Emotion Analysis from Dance Performance Using Time-delay Neural Networks

Woontack Woo†, Jong-Il Park‡ and Yuichi Iwadate†

† ATR MIC Labs
Advanced Telecommunications Research
Kyoto 619-0288, Japan
E-mail:{wwoo,yiwadate}@mic.atr.co.jp

‡ Electrical and Computer Engineering
Hanyang University
Seoul 133-791, Korea
jipark@email.hanyang.ac.kr

Abstract

In this paper, we propose a novel neural networks-based scheme to extract emotional information from dynamic gestures of dance performance. The extracted information is used as a part of ATR-MIDAS, MIC Interactive Dance System, to control and composite audio-visual data such as sound, music, image, graphics and video. Though over last few years many research works on gesture analysis have been reported to help human-machine interface, conventional approaches have limitations in applying to dancing sequence analysis, because dancing is dynamic, rather than static, and human emotion is too abstract to analyze. Such well-known limitations of gesture analysis can be alleviated by adopting neural networks to learn the dancer's emotional intention through learning. To address these problems, we propose a time delay multilayer perceptron-based emotion analysis scheme. Our experimental results demonstrate that consistent emotional analysis can be achieved using the proposed scheme, which maps between local features and symbolic representation of emotion in realtime.

1 Introduction

In general, nonverbal, as well as verbal, expression plays a key role in communications between humans. Especially, various gestures are one of important means to deliver human emotion. Therefore, over last few years many research has been focused on gesture recognition to support more efficient human-machine interface. However, such efforts has limitations because many realtime applications, *e.g. emotion analysis from dance performance*, which require "dynamic", rather than "static", gesture analysis. Note that dance performance is a representative target of emotion analysis because a dancer tries to convey various emotions through his/her dynamic but distinctive body movements [1–3].

Therefore, we propose a robust way to extract intentional emotion of dancer from dancing sequences. The aim of this research is to examine the physical, emotional and cognitive aspects of body movement, *i.e.* how the human emotion and physical laws interact with each other. The relationship might be a mean to create tools for the interface between human and machine. To show the effectiveness of our approach, we build a ATR MIDAS (ATR MIC interactive dance system), which allows to interactive composition of audio-visual data corresponding to the analyzed intentions of the dancer.

As shown in Figure 1, the ATR-MIDAS is a non-contact vision-based system. First, we ask a dancer to express his/her emotion, what is called 7 motives, through dance performance. The performance is captured using multi cameras, *e.g.* front and top cameras are used to trace personal and general spaces, according to Laban's theory [4]. Then, corresponding features are extracted from the sequences. The extracted features are used as input of a time delay multilayer perceptron (TD-MLP). Thus, The proposed TD-MLP maps the extracted features onto emotional representation, i.e. 7 motives. As a result, TD-MLP tries to mimic dancer's intentions through backpropagation learning algorithm. Finally, given the trained TD-MLP, anyone can composites audio-visual data, echoing the dancer's emotion.

In this scenario, the main advantage of the proposed approach is that it is easy to map nonlinear relationships without evaluating features. According to our experimental results, the proposed novel neural networks-based scheme, TD-MLP, achieves consis-

Figure 1: Basic structure of ATR MIDAS.

tent interpretation of dancer's intentions through the learning of body movement or dynamic gestures.

This paper organized as follows. In Section 2, we briefly explain about features corresponding to Laban's theory. In Section 3, the proposed scheme based on TD-MLP is described. In Section 4, we provide some experimental results to compare the effectiveness of the proposed scheme. Brief discussion and future works are also given in Section 4.

2 Features for Emotion Analysis

The language of movement is subjective and objective: every movement we make contains psychological meaning and physical laws. R. Laban, a choreographer and movement theorist as well as a dancer, believed that movement of the body and the mind is the basis of all human activity. Laban's broad vision revealed the general laws of human movement as they occur in work and at play, in expression and relationships, and in everyday life. Laban Movement Analysis (LMA) is a system of observing, analyzing, and classifying movement. As a result, the possible applications of LMA has been used by a wide and diverse range of groups and individuals including researchers, dancers, athletes, actors, therapists and educators.

The LMA encompasses four main categories: Body, Effort (dynamics), Shape, and Space [4]. The Body aspect deals with principles such as the initiation and sequencing of movement from different parts of the body, and the connection of body parts to each other. The Effort dimension is concerned with movement qualities and dynamics, and is subdivided into weight (energy), space, time and flow factors. Shape is about the way the body interacts with its environment. There are three types of shape change: shape-flow (growing and shrinking, folding and unfolding, etc.), directional (spokelike or arclike) or shaping (molding, carving and adapting). Space involves the study of moving in connection with the environment and is based on spatial patterns, pathways and lines of spatial tension. The theory and practice of Space Harmony acts as a framework for Space, Effort and Shape in the form of established scales of movement within geometric forms.

In this paper, to analyze dancer's intentions in an efficient way, we focus on the Effort aspect and then extract some features from dancing sequences based on Laban's Theory [4]. As shown in Figure 2, we formalized dynamic nature by the relationships among space, time and energy, which can be described by openness, velocity and acceleration of the movement. The principles of space, time and energy are no less important for the movement than color, line and form for a painter.

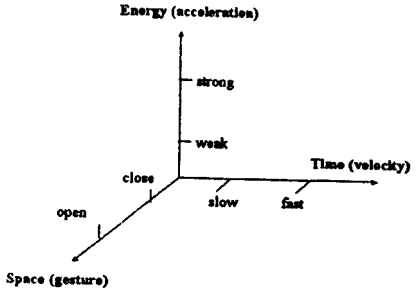

Figure 2: Emotion Space. The emotion space consists of space, time and energy, which can be described by openness, velocity and acceleration of the movement.

To capture body movements, we need to use multi cameras, *e.g.* one is on top of the studio and the other is in front of the studio. In general, while top camera is good for tracking the features of general space, front camera is useful for extracting features of personal space. However, in this paper, as the first stage to this research, we use a camera in front of the stage. Using the front camera, we trace the width and height of the box which warp the moving object, *i.e.* human body, as well as its center. Then, we calculate the ratio of object within the traced box, which is used as space information. In the same time, the change of centroid (*i.e.* velocity and acceleration) are used as time and energy information, respectively. The extracted features are used to trace dancer's intended emotion, which is represented by his/her body movement or rhythm information.

3 Analysis Using Neural Networks

It has long been postulated that neural networks might provide the most sound basis for approximating any (linear or nonlinear) function with a finite number of discontinuities, [5]. Especially, MLP might provide the most valid way to map any nonlinear relation, given sufficient neurons in the hidden layers. After proper training on a representative set of input and output vectors, MLP tends to lead a new input vector (that the MLP has never seen) to a similar output

(to the correct output for the close input vector used in training). However, the MLP may have limitations in applying for applications requiring time-dependent nonlinear mapping such as emotion analysis.

Therefore, we extend the MLP to TD-MLP to learn emotion from dancing sequences. As shown in Figure 3, the proposed TD-MLP consists of three layers. In this framework, a set of features extracted based on Laban's theory and its time delay set are used as input vectors. Time-delayed features are adopted for TD-MLP to learn dynamic movement using time-dependent local features. A set of symbolic representation of emotion is used as output vectors to train the TD-MLP. Finally, TD-MLP approximates the nonlinear mapping, *i.e.* associate the time delay input feature vectors with specific emotional category.

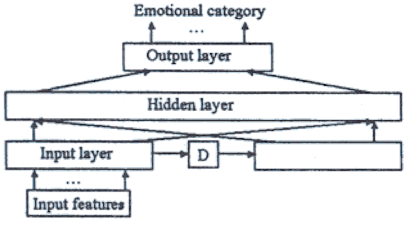

Figure 3: Time-delayed Multilayer Perceptron. It consists of 3 layers including input, hidden and output layers. Time-delayed features are used as an input to generalize a nonlinear mapping. A set of input and output vectors and backpropagation algorithm is used to train the TD-MLP.

The most common learning rule for training the networks is the error backpropagation rule, where the term backpropagation refers to the manner in which the gradient is computed for nonlinear multilayer networks[1]. There are many variations of the backpropagation algorithm. The simplest implementation of backpropagation learning updates the network weights and biases in the direction in which the performance function decreases most rapidly the negative of the gradient. There are two ways in which the gradient descent algorithm can be implemented: incremental mode and batch mode. For faster convergence, the gradient descent algorithms are equipped with momentum. In addition to these, there are other high performance algorithms which fall into two main categories. The first category uses the heuristic techniques such as variable learning rate backpropagation and resilient backpropagation. The second category uses standard numerical optimization, *e.g.* conjugate gradient method, quasi-Newton method and Levenberg-Marquardt algorithm. In this paper, among many variations of the backpropagation algorithm, we adopt an adaptive learning algorithm, which is capable of speeding up the learning process in batch mode training.

[1]The backpropagation algorithm was created by generalizing nonlinear differentiable transfer functions and the Widrow-Hoff learning rule to multilayer networks such as MLP [5].

4 Experimental Results

We take dancing sequences in the studio with multiple cameras. Note however that in the experiments, we only use a camera in front of the studio. In learning phase, we prepare 21 segments, where each motive contains 3 segments (walk, turn, jump). As a result, 21 segments of training sequence are used. After training, we used 7 segments, corresponding to 7 motives (*e.g.* natural, happy, fluent, lonely, lively, sharp and solemn), in which each segment contains several motions such as walk, turn and jump.

The first step is to extract the input features from the dancing sequences. The features are extracted from the sequence of one second, *e.g.* in case of 10 fps case, 10 frames are used to calculate features. Then, the features are remained in the input layer for a few second. Figure 4 shows the emotional space with features extracted from the dancing sequences based on the Laban's theory. Each space corresponds to 7 motives and the last plot shows the emotion space containing whole data of training sequences. As can be seen in Figure 4, the problem to solve using the TD-MLP is nonlinear separation of the emotional space.

The second step is emotional mapping between 7 input features and target 7 motives using the proposed TD-MLP. The TD-MLP consists of 3 layers, *i.e.* input, hidden and output layers. Input layer has 2 time delay components and thus each layer has 21 (*i.e.* 3x7), 147 (*i.e.* 3x7x7) and 7 nodes, respectively. As shown in Figure 4, the learning of emotional mapping requires about 1000 iterations to achieve learning goal. Figure 4 shows how well the TD-MLP learn the training sequences after finishing learning. Note that, as explained, we use no special knowledge on emotional mapping.

The third step is to consider the case of new data to test the generalization capacity of the TD-MLP. After the learning, the 7 segments of dancing sequences, which are not used in learning, are used in test. As shown in Figure 4, the proposed TD-MLP is capable of generalizing the emotional mapping, though we fail to get continuous mapping in some parts of the test

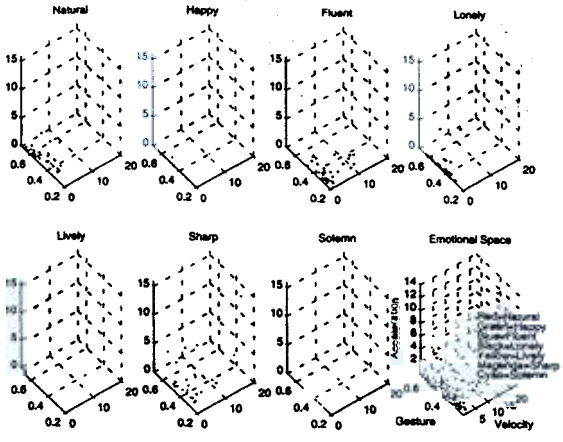

Figure 4: Projection of features onto emotion space. The axis represents space, time and energy, respectively. Each space corresponds to 7 motives: natural, happy, flowing, lonely, dynamic, sharp and solemn. The last plot shows the emotion space containing whole data of training sequences.

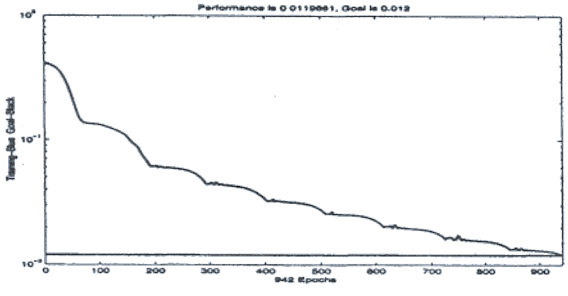

Figure 5: Learning curve of TD-MLP. The learning of emotional mapping requires about 1000 iterations to achieve learning goal.

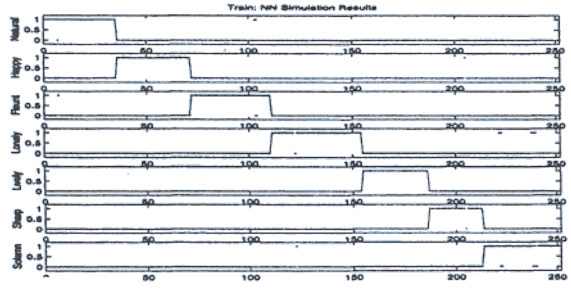

Figure 6: Output of TD-MLP for training sequence.

sequences.

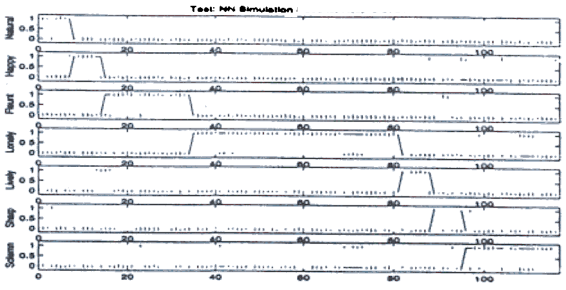

Figure 7: Output of TD-MLP for test sequence.

In this paper, we showed that consistent emotion analysis can be achieved using the proposed TD-MLP, which has availability to map from local features to an abstract symbolic representation of emotion. Note however that, if we only obtain features from only few seconds of segments, there is some limitations in reducing error rates (currently about 20 % of the test sequences). Especially, according to the experimental results, classification between lonely and solemn has failed, even in the training process, because both have similar movements in test as well as training sequences. Therefore, to get a more consistent emotional mapping, we will extend the proposed framework to incorporate a more stronger time-domain constraint into the proposed TD-MLP. The results of the proposed scheme also can be applied into emotion analysis of a group dancing without the loss of generality.

References

[1] J. Abouaf, "Biped: A dance with virtual and company dancers," *IEEE Multimedia*, vol. 6, no. 3, pp. 4–7, 1999.

[2] A. Camurri, M. Ricchetti, and R. Trocca, "Eyesweb-toward gesture and affect recognition in dance/music interactive system," in *Proc. IEEE Multimedia Systems*, June 1999, pp. –.

[3] M. Csaky, R. Suzuki, J. Park, and M. Inoue, *Video Rhythm and Motion Analysis*, ATR Technical Report, TR-M-0043, 1999.

[4] R. Laban, *Modern Educational Dance*, Hyundai Mihak-sa, Korea, 1988.

[5] D.E. Rumelhart and J.L. McClelland, *Parallel distributed processing: Explorations in the microstructure of cognition*, MIT Press, 1999.