MIDAS : MIC Interactive Dance System

Ryotaro Suzuki
ATR Media Integration & Communications Research Laboratories
Kyoto, Japan 619-0288

Yuichi Iwadate
NHK Science Technical Research Laboratories
Tokyo, Japan 157-8510

Masayuki Inoue
Hiroshima Institute of Technology
Hiroshima, Japan 731-5193

Woontack Woo
ATR Media Integration & Communications Research Laboratories
Kyoto, Japan 619-0288

Abstract

We have been studying how to establish a method for extracting human emotion in order to express emotional images by utilizing multimedia such as video and sound. Human body motion is the most basic essence in expressing human emotion. MIDAS (MIC Interactive Dance System) is an application of the emotional information extracting/expressing method in a dance system. In this system, we applied Rudolf Laban's dance theory to extract physical characteristics of dance motion from real-time video sequences and mapped this information to categories of emotional image expression. In this way, we can relate the motion with emotional representation using multimedia.

1 Introduction

In our laboratory, we are researching various possibilities for image expression in communications utilizing multimedia technology. This research has three major factors, that is, emotional information processing, communication, and multimedia. For multimedia, we are pursuing the integration of various visual and audio media such as 2D & 3D CG animation, real video, still image, MIDI, sampling audio, etc. In addition to the issues of how to present each medium and how to relate them, another important issue is how to control these media as a human interface. This is also a communication problem. In particular, by paying attention to the emotional aspect of communication, multimedia technology can be utilized much more effectively in human life. If we can propose a prototype model of such emotional communication using multimedia, then we will answer to above mentioned three issues.

Therefore, we selected a dance system as our target. We chose dance because human motion is the most related to human emotion and dancing is the most basic way to represent emotional images. Furthermore, if we analyze human motion in dancing, we can expect to be able to extract human emotion more directly than by using drawings, musical pieces, etc.

We started our experiment of making a multimedia dance system a few years ago. We started from a very primitive stage where a simple CG animation's motion is related to a human posture. We then moved to a second stage where some simple motions such as standing up and squatting down can be detected.

At this point we felt that we were standing at the crossroads. Today, most computer systems that detect human motion are based on so-called motion capturing technology and they typically detect the three dimensional motion of the body and the limbs as precisely as possible in real time. Such information is very useful for direct manipulation. However, we do not think that direct manipulation should be the only communication method between a human and a machine that makes much of human's feeling. We think that there is some other possibility of an emotional relationship between a human and a machine and it can be integrated with direct manipulation in a suitable way.

As one experimental study, we tried to make a time frequency analysis using MEM (Maximum Entropy Method) against repetitive dance motion that corresponds to a pop music's steady beat. We confirmed that it works well. This rhythmic detection seemed to give us one possibility of detecting emotion. However, it is too difficult to directly analyze frequency pattern for emotion detection, though it can be a part of the information.

Finally, we reviewed a standard dance theory that we had learned before from our cooperative dance researcher, and decided to introduce Rudolf Laban's dance theory as a base for our emotional information extraction method. We developed the first version dance system called "MIDAS"(MIC Interactive Dance System) using this method in 1999.

In this paper, we first show the essence of Laban's theory. Next, we present the dance motion analysis and the motion-emotion mapping mechanism based on this theory. We then introduce MIDAS. Finally, we review the performance of MIDAS and discuss our future work.

2 Overview of Dance Research

Choreographers have been studying human motion in order to understand its physical features and to find an effective way to express emotional images through human motion. A famous Austrian choreographer, Rudolf Laban (1879-1958), proposed three kinds of descriptions for human motion : Motif description, Effort-shape description, and Structural description [1]. The Motif description provides the most salient feature of a motion. The

0-7803-6583-6/00/$10.00 © 2000 IEEE

Effort-Shape description can describe human motion in terms of its quality and expression. The third kind is well known as Labanotation. Labanotation is used to systematically record human motion. It is useful in describing the human motion in the field of technical research [2].

The Effort-Shape description is adequate for categorizing human motion with emotional features. There are four parameters in this description, space, weight, time, and flow. Time has categories of sudden movement and sustained movement. Space stands for a design of posture and is the directivity of posture and movement. Weight stands for strength and power of movement. Weight is categorized into firm movement and fine touch movement. Flow stands for the carefulness versus easiness that can be seen in movement.

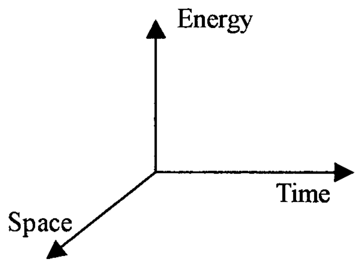

Figure 1. Time-Space-Energy.

In Japan, a study based on the Laban's theory by dance researchers has shown that the image patterns expressed by dance motion can be categorized into seven typical motives called 7 Motives [3][4]. The relationship between the Effort-Shape description and the 7 Motives is given as shown in Table 1. The 7 Motives represent the emotional image expressions of happy, solemn, lonely, natural, sharp, dynamic and flowing.

The Effort-Shape description and the 7 Motives provide us with background knowledge for our research. We expect that they will give us a base for establishing an emotional information extraction method that uses dance motion.

Table 1. Relationship between 7 Motives and Time-Space-Energy.

7 Motives	Time	Space	Energy
Happy	speedy	light	
Flowing	soothing		rounded
Lonely	slowly	sustained	
Natural	regular	sustained	balanced
Solemn	gradual	massive sustained	
Sharp	accented	firm sudden	straight
Dynamic	speedy accented	light	spread

3 Method of Extracting Emotion from Human Body Motion.

Based on Laban's concept of Time-Space-Energy, we extract the features of dance motion as the time, space, and energy parameters of the variance of the dance video image; we then map these parameter values to a certain category of emotion in our dance system. The details of the motion extraction and motion-emotion mapping are shown below.

3.1 Motion Extraction by Time, Space, Energy

The Time-Space-Energy concept introduced by Laban is a very general concept, and anyone can guess what it is. Its origin lies in physics and it is closely related with the world where we live. Laban thought that all the human motion can be characterized by this concept, which would include all kinds of dance in the whole world throughout human history. Looking back to this origin, we decided to concentrate on this concept and redefined Time-Space-Energy features to be as generally and easily extractable from video images as possible.

Time : Human motion speed
Space : Openness of human body
Energy : Acceleration of human motion

In this stage, we omitted Flow features and redefined Space as a wideness or an amount of space instead of design or form features. For example, the difference between a straight line and a curved line in dance motion is very important in the image expression of dancing. Such a feature is closely related to the Flow feature and the Space feature as design. However, we think that it is a minor feature that can be derived from these redefined Time-Space-Energy features. Laban said that the origin of all the motions is the combination of the motion into the inner direction and the outer direction in his theory of forms of movement in space "Choreutics" [5]. This can be recognized as a model of motion for an "open-close" model or a "breathing" model. We expect that our redefinition reflects Laban's original basic idea.

In this sense, what is essentially important as a physical information of human motion feature is the center of gravity of a human image. In the observation of a dance video, the motion of the center of gravity, the wideness of body around the center of gravity, and the position of the center of gravity in the body area are the actual features that correspond to the Time-Space-Energy features.

In our previous frequency analysis of a dancer's motion, we first used the difference of neighboring video frames as the number of changing pixels in silhouette images. The performance of this method was not so steady as we expected. The information had too much noise, even though it is often used in image processing. By introducing the distance of the center of gravity instead of the difference of images, we succeeded in improving the performance to a feasible level. This is because the center of gravity better reflects the feature of human body motion as a whole.

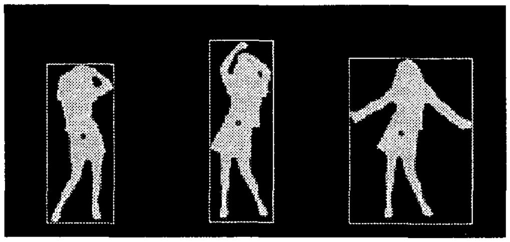

Figure 2. Silhouette Images of Reference Dancer

We use the center of gravity and circumcised rectangles of a dancer' silhouette images as shown in Figure 2. The center of gravity represents a dancer's motion over the dancing space. The circumcised rectangle represents human local motion [6]. The present implementation of the motion extraction is as follows.

R_n, S_n, Un and (X_n, Y_n) define the area of the circumcised rectangle, area of silhouette, aspect ratio of the rectangle, and coordinates of the center of gravity, respectively. n represents the present video frame. The ratio of the silhouette area and rectangle area is defined with $W_n=S_n/R_n$. We like to have human motion parameters so that alterations of every physical parameter will be calculated. For instance, change in the speed of R_n and change in the acceleration of R_n can be calculated using R_n-R_{n-1} and $R_n-2*R_{n-1}+R_{n-2}$, respectively.

3.2 Mapping to Emotional Categories

While Time-Space-Energy are the very general features that are assumed to be applied to any human motion, 7 Motives are the categories that are limitedly used as the categories of forms in original dance motion. Suitable categories of emotion or emotional expression depend on what the target field of the dance system is and what the dance system tries to express. Any category such as 6 basic emotions can be used depending on its purpose.

In the first implementation of our dance system MIDAS, we decided to introduce the 7 Motives into a sort of amusement system that had dancing in a disco club style. Actually, there were some problems with this implementation. For example, a disco dance is different from an original dance, and it is too difficult for novice dancers to distinguish 7 Motives in dancing, etc. However, 7 Motives have a certain generality that can be effectively utilized in the system's visual & audio expression. Moreover, it is possible to redefine the categories and to improve the system after the first implementation.

A professional dancer's scene captured by a camera is used for the emotion analysis. The dancer performs in accordance with the 7 Motives. Simple image processing is used to extract physical features from the dancer's sequences.

There are several ways to map Time-Space-Energy parameter values to 7 Motives. At first we tried to use a neural network. Figure 3 shows a result of 7 Motives mapping made by a neural network that was trained by some dance motion video data based on 7 Motives.

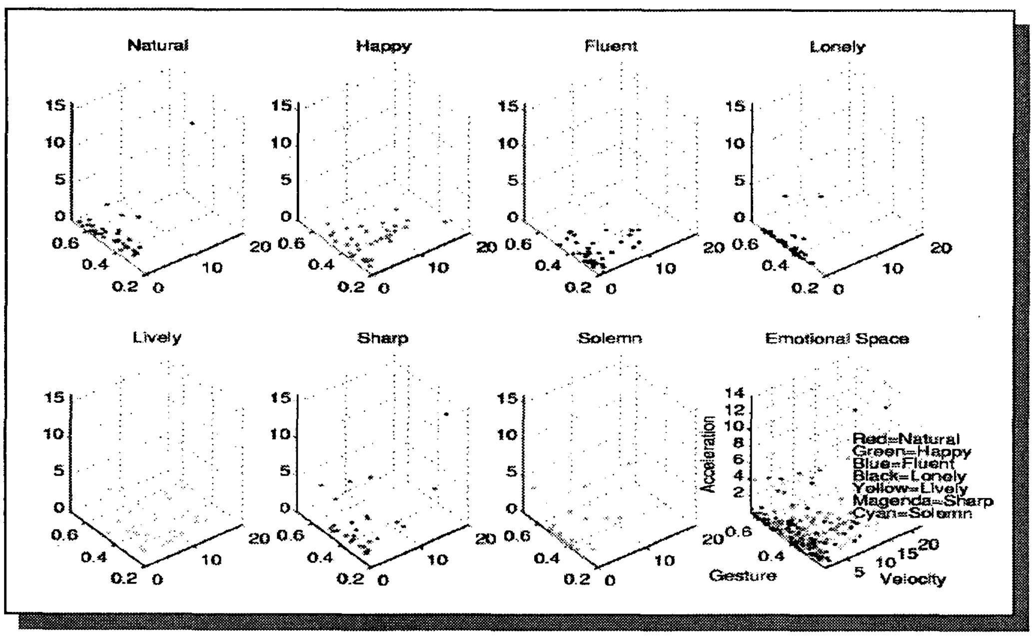

Figure 3. 7 Motives Mapping by Neural Network

Although the distribution patters are non-linear, linear methods are thought to be more suitable for Time-Space-Energy based mapping, compared with a case where a neural network is used for a pattern recognition of more direct body motion forms. Even if it is still difficult to distinguish 7 Motives by linear mapping, easier and steadier controllability is expected. For this reason, we gave up on using the neural network and introduced a multiple regression analysis that is often used for mapping between physical data and emotional data.

When we applied the multiple regression method to the mapping, we also changed our evaluation of emotional values. In our first trial using neural network, 7 Motives values were directly gained from dance video sequences, because we know the dancer's intention of a certain motive expression in each video sequence. This means that a dancer's intention of emotional expression is to be reflected in the visual & audio expression of the dance system. However, many dancers must dance to make the general dance motion data in this case. Another possibility is to introduce training sequence of the dance system for each user. Since neither way is feasible in our experiment, we chose to make a subjective assessment test of 15 subjects using those dance video sequences.

With the subjective assessment test, we can get some principal components, such as P_1, P_2, and P_3. As a result, we obtain coefficients of α_n for each physical parameter such as $f(R)$ and $f(U)$ as follows by applying multiple regression method against these components.

$$\left.\begin{array}{l} P_1=\alpha_1*f(R)+\alpha_2*f(U)+\alpha_3*f(W) \\ \qquad +\alpha_4*f(X)+\alpha_5*f(Y) \\ P_2=\alpha_6*W_n+\alpha_7*U_n \\ P_3=\alpha_8*g(R)+\alpha_9*g(U)+\alpha_{10}*g(W) \\ \qquad +\alpha_{11}*g(X)+\alpha_{12}*g(Y) \end{array}\right\} \quad (1)$$

These equations show the relationship between emotional values and physical values. Here f( ) and g( ) are defined with $f(x)=x_n-x_{n-1}$, and $g(x)=x_n-2*x_{n-1}+x_{n-2}$.

Since emotional values P_1, P_2, and P_3 are abstract values in a virtual emotional space, we have to relate them to 7 Motives. Anther psychological experiment is held to categorize the dance video sequences with the same 15 subjects. In this experiment, the subjects vote for one Motive for each sequence. We assume that 7 Motives can be represented with linear combinations of P_1, P_2, and P_3 as follows.

$$V_k=\beta_{k1}*P_1+\beta_{k2}*P_2+\beta_{k3}*P_3 \qquad (2)$$
$$k=\{1,2,3,4,5,6,7\}.$$

V_k represents each 7 Motive value. The multiple regression method is used to find adequate coefficients of β_{nk}. The Motive decision is made by selecting the maximum value of V_k. With these functions, we can determine one motive among 7 Motives. The highest value is chosen and gives one motive.

4 MIDAS

4.1 The Outline of MIDAS

We developed the first implementation of MIDAS based on the Time-Space-Energy concept in 1999. The target of MIDAS is an amusement system utilizing dance motion and multimedia. The above mentioned emotional information extraction mechanism is implemented into MIDAS to detect 7 Motives from real time dance sequence captured by a video camera. The extracted emotional information is sent to a multimedia controller. The multimedia controller manages the video switcher, real-time disk system, and sound system. The multimedia controller interprets the received emotional information and selects adequate video and sound clips corresponding to 7 Motives. Video clips are displayed on a 120-inch projection monitor. Sound clips are replayed simultaneously. Figure 4 shows a schematic of MIDAS.

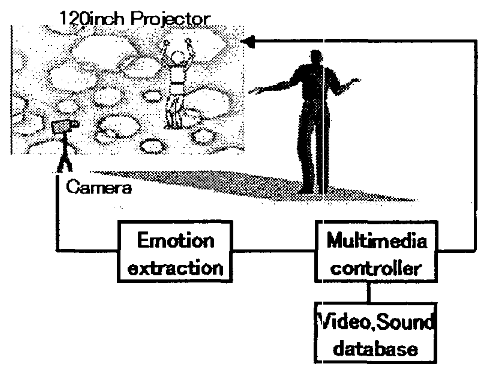

Figure 4. Schematic of MIDAS

Table 2. Video and Sound Clips for 7 Motives.

7 Motives	Visual Expression	Rhythm
Happy	2D-CG animation	Reggae
Flowing	Colored real picture with video effect	Bossa Nova
Lonely	Monochrome real picture	Fusion
Natural	Colored real picture	Jazz Funk
Solemn	3D-CG animation	Ethnic
Sharp	Motion graphics	Techno
Dynamic	Combination	House

Table 2 shows the relationship between 7 Motives and video & sound clips. Each motive has its own rhythm and its own expression style, which makes the dancer recognize the present motive and control the system easily. The clips are designed by graphic artists and a sound creator along

with the images of 7 Motives [7]. For instance, when the system recognizes a "Happy" motive, a funny 2D-CG animation video clip is displayed with reggae music.
A performer's image is synthesized on video clips. This presentation also differs in relation to 7 Motives. For example, a colored silhouette is used for the "Solemn" motive, while monochrome image is used for the "Lonely" motive. In addition to 7 Motives information, values of Time-Space-Energy parameters are used to add some real time effect to the play of video & sound in each motive. In this way, more direct information is integrated with the category information to improve the controllability of the system. Figure 5 shows an appearance of MIDAS.
The dance motion frequency analysis mechanism that we previously developed is also implemented in MIDAS. We do not positively utilize this function in the first implementation, but the mechanism has a wide potential for improving the system in the future. Its typical example is to synchronize the system's beat (BPM) to the dancer's beat extracted by this frequency analysis.

Figure 5. Appearance of MIDAS

4.2 The Performance of MIDAS

It is not very easy to control the motives of MIDAS according to a dancer's will. It becomes easier, if the dancer trains a lot with MIDAS,. If he/she is a good dancer, it takes less than one hour to become able to control the system. If he/she is an awful dancer, he/she will never be able to control the system. This kind of naive balance of controllability makes MIDAS an attractive amusement game, and this aspect seems to be closely related with what the emotion is.
As a performance of emotional information extraction, we still have much to do. Since MIDAS is too sensitive to noise, its reaction to a dancer's motion is not steady enough. Furthermore, we do not have an output that proves that the emotional information extraction ability of MIDAS exceeds a certain required value, even though many people have enjoyed dancing with MIDAS.

5 Conclusion and Future work

We developed an emotional information extraction method of dance motion based on Laban's theory. We successfully applied this method to a dance amusement system called MIDAS. Using MIDAS, a dancer can communicate with the system according to his/her emotion, and thus the dancer can express his/her emotional images.
We are continually improving of the extraction ability, and we are also trying to apply this method to many kinds of dance applications such as a networked communication system, an exercise system, an education system, and a VJ(Video Jockey) system. We developed a second version of MIDAS in 2000 based on the same extraction method (Figure 6). In this version, a dancer can dance with CG characters that change according to emotional categories. The categories are summarized into five based on the 7 Motives in order to realize more steady controllability. We are convinced that this method using Time-Space-Energy features is a very general method of extracting emotions by human motion.

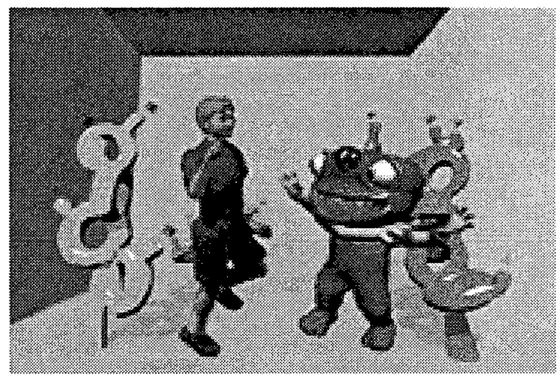

Figure 6. MIDAS for Amusement Application

Acknowledgement

The authors wish to thank Professor Shiba of Kobe university for her suggestions on Laban's theory and 7 Motives.

References

[1] A. Hutchinson, "Labanotation", Dance Books, 1996.
[2] T. Nakata, T. Sato and T. Mori, "Expression of Emotion and Intention by Robot Body Movement", IAS-5, IOS Press, pp. 352-359, 1998.
[3] C. Matsumoto, "Dance Research : Problem Situation and Learning of Problem Solving II – Qualities of Movements and Feeling Values", Proc. Japanese association of physical education and sports for girls and

women, pp. 53-89, 1987 (in Japanese).
[4] M. Shiba, "Extraction of Kansei Information in Dance Movements", Proc. ATR Workshop on Virtual Communication Environments, pp. 70-85, 1998.
[5] T. Ishifuku, "The History of Dance", Kinokuniya-shoten, 1974 (in Japanese).
[6] Y. Iwadate, "Study on Image Expression", IEICE HCS99-53, pp. 87-94, 1999 (in Japanese).
[7] http://www.mic.atr.co.jp/organization/dept3 /index.html