Real-time Hand Shape Recognition for a Human Interface

Horimoto, Shuhei
Department of Intelligent Systems, Kyushu University

Arita, Daisaku
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

出版情報：Proc. of the 9th Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp.106-111, 2003-02
バージョン：
権利関係：

Real-time Hand Shape Recognition for a Human Interface

Shuhei HORIMOTO, Daisaku ARITA and Rin-ichiro TANIGUCHI
Department of Intelligent Systems, Kyushu University
{horimoto,arita,rin}@limu.is.kyushu-u.ac.jp

Abstract Recently, a natural and user-friendly human interface has been actively researched. Although conventional user interface is based on keyboards and mice, human body actions can be used for human machine interface. From this point of view, we have been researched into vision-based real-time human motion analysis. The important point is that we have employed vision-based methods, because they do not impose any physical restrictions on users. In this paper, we present a robust hand shape recognition system based on an eigenspace method, in which we employ a simple additional learning feature, and we show its performance through some experiments.

1 Introduction

Recently, a natural and user-friendly human interface has been actively researched. Although conventional user interface is based on keyboards and mice, human body actions can be used for human machine interface. From this point of view, we have been researched into vision-based real-time human motion capturing [1, 2], which is aiming to analyze full-body human motion. Our system can not recognize detailed parts of human body such as a hand shape due to the image resolution. However, hand shapes are often used for human machine interface and nonverbal communications as well, and we have started to construct a recognition system of hand shapes for human machine interface. The important point is that we have employed vision-based methods, because they do not impose any physical restrictions on users.

Essentially, there are two approaches for hand shape recognition: one is rather hand shape measurement not hand shape recognition, which measures hand shape parameters, such as finger joint angles, as precise as possible; the other is not measurement but classification of hand shapes into predefined categories. The first approach has a computational cost problem, because a hand has geometrically high DOFs and it is very difficult to analyze. In addition, even with the measurement approach, classification procedure is required to a certain extent. From these considerations, we have adopted the latter approach, classification of hand shape images into predefined categories, assuming that users show their specific hand shapes toward a camera consciously. In this paper, we present a robust hand shape recognition system using eigenspace method, in which we employ a simple additional learning feature, and show its performance through some experiments.

2 Eigenspace Method

The recognition method based on eigenspace method is a statistic pattern matching method, and is famous as "eigenface" for face recognition. Information of input image patterns is compressed by principal component analysis, and their meaningful features can be extracted. Thus, robust and fast recognition of various image patterns can be achieved [3]. The recognition algorithm is summarize as follows.

First, we define an image $u = \{u_1, u_2 ..., u_n\}$ in the n-dimensional space where n is the number of pixel in the image, and we suppose that we have a group of template images $u^i = \{u_1^i, u_2^i ..., u_n^i\}$ $(i = 1..., m)$. Then, we calculate an averaged image $\bar{u} = \{\bar{u}_1, \bar{u}_2 ..., \bar{u}_n\}$ and a covariance matrix $C = |c_{ij}|$.

$$c_{ij} = \sum_{l=1}^{n} (u_l^i - \overline{u}_l) \cdot (u_l^j - \overline{u}_l) \quad (1)$$

Eigen image bases $e^i = \{e_1^i, e_2^i, ...e_n^i\}(i = 1, ..., m_1 \leq m)$ of the template images are established by choosing m_1 eigenvectors of the matrix C. e^i is calculated as follows:

$$e_l^i = \frac{1}{\sqrt{\lambda_i}} \sum_{k=1}^{m} v_k^i \cdot (u_l^k - \overline{u}_l) \quad (2)$$

where λ_i and $v^i = \{v_1^i, v_2^i ..., v_m^i\}$ is the i-th largest eigenvalue and its corresponding eigenvector of the matrix C.

Finally, an input image is projected into the eigenspace based on the established eigen bases.

$$\tilde{u}_l = \sum_{k=1}^{m_1} w_k e_l^k + \overline{u}_l \quad (3)$$

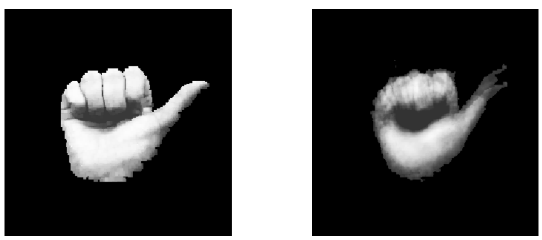
where w_k is the weight for the eigen basis e^k, and we call $\{w_k\}$ eigen coefficients here after.

As a result, we can express the input image as the eigen coefficients, or an compressed representation of the image, whose dimensionality is much lower than that of the image(See Fig.1(b)).

In matching, an input image is converted to eigen coefficients $\{w_k^{in}\}$ with the same eigen bases and then matched with that of templates $\{w_k^j\}$ $(j = 1, 2, ..., n)$ based on the distance:

$$dis\{j\} = \sum_{k=1}^{m_1} \frac{|w_k^j - w_k^{in}|}{|e^k|} \quad (4)$$

where $|e^k|$ is the size of the eigen basis e^k. The best matched template, whose distance $dis\{j\}$ is the smallest, is as the result of recognition of the input image.



(a)Normalized Image (b)Eigen Image

Fig. 1: Convert a Normalized Image to an Eigen Image

3 Hand Shape Recognition System using Eigenspace Method

3.1 Hand Shape Models

We try to recognize 20 kinds of hand shapes, which are used to indicate a finger alphabet of the Japanese sign language (see Fig.2). Originally, the finger alphabet contains rotational variations of the same hand shape, but we exclude the rotational variations, because, here, we pay attention only to the shape not to the alphabet. The number of the hand shapes to identify is set to be 20 here, which seems to be small compared with the number of possible hand shapes. However, for human interface, we do not need to distinguish so many hand shapes, because if there are too many different hand shapes to identify, we may feel difficulty to remember them, which leads to "uneasy" interface. In this sense, we think those 20 hand shapes, which are distinctive one another, are enough for most hand-based human interfaces.

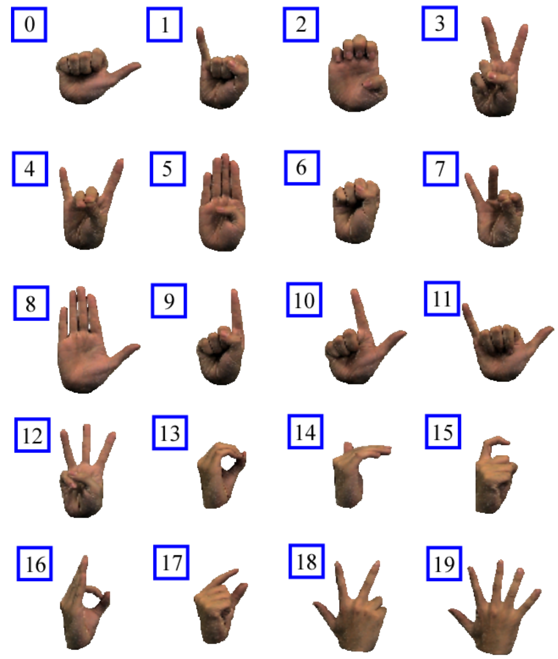


Fig. 2: Hand Shape Models

3.2 Establishment of Eigen Bases

To deal with the rotation of a hand, we took images of rotational variations of the hand shapes as template images. For each hand shape, we took 20 rotational variations. They are 7 variations around x-axis, 6 around y-axis, 7 around z-axis, where we define y-axis as the axis of an arm, z-axis as the direction of the camera, and x-axis as a cross product of y and z axes(see Fig.3). The rotation step is 10 degrees here. Therefore, 400 template images were taken for one person. In addition, to deal with personal differences, we took template images of 10 persons, 5 men and 5 women. As a result, total template hand images amount to 4000 (400×10 persons). Then we calculated 40 eigen bases from them and converted the template images to eigen coefficients. The number of eigen bases were decided based on experiments, which is discussed in a later section.

3.3 Nearest Neighbor Search

For shortening the recognition time, we adopt an efficient algorithm, the nearest neighbor search in high dimension space[4]. The algorithm is outlined as follows.

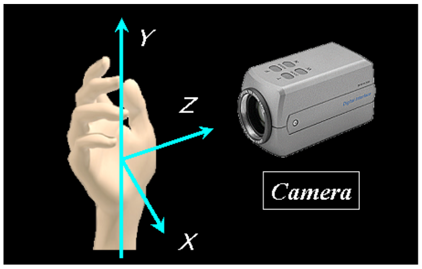


Fig. 3: Coordinate Axes

1. We assume that, in d-dimension space, there are many points, each of which corresponds to a template. The problem is to find the closest point to a newly given point $Q(Q_1, Q_2, ..., Q_d)$, or an input image.

2. Referring to the first coordinate, we look for points whose first coordinate values are lying between $Q_1 - \epsilon$ and $Q_1 + \epsilon$ (ϵ is a small offset), and add them to a list, which we call a candidate list.

3. In the candidate list, we select points whose k-th coordinate values are lying between $Q_k - \epsilon$ and $Q_k + \epsilon$ and discard other points from the list. We sequentially apply this operation to all the coordinates, $k = 2, 3, ..., d$.

4. The candidate list contains only the points inside a cube with the size of 2ϵ around the novel point Q, shown in Fig.4. Generally, ϵ is very small, and the number of points inside the cube is also small.

5. The nearest neighbor point is found by performing an exhaustive search within the cube.

3.4 Additional Learning

The main reasons of incorrect recognition are a shortage of templates and an ambiguity of border surfaces among hand models in a feature space. To solve the problems, we have adopted a very simple approach: when incorrect recognition occurs, the input image is registered as an additional template. As a result, the number of the templates gradually increases. The important issue here is computational cost. For example, in calculating 40 eigen bases from 4000 template images by a PC with Celeron 1.3GHz CPU, it takes 15 hours or more. We have to spend such a long time whenever we have new templates, but it does not seem practical.

A possible alternative is learning an image itself, not a set of eigen coefficients, as a template. However, this approach not only abandons the merits of the eigenspace method but also it is not practical. An image is so high dimensional that it takes much time to search for a hand model with the nearest distance, and memory space for storing templates images become larger.

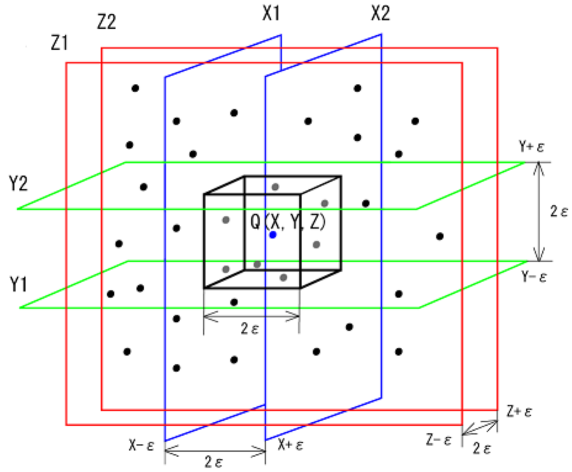


Fig. 4: Nearest Neighbor Search

Consequently, we have adopted a suboptimal approach: the online learning system learns the eigen coefficients of a new image on the already established eigenspace. We suppose that any 20 hand model images may be similarly projected into the eigenspace by using the 40 eigen bases calculated from the original 4000 rotated hand images. As a result, we can save the calculation time of eigen bases and can treat an image in a lower dimensions. We can learn a pile of templates online.

3.5 Evaluation for Continuous Frames

When the system is used online and recognizes a hand shape in every frame, there are some flickering incorrect recognitions. This is because the system recognizes a hand shape when a user does not show his/her hand shape consciously or when he/she is changing his/her hand shape to present another hand shape. By evaluating the hand shape in continuous frames, the system reduces the flickers and recognizes the hand shape almost only when a user shows his/her hand shape intentionally. Currently, we have used a very simple tactics: only when the same hand model is continuously recognized for a certain number of frames, the system judges that an user may show the hand shape consciously.

4 Experiment

We use one digital camera and one PC(Celeron 1.3 GHz) in this experiment. The camera used here is an IEEE1394-based digital camera, Sony DFW-VL500.

4.1 Normalization of Images

First, background subtraction is applied to an RGB color image(160×120), which is taken with the camera. Then, the color image is converted to a HSV color image. Second, a skin color region is detected based on H(hue). In order to make the system robust against a parallel translation, an apparent difference of size and a lightning environment, we move the hand region to the center of an image, and scale it linearly to a 100×100 sized image, and normalize the contrast. Finally, we can get a normalized hand image (100×100), shown in Fig.1(a).

4.2 Experiment1: Matching with the Eigen Coefficients

In order to make our system person-independent, we took rotated hand images of 10 persons (5 men and 5 women, whose ages were from 20 years old to 30 years old). The number of template images were 400 per one person, and totally 4000 images were taken. The number of target images for recognition test were 6000 per one person , i.e., 300 images for each hand model. In total, 60000 images were taken, which were classified into the 20 hand models according to human eyes beforehand (See Fig.9). In this experiment, hand images of the 10 persons, from who training images were acquired. When the number of eigen bases was 40, which provides good performance, the recognition rate was 83.7% on the average.

The Dimension of Eigenspace	40
Experiment Subjects	10(5 men,5 women)
Templates	4000
Test Data	60000
Average Recognition Rate	83.7

Table 1: Experiment1

Fig.5 shows the change of the recognition rate and the recognition time when the dimensionality of the eigenspace is changed. As for the recognition rate, it goes up to 40, but when the dimensionality exceeds 40, it is almost saturated. As for the recognition time, it monotonously increases in proportion to the dimensionality of the eigenspace. In the eigenspace of 40 dimensions, the average recognition time using 4000 templates is 54[msec](18fps), which shows our system can recognize hand shapes in real time. The recognition time consists of the time of image normalization, eigen coefficients calculation and searching for the nearest hand model. The time required at each step when an eigenspace of 40 dimensions is used is shown in Table 2.

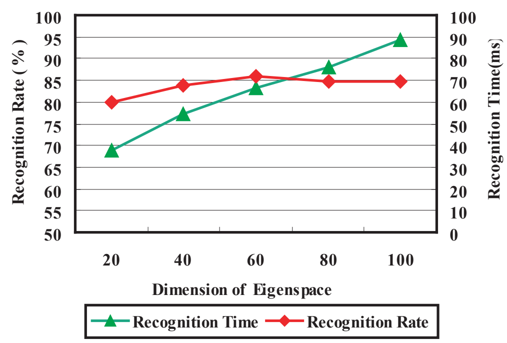


Fig. 5: The Dimension of Eigenspace and the Recognition Rate and the Recognition Time

Normalization	25ms
Eigen Coefficients Calculation	20ms
Search	9ms
Total	54ms

Table 2: The Recognition Time with an Eigenspace of 40 Dimensions

4.3 Experiment2: Additional Learning

In the proposed online additional learning system, eigen bases are not calculated again. We assume that any 20 hand model images may be projected similarly into the eigenspace based on the previously calculated eigen bases, or ones calculated from 4000 rotated hand images. In order to check whether the assumption is appropriate or not, in view of the recognition rate, we compared the case when we increased the number of templates up to 4500 (basic rotated hand templates were 4000, and added templates were 500) and the case when we recalculated the eigen bases from the same 4500 images. Here, the dimensionality of eigenspace was 40, and data for recognition test and experiment subject were the same in Experiment1. The result is shown in Fig.6, where we make a comparison between the two cases of each experiment subject. Judging from the result, the recognition rates were almost the same (84% on the average), and we find our assumption is almost appropriate. In addition, the recognition time was almost the same, too.

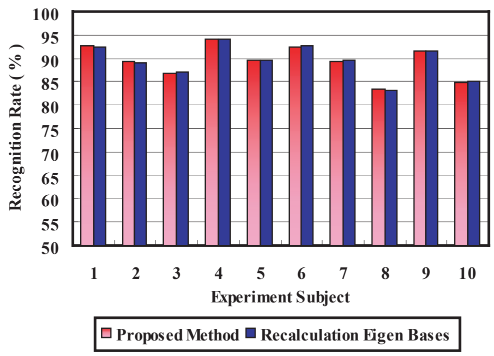


Fig. 6: Comparison Proposed Method with Recalculation Eigen Basis

Then, we investigated the recognition rate after learning enough templates in the eigenspace about the same 10 experiment subjects. Here, the dimensionality of eigenspace was 40, and data for recognition test were the same in Experiment1 ,but data for additional learning were different from that for the recognition test. The recognition rates about each experiment subject before (same as Experiment1) and after the learning are shown in Fig.7. We have achieved about 95% recognition rate on the average by additional learning. The number of templates was about 6500(the basic rotated hand templates were 4000, and the learned templates by using our system were 2500).

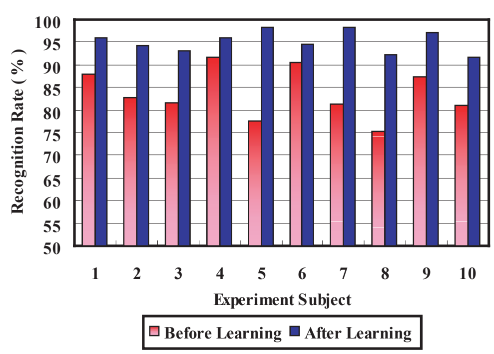


Fig. 7: Recognition Rate Before or After Learning

4.4 Experiment3: Person Independency

In Experiment1 and 2, the recognition objects were hands of the 10 experiment subjects whose hand images were included in the templates, which means the system is person dependent. In this experiment, we examined whether the system could recognize hands of non-registered persons whose hand images were not included in the templates. We took rotated hand images of other 6 persons (4 men and 2 women, whose ages were from 20 to 30 years old). Image data for recognition test were 6000 per one person as Experiment1. Here, again, the dimensionality of the eigenspace was 40, and the same templates as Experiment2 (after learning) were used. The result was shown in Fig.8. Even for non-registered persons, we had also achieved high recognition rate, too (93% on the average), ant this result indicate that our system has a person independent feature.

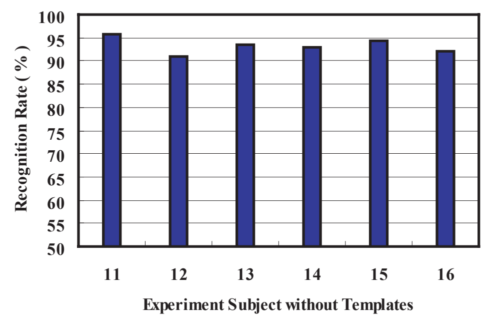


Fig. 8: Experiment Subject without Templates

5 Conclusion

In this paper, we proposed real-time hand shape recognition system for a human interface and have shown its good performance through some experiments. In order to make the system robust against the change of its appearance, we took a lot of rotated hand images of multiple people and used them as templates for recognition. To shorten the processing time and to extract essential information in the templates, we represented the templates in the eigen-subspace. We have achieved real-time recognition (18fps by 1.3 GHz Celeron CPU). In addition, we have achieved high recognition rate by online additional learning in the eigenspace (93% or more). Here, new templates are added which are represented in the initially established eigenspace, which makes online incremental learning possible. For future works, we should analyze the characteristics of the additional learning more precisely, and we should improve the nearest search method to achieve better performance. Improvement of segmenting a hand region from a cluttered background is also important to make the system more robust.

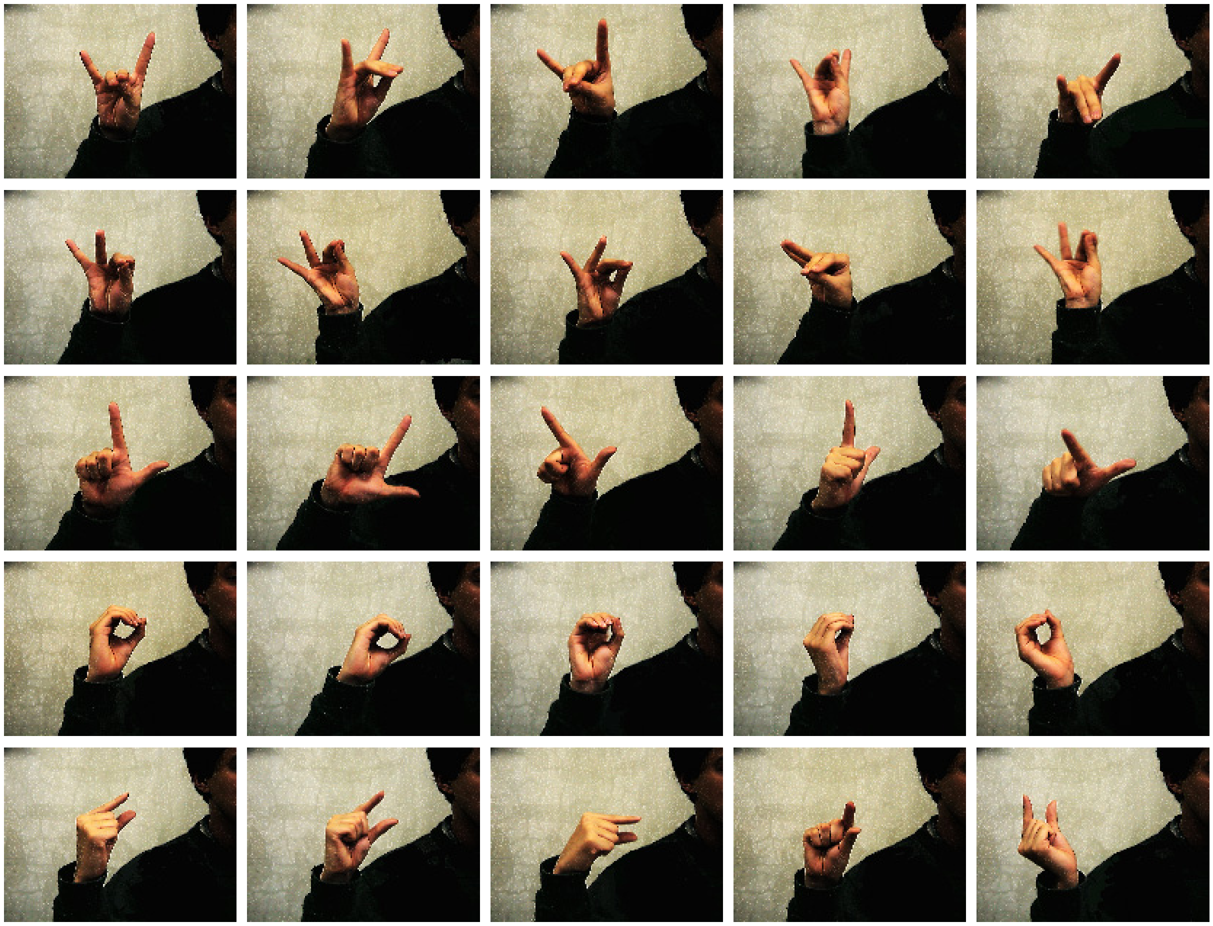


Fig. 9: Samples of Data for Recognition Test

References

[1] R.Taniguchi and S.Yonemoto and D.Arita and R.Hoshino. Real-time Human Motion Analysis for Human Machine Interface. Proc.of Working Conference of Advanced Visual Interfaces, pp.195-202, May 2002.

[2] R.Hoshino and D.Arita and S.Yonemoto and R.Taniguchi. Real-time Human Motion Analysis based on Silhouette Contour and Color Blob Analysis. Proc.Second International Workshop on Articulated Motion and Deformable Objects, pp.92-103, Nov.2002.

[3] H.Murase and S.K.Nayar. Visual learning and recognition of 3-d objects from appearance. International Journal of Computer Vision, Vol.14, No.1, pp.5-24, 1995.

[4] S.A.Nene and S.K.Nayar. A Simple Algorithm for Nearest Neighbor Search in High Dimensions. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 17, No.9, pp.989-1003, Sep.1997.