PAPER

Hand-Shape Recognition Using the Distributions of Multi-Viewpoint Image Sets

Yasuhiro OHKAWA[†a)], ***Student Member and*** **Kazuhiro FUKUI**[†b)], ***Member***

SUMMARY This paper proposes a method for recognizing hand-shapes by using multi-viewpoint image sets. The recognition of a hand-shape is a difficult problem, as appearance of the hand changes largely depending on viewpoint, illumination conditions and individual characteristics. To overcome this problem, we apply the Kernel Orthogonal Mutual Subspace Method (KOMSM) to shift-invariance features obtained from multi-viewpoint images of a hand. When applying KOMSM to hand recognition with a lot of learning images from each class, it is necessary to consider how to run the KOMSM with heavy computational cost due to the kernel trick technique. We propose a new method that can drastically reduce the computational cost of KOMSM by adopting centroids and the number of images belonging to the centroids, which are obtained by using k-means clustering. The validity of the proposed method is demonstrated through evaluation experiments using multi-viewpoint image sets of 30 classes of hand-shapes.

key words: *hand-shape recognition, multi-viewpoint, kernel orthogonal mutual subspace method*

1. Introduction

In this paper, we propose a novel hand-shape recognition method that uses sets of hand-image patterns inputted through a multi-camera system.

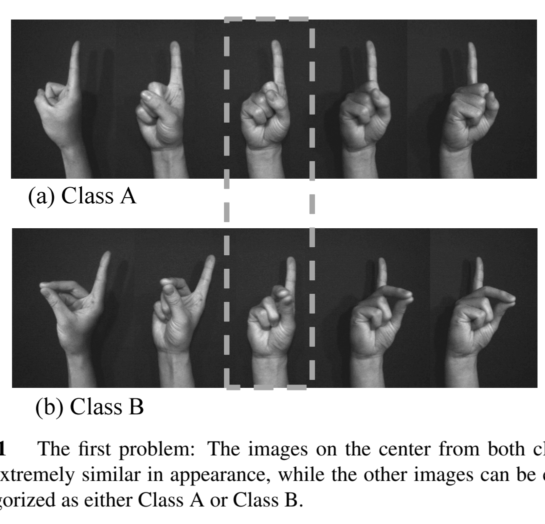
Hand gestures are often used in our daily life to achieve smooth communication with other people. Therefore, it is expected that hand gestures can also be utilized for achieving more natural interaction between humans and computer systems. The recognition of the 3D shape of a hand is the most elementary technique for recognizing hand gestures automatically. Many types of hand-shape recognition methods have been proposed. They can be divided into two categories: model-based methods and appearance-based methods [1], [2].

Model-based methods use a 3D hand model for recognition [3]–[5]. They extract feature points such as edges and corners from a hand-image and match them to a 3D hand model. For example, Imai has proposed a method for estimating 3D hand postures by matching the edges extracted from a hand-image to the silhouette generated from a typical hand model [5]. Although model-based methods are widely used in various trial systems, they often suffer from instability of matching and high computational complexity since a hand is a complex 3D object with 20 degrees of freedom [1].

Appearance-based methods [6]–[9] classify a hand-shape from its appearance, where an $n\times n$-pixel pattern is treated as a vector $\mathbf{x}$ in n^2-dimensional space. Such models can account for variations in appearance due to changes in viewpoint, illumination and differences between individual characteristics by preparing a statistical model representing these variations. There are two problems in appearance-based hand-shape recognition.

The first problem is that the appearance of a hand-shape from one class can be highly similar to that of other classes depending on viewpoint, as shown in Fig. 1. This often leads to false identification in a method using a single-viewpoint image. The problem can be partially avoided by choosing an appropriate camera position and selecting image features. However, the problem may not be solvable with the above techniques since it becomes highly complex when the number of classes is large.

When recognizing complex 3D objects, humans examine them from various viewpoints and make a comprehensive decision. This fact implies that one method for solving the above problem is to measure the similarity of the hand-shapes by using multi-viewpoint image sets. As methods that can effectively classify multiple sets of images, such as multi-viewpoint image sets and image sequences, in this paper, we focus on MSM (Mutual Subspace Method) [10], CMSM (Constrained MSM) [11] and OMSM (Orthogonal MSM) [12]. A conceptual diagram of these methods is shown in Fig. 2. In these methods, a set of images is repre-



(a) Class A

(b) Class B

Fig. 1 The first problem: The images on the center from both classes are extremely similar in appearance, while the other images can be easily categorized as either Class A or Class B.

Manuscript received August 8, 2011.
Manuscript revised January 6, 2012.

†The authors are with the Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba-shi, 305–8577 Japan.
a) E-mail: ohkawa@cvlab.cs.tsukuba.ac.jp
b) E-mail: kfukui@cs.tsukuba.ac.jp

DOI: 10.1587/transinf.E95.D.1619

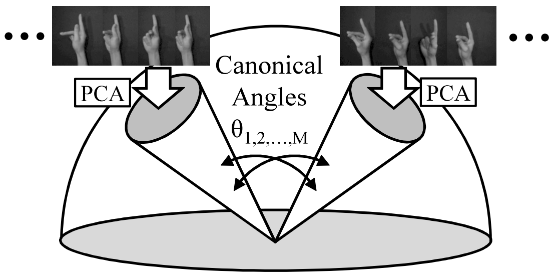


Fig. 2 Conceptual diagram of the subspace-based method. The distributions of multi-viewpoint image sets of hands are represented by linear-subspaces, which are generated by PCA. The canonical angles between two subspaces are taken as a measure of the similarity between the distributions.

(a) Differences of appearance depending on a viewpoint.

(b) Differences of appearance due to an individual characteristics.

Fig. 3 The second problem: Images of hand-shapes. Although appearance differs greatly, these represent the same shape altogether.

sented by a linear subspace which is generated by applying PCA (Principal Component Analysis). The canonical angles between different image sets are used to calculate their similarity. The classification of the set of input images is based on this similarity. The effectiveness of these methods has been demonstrated in various applications, such as frontal face recognition [10]–[12] and apple identification [13].

The second problem is that the appearance of a hand changes considerably depending on the viewpoint, as shown in Fig. 3 (a). In addition, changes in illumination conditions and individual characteristics (Fig. 3 (b)) induce further changes in appearance. Such large variation produces strong nonlinearity. To solve this problem, MSM, including its extensions (CMSM and OMSM), has been extended to nonlinear kernel methods [14], namely KMSM (Kernel MSM) [15], KCMSM (Kernel CMSM) [16] and KOMSM (Kernel OMSM) [17], respectively. In this paper, we apply KOMSM to hand-shape recognition, considering its powerful classification capabilities and compactness of the algorithm which only requires few a parameters to be adjusted.

However, as mentioned previously, the changes in the appearance of a hand are notably greater than those of a frontal image of a face. Therefore, to achieve high recognition performance in hand-shape recognition, KOMSM requires more learning patterns for each hand-shape class as compared with frontal face recognition. The need for large learning pattern sets leads to the following serious problem. The problem is the difficulty in running the algorithm of KOMSM in practical processing time limits performance since the processing time of KOMSM depends on the number of learning patterns due to the use of the kernel trick. In practical terms, if the number of learning patterns is large, the learning process would be impossible to conduct due to limitations imposed by memory size for computing the eigenvalues of the resulting large matrix.

To solve the above problem, we introduce a novel method to reduce the number of learning patterns. The proposed reduction method approximates the distribution of a set of learning image patterns with the centroids of clusters and the number of the patterns belonging to these clusters, which are obtained by applying k-means clustering to the set of the learning image patterns. Conventional reduction methods [18] use only the centroids of clusters. In contrast, by using the number of patterns retaining information on the original distribution of learning patterns, the proposed method can generate a class subspace that approximates the original distribution with higher accuracy. As a result, we can achieve the maximum performance of KOMSM which can be obtained in case all of the learning patterns are used, while reducing the computational complexity of KOMSM.

The contributions of this paper are as follows.

(1) We propose a framework for a hand-shape recognition method based on KOMSM with multi-viewpoint images, and demonstrate its effectiveness through experiments with a large-scale database.
To implement the above framework,
(2) we propose a method for reducing the computational complexity of KOMSM on the basis of the information about clusters obtained by using k-means clustering.

The rest of the paper is organized as follows. In Sect. 2, we discuss the effectiveness of using multiple viewpoints and review the algorithm of KOMSM. Section 3 presents the proposed complexity reduction of KOMSM. Section 4 describes the process flow of the proposed method. In Sect. 5, the validity of the proposed method is demonstrated through experiments using 30 classes of images of hand-shapes captured form 40 subjects. The final section concludes the paper.

2. Hand-Shape Recognition Based on KOMSM with Sets of Multi-Viewpoint Images

First, we discuss the effectiveness of using sets of multi-viewpoint images and examine what types of similarity is suitable for hand-shape recognition by considering example image sets. Then, we explain the algorithm of KOMSM.

2.1 Similarity between Sets of Images

Figure 4 (a) shows the relation between single-viewpoint images from four classes on the basis of multi-dimensional scaling [19]. It is difficult to distinguish the four classes by considering the relation between single-viewpoint images,

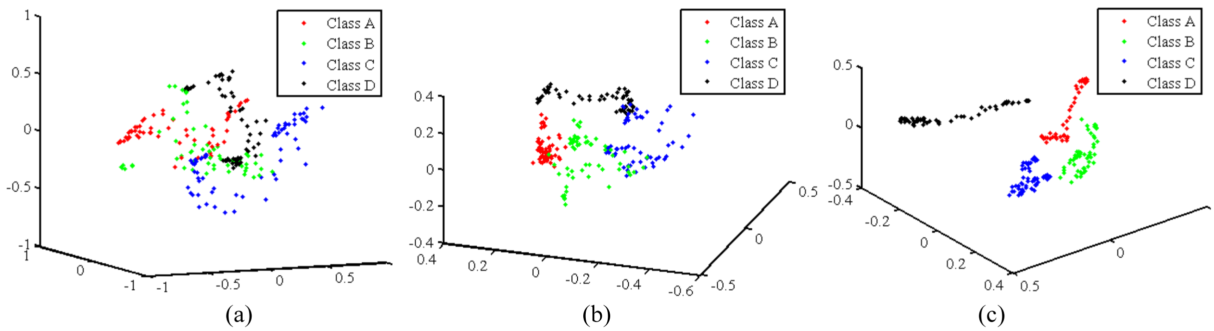


Fig. 4 (a) The relation between single-viewpoint image. (b) The relation between subspaces based on the first canonical angle, where each subspace is generated from multi-viewpoint image sets. (c) The relation between subspaces based on all canonical angles, where each subspace is generated from multi-viewpoint image sets.

as shown in Fig. 4 (a).

In the framework of the proposed method, the technique for measuring the similarity between distributions is an important issue. We define the similarity between two distributions by using canonical angles [20], which determine strictly the structural relationship between two subspaces in high-dimensional vector space.

M canonical angles ($0 \leq \theta_1 \leq \ldots \leq \theta_M \leq \frac{\pi}{2}$) between M-dimensional subspace $\mathcal{U}$ and N-dimensional subspace $\mathcal{V}$ ($M \leq N$) are defined. The i-th canonical angle θ_i is defined as follows:

$$\cos\theta_i = \max_{\boldsymbol{u}_i \in \mathcal{U}} \max_{\boldsymbol{v}_i \in \mathcal{V}} \boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{v}_i \tag{1}$$

subject to $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_i = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_i = 1, \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j = 0, i \neq j$.

Various similarities between subspaces based on canonical angles have been proposed. Two of the most widely used similarities are as follows. The first one is $Sim_1 = \cos^2\theta_1$, which uses only the first canonical angle. The most similar appearance between 3D objects reflects a value into Sim_1. The second one is $Sim_{all} = \frac{1}{M}\sum_i^M \cos^2\theta_i$ which uses all canonical angles. Sim_{all} is based on the overall structural resemblance between two distributions. Recognition using the appearance from various viewpoints is conducted more comprehensively, similarly to the recognition approach naturally adopted by humans. Thus, we adopt Sim_{all} as the similarity between two distributions.

Figure 4 (b) and (c) show the relation between sets of multi-viewpoint images of four classes of hand-shapes. In case (a), each point corresponds to one image. However, in cases (b) and (c), each point contains seven-viewpoint images, which are represented by nonlinear subspace generated from them by using kernel PCA [21]. Figure 4 (b) and (c) represents calculations based on Sim_1 and Sim_{all}, respectively. We can see from the figures that by using a multi-viewpoint image set, the classes can be distinguished more accurately, and Sim_{all} achieves higher performance than Sim_1. The recognition method using these similarities is called Kernel Mutual Subspace Method (KMSM).

2.2 Algorithm of KOMSM

The framework of KMSM mentioned in the previous section is able to compute the similarity between two nonlinear distributions. However, high recognition performance can not be achieved since KMSM does not consider the relations between classes. In KOMSM [17], nonlinear subspaces are orthogonalized by using the framework of the transform proposed by Fukunaga and Koontz [22] before measuring the canonical angles between them. This transform emphasizes the differences between classes and significantly improves the recognition performance of KMSM.

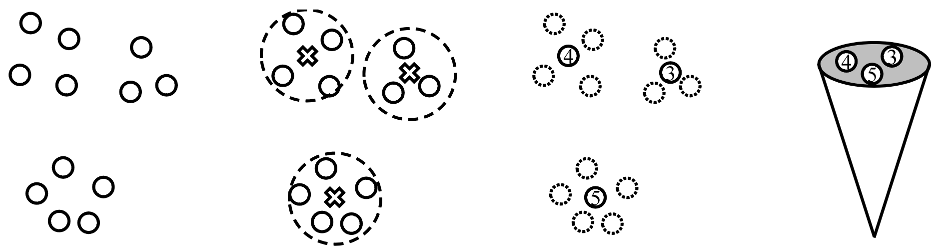
The whitening transform $\mathbf{W}$ for orthogonalization is calculated as follows: $\mathbf{W} = \mathbf{C}^{-1/2}\mathbf{B}^{\mathrm{T}}$, where $\mathbf{C}$ is a diagonal matrix with the i-th highest eigenvalue of the matrix $\mathbf{P}$ as the i-th diagonal component, and $\mathbf{B}$ is a matrix whose i-th column vector is the eigenvector of the matrix $\mathbf{P}$ corresponding to the i-th highest eigenvalue. Here, the matrix $\mathbf{P} = \sum_{i=1}^{C}\mathbf{P}_i$ is the sum of the projection matrices corresponding to the projection onto class i subspace. For details regarding KOMSM, please refer to [17].

3. Computational Cost Reduction of KOMSM

In hand-shape recognition, a large number of learning patterns is required to achieve highly accurate recognition. However, since KOMSM uses KPCA, the computational complexity depends on the number of learning patterns. To solve this problem, various computational cost reduction methods for KPCA have been proposed [18], [23]–[25]. In [18], eight types of reduction methods are applied to KMSM. According to the results, the k-means based method [18] outperforms the other seven reduction methods.

In the method in [18], all N patterns of each class are replaced with K centroids extracted by k-means clustering. The patterns belonging to a centroid are represented by the centroid. As a result, the computational complexity is greatly reduced by applying KPCA to only K centroids ($K \leq N$).

However, although the method is capable of reducing



(a) Patterns in input space (b) Clustering result (c) Patterns are moved to belonging centroids (d) KPCA is applied to moved patterns

Fig. 5 The proposed complexity reduction method of KPCA.

the computational complexity, there is a disadvantage which must be eliminated. The problem is that the information about the number of the patterns belonging to each centroid is discarded, even though it is important for approximating the pattern distribution.

In this paper, we propose a novel reduction method considering the number of patterns belonging to a centroid. Figure 5 shows a conceptual diagram of the method. The key idea of the proposed method is to move all patterns belonging to a centroid into the position of the centroid. KPCA is applied to all of the moved patterns, as shown in Fig. 5 (c) and (d). The size of a Gram matrix to be computed in the proposed method is $K \times K$, which is the same as that of the method only using K centroids.

The concrete flow of the proposed method is as follows: First, N patterns are partitioned into K clusters ($K \leq N$) by using k-means clustering in input space (Fig. 5 (b)). Next, patterns are moved to their corresponding centroids (Fig. 5 (c)). Kernel PCA is applied to these N moved patterns (Fig. 5 (d)). $N(k)$ is the number of patterns that belong to the k-th centroid $\boldsymbol{x}_k$. $\boldsymbol{f}_k$ is a nonlinear map of $\boldsymbol{x}_k$ and $\mathbf{R}_{\mathcal{F}}$ is an autocorrelation matrix in feature space $\mathcal{F}$.

$$\mathbf{R}_{\mathcal{F}} = \sum_{k=1}^{K} \sum_{i=1}^{N(k)} \boldsymbol{f}_k \boldsymbol{f}_k^{\mathrm{T}} \tag{2}$$

$$= \sum_{k=1}^{K} N(k) \boldsymbol{f}_k \boldsymbol{f}_k^{\mathrm{T}} \tag{3}$$

$$= \sum_{k=1}^{K} \hat{\boldsymbol{f}}_k \hat{\boldsymbol{f}}_k^{\mathrm{T}}, \tag{4}$$

where $\hat{\boldsymbol{f}}_k = \sqrt{N(k)} \boldsymbol{f}_k$. The kernel Gram matrix $\mathbf{G}$ corresponding to $\mathbf{R}_{\mathcal{F}}$ is as follows:

$$\mathbf{G} = \{g_{k,l} | k, l = 1, \ldots, K\} \tag{5}$$

$$\begin{aligned} g_{kl} &= \hat{\boldsymbol{f}}_k \cdot \hat{\boldsymbol{f}}_l \\ &= (\sqrt{N(k)} \boldsymbol{f}_k) \cdot (\sqrt{N(l)} \boldsymbol{f}_l) \\ &= \sqrt{N(k)} \sqrt{N(l)} \mathrm{k}(\boldsymbol{x}_k, \boldsymbol{x}_l), \end{aligned} \tag{6}$$

where $\mathrm{k}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a kernel function. The s-th principal component vector $\boldsymbol{v}_s$ is calculated as follows:

$$\boldsymbol{v}_s = \frac{1}{\sqrt{\lambda_s}} \sum_{k=1}^{K} \alpha_{sk} \hat{\boldsymbol{f}}_k \tag{7}$$

$$= \frac{1}{\sqrt{\lambda_s}} \sum_{k=1}^{K} \alpha_{sk} \sqrt{N(k)} \boldsymbol{f}_k, \tag{8}$$

where λ_s is the s-th highest eigenvalue of the matrix $\mathbf{G}$ and α_{sk} is the value of the k-th element of the eigenvector of the matrix $\mathbf{G}$ corresponding to the s-th highest eigenvalue. In the proposed method, the basis vectors of a nonlinear subspace of a class are represented by a weighted linear combination of mapped K-centroids, where the weight is the square root of the number of patterns belonging to the respective centroid.

The term of $\alpha_{sk} \sqrt{N(k)}$ in Eq. (8) can be calculated in advance. Therefore, the computational complexity of the proposed method and the conventional method [18] are equal in recognition phase. In learning phase, the computational complexity of the proposed method increases slightly to calculate the weighted kernel Gram matrix $\mathbf{G}$.

4. Process Flow of Proposed Hand-Shape Recognition

The flow of the proposed method consists of a Learning Phase and a Recognition Phase, as shown in Fig. 6.

Learning phase

1. Hand-shape images of each class are collected from various viewpoints for learning with a multi-camera system.
2. Image features are extracted from the learning images.
3. k-means clustering is performed on the learning patterns of each class.
4. A kernel whitening matrix is calculated by using the proposed complexity reduction method.
5. The learning patterns are transformed by the kernel whitening matrix.
6. The reference subspace of each class is calculated by applying PCA to the transformed patterns.

Recognition phase

1. Multi-viewpoint hand-shape image sets are collected with a multi-camera system.
2. Image features are extracted from the input images in a manner similar to the learning phase.
3. The input patterns are transformed by the kernel whitening matrix.

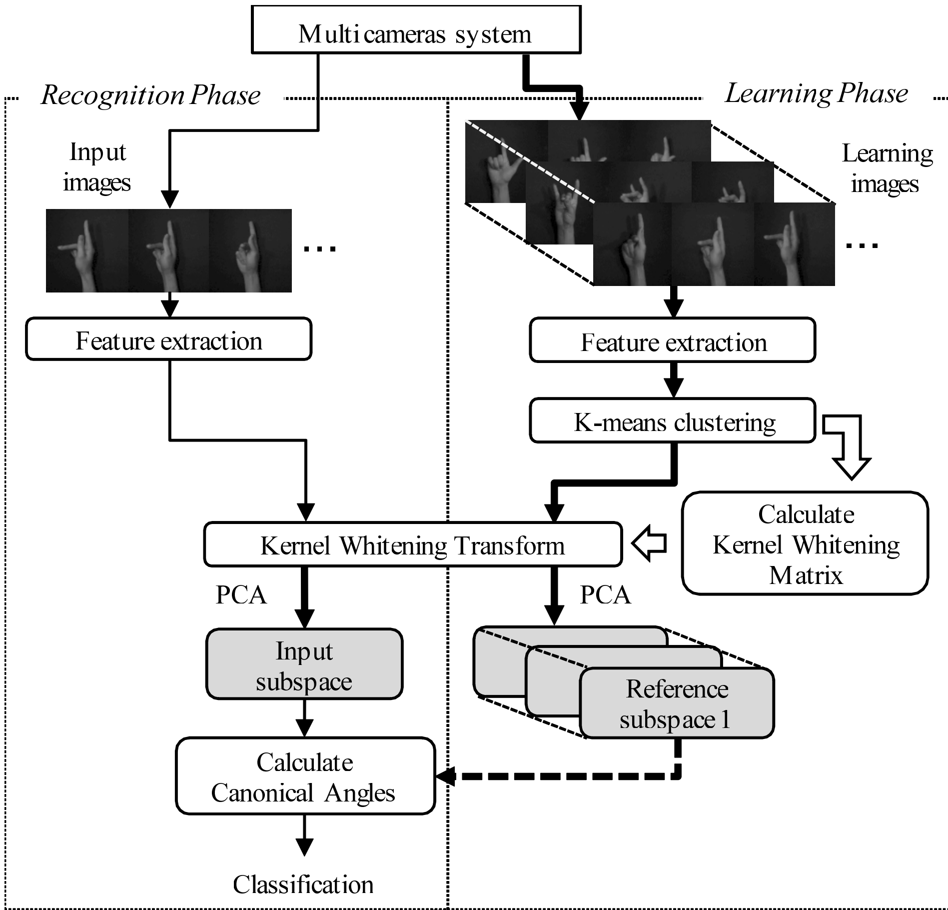


Fig. 6 Flow of the proposed method consists of a Learning Phase and a Recognition Phase.

4. An input subspace is calculated by applying PCA to the transformed input patterns.
5. The canonical angles between the input subspace and the reference subspace are calculated. This process is executed for all classes.
6. The similarity with each class is calculated by using the canonical angles. The set of input images is classified into the class with the highest similarity.

5. Experiments and Considerations

5.1 Experimental Settings

We constructed a multi-camera system which consists of seven IEEE1394-b cameras connected to a Host PC with 3.0 GHz quad-core Xeon and 16 GB RAM, as shown in Fig. 7, to collect evaluation images. The system was designed to capture hand-shape images from various viewpoints with a same distance. The angle between adjacent cameras was 10 degrees, and the distance between camera D and the hand to be recognized was about 36 cm. The images as shown in Fig. 3 (a) were synchronously captured with the seven cameras at a speed of 15 fps.

Gray-scale images with a size of 400×400 pixels were collected from 40 subjects (Male: 29, Female: 11) and categorized into 30 classes, as shown in Fig. 8. To increase the diversity of viewpoints, we asked subjects to rotate their

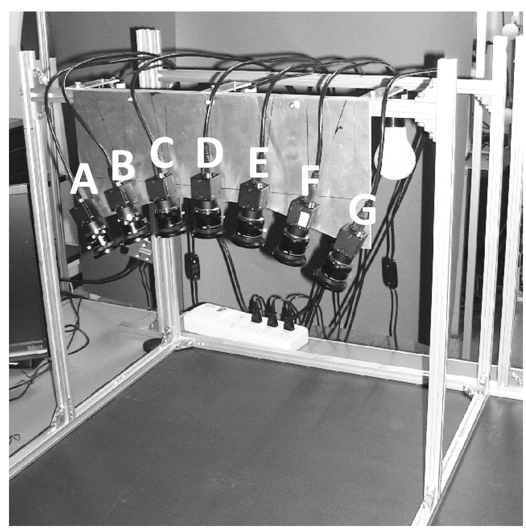


Fig. 7 Multi-camera system for evaluation. The angle between two adjoining cameras is 10 degrees, and the distance between the cameras and the hand is about 36 cm.

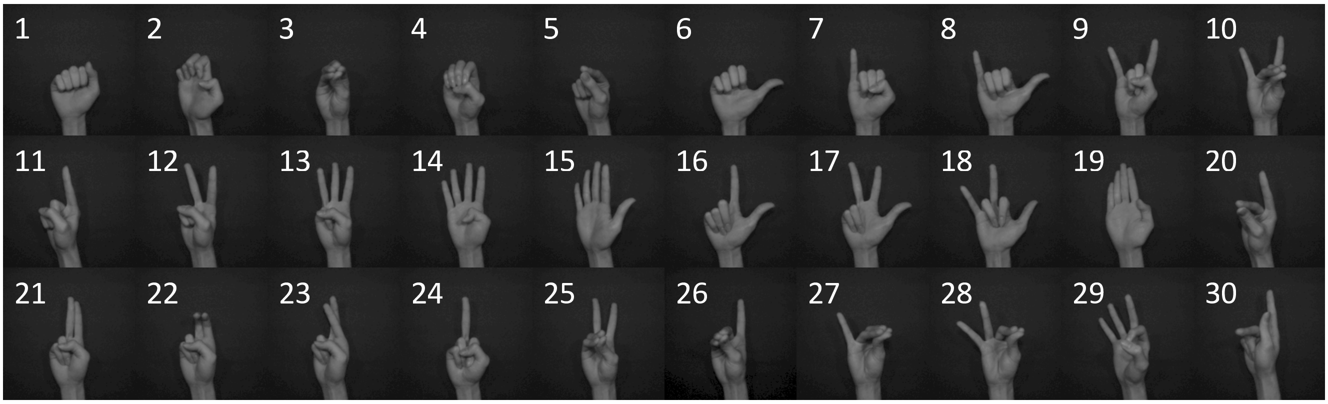


Fig. 8 The 30 classes of hand-shape images for evaluation. The number in the upper left corner indicates class number.

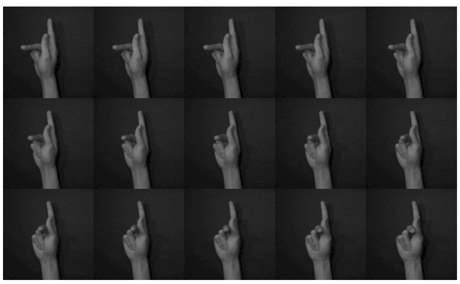


Fig. 9 Sequential input images captured with camera D (Fig. 7) at a speed of 15 fps.

hand at a constant speed in order to obtain image sequences as shown in Fig. 9. As a result, the total number of collected hand-shape images was 504, 000 (= 60-images×7-cameras×30-classes×40-subjects). The size of the images was reduced from 400 × 400 to 64 × 64 pixels. HLAC (Higher-order Local Auto-Correlation) [26] is one of the well-known shift-invariance features. Using HLAC enables us to greatly reduce the number of learning patterns since two patterns of the same hand-shape appearing at different positions in an image are represented by the same HLAC features. We extracted 35-dimensional HLAC features from seven-level pyramid structures of the input images after applying the Roberts edge extraction filter, and we joined all features into a single 245-dimensional HLAC feature.

In all experiments, the Gaussian kernel

$$\mathrm{k}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp\left(-\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}{2\sigma^2}\right) \tag{9}$$

was used as a kernel function, and the parameter $2\sigma^2$ of the Gaussian kernel was set to 7.0. The dimensionality of the reference subspaces was fixed to 145, and the dimensionality of the input subspace was varied between 1 and 10.

We conducted three types of evaluation experiments to confirm the validity of the proposed method. In experiment-I, we evaluated the effectiveness of using multi-viewpoint image sets by comparing it with a method using a single image. In experiment-II, we compared the proposed method for reducing the computational cost of KOMSM with several other reduction methods. In the final experiment-III, we evaluated the overall performance of the proposed framework.

Table 1 Recognition rate against the number of viewpoints. The letters (A~G) in the first column indicate the indexes of the cameras used for recognition as shown in Fig. 7.

	Number of sequential images			
Number of cameras	1	5	10	15
1 (D)	85.8	86.5	88.3	89.9
3 (CDE)	88.7	89.8	91.0	92.0
5 (BCDEF)	91.7	92.1	92.8	93.6
7 (ABCEFDG)	94.2	94.8	95.2	95.3

When the total number of learning patterns was larger than 100,000, KOMSM without the reduction could not conduct its learning process, since the size of the kernel Gram matrix for generating each class subspace became too large. Therefore, in experiment-I and experiment-II, we only used 10 classes selected from class 1 to class 10 in Fig. 8. The images obtained from odd numbered subjects were used as learning patterns, and the remaining images were used as testing patterns. All the experiments were implemented by using MATLAB.

5.2 Experiment I: Effectiveness of Using Multiple Viewpoints

In this experiment, we evaluated the effectiveness of using multi-viewpoint image sets while changing the respective numbers of cameras used and sequential input images inputted from each camera. The number of sequential input images was changed as follows: 1,5,10,15.

Table 1 summarizes the experimental results. From this table, we can see that the performance increased as the number of the cameras and input patterns increased. In particular, the increase of the number of cameras contributed towards a significant improvement of the performance. This result indicates clearly that images taken from various viewpoints can boost the performance of hand-shape recognition.

Table 2 Comparative experiment of complexity reduction method.

	KOMSM			KMSM		
	Proposed	*k*-means based [18]	Greedy KPCA [25]	Proposed	*k*-means based [18]	Greedy KPCA [25]
200	95.0	94.9	94.3	92.9	92.8	92.4
400	95.0	95.0	94.8	93.6	93.4	93.3
800	95.2	94.8	94.9	93.7	93.5	93.5
1600	95.2	94.6	-	93.7	93.6	-
3200	95.3	94.7	-	93.7	93.7	-
ALL (8400)	95.3			93.6		

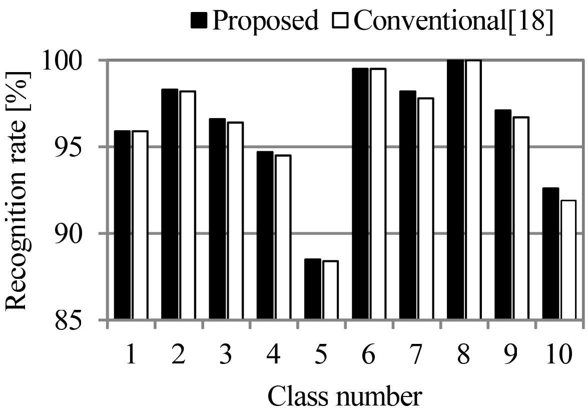


Fig. 10 Recognition rate for each class.

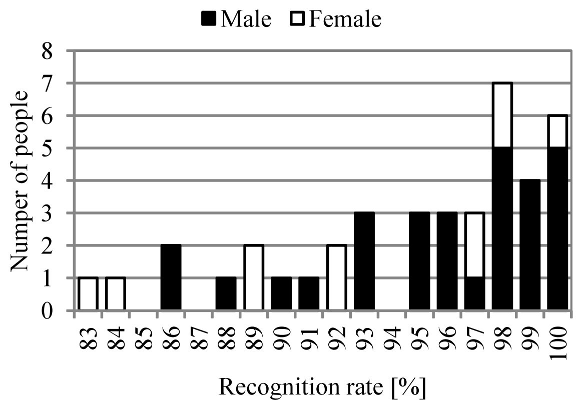


Fig. 11 Histogram of recognition rate.

5.3 Experiment II: Reduction of the Computational Cost of KOMSM

In this experiment, we compared the performance of the proposed reduction method with the *k*-means based method [18] and Greedy KPCA [25], [27]. We used the Greedy KPCA algorithm provided with the Statistical Pattern Recognition Toolbox [25], [27].

We evaluated the performance of all of the above methods while changing the number of patterns to be selected for each class as follows: 200, 400, 800, 1600 and 3200. The numbers of cameras used for input and sequential input images were set to 7 and 15, respectively. Since *k*-means clustering depends on the initial conditions, we conducted this evaluation 15 times, and the average of all obtained recognition rates was used as the final rate.

Table 2 shows the experimental results. In the case of using over 1600 patterns, the rate column is blank since the process of Greedy KPCA had not completed even after 10 days from the starting time. The experimental results show that the proposed method outperforms the conventional methods in the case of both KOMSM and KMSM.

Figure 10 shows the recognition rates for each class when using KOMSM (800 centroids) with the proposed method and the conventional *k*-means based method [18]. The results were obtained by applying a Leave-one-subject-out Cross-Validation. Since the performance of *k*-means clustering depends on its initial values which are determined randomly, we conducted the evaluation 10 times at each subject to use the average value of the 10 recognition rates for the evaluation. The performance of the proposed method equals or even exceeds that of the conventional method for all the classes. From the results, we can see that the proposed method which considers the numbers of patterns belonging to the centroids is working effectively.

5.4 Experiment III: Performance of the Proposed Framework

In this experiment, we evaluated the recognition performance of the proposed method by applying Leave-one-subject-out Cross-Validation. The hand-images collected from 39 subjects were used as learning patterns, and those from the remaining one subject were used as testing patterns. The evaluation was conducted 40 times. All 30 classes of hand-shapes shown in Fig. 8 were classified.

In the proposed reduction method, a set of 16380 learning patterns from each class was compressed to 800 centroids by *k*-means. Since the *k*-means clustering depends on the initial conditions, we conducted the evaluation 5 times. The average of all obtained performance results was used as the final performance score. The number of cameras used for input and the number of sequential input images were set to 7 and 15, respectively.

Figure 11 shows the experimental results. The average recognition rate for all subjects was 95.3% (Male: 96.1%, Female: 93.2%). The average recognition rate for female subjects was lower than that for male subjects. This may be due to the fact that the number of female subjects was less than that of male subjects. It should be noted that the recognition rate was 100% for 6 out of the 40 subjects, even though the set included female subject. From this result, we could improve the performance of the proposed method

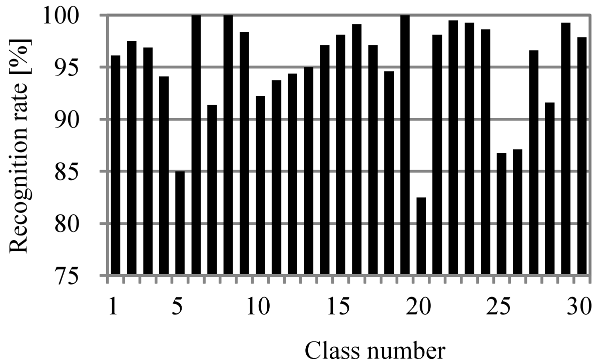


Fig. 12 Recognition rate for each class.

even further by collecting learning images from more subjects.

Figure 12 shows the recognition rate for each class. It should be noted that 3 out of 30 classes were classified at 100%. Although most of the recognition rates are acceptable, several classes must be improved. For example, the recognition rate of 82.5% for class 20 was not sufficiently high as compared with other classes since the shape of the images in class 20 is highly similar to that in class 26, as can be seen in Fig. 8, and inevitably class 20 was often misclassified as class 26. To avoid the misclassification, it is necessary to consider how to collect more learning patterns. In addition, more effective features should be considered as well.

The computational time of the recognition phase was 0.57 seconds. The process speed is sufficiently high to construct a real-time system of hand-shape recognition.

6. Conclusion

In this paper, we proposed the framework for recognizing hand-shapes by using multi-viewpoint image sets on the basis of KOMSM. To implement the above framework, we proposed a method for reducing the computational complexity of KOMSM based on information about clusters obtained by using *k*-means clustering. The experimental results demonstrated that the proposed method is able to recognize 30 classes of complex hand-shapes effectively at a speed suitable for practical application.

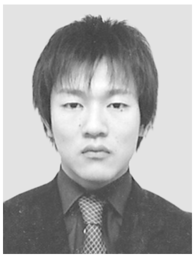
Acknowledgment

This work was supported by KAKENHI (22300195).

References

[1] A. Erol, G. Bebis, M. Nicolescu, R. Boyle, and X. Twombly, “Vision-based hand pose estimation: A review,” Comput. Vis. Image Understand., vol.108, no.1-2, pp.52–73, 2007.
[2] S. Ong and S. Ranganath, “Automatic sign language analysis: A survey and the future beyond lexical meaning,” IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.6, pp.873–891, 2005.
[3] B. Stenger, P. Mendonça, and R. Cipolla, “Model-based hand tracking using an unscented Kalman filter,” British Machine Vision Conference, vol.1, pp.63–72, 2001.
[4] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla, “Model-based hand tracking using a hierarchical Bayesian filter,” IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.9, pp.1372–1384, 2006.
[5] A. Imai, N. Shimada, and Y. Shirai, “Hand posture estimation in complex backgrounds by considering mis-match of model,” Asian Conference on Computer Vision, pp.596–607, 2007.
[6] J. Martin and J. Crowley, “An appearance-based approach to gesture-recognition,” Image Analysis and Processing, pp.340–347, 1997.
[7] H. Birk, T. Moeslund, and C. Madsen, “Real-time recognition of hand alphabet gestures using principal component analysis,” Scandinavian Conference on Image Analysis, vol.1, pp.261–268, 1997.
[8] Y. Cui and J. Weng, “Appearance-based hand sign recognition from intensity image sequences,” Comput. Vis. Image Understand., vol.78, no.2, pp.157–176, 2000.
[9] K. Yasumoto and T. Kurita, “Improvement of generalization ability of kernel-based fisher discriminant analysis for recognition of japanese sign language hand postures, “yubi-moji”, using k-means method,” IAPR Workshop on Machine Vision Applications, pp.269–272, 2002.
[10] O. Yamaguchi, K. Fukui, and K. Maeda, “Face recognition using temporal image sequence,” Automatic Face and Gesture Recognition, pp.318–323, 1998.
[11] K. Fukui and O. Yamaguchi, “Face recognition using multi-viewpoint patterns for robot vision,” 11th International Symposium of Robotics Research, pp.192–201, 2003.
[12] T. Kawahara, M. Nishiyama, T. Kozakaya, and O. Yamaguchi, “Face recognition based on whitening transformation of distribution of subspaces,” Workshop on Asian Conference on Computer Vision, Subspace, pp.97–103, 2007.
[13] H. Niigaki and K. Fukui, “Classification of similar 3D objects with different types of features from multi-view images –An approach to classify 100 Apples,” Pacific-Rim Symposium on Image and Video Technology, pp.1046–1057, 2009.
[14] A. Aizerman, E. Braverman, and L. Rozoner, “Theoretical foundations of the potential function method in pattern recognition learning,” Automation and Remote Control, vol.25, pp.821–837, 1964.
[15] H. Sakano, N. Mukawa, and T. Nakamura, “Kernel mutual subspace method and its application for object recognition,” Electron. Commun. (in Japanese), vol.88, no.6, pp.45–53, 2005.
[16] K. Fukui, B. Stenger, and O. Yamaguchi, “A framework for 3d object recognition using the kernel constrained mutual subspace method,” Asian Conference on Computer Vision, pp.315–324, 2006.
[17] K. Fukui and O. Yamaguchi, “The kernel orthogonal mutual subspace method and its application to 3d object recognition,” Asian Conference on Computer Vision, pp.467–476, 2007.
[18] M. Ichino, H. Sakano, and N. Komatsu, “Reducing the complexity of kernel mutual subspace method using clustering,” IEICE Trans. Inf. & Syst. (Japanese Edition), vol.J90-D, no.8, pp.2168–2181, Aug. 2007.
[19] T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning: data mining, inference, and prediction, Springer Verlag, 2009.
[20] J. Camille, “Essai sur la géométrie à *n* dimensions,” Buletin de la Société Mathématique de France, vol.3, pp.103–174, 1875.
[21] B. Schölkopf, A. Smola, and K. Müller, “Kernel principal component analysis,” Artificial Neural Networks, pp.583–588, 1997.
[22] K. Fukunaga, Introduction to statistical pattern recognition, Academic Press Professional, 1990.
[23] M. Tipping, “Sparse kernel principal component analysis,” Advances in Neural Information Processing Systems, pp.633–639, 2001.
[24] A. Smola, O. Mangasarian, and B. Schölkopf, “Sparse kernel feature analysis,” Technical Report, Data Mining Institute, University of Wiscosin, Madison, pp.99–104, 1999.
[25] V. Franc and V. Hlaváč, “Greedy kernel principal component analysis,” Cognitive Vision Systems, pp.87–105, 2006.

[26] N. Otsu and T. Kurita, "A new scheme for practical flexible and intelligent vision systems," IAPR Workshop on Computer Vision, pp.467–476, 1988.
[27] http://cmp.felk.cvut.cz/cmp/software/stprtool/ (May 29, 2011).



Yasuhiro Ohkawa received the B.E degree and the M.E degree from the University of Tsukuba, in 2008 and in 2010, respectively. He is currently a Ph.D. candidate at the University of Tsukuba.

Kazuhiro Fukui received his B.E and an M.E (Mechanical Engineering) from Kyushu University, in 1986 and, 1988, respectively. In 1988, He joined in Toshiba corporation. In 2002, he became a senior research scientist at Multimedia Laboratory, Toshiba Corporate Research and Development center, where he had been researched and developed various image recognition systems, such as face recognition system (FacePass). In 2003 he obtained Ph.D. (Engineering) from Tokyo Institute of Technology. From 2004, he is an associate Professor, Graduate School of Systems and Information Engineering, University of Tsukuba.