MVA'98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

3—20 Stable Gesture Verification in Eigen Space

Shigeyoshi Hiratsuka *
Graduate School of Information Sciences / Mechanical Engineering Laboratory
Tohoku University / AIST, MITI

Kohtaro Ohba †
Mechanical Engineering Laboratory
AIST, MITI

Shinya Kajikawa ‡
Faculty of Mechanical Engineering
Miyagi National Collage of Technology

Hikaru Inooka §
Graduate School of Information Sciences
Tohoku University

Kazuo Tanie ¶
Mechanical Engineering Laboratory
AIST, MITI

Abstract

This paper mainly discusses the stable gesture classification criteria, which can identify several human motions from a sequence of images with the eigen space method. We have already proposed the gesture classification technique with the silhouette images and the PCA method, which includes the difficulty in the correspondence between two sets of eigen points, i.e. dictionary and input sequences, which are taken with the different time step. In this paper, a new criteria has been proposed to achieve a stable gesture verification technique with the spline approximation of the input sequence in the eigen space. Experimental results show the validity of this proposed method.

1 Introduction

Recently, many studies on the man-machine interfaces with vision systems have been proposed, because of the simpleness to use without any wareable sensors, such as data-glove. Especially in human motion understanding, vision is strongly recommended for its variety of applications in man-machine interfaces, such as for the robot in hospital, the helper-robot in house, and so on.

Up to now, there are many approaches to recognize human motions from image sequences[1] - [3]. One of the most reliable technique might be a geometric based approach, in which several feature points on the human body, such as hands, knees, foots, and head, had been detected to calculate the angle of each joint. But, to detect the feature points, several marks on the body are required, or huge calculation is necessary to make correspondences with the template matching technique.

In our previous research [7], one of the appearance based approach, PCA, had been proposed to identify the human gestures, which can reduce the dimension of images and project them into the low-dimensional eigen space without any consideration of image features [4] - [7]. The process of this identification method is as follows. At first, a gesture space (GS) and a set of template gesture points have been obtained from a dictionary image sequence with the PCA criteria. Here, a sequence of silhouette images has been used, instead of original black/white image sequences, to eliminate the background noise and the difference between each person. Secondly, the input image sequence had been projected into the GS with the several significant eigen values. Finally, with the comparison between two sets of a template gesture eigen points and an input eigen points, the gesture can be identified in real time. Then we come up to a problem in this corresponding process, which is that the error between two set of points becomes noisy, even though the gesture is exactly same.

In this paper, we discuss the several factors of the unreliable correspondence, and propose the new criteria with the spline approximation to make a stable correspondence. Several experimental results have been evaluated in real-time to show the validity of this proposed method.

2 Gesture Space with PCA

2.1 PCA

In this section, a brief overview of the PCA criteria has been reviewd.

Let M be the number of the images in a training

*Address: 1-2 Namiki, Tsukuba, Ibaraki 305-8564 Japan E-mail: hiratuka@mel.go.jp

†E-mail: kohba@mel.go.jp

‡Address: 48 Aishimashiote-Aza-Nodayama, Natori, Miyagi 981-1239 Japan E-mail: kajikawa@miyagi-ct.ac.jp

§Address: Aramaki-Aza-Aoba, Aoba-Ku, Sendai, Miyagi 980-8579 Japan E-mail: inooka@control.is.tohoku.ac.jp

¶E-mail: tanie@mel.go.jp

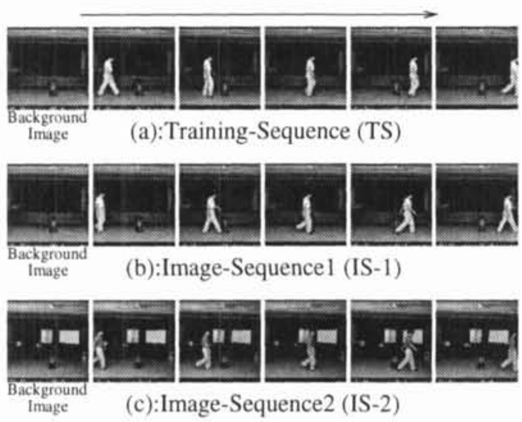


Figure 1: Original Image Sequences of Walking Motion.

Table 1: Types of Gesture Sequences.

Image sequence	Experimental condition
Training-Sequence (TS)	Motion A
Image-Sequence1 (IS-1)	Motion A (same image sequence as TS, but different sampling step)
Image-Sequence2 (IS-2)	Motion A (different background, and different person)

set. Each image has been scanned into a column vector $\mathbf{z}_i$ of length N. By subtracting the average image $\mathbf{c}$ of the all images, we obtain the training matrix $\mathbf{Z} = [\mathbf{z}_1 - \mathbf{c} \;\; \mathbf{z}_2 - \mathbf{c} \cdots \mathbf{z}_M - \mathbf{c}]$, where the size of the matrix $\mathbf{Z}$ is $N \times M$.

The sample covariance matrix $\mathbf{Q}$, $N \times N$, is obtained as: $\mathbf{Q} = \mathbf{Z}\mathbf{Z}^T$. This matrix provides a series of eigenvalues λ_i and eigenvectors $\mathbf{e}_i (i = 1, \cdots, N)$, where each corresponding eigenvalue and eigenvector pair satisfies: $\mathbf{Q}\mathbf{e}_i = \lambda_i \mathbf{e}_i$.

From this significant set of eigenvectors, the matrix $\mathbf{E} = [\mathbf{e}_1 \;\; \mathbf{e}_2 \cdots \mathbf{e}_k]$ is constructed to project an image, $\mathbf{z}_i$ (dimension N) into the eigen space as an eigen point, ζ_i (dimension k): $\zeta_i = \mathbf{E}^T(\mathbf{z}_i - \mathbf{c})$. This eigen space analysis can drastically reduce the dimension of the images (N) to the eigen space dimensions (k) while keeping several of the most effective features of the original images.

2.2 Gesture Space(GS)

We define this eigen space, which has been obtained in the previous section, as "**G**esture **S**pace". This gesture space can show the movement of human motion as the trajectory in the low dimensional space without any consideration of the moving parts

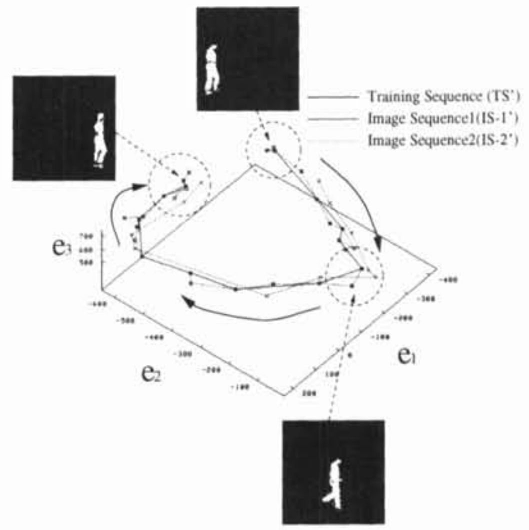


Figure 2: Gesture Trajectory in GS with Silhouette Images.

Table 2: Types of Gesture Images.

Image sequence	Types of motion
Training-Sequence (TS)	Motion A
Image-Sequence1 (IS-1)	Motion A
Image-Sequence3 (IS-3)	Motion B

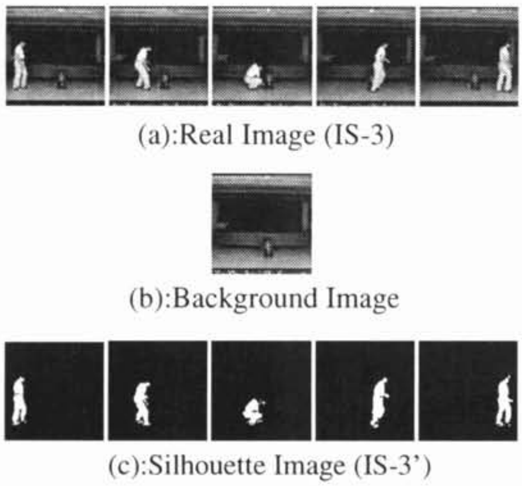
of human body.

To show the validity of this eigen space technique, several image sequences are taken. Figure 1 (a), (b) and (c) depict a training sequence and two input sequences, respectively. Table 1 shows experimental conditions for each image sequences.

Here, to reduce the drawback of original eigen space technique, such as disturbance effects in the eigen space written in our previous work [7], we apply silhouette images instead of original black/white images. The silhouette images can make the disturbance effects small under the detection of the moving objects only, which is obtained by subtraction of background image from input images, and binalization with pre-defined threshold value. TS', IS-1' and IS-2' depict silhouette image sequences of TS, IS-1 and IS-2, respectively.

First, the template silhouette image sequence derives the significant eigen vectors to make a template gesture trajectory in the gesture space with the KL expansion criteria.

With the evaluation of weight of the obtained eigen values, only three dimensions are enough to reconstruct the 75% of original images. From now on in this paper, three dimensions of eigenvectors are taken as the gesture space to calculate the principle components. Then, the template training curvature can be obtained with the projected eigen points in



(a):Real Image (IS-3)

(b):Background Image

(c):Silhouette Image (IS-3')

Figure 3: Motion Images(Motion B).

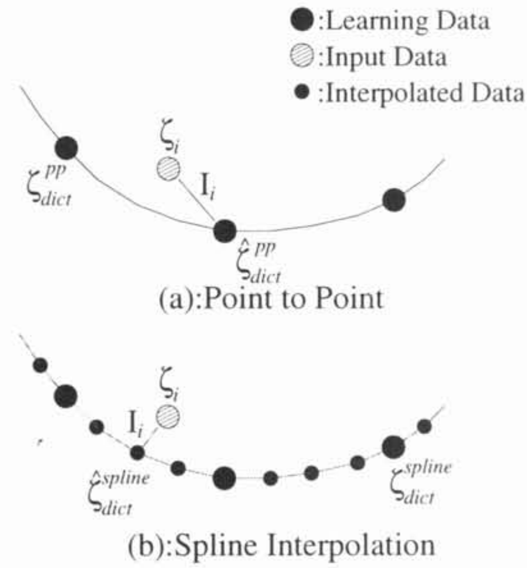


Figure 4: Matching Methods.

GS. This curvature is shown in figure 2 by a solid line.

Next, each input silhouette image sequence in figure 1 (b) and (c) have been projected into GS, and made input curvatures in figure 2 by square marks - a dotted line, and by 'x' marks - a dotted line, respectively.

In this case, each curvature is almost same, in spite of the different experimental conditions, such as difference backgrounds and different persons.

3 Classification of Human Gestures

3.1 Projection of other Gesture

To evaluate the several human gestures in GS, another human motion for input image sequence has been taken as shown in figure 3. Figure 3 (a), (b) and (c) show original image, background image and silhouette image of motion B, respectively. Table 2 shows conditions of motion.

An another input image sequence IS-3' which is coming up with another gesture, are also projected and make another curvature as shown in figure 6 by 'x' marks - dotted line.

The input curvature of IS-3' is obviously different from a template gesture curvature in GS.

3.2 Gesture Classification

We have already proposed a simple gesture classification technique, which evaluate the similarity with the correspondence of the two set of eigen points, i.e. a dictionary and an input image sequence. But it includes a difficulty to make a good correspondence between two sets of sequences on the difference of time step.

In our previous approach, similarity had been evaluated with L1-norm, which is obtained by disparity between an input GS point and a nearest template GS point on learning locus in the gesture space, as shown in figure 4 (a). A nearest GS point $\hat{\zeta}^{pp}_{dict}$ in dictionary points from input point ζ_i is obtained with $\hat{\zeta}^{pp}_{dict} = \arg\min_{\forall \zeta^{pp}_{dict}} (\|\zeta_i - \zeta^{pp}_{dict}\|)$. Then, the disparity I_i is obtained as: $I_i = \|\zeta_i - \hat{\zeta}^{pp}_{dict}\|$, where $\|x\|$ depicts the norm of x.

3.3 Discussion

Even though, the proposed criteria works well with some particular image sequences, several erroneous results had been observed in the several situation on other sequence of images.

One of the reasons of this error might be that we are trying to identify the location and the pose on the human gesture at same time. Actually, other researchers had proposed the human gesture recognition criteria, which includes two phases, the localization technique with the moving object and the PCA criteria only on the moving object's template. Further more, the stochastic modeling technique, HMM, had been applied to realize the stable recognition. But this kind of process makes system complex, and finally the real-time sytem could not be realized.

Then, we would like to carefully observe the movement of eigen points. Figure 5 shows detail of eigen points around '**C**'-point in figure 6. This figure shows that the erroneous results are caused by the difference of the sampling step, such as in the case of that an input eigen points are projected into the almost middle points of the two dictionary eigen points, even though it is almost located on the line of two points.

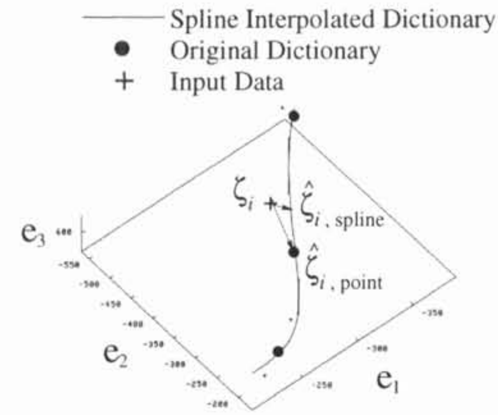


Figure 5: The Movement of Eigen Points.

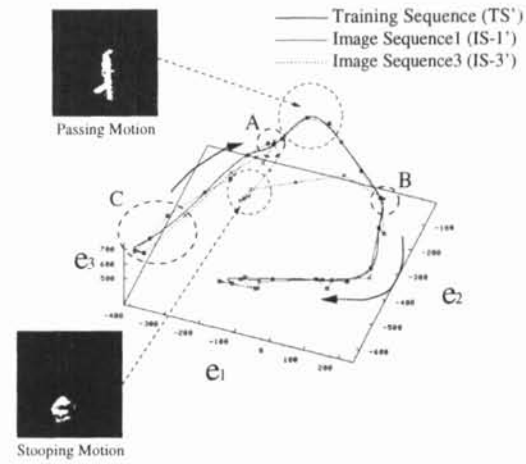


Figure 6: Classification of Gestures.

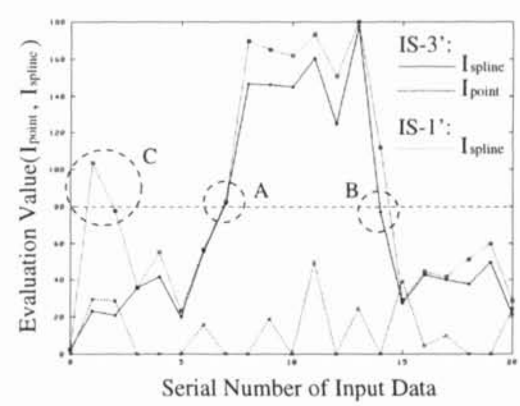


Figure 7: Evaluation of Gesture Classification.

3.4 Spline Approximation

To get more stable correspondence between two point sets, a new evaluation technique has been proposed, which define the disparity as the distance between input GS point and the template GS spline curvature besides the GS point: $I_i = \|\zeta_i - \hat{\zeta}_{dict}^{spline}\|$, as shown in figure 4 (b). This spline curvature interpolate the template GS points, and make the similarity between input GS points and template GS curvature more stable.

Figure 6 also shows a learning locus on eigen space by solid line. The smooth learning locus is obtained with spline approximation. Still more, in figure 6, projection results are shown.

Figure 7 shows the result of classification. The spline method is more precise than the previous method, and derive more exact gesture classification, as shown in figiure 7 '**C**'-point.

4 Conclusions

This paper discussed how to obtain a stable gesture verification in the eigen space.

We have discussed to utilize a spline curvature for making a more stable correspondence in the eigen space. Then, we have proposed a new evaluation criteria to achieve a stable gesture verification with the spline approximation criteria. Finally, we have applied this technique to classify several human motions. Several experimental results have been evaluated in real time to show the validity of this proposal method. And this proposed criteria has efficacy for the effect of the difference of the sampling step.

References

[1] Y. Yacoob, M. J. Black, "Parameterized Modeling and Recognition of Activities", 6th International Conference on Computer Vision, pp. 120-127, 1998

[2] A. D. Wilson, A. F. Bobick, "Recognition and Interpretation of Parametric Gesture", 6th International Conference on Computer Vision, pp. 329-336, 1998

[3] C. Vogler, D. Metaxas, "ASL Recognition Based on a Coupling Between HMMs and 3D Motion Analysis", 6th International Conference on Computer Vision, pp. 363-369, 1998

[4] H. Murase, S. K. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance," International Journal of Computer Vision, 14-1, pp. 5-24, 1995

[5] K. Ohba, K. Ikeuchi, "Detectability, Uniqueness, and Reliability of Eigen Windows for Stable Verification ofPartially Occluded Objects," IEEE Transactions on Pattern Analysis and Machine Intelligence, 19, 9, pp. 1043-1048, 1997

[6] T. Watanabe, M. Yachida, "Real-Time Gesture Recognition Using Eigenspace from Multi Input Image Sequence," IEEE Proc. of the 3rd International Conference on Automatic Face and Gesture Recognition, pp. 428-433, 1998

[7] S.Hiratsuka, K. Ohba, S. kajikawa, H. Inooka, K. Tanie, "Real-Time Classification on Human Motions," IEEE Proc. of the 8th International Workshop on Robotics and Human Communications, 1998