We have two ways to remove outliers: one is to detect and remove outliers before further analysis are carried out. The other is to use all data in our dataset but employ robust statistical analysis that is not sensitive to outliers. However, the distribution for angry and happy motion are unknown and just a few samples available, we decided to remove outliers based on psychological experiments.

Design and Procedure

Movements were shown to observers with two different tasks. In first task, they were asked to correctly identify the emotion depicted in the displays in a two-alternative forced choice (2AFC) task. Pilot work had, however, suggested that while observers were relatively accurate at recognising angry movements, they sometimes used the “happy” category to mean “not angry”. In order to take account of this, we introduced a second task in which observers rated both angry and happy movements on a six-point angry and happy scale, with 0 depicting the absence of the emotion and 5 depicting the definite presence of the emotion. In the experiments, the order of the two tasks and the emotions for the rating tasks was counterbalanced.

For the 2AFC task all angry and happy movements were shown in a random order. After each presentation the observer had to make a response that indicated whether they thought the depicted emotion was either angry or happy. For the rating task observers saw all movements presented in a random order with the task of rating one emotion and then again with the task of rating the other. There were 2 presentations of each movement for each of the tasks. Hence observers saw each movements 6 times under the 2 different tasks.

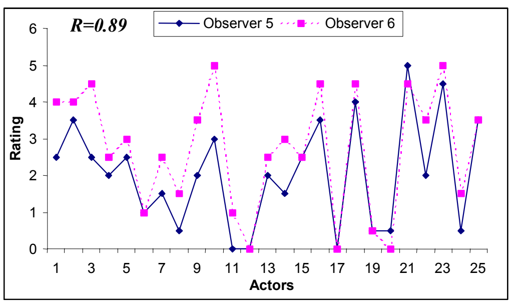
2.3 Results

From the psychophysics, we obtained discrimination results from the 2AFC tasks and rating results from the other tasks. Overall, the average percentage correct of classification for the movements from 25 actors was 68.1%. The details of the rating data and percent correct averaged over all observers and actors are summarised in table 1

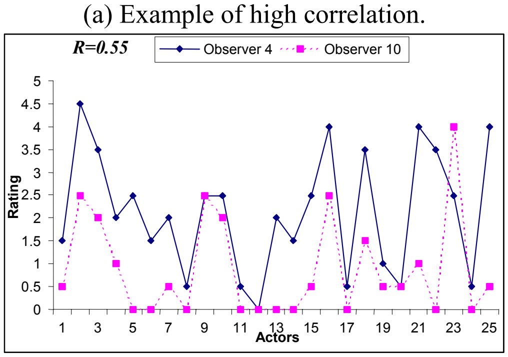
		Movement	
		Angry	Happy
Tasks	% Correct	64.87	71.25
	Rating as Angry"	2.46	1.318
	Rating as "Happy"	1.58	2.36

Table 1. Data form human psychophysics

Three atypical human observers were removed based on their rating results compared with that of other observers hence the results from 13 human observers are used in the subsequent analysis. To remove the atypical observers, the standard correlation coefficient R of pair (x, y) was computed, where x and y were the rating data for 25 actors by two individual observers. Note that the minimum absolute value of R is 0.0 and maximum is 1.0. Figure 2 gives the examples of high and low correlations. We only removed the observers whose correlation values were all below 0.6.



(a) Example of high correlation.



(b) Example of low correlation

Figure 2. Correlation test between observers

By using the remaining 13 observers data, we also identified actors whose movements were ambiguous i.e. actors whose happy or angry movement was classified at around chance (40-60%). In this way we ensured that the movements performed by remaining actors were confidently identified as angry or happy movements. In some cases, one actor's happy movement was classified as angry and his/her angry movement was classified as happy. Therefore, we use the perception label (observer's classifications) to label actual movements. After removing ambiguous actors, 14 actors from the original 25 remained. After the removal of atypical observers and ambiguous actors, and switching to perceptual labels the average percentage correct of classification for the movements from 14 remaining actors rose to 90.9%. The average rating for angry movements as "angry" rose to 3.75 out of 5 and the average rating for happy movements as "happy" rose to 2.76 out of 5.

3. COMPUTATIONAL ANALYSIS OF MOVEMENTS

In the previous section, we used human visual psychophysics to obtain a rating data for 14 typical actors' movements. In this section we explore a computational analysis of these movements with the goal of finding the features which drive human perception of angry versus happy intent. Our basic approach is to first perform principal component analysis (PCA) separately on the angry and happy movements and find the principal components which correlate highly with human ratings of anger and happiness respectively. Next, to verify the effectiveness of these select principal components in representing anger and happiness we obtain reconstructed movements using a limited set of principal components and examined whether the resulting movements could be discriminated using a Parzen Window.

3.1 Finding the features by using PCA

Human movements can be described as the change in body pose over time. In order to perform PCA, we arranged pose (arm joints) and time into one dimension and actors in the other dimension (Figure 3).

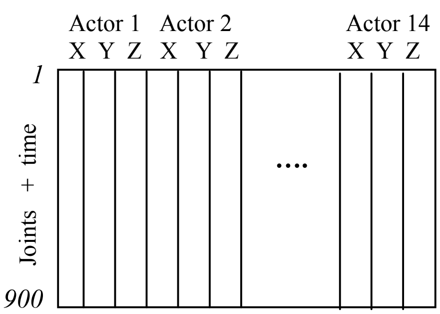
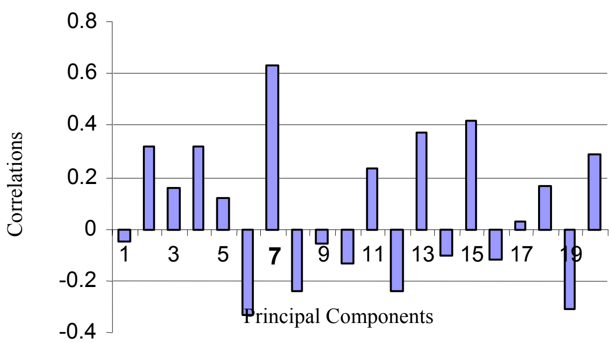


Figure 3. Movement matrix

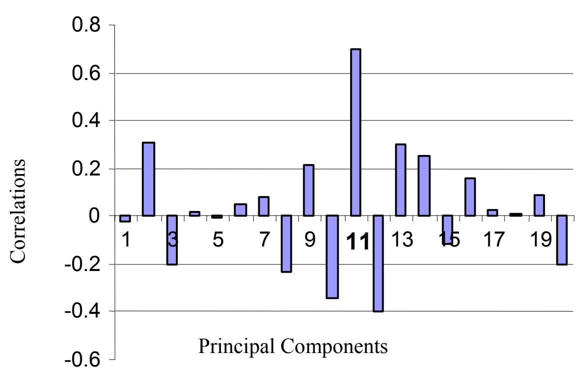
The data for angry movements were put in a matrix $\boldsymbol{A}$ (angry matrix) and the data for happy movements in a matrix $\boldsymbol{H}$ (happy matrix). Before decomposition of the movement matrix, we first centred the motions by mean-subtraction of the trajectories along X, Y and Z coordinates. Then we computed the covariance matrix $\boldsymbol{B}$ from the zero-mean matrix $\boldsymbol{W}$. Finally, we performed Singular Value Decomposition (SVD) on the covariance matrix to obtain eigen values and vectors. Recall that the SVD of the covariance matrix $\boldsymbol{B}$ represents it as a product of three matrices, $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$, where $\boldsymbol{U}$ is the orthogonal matrix of the eigenvectors, $\boldsymbol{S}$ is the diagonal matrix whose diagonal elements are the eigenvalues of $\boldsymbol{B}$ and $\boldsymbol{V}$ is an orthonormal matrix whose columns are the right singular vectors. The PCs (feature vectors) can be obtained by $\boldsymbol{F} = \boldsymbol{U}^T\boldsymbol{W}$, where $\boldsymbol{F}$ is PCs and $\boldsymbol{W}$ is the zero-mean matrix.

PCA produced 900 principal components. Because during SVD, we sort out eigenvalues by value from high to low, only the first few PCs count for the major variance of original dataset. In our case, the first 20 PCs count for 99.2% variance and we examined how well these 20 PCs related to the human rating data. To do this we took each principal component and found the projection of the entire set of angry or happy movements onto this component. Next, we took these projections and using the standard correlation coefficient we calculated their correlation to the human rating data. In Figure 4 we show these results, where large positive correlations indicating that the more this component was present the higher the human rating of emotional intensity.

From figure 4, we can see that PC7 has the highest expressive value for the angry movements and PC11 has the highest expressive value for the happy movements. Thus, for *the current movement set*, we can consider PC7 and PC11 to be essential features of the movement in conveying affect.



(a) The correlation between PCs of the angry matrix and human angry rating data



(b) The correlation between PCs of the happy matrix and human happy rating data

Figure 4

3.2 Discrimination of Reconstructed Movements

In order to examine the expressiveness of the selected principal components we examined the ability to discriminate between reconstructed angry and happy movements. In the following we first present our techniques for reconstructing movements and then follow with the techniques used for discriminating between the reconstructed angry and happy movements.

There are a number of possible principles we could use in selecting components for movement reconstruction via the Karhunen-loeve transform. Foremost, we wish to examine the expressiveness of PC 7 for the angry movement and PC 11 for the happy movement, however since these two components do not by themselves account for a substantial part of the variance of the movements we considered it best to consider these expressive components in conjunction with other components that are important in the reconstruction of the movement. Thus we choose the first two principal components from each matrix. This resulted in the four different sets of movement: “two-PC” angry movements from angry matrix ***A*** using only first two PCs, “two-PC” happy movements from happy matrix ***H*** by using only first two PCs. “expressive” angry movements from angry matrix ***A*** by using the first two PCs along with PC7, “expressive” happy movements from happy matrix ***H*** by using the first two PCs and PC11. Figure 5 provides a comparison of three kinds of reconstructed movements. The three curves are shown in figure 5 are the trajectories of the wrist joint along x coordinate. In the following we discuss how a Parzen window was used to discriminate between the reconstructed “two-PC” angry and happy movements as well as the reconstructed “expressive” angry and happy movements.

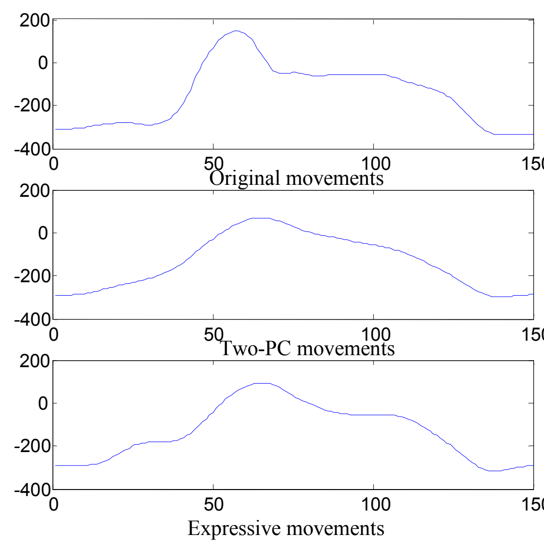


Figure 5. Comparison of the types of reconstructed movements

In order to discriminate between the reconstructed angry and happy movements we used a nonparametric density estimation approach based on Parzen window [12][13][14] as follows:

$$\hat{p}(\boldsymbol{x}|angry) = \frac{1}{N_2 h^d} \sum_{n=1}^{N_2} \kappa\left(\frac{\boldsymbol{x} - \boldsymbol{x}_n^{angry}}{h}\right) \quad (1)$$

$$\hat{p}(\boldsymbol{x}|happy) = \frac{1}{N_1 h^d} \sum_{n=1}^{N_1} \kappa\left(\frac{\boldsymbol{x} - \boldsymbol{x}_n^{happy}}{h}\right) \quad (2)$$

where $\hat{p}(\boldsymbol{x}|angry)$, $\hat{p}(\boldsymbol{x}|happy)$ are estimates of likelihood of measurement $\mathbf{x} \in R^d$ given angry or happy movements, respectively; $\kappa(\cdot)$ is the weighted (kernel) function with such a constrains that it should be probability density function (PDF) itself, e.g., Gaussian density; d is the dimensionality (in our experiment, $d=900$); N_1 (N_2) is the number of samples in angry (happy) movement datasets, respectively (in our experiment $N_1 = N_2 = 42$); h is the smoothing parameter; $\boldsymbol{x}_n^{angry} \in R^d$ is the n-th column of matrix $\boldsymbol{A}$; and $\boldsymbol{x}_n^{happy} \in R^d$ is the n-th column of matrix $\boldsymbol{H}$. In order to optimise the performance of Parzen window, we have to optimise the h. When the number of samples is few, the common way to find the best h is *jackknifing*. It holds back one known sample from the training dataset and then tests the performance of the Parzen window. Then we put this sample back into training dataset and remove a different sample. By doing this for every case, we are able to evaluate the efficacy of a value h. Thus, to optimise h, we simply select many values (from 0.1 to 20.0) and choose one that performs best on the jackknife test. The best value for h we obtain in this experiment is 0.99.

There are 2 main reasons why we employed Parzen window estimates: 1) it is difficult to assume *a priory* a particular functional form of the PDF of underlying distributions related to sensor measurements obtained from actors; 2) one of the main advantage of Parzen window is a little or no training time is required.

But unlike parametric or semi-parametric approaches to probability density estimation, Parzen density estimates employ the entire training datasets in defining density estimates for new observations. Because it is expensive and time consuming to perform psychology experiments there are only a few training sample available and storage requirements are reasonable.

Using maximum-likelihood approach we discriminate between angry and happy movements as following

$$\boldsymbol{C}_{\boldsymbol{ML}} = \underset{\boldsymbol{C}_{\boldsymbol{k}} \in \{\boldsymbol{angry}, \boldsymbol{happy}\}}{\arg\max} \hat{\boldsymbol{p}}(\boldsymbol{x}|\boldsymbol{C}_{\boldsymbol{k}}) \quad (3)$$

where C_{ML} is the result of classification by an artificial ideal observer based on the $\hat{p}(\boldsymbol{x}|angry)$ and $\hat{p}(\boldsymbol{x}|happy)$ estimates. The average of correct classifications for original angry and happy movement datasets (using matrices $\boldsymbol{H}$, $\boldsymbol{A}$ and Eq. (1), (2), (3)) equals to 81%. But the probability of correct classification was improved to 88% if we used the "two-PC" movements, and to 96% if we used the "expressive" movements (first two PCs and feature PC).

However, we are not only interested in finding features that give us the best results of classifier performance, but also we want to compare the performance of human observers and the performance of artificial observers (Parzen window). That is why, it is necessary compare the correlation between the output of Parzen window $\hat{p}(\boldsymbol{x}|angry)$, $\hat{p}(\boldsymbol{x}|happy)$ and human rating data of the angry and happy movements. Table 2 give the results of using the standard correlation coefficient to find the correlation between rating data from human observers and output probabilities from the artificial observer for the reconstructed 'expressive' movements. A high positive correlation indicates that when the artificial observer found a movement to be highly probable in expressing a certain emotion, that human observers rated this movement as highly expressive in conveying the emotion.

Table 2: The correlation results between outputs of Parzen window and human rating data.

		Human observer	
		Angry rating	Happy rating
Parzen window	Angry	0.807	-0.651
	Happy	-0.617	0.626

4. CONCLUSIONS

The goal of this research was to reveal features of human movement that could be diagnostic for the recognition of affect. To achieve this goal we began with a motion capture database of arm movements and performed psychophysical experiments to provide us with a subset of these original movements that could be reliably labeled by human observers. Analysis of this subset of movements using principal component analysis revealed components that correlated well with human judgements and could be considered as expressive components. Further analysis of the classification of reconstructed movements containing these expressive components were consistent with our interpretation that we had uncovered features that are diagnostic for human recognition of affective movement.

Although the current findings are limited in that they apply to only the recognition of angry and happy lifting movements of the arm, the methods developed can be applied to other actions and affects. The significance of this result being that it provides a mean for further study to obtain robust definitions of the information within movements that indicate the expression of affect. Given the complexity of the human movement signal and the current limited understanding of the precise features that lead to the perception of affect; the integration of psychological data within the pattern recognition approach provides a potentially powerful method to reveal the biological basis of affection recognition.

In the future we plan to apply this method to other actions using 3D motion capture recordings of the entire body. By examination of other actions and body segments we hope it will be possible to generalize the features used for affect recognition. In addition to this we hope to carefully investigate the visual recognition of the reconstructed movements in comparison to the original movements. Finally, it is also of interest for engineering applications like motion compression to utilize our results to achieve high rate compression for human motions without losing the expressive quality of the data.

ACKNOWLEDGMENTS

This work is supported by Engineering and Physical Science Research Council Grant GR/M30326, a Royal Society UK-Korea Joint Project Grant and a British Academy postdoctoral fellowship to the second author. The support of the Engineering and Physical Science Research Council Grant GR/R61215 is gratefully acknowledged. Alexander N. Dolia is supported by supported by Data and Information Fusion, Defence Technology Center Grant, UK.

REFERENCES

[1] F. E. Pollick, H. M. Paterson, A. Bruderlin, and A. J. Sanford, "Perceiving affect from arm movement," *Cognition*, vol. 82, pp. B51-B61, 2001.

[2] F. E. Pollick, H. Hill, A. Calder, and H. Paterson, "Recognizing Facial Expressions from Spatially and Temporally Modified Movements," *Perception*, vol. 82, pp. 813-826, 2003.

[3] J. W. Davis and H. Gao, "An expressive three-mode principal components model of human action style," *Image and Vision Computing*, vol. 21, pp. 1001-1016, 2003.

[4] M. Unuma, K. Anjyo, and R. Takeuchi, "Fourier principals for emotion-based human figure animation," presented at International Conference on Computer Graphics and Interactive Techniques, 1995.

[5] K. Amaya, A. Bruderlin, and T. Calvert, "Emotion from motion," presented at Proceedings Graphics Interface '95., Toronto, Ont, 1996.

[6] A. Bruderlin and L. Williams, "Motion signal processing," presented at Computer Graphics Proceedings. SIGGRAPH 95., Los Angeles, CA, USA, 1995.

[7] W. H. Dittrich, T. Troscianko, S. Lea, and D. Morgan, "Perception of emotion from dynamic point-light displays represented in dance," *Perception*, vol. 25, pp. 727-738, 1996.

[8] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception and Psychophysics*, vol. 14(2), pp. 201-211, 1973.

[9] H. M. Paterson, F. E. Pollick, and A. J. Sanford, "The role of velocity in affect discrimination," presented at Twenty-Third Annual Conference of the Cognitive Science Society, Edinburgh, 2001.

[10]L. T. Kozlowski and J. E. Cutting, "Recognizing the sex of a walker from a dynamic point-light display," *Perception and Psychophysics*, vol. 21, pp. 575-580, 1977.

[11] W. A. Sparrow and C. Sherman, "Visual expertise in the perception of action," *Exercise and Sport Science Reviews*, vol. 29, pp. 124-128, 2001.

[12] E. Parzen, “On estimation of a probability density function and mode”, Annals Mathematical Statistics, 33, 1065-107, 1962.

[13] C. M. Bishop, “Neural Networks for Pattern Recognition”, Oxford University Press, Inc., New York, NY, 1995.

[14] K. Fukunaga, “Introduction to statistical Pattern Recognition”, 2nd Edition, Academic Press, New York, 1990.