Hand Shape Recognition based on Kernel Orthogonal Mutual Subspace Method

Yasuhiro Ohkawa, Kazuhiro Fukui
Graduate School of Systems and Information Engineering,
University of Tsukuba, JAPAN
ohkawa@cvlab.cs.tsukuba.ac.jp, kfukui@cs.tsukuba.ac.jp

Abstract

This paper proposes a method of recognizing a hand shape using multiple view images. The recognition of a hand is a difficult problem, as its appearance changes largely depending on view point, illumination condition and individual characteristics. To overcome this problem, we apply the Kernel Orthogonal Mutual Subspace Method to shift invariant features, HLAC (Higher-order Local Auto-Correlation), from the multiple view images of a hand. The validity of the proposed method is demonstrated through the evaluation experiments using the multiple view images of ten kinds of hands.

1 Introduction

The hand shape recognition is one of the attractive research topics in computer vision, as the recognition results are available for improving a usability of human interface in various existing systems[1, 2, 3, 4]. Recently, many types of view based methods have been proposed and succeeded in object recognition. However, the following problems should be considered when they are applied to hand recognition.

Firstly, it should be noted that an appearance of a hand changes largely depending on a view point as shown in Fig.1. In addition, the changes of illumination conditions and individual characteristics also make the changes of the appearance larger. Secondly, a stable segmentation of a hand from an input image is often unstable, as a hand is a complicated object, which consists of five fingers and a wrist, and its position is variable in an image. Thus, an explicit segmentation of a hand from an input image should be avoided.

Regarding the first problem, Fig.1 suggests that only a single view image is not enough for achieving a high performance of recognition. How to use rich information of multiple view images is an important issue to be considered. Thus, to handle efficiently multiple view images with a nonlinear structure, we will introduce the Orthogonal Kernel Mutual Subspace Method (KOMSM)[10], which can classify such complicated distributions of images.

KOMSM is based on KMSM, in which distribution of each class is represented by nonlinear subspace generated from image patterns of each class by kernel PCA[7], where an $n \times n$ image pattern is considered as a vector in $n \times n$ vector space. Then, KMSM classifies nonlinear subspaces by using the canonical angles θ between them as similarity, as shown in Fig.2. KOMSM is an extensional method of KMSM with a function of orthogonalization of nonlinear subspaces of all classes.

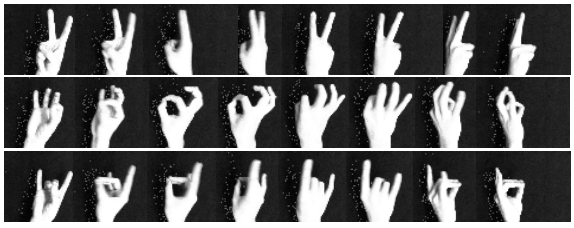

Figure 1: Multiple view images of a hand.

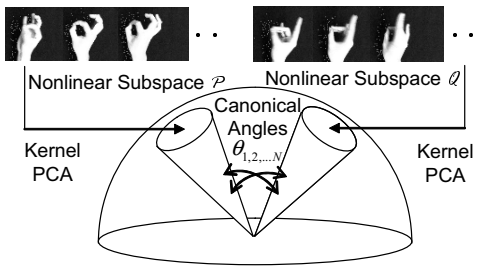

Figure 2: The similarity between distributions of patterns.

The orthogonalization enhances the ability of classification of KOMSM comparing with that of KMSM.

Although KOMSM can realize a high performance, comparing with the conventional linear mutual subspace method (MSM)[8] and KOMSM, it has a serious drawback that the computation time increases in proportion to the number of learning patterns, as the kernel trick is used for nonlinear mapping. Thus, we have to deal with this additional problem of how to reduce the computing time in order to construct a real time system with the large number of learning patterns. For solving this problem, the compression of a set of learning patterns by k-means has been experimentally shown to be valid for KMSM[6]. Motivated by this result, we will apply the compression by k-means to KOMSM.

To overcome the second problem, we will introduce HLAC (Higher-order Local Auto-Correlation) features[5] obtained from an input image. Using HLAC feature as input feature for KOMSM can realize the recognition without an explicit segmentation of a hand from an input image, as HLAC feature is shift invariant to the position of a hand in an input image.

The rest of the papers are organized as follows. In Section 2, we outline the framework of the proposed method based on KOMSM. Section 3 describes the

process flow of the proposed method. In Section 4, the validity of the proposed method is demonstrated through the evaluation experiments using ten kinds of hands. Final section concludes.

2 The proposed method

In this section, firstly we outline how to handle the multiple views by KOMSM. Then, we describe the extraction of the HLAC features and the reduction of computing time by k-means.

2.1 Recognition based on KOMSM

Measurement between two subspaces
We firstly outline an algorithm of KMSM, which is the basis of KOMSM. In KMSM, the distributions of reference patterns and input patterns are represented by nonlinear subspaces, which are generated by Kernel PCA. Then, the canonical angles between two nonlinear subspaces are measured as the similarity between the distributions[9, 10].

The canonical angles can be calculated as follows. Given an m_p-dimensional nonlinear input subspace $\mathcal{P}$ and an m_q-dimensional nonlinear reference subspace $\mathcal{Q}$ in high dimensional feature space, the m_p canonical angles $\{0 \leq \theta_1, \ldots, \theta_{n_p} \leq \frac{\pi}{2}\}$ between $\mathcal{P}$ and $\mathcal{Q}$ (for convenience $m_p \leq m_q$) are uniquely defined.

Suppose that $\mathbf{\Phi}_i$ and $\mathbf{\Psi}_i$ denote the i-th n-dimensional orthonormal basis vectors of the nonlinear subspaces $\mathcal{P}$ and $\mathcal{Q}$. A practical method of finding the canonical angles is by computing the matrix $\mathbf{X}$=$\mathbf{A}^\top\mathbf{B}$, where $\mathbf{A} = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_{m_p}]$ and $\mathbf{B}$=$[\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_{m_q}]$. These orthogonal basis vectors can be obtained from r learning patterns $\{\mathbf{x}\}$ of each class by Kernel PCA. Let $\{\kappa_1, \ldots, \kappa_{m_p}\}$ be the singular values of the matrix $\mathbf{X}$. Finally, the canonical angles can be obtained as $\{cos^{-1}(\kappa_1), \ldots, cos^{-1}(\kappa_{m_p})\}$.

Orthogonalization of nonlinear subspaces
In KOMSM, the nonlinear subspaces are orthogonalized in the framework of Fukunaga and Koontz's method orthogonalization before measuring the canonical angles between them. The orthogonalization can be achieved by using the whitening matrix $\mathbf{O}$ defined as $\mathbf{O} = \mathbf{\Lambda}^{-1/2}\mathbf{H}^\top$, where $\mathbf{\Lambda}$ is the diagonal matrix with the i-th highest eigenvalue of the matrix $\mathbf{G}$ as the i-th diagonal component, and $\mathbf{H}$ is the matrix whose i-th column vector is the eigenvector of the matrix $\mathbf{G}$ corresponding to the i-th highest eigenvalue. Where, the matrix $\mathbf{G}$=$\sum_{i=1}^{r}\mathbf{P}_i$ is a sum matrix of the projection matrix corresponding to the projection onto the class i nonlinear subspace $\mathcal{P}_i$. For details of KOMSM, refer to [10].

Definition of similarity
In practice, the value, $S[t] = \frac{1}{t}\sum_{i=1}^{t}\cos^2\theta_i$, is used as the similarity between two nonlinear subspaces. The value S reflects the structural similarity between two nonlinear subspaces.

2.2 Extraction of HLAC feature

We describe in belief the algorithm of extracting HLCA feature from an input image. HLAC is defined as follows.

$$X(a_1, ...a_n) = \int I(r)I(r+a_1)....I(r+a_n)dr \quad , \quad (1)$$

where the n th-order autocorrelation functions with n displacements $\{a_1, ..., a_n\}$. $I(r)$ denotes the pixel value of image. Here, we restrict the order n up to the second and restrict the range of displacements within a local 3×3 window.

To restrain the influence of the changes in size of a hand, we extract five 35 dimensional HLAC feature from the five level pyramid structure of the input images, and then combine them to 175 dimensional HLAC feature vector. Since the dimension of HLAC feature is generally smaller than that of an original image, the computing time decreases.

2.3 Reduction of computing time by k-means

As previously mentioned, KOMSM requires the large computing time in proportion to the number of learning patterns. To reduce the number of learning patterns, we apply the k-means, and then calculate the kernel orthogonalization matrix $\mathbf{O}_\Psi$ from k centroids of clusters obtained by k-means. If k is set to a smaller value than the number of learning patterns, the computing time could be remarkably reduced.

3 The flow of the proposed method

Fig.3 shows the flow of the proposed method, which consists of learning phase and recognition phase.

Learning phase

1. HLAC features $\{\boldsymbol{x}^l\}$ are extracted from reference images of class l.
2. The distribution of $\{\boldsymbol{x}^l\}$ is approximated by k cluster centroids obtained by k-means.
3. The nonlinear subspace of class l is generated from k centroids of class l. This generation is executed on all classes.
4. The kernel whitening matrix $\mathbf{O}_\Psi$ is calculated from the projection matrices of all the class nonlinear subspaces.
5. The nonlinear mapped patterns $\{\Psi(\boldsymbol{x}_i^l)\}$ are transformed to $\{\chi(\Psi(\boldsymbol{x}^l))\}$ by the kernel whitening matrix $\mathbf{O}_\Psi$.
6. The linear subspace $\mathcal{P}^l_{\mathbf{O}_\Psi}$ of class l is generated from $\{\chi(\Psi(\boldsymbol{x}^l))\}$ by applying the linear PCA.

Recognition phase

1. HLAC features $\{\boldsymbol{x}^{in}\}$ are extracted from input images.
2. The nonlinear mapped patterns $\{\Psi(\boldsymbol{x}^{in})\}$ are transformed to $\{\chi(\Psi(\boldsymbol{x}^{in}))\}$ by the kernel whitening matrix $\mathbf{O}_\Psi$.
3. The linear subspace $\mathcal{P}^{in}_{\mathbf{O}_\Psi}$ is generated from $\{\chi(\Psi(\boldsymbol{x}^{in}))\}$ by the linear PCA.
4. The canonical angles between the input subspace $\mathcal{P}^{in}_{\mathbf{O}_\Psi}$ and the reference subspaces $\mathcal{P}^l_{\mathbf{O}_\Psi}$ of class l are calculated. This calculation is executed on all the classes.

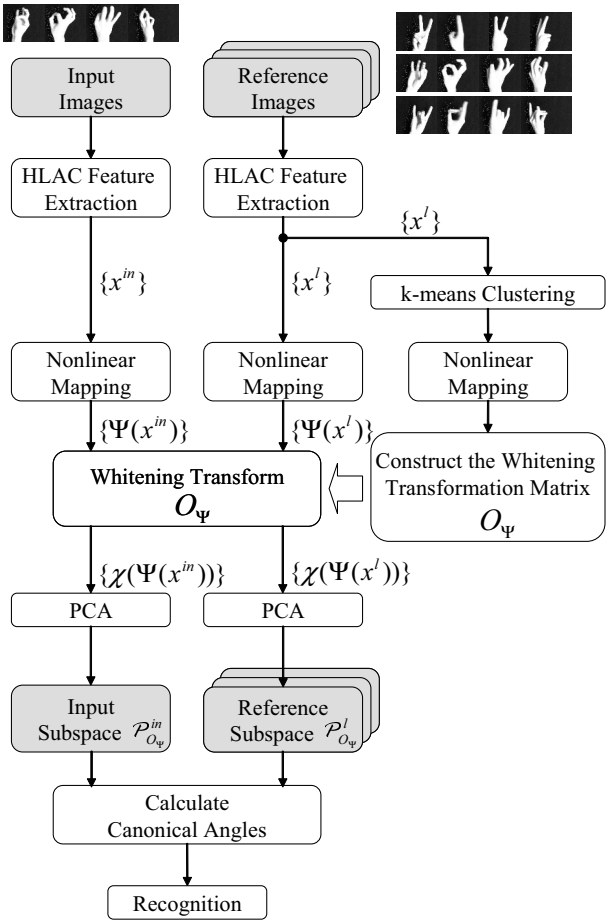

Figure 3: The flow of the proposed method.

5. The similarity S is calculated from the canonical angles. The set of input patterns is classified as the class with the highest similarity.

4 Experiments and considerations

4.1 Evaluation system

We constructed a multiple cameras system, which consists of three IEEE1394 cameras (Flea) with a lens (Focal length=8mm) and a Host PC (Dell Precision T7400) as shown in Fig.4, to collect evaluation images from eight subjects. The angle between the optical axes of right and left cameras was set to about 30 degrees. The distance between the left camera and the right camera was set to 21cm, and the distance from the center camera to a hand was about 40cm. Fig.4 shows the three view images captured by the three cameras at the same time. Fig.5 shows a sequential images captured from the center camera at the speed of 30fps. To collect various view images, we made subjects by rotating his/her hand at a constant slow speed while capturing. After all, 210000 (=8people×10classes×3cameras ×1000) learning patterns were collected. Fig.6 shows the evaluation images of ten kinds of hands collected from eight subjects.

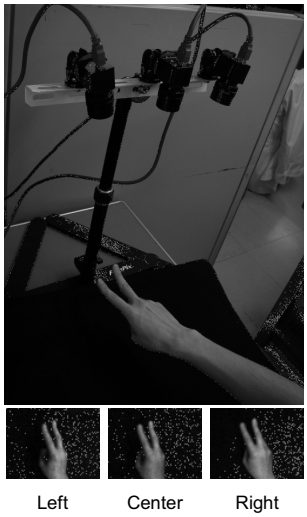

Figure 4: Multiple cameras system and the multiple view images captured from the three cameras.

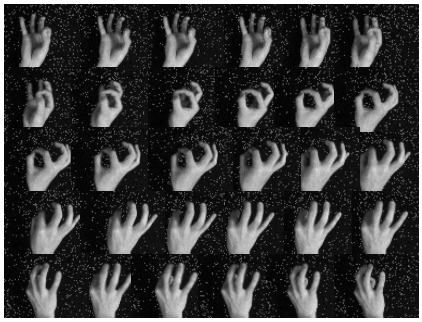

Figure 5: Sequential images captured from the center camera.

4.2 Experiment I: Validity of multiple view images

The value, σ^2 of the Gaussian kernel function used in KOMSM was set to 0.02 for all the experiments. The dimensions of nonlinear reference subspaces and nonlinear input subspace were set to 100 and 2. The average of two canonical angles was used as the similarity. A set of 3000 learning images of each class was compressed to a set of 300 images by k-means. The whitening matrix $\mathbf{O}$ was generated from the above 300 images of a subject, and the evaluation was conducted using the 2100 images of other subjects.

We compared the performances of the methods using the center camera with that using the three cameras, while changing the number of input patterns. Fig.7 shows the changes of the recognition rate. From this figure, we can see the performance was more improved as the number of input patterns increased. This indicates that various view images worked well for achieving high performance recognition. Regarding the number of the cameras, the performance with the three cameras was better than that with the single camera. We expect that the validity of multiple cameras becomes more obvious if the alignment of cameras is optimized.

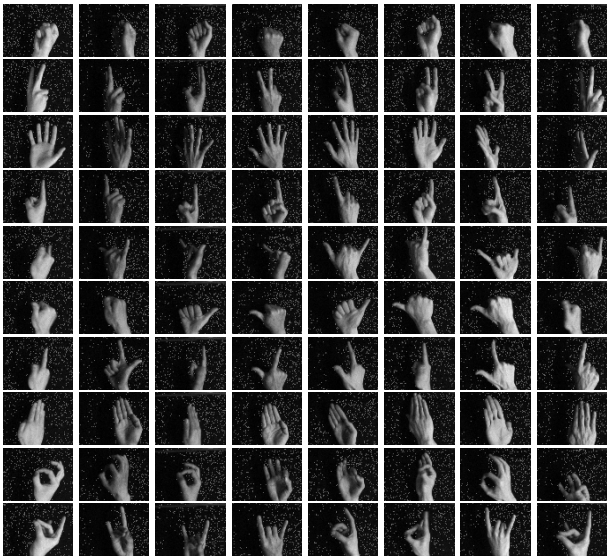

Figure 6: Images of 10 categories used in experiment, column: a same category, row: a same person.

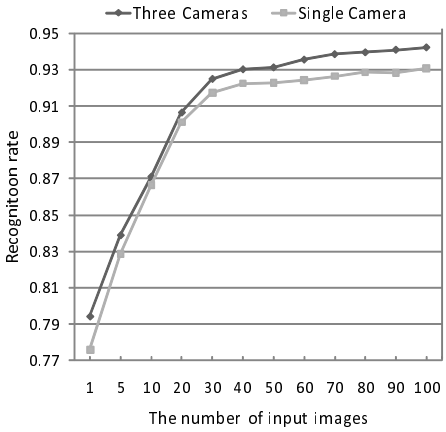

Figure 7: Changes of recognition rate in terms of the number of input patterns.

4.3 Experiment II: Reduction of computing time by k-means

We evaluated the validity of the compression of the learning data in terms of the computing time and the recognition rate. In this experiment, we will evaluate the validity of k-means. 200 images were collected for each class from one subject. The number of clusters was varied in the following values: 10, 20, 40, 60, 80, 100, 120 and 200. The number of input patterns was set to 30. Table 1 shows the experimental results measured by 3.0GHz CPU. The computing time was reduced largely, remaining high performance, by k-means. The recognition rate with 60 clusters was almost equal to that with all the 200 data, while reducing the computing time by 95 percent. From this result, we can confirm that the data compression by k-means is effective also for KOMSM as well as KMSM.

Table 1: Reduction of computing time.

Num. of clusters	Recog.rate[%]	Recog.time[ms]
10	85.7	0.26
20	94.2	0.46
40	98.1	0.96
60	99.2	2.54
80	99.5	8.81
100	99.5	12.85
120	99.5	18.23
200	99.5	45.84

5 Conclusion

In this paper, we have proposed the method of recognizing a hand using multiple view images, where KOMSM was applied to shift invariant features, HLAC, which were obtained from multiple view images. The experimental results demonstrated that the proposed method has high ability enough to recognize well ten kinds of hands with complicated shapes. In future work, we will evaluate our method using larger number of categories of hands.

References

[1] E. Ueda, Y. Matsumoto, M. Imai, T. Ogasawara: "Hand Pose Estimation Using Multi-Viewpoint Silhouette Images," Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, vol.4, pp.1989-1996, 2001.

[2] T.Campos, D. Murray: "Regression-based Hand Pose Estimation from Multiple Cameras," Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp.782-789, 2006.

[3] T.Ike, N.Kishikawa, B.Stenger: " A Real-Time Hand Gesture Interface Implemented on a Multi-Core Processor," Proc. IAPR Conference on Machine Vision Applications, pp.9-12, 2007.

[4] H. Shuhei, A. Daisaku, T. Rinichiro: "Real-time Hand Shape Recognition for Human Interface," Proc. International Conference on Image Analysis and Processing, pp.20-25, 2003.

[5] T. Kurita, N. Otsu, T. Sato: " A face recognition method using higher order local autocorrelation and multivariate analysis, " Proc. International Conference on Pattern Recognition, vol.2, pp.213-216, 1992.

[6] M. Ichino, H. Sakano, and N. Komatsu: "A study on speed up of kernel mutual subspace method," in Proc. Meeting on Image Recognition and Understanding, pp.1035-1042, 2005. (in Japanese)

[7] B. Schölkopf, A. Smola, K.-R. Müller: "Nonlinear principal component analysis as a kernel eigenvalue problem, " Neural Computation vol.10, pp.1299-1319, 1998.

[8] O. Yamaguchi, K. Fukui, K. Maeda: "Face recognition using temporal image sequence," Proc. International Conference on Automatic Face and Gesture Recognition, pp.318-323, 1998.

[9] H. Sakano, N. Mukawa, T. Nakamura: "Kernel Mutual Subspace Method and its Application for Object Recognition," Electronics and Communications in Japan, vol.88, pp.45-53, 2005.

[10] K. Fukui, O. Yamaguchi: "The Kernel Orthogonal Mutual Subspace Method and its Application to 3D Object Recognition, " Proc. Asian Conference on Computer Vision 2007, pp.467-476, 2007.