# Gesture Recognition using Character Recognition Techniques on Two-dimensional Eigenspace

Hiroshi Ohno†and Masanobu Yamamoto
Department of Information Engineering, Niigata University
8050 Ikarashi 2-nocho, Niigata-city, 950-2102, Japan
Email†: ohno@vision.ie.niigata-u.ac.jp

## Abstract

*This paper describes a novel method for gesture recognition using character recognition techniques on two-dimensional eigenspace. An image-based approach can capture human body poses in 3D motion from multiple image sequences. The sequence of poses can be reduced into a trajectory on the two-dimensional eigenspace with preserving the main features in gesture, so that the gesture recognition equals the character recognition. Experiments for the gesture recognition using some character recognition techniques show our method is useful.*

## 1 Introduction

The gesture recognition has played an important role in various applications such as human-computer interaction and security systems that can detect peculiar behavior. This is one of the most active research areas with the improvement of computer. Existing methods are classified into an appearance-based approach[1,2,3,4] and a motion parameter-based approach [5,6].

For using a gesture image sequence, hidden Markov models (HMMs), dynamic time warping (DTW) and parametric eigenspace were applied. Yamato et al.[1] employed HMMs for the recognition of actions in tennis. Starner et al.[2] put a camera on the cap worn by the user and applied HMMs for real-time recognition of American Sign Language. Takahashi et al.[3] used DTW for the recognition of dexterous manipulation. Murase and Sakai[4] extended the parametric eigenspace method for identification of pedestrian. In these methods, the features in gesture are detected from gesture image sequences, so that the dimension of the feature space is very large. Additionally, the pose of the person relative to camera is strongly restricted.

For using motion parameters, Campbell and Bobick[5] used a "phase space" for representation of the ballet steps. The axes of the phase space correspond to a few motion parameters measured by the motion captures. A problem is left. That is how to select particular parameters among many motion parameters to preserve main features according to each step.

This paper proposes a new parameter-based method. Using the KL transform, a sequence of human poses can be reduced into a trajectory on a two-dimensional eigenspace with preserving the main features in gesture. That is, a gesture is drawn on the two-dimension plane, so that the gesture recognition equals the character recognition. This reduction makes the gesture recognition very simple by using the character recognition techniques. We will apply the proposed method to the basic gymnastic exercises.

The next Section describes the representation of human poses and the image-based tracking of human action. Section 3 explains a representation of the gesture as the trajectory on the two-dimensional eigenspace. Section 4 deals with the gesture recognition. Experiments in Section 5 refer to the gesture recognition of basic gymnastic exercises.

## 2 Representation of Human Poses and Image-based Tracking

We represent a human body by an articulated structure consisting of 11 rigid parts corresponding to head, trunk, waist, upper arms, forearms, thighs and shins, respectively. Each part of the body is approximated by a polyhedron which is made by a CAD modeler[7]. Fig.1 denotes the human body model. Each part of the body has a local coordinate system in which the origin is located at a joint, and a unique label to discriminate from each other. The model has a tree structure with a root at the trunk. The poses of parts calculated by image-based tracking denote rotational angles at the joint with respect to the straightened pose. The motion parameters of this model are rotational and translational quantities of the root and rotational quantities of the other parts. Even a single camera can capture the human body pose in 3D motion. However, there are a few problems. First, an occlusion makes it difficult to capture entire body movement, second, a visual degenerate in motion measurement may occur when the body moves toward or away from the camera. To overcome these problems, we use multiple camera views[8].

The model fitting is carried out for the multiple camera views at the initial frame to start tracking. Then the model represents the 3D pose of the human body. The motion parameters are estimated using the spatial and temporal gradient method, and the pose of human body at each frame is obtained from integration of a sequence of motion parameters onto the pose at the initial frame[8]. Fig.2 shows multiple images of

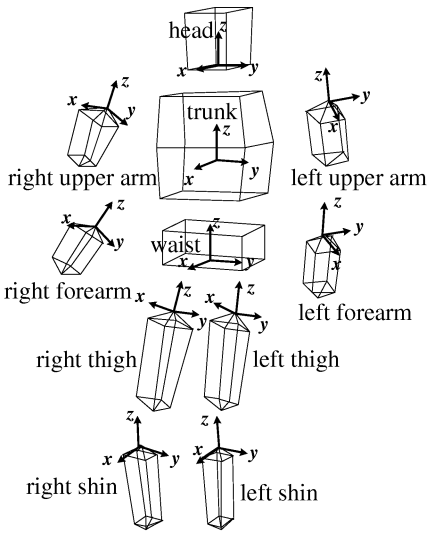


Figure 1: CAD modeling of parts of human body

basic gymnastic exercises from four camera views, on which the body model overlaps. Our gesture recognition method uses a sequence of the poses of the ten parts; i.e. trunk, waist, upper arms, forearms, thighs and shins.

The poses of each part obtained from the image-based tracking indicate rotational angles around the x, y and z axis with respect to the parent part coordinate system. Each axis of the part-oriented coordinate system is shown in Fig.1, where the z axis denotes a spine of each part. It is difficult to measure the rotational angles around the z axis in this image-based tracking because the area of parts, e.g. arms and legs, is observed to be very thin. To avoid this problem, we represent the pose by a unit vector $\boldsymbol{n}$ aligning the z axis of the part with respect to the parent part coordinate systems as shown in Fig.3. Supposing $\psi$ and $\varphi$ to be rotational angles around the x and y axis, respectively, the $\boldsymbol{n}$ is described by

$$\boldsymbol{n} = (-sin\psi, cos\psi sin\varphi, cos\psi cos\varphi)^T. \quad (1)$$

Fig.4 shows the measured components of $\boldsymbol{n}$ of body parts in one of the basic gymnastic exercises.

## 3 Representation of Gesture on Low-Dimensional Eigenspace

A human body consists of many parts, and each has many pose parameters. If these parameters are expressed in the low-dimensional feature space, gesture recognition becomes very simple. Here we will propose a representation of the gesture on the low-dimensional eigenspace.

Since parts of the human body constrain each other, the number of pose parameters can be reduced. The reduction can be performed by the KL (Karhune-Loève) transform.

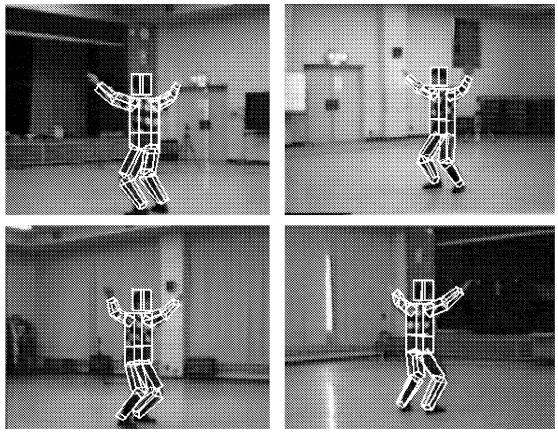

Figure 2: Images from four cameras, on which the body model overlaps

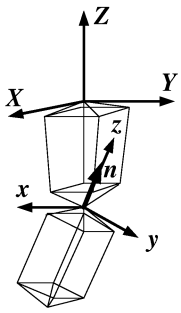


Figure 3: Part pose

We will represent the gesture on the eigenspace. Let a set of pose parameters of the training gestures be

$$\boldsymbol{x}_i = [x_{i1}, x_{i2}, \cdots, x_{im}]^T \quad (2)$$

where $i$ is the number of frames of training gesture and $m$ is the total number of the parameters($m$=30 in Fig.4). The pose parameters are normalized so that the range of every parameter could be from -1 to 1. The average of pose parameters, $\boldsymbol{c}$, is defined by

$$\boldsymbol{c} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i \quad (3)$$

where $N$ is the total number of frames. Subtracting the bias, $\boldsymbol{c}$, from pose parameters results into the following pose matrix $\boldsymbol{X}$.

$$\boldsymbol{X} = [\boldsymbol{x}_1 - \boldsymbol{c}, \boldsymbol{x}_2 - \boldsymbol{c}, \cdots, \boldsymbol{x}_N - \boldsymbol{c}]^T \quad (4)$$

The covariance matrix $\boldsymbol{Q}$ of pose matrix $\boldsymbol{X}$ is represented by

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{X}^T. \quad (5)$$

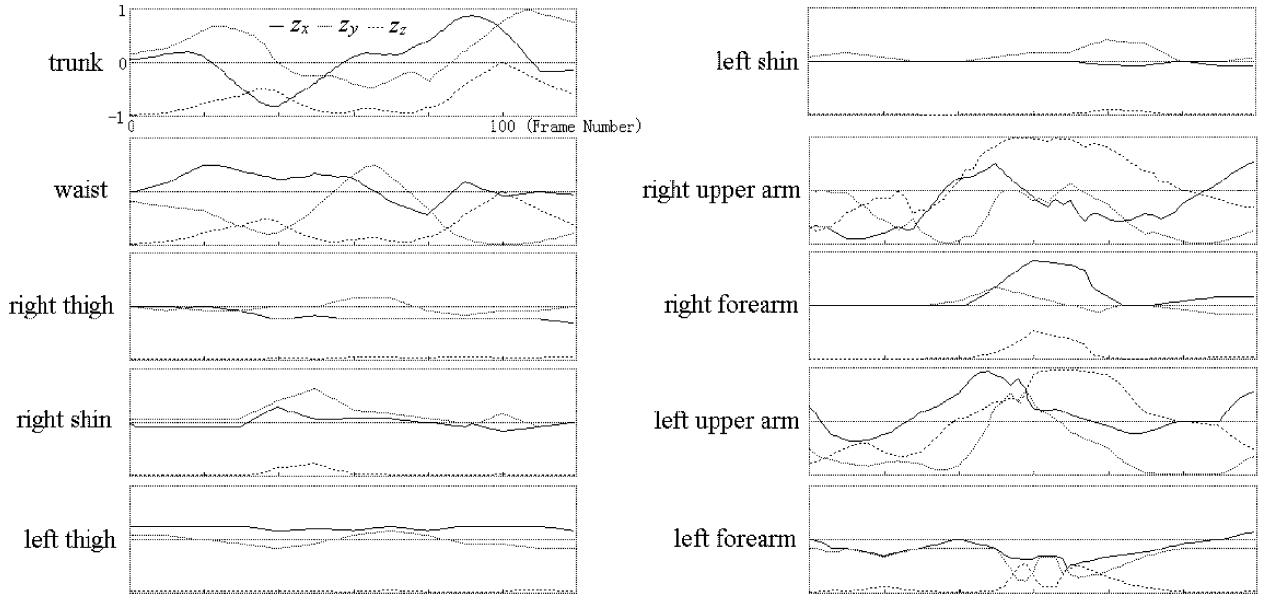


Figure 4: Components of unit vectors representing part pose

An eigenvector $\boldsymbol{e}_i$ and the corresponding eigenvalue $\lambda_i$ of $\boldsymbol{Q}$ are computed by solving the eigenvector decomposition problem:

$$\lambda_i \boldsymbol{e}_i = \boldsymbol{Q}\boldsymbol{e}_i \tag{6}$$

The $k$-dimensional eigenspace is spanned by the $k$ eigenvectors $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k)$, which correspond to the largest $k$ eigenvalues $(\lambda_1 \geq \cdots \geq \lambda_k)$, as basis vectors.

A training gesture is represented on the eigenspace. A set of pose parameters, $\boldsymbol{x}_i$, is projected onto the eigenspace by

$$\boldsymbol{g}_i = [\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_k]^T (\boldsymbol{x}_i - \boldsymbol{c}). \tag{7}$$

The set of pose parameters is projected as one point onto the eigenspace. Assuming the pose displacement between two successive frames to be very small, the two projections onto the eigenspace are close to each other. So the gesture is represented as a continuous trajectory on the eigenspace.

Now, we consider dimension of the eigenspace, which is sufficient to represent the gesture. One approach is to select the first $n$ eigenvectors representing the important gesture variations. The cumulative proportion

$$SC_n = \frac{\sum_{j=1}^{n} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \tag{8}$$

can be often referred to determine the dimension of eigenspace enough to represent the gesture. If it is close to unity, the $n$-dimensional subspace represents an original gesture very well. In this paper, we represent the gesture on the two-dimensional eigenspace, which will be found to be adequate. We call the two-dimensional eigenspace an eigenplane. Fig.5 shows a gesture on the eigenplane. Poses from 1 to 4 are projected as the points from $P1$ to $P4$, and the gesture is represented as a trajectory on it.

## 4 Gesture Recognition

An unknown gesture can be projected onto the eigenplane as a trajectory using eigenvectors obtained from training gestures. A sequence of pose parameters is projected onto this space by

$$\boldsymbol{g}_i = [\boldsymbol{e}_1, \boldsymbol{e}_2]^T (\boldsymbol{x}_i - \boldsymbol{c}). \tag{9}$$

The gesture recognition is performed by comparing a similarity between the trajectory of the unknown gesture and the trajectory of the training gestures on the eigenplane. Since the gesture trajectory is represented as a curve on the eigenplane, one can regard the gesture recognition as the character recognition. Many methods have been proposed for the character recognition, where the selection of the feature vector is very important.

### 4.1 Selection of Feature Vectors

This section shows the typical feature vectors used in the character.

(1) Centroid and length of trajectory

A different position and shape of the trajectory on the eigenplane corresponds to the different gesture.

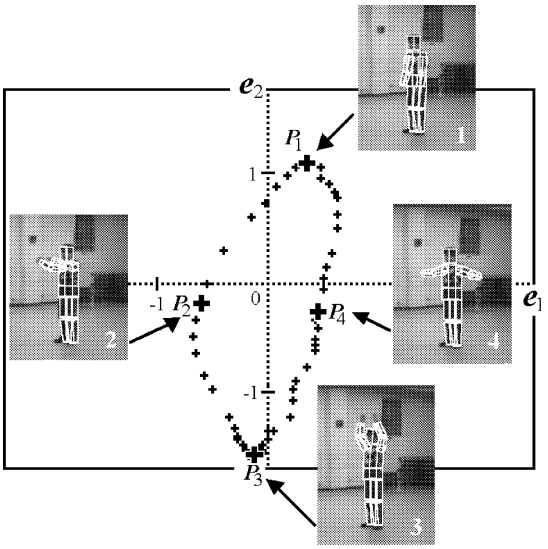


Figure 5: Gesture on eigenplane

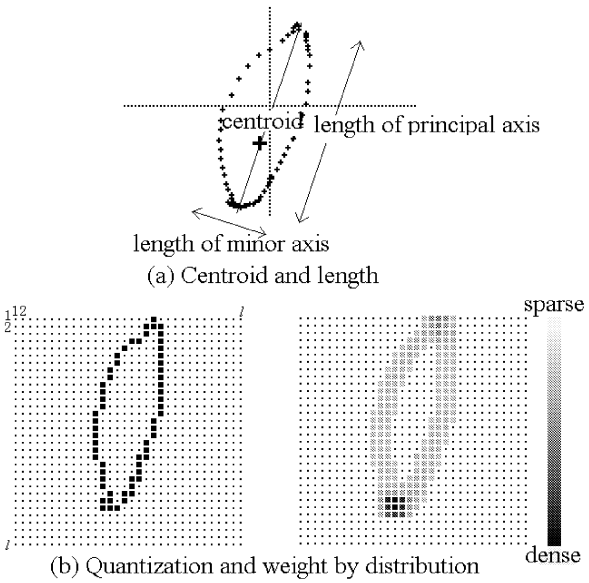


Figure 6: Gesture trajectory on eigenplane

So, we select the centroid, the length of the principal axis of the trajectory and the length of the minor axis of the trajectory as the features shown in Fig.6(a). We set these as components of the feature vector. The gesture trajectory is projected as one point onto the feature space. The points in the space are scaled so that average distance from the centroid is equal to unity.

(2) Subspace eigenvector

The subspace method has been used as a typical technique for the character recognition[9]. We apply this method to the gesture recognition. When a gesture trajectory is drawn on the $l \times l$ eigenplane, it is quantized at $l^2$-meshes shown in the left of Fig.6(b) and expressed by $l^2$-dimensional vectors

$$\boldsymbol{x} = [x_1, x_2, \cdots, x_{l^2}]^T. \tag{10}$$

The covariance matrix and its eigenvectors are calculated by the same way in Section 3. The subspace is spanned by some eigenvectors, which corresponded to some large eigenvalues. The gesture is projected as one point onto the subspace.

In the character recognition, the subspace method usually normalizes the scale and position of the character, while, in the gesture recognition, the scale and position should not be normalized, since the different scale and position of the trajectory corresponds to the different gesture. However, a small change of the scale and position of the trajectory may mean the same gesture. Therefore, the trajectory is thickened so that the trajectories having a small change are recognized as the same. The meshes including the trajectory are assigned to 1, and the other meshes are 0. We lose a distribution density of the points on the eigenplane, which is one of the features in gesture. So, we weight the trajectory with the distribution density as shown in the right of Fig.6(b).

## 4.2 Linear Discriminate Method

A gesture expressed as a trajectory on the eigenplane is projected as one point onto the feature space that is spanned by the feature vectors. The training gestures consist of many samples of performers, so that the gesture is recognized by the linear discriminate method.

Suppose the dimension of the feature space to be $d$, and that the feature space contains two classes. The hyperplan which discriminate two classes well is decided by the Fisher's linear discriminate method[10].

A criterion of separation of two classes is to increase a within-class scatter and to decrease a between-class scatter. The scatter matrix $\boldsymbol{S}_i$ which means the scatter of classes $\omega_i (i = 1, 2)$ is defined by

$$\boldsymbol{S}_i = \sum_{y \in y_i} (\boldsymbol{y} - \boldsymbol{m}_i)(\boldsymbol{y} - \boldsymbol{m}_i)^T \tag{11}$$

where $\boldsymbol{y}$ is a feature vector of class $\omega_i$ and $\boldsymbol{m}_i$ is the average of $\boldsymbol{y}$. The within-class scatter $\boldsymbol{S}_W$ and the between-class scatter matrix $\boldsymbol{S}_B$ are defined by

$$\boldsymbol{S}_W = \sum_{i=1,2} \sum_{y \in y_i} (\boldsymbol{y} - \boldsymbol{m}_i)(\boldsymbol{y} - \boldsymbol{m}_i)^T \tag{12}$$

$$\boldsymbol{S}_B = \sum_{i=1,2} n_i (\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^T \tag{13}$$

where $\boldsymbol{m}$ is an average of all patterns and $n$ is the number of patterns. $\boldsymbol{J}_S(\boldsymbol{A})$ is a ratio of the transformed within-class scatter $\tilde{\boldsymbol{S}_W}$ to the transformed

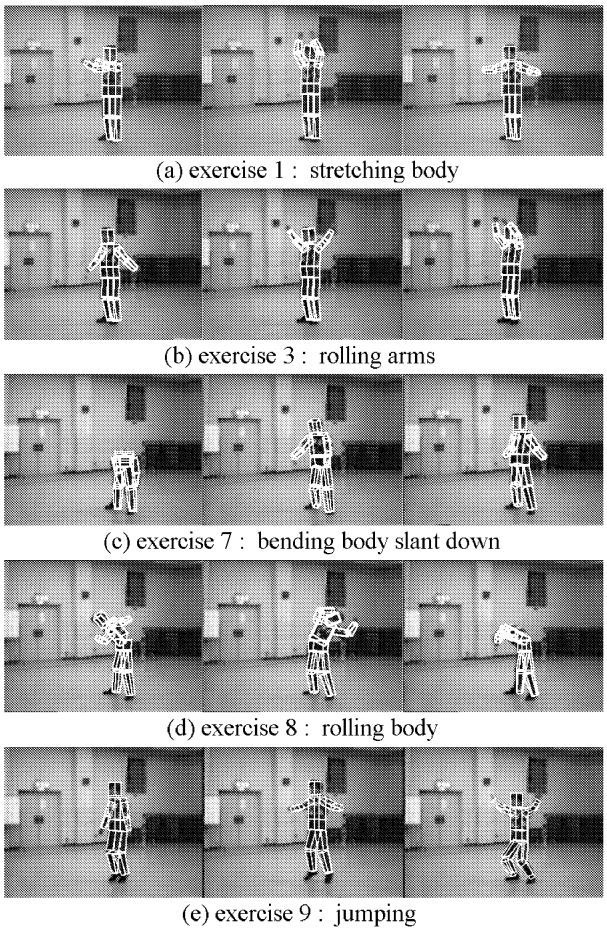


Figure 7: Tracking results of the basic gymnastic exercises

within-class scatter $\tilde{\boldsymbol{S}_B}$

$$\boldsymbol{J}_S(\boldsymbol{A}) = \frac{\tilde{\boldsymbol{S}_B}}{\tilde{\boldsymbol{S}_W}} = \frac{\boldsymbol{A}^T \boldsymbol{S}_B \boldsymbol{A}}{\boldsymbol{A}^T \boldsymbol{S}_W \boldsymbol{A}}. \tag{14}$$

Calculating $\boldsymbol{A}$ which maximizes $\boldsymbol{J}_S(\boldsymbol{A})$ equals maximizing

$$\tilde{\boldsymbol{S}_B} = \boldsymbol{A}^T \boldsymbol{S}_B \boldsymbol{A} \tag{15}$$

subject to

$$\tilde{\boldsymbol{S}_W} = \boldsymbol{A}^T \boldsymbol{S}_W \boldsymbol{A} = 1. \tag{16}$$

The $\boldsymbol{A}$ is a normal vector of the boundary of the two classes for discrimination. Therefore, we decide the position of the boundary, which could pass through the center of the centroids of two classes. According to more classes, we need to calculate more boundaries.

# 5 Experimental Results

## 5.1 Set up Experimental Conditions

We try to recognize nine classes in the basic gymnastic exercises (exercise 1: stretching. exercise 2: swinging arms while making shallow knee bends. exercise 3: arm rolling. exercise 4: side bending. exercise 5: back and forth torso bending. exercise 6: full body stretching, arms above head to the floor. exercise 7: angled toe touching. exercise 8: torso rotation, exercise 9: jumping). Eight performers played the exercises. They are named as A to H, respectively, and are males between twenty and forty years old.

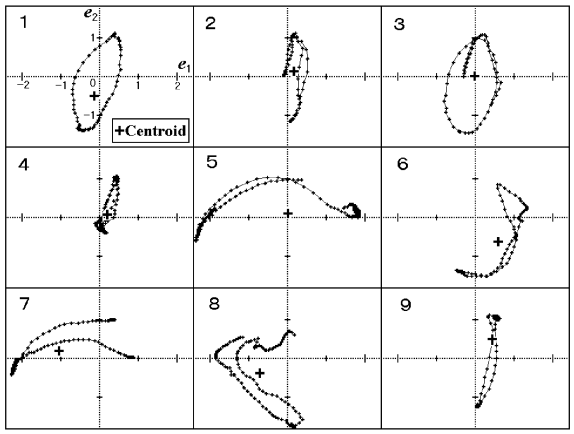


Figure 8: Nine gestures on the eigenplane in the basic gymnastic exercises (performer A)

An image from the camera has 8-bits gray scales in $320 \times 240$ pixels. A gesture is tracked from multiple camera views. One cycle time of exercise is varying from 48 to 200 frames. Fig.7 shows some examples of the tracking results.

In the subspace method, a trajectory on the eigenplane is quantized by the $32 \times 32$ meshes.

## 5.2 Results and Discussion

Eight performers are divided into two groups, six and two. Six performers from A to F acted training gestures and the others G and H acted unknown gestures. When the number of training gestures is (1) 4 from A to D, (2) 5 from A to E, (3) 6 from A to F, eighteen gestures of nine exercises acted by the two, G and H, are tested. Fig.8 shows gestures on the eigenplane in the basic gymnastic exercises acted by the performer A. Fig.9 shows one gesture on the eigenplane played by the eight performers. Table 1 shows the recognition rate when the feature vectors are composed of centroid and length. Table 2 shows the recognition rate using subspace method.

According to the experimental results in Table 1 and Table 2, the recognition rate increases when the number of training samples increases. However, the recognition rate decreases when the training data and unknown data are interchanged. For example, when the performer D taking a unique action is included in training samples, the recognition rate decreases.

In the subspace method, the recognition rates are calculated when the cumulative proportion is 60, 80 and 90 % which correspond to feature space with 4,

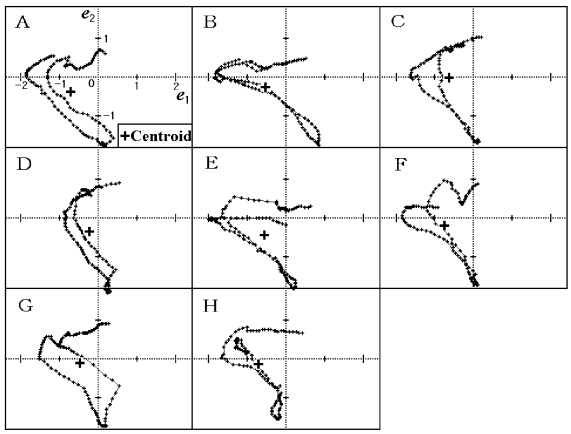


Figure 9: One gesture on the eigenplane played by eight performers (exercise 8: torso rotation)

13 and 22 dimensions, respectively. The recognition rate increases when the dimension of the feature space increases. However, the number of training sample has to increase according to the dimension of the feature space.

Comparing the subspace method with weight to one without weight, weighting is effective. In the case without weight it is difficult to discriminate exercise 1 from 3, and exercise 2 from 9, because these trajectories have the similar shapes on the eigenplane. But the distribution density of the points on the eigenplane depends on the gesture. So weighting by the distribution density of the points makes the recognition rate increase.

The pose parameters in all parts of the body model are reduced using KL transform and the gesture is expressed by a trajectory on two-dimensional eigenspace(eigenplane). When the number of training samples is 4, 5 and 6, the cumulative proportion is about 58 % and it is not so high. However, the high recognition rate means that the two-dimensional eigenspace is sufficient to represent the main features in gesture.

## 6 Conclusion

This paper described a new method for gesture recognition using character recognition techniques on a two-dimensional eigenplane. Experiments for the gesture recognition using some character recognition techniques show that our method is useful. The human motion is complicated, but the expression of gestures on the eigenplane makes the gesture recognition very simple. This means that the gesture recognition equals the character recognition and we can get high recognition rate using some character recognition techniques.

As for future works, we have to make more experiments with many other gestures.

Table 1: Recognition rate (%)(centroid and length)

| number of training samples | | |
|---|---|---|
| 4 | 5 | 6 |
| 88.9 | 94.4 | 100.0 |

Table 2: Recognition rate (%)(subspace method)

| | dimension | number of training samples 4 | 5 | 6 |
|---|---|---|---|---|
| without weight | 4 | 72.2 | 77.8 | 83.3 |
| | 13 | 83.3 | 83.3 | 83.3 |
| | 22 | 83.3 | 83.3 | 83.3 |
| with weight | 4 | 83.3 | 83.3 | 88.9 |
| | 13 | 83.3 | 88.9 | 88.9 |
| | 22 | 88.9 | 88.9 | 94.4 |

## References

[1] J.Yamato, J.Ohya and K.Ishii : Recognition human action in time-sequential images using hiddenMarkov model, Proc. CVPR'92, pp.379-385, 1992.

[2] T.Starner, J. Weaver and A.Pentland : Real-time American Sign Language recognition using desk and wearable computer based video, IEEE Trans. Pattern Anal. And Mach. Intell., Vol.20, No.12, pp1371-1375, 1998.

[3] K.Takahashi, S.Seki, H.Kojima and R.Oka : Recognition of dexterous manipulation from time-varying iamges, Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects, Austin, pp.23-28, 1994.

[4] H.Murase and R.Sakai:Moving object recognition in eigenspace representation: gaint analysis and lip reading, Pattern Recognition Letters, Vol.17, No.2, pp.155-162, 1996.

[5] L.Campbell and A.Bobick : Recognition of human body using phase space constraints, Proc. of 5th ICCV, pp.624-630, 1995.

[6] A.Bobick and A.Wilson : A state-based approach to the representation and recognition of gesture,IEEE Trans. Pattern Anal. and Mach. Intell., Vol.19, No.12, pp1325-1337, 1997.

[7] K.Koshikawa and Y.Shirai : A 3-D modeler vision research. Proc. of Int. Conference on Advanced Robotics, pp.185-190, 1985.

[8] M.Yamamoto, A.Sato and S.Kawada : Increment tracking of human action from multiple views, Proc. of CVPR'98, pp.2-7, 1998.

[9] E.Oja : Subspace methods of pattern recognition, Research Studies Press, Hertfordshire, 1983.

[10] R.Ouda and P.Hart : Pattern classification and scene analysis, Wiley-interscience, 1973.