J-014

Hand Shape Recognition using Position Invariant Feature from the Sequence of Depth Images

Yang Zhang † Chendra Hadi Suryanto† Kazuhiro Fukui†

Abstract

The technique of hand shape recognition is very essential in developing human friendly interface systems. In this paper, we introduce the usage of position invariant feature from the depth images for hand shape recognition using subspace based method. We consider Higher-order Local Autocorrelation (HLAC) features as one of such useful effective position invariant features and compared its effectiveness with that of naive raw depth images. The experiments on 10 classes of Japanese finger alphabets showed that HLAC features can improve the recognition performance of the conventional method which uses raw depth images.

1 Introduction

Researches on hand shape recognition has become an active field, as many interactive applications can take the advantages of the hand shape recognition system, such as for human-machine interaction system and interactive software for entertainment or education. For example, in [1], hand shape recognition system using multi-viewpoint depth images is used for developing an interactive learning system for Japanese finger alphabet. The system can recognize 41 kinds of static Japanese finger alphabets with high accuracy. However, the system in [1] has difficulties in recognizing some similar finger alphabets. For examples, the pair of finger alphabet of 'I" and 'Chi' or 'So' and 'Ha' shown in Figure 1 were frequently misclassified because they are very similar each other and it is difficult to describe the local detailed information between the fingers using the depth images. Besides, [1] used only raw values of the depth images which suggests that the system is sensitive to the variation of the hand position.

In this paper, we consider Higher-order Local Autocorrelation (HLAC) [2] features as one of position invariant features which can deal with the variation of hand position. Firstly, the HLAC feature is extracted from the multi-view depth images. Then, a nonlinear orthogonal subspace is generated from each set of the HLAC features. Finally, the classification of the hand shape is performed by calculating the similarity defined by canonical angles between the input and reference subspaces, by using the framework of KOMSM [3].

The rest of this paper is organized as follows. In Section 2, we describe how to use the HLAC features with KOMSM. In Section 3, we evaluate the performance of using the HLAC features and discuss the results. Finally, in Section 4, we provide the conclusions and the direction for our future work.

†University of Tsukuba, Graduate School of Systems and Information Engineering, Department of Computer Science

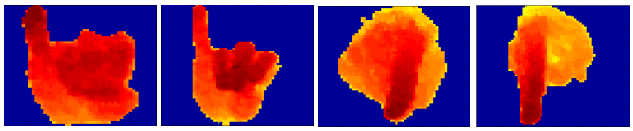

Figure 1: 'I' and 'Chi, 'So' and 'Ha' are frequently misclassified to each other in [1].

2 Basic Idea

In hand shape recognition, the variations of the viewpoints and hand position affect the classification performance significantly. To deal with the variation depending on the viewpoints, multi-viewpoints images are used [4, 1]. To deal with the variation of the hand position, position invariant feature extraction is required. Therefore, we proposed the use of position invariant feature using HLAC for the raw depth images of the hand shape.

2.1 HLAC feature

The HLAC feature is a derivative of a higher-order autocorrelation function, which is defined as follows:

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

where the $I(r)$ represent an image; r is a reference point in the image; $\{a_1, ..., a_n\}$ is an n displacement with the n-th order of the autocorrelation function. We restrict the order n up to the second order ($n = 0, 1, 2$) and remove the duplicates, resulting 35 types of local displacement pattern (called masks) shown in Figure 2 with the window size of 3×3 ($r = 1$).

2.2 Flow of the classification framework

The process flow when using HLAC features with KOMSM are as follows:

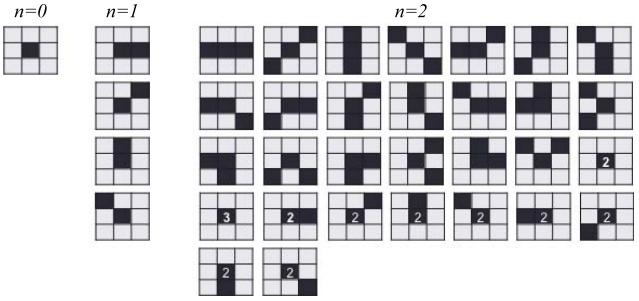

Figure 2: The 35 HLAC mask patterns.

Step 1: HLAC feature extraction of raw depth image. To obtain more distinctive features, we used five different scales for the window size, by changing the HLAC parameter $r = 2, 4, 8, 16, 32$, and combined them, to produce a 175-dimensional HLAC feature vector for each depth image.

Step 2: Subspace construction. In the learning phase of KOMSM, we take the 175-dimensional HLAC features belonging to each class and generate a nonlinear subspace using Kernel PCA with Gaussian kernel and apply whitening transform to the nonlinear subspaces of all the classes [3].

Step 3: Similarity computation based on canonical angles. The similarity between an input hand shape and the reference hand shapes is defined by the canonical angles between the two corresponding subspaces.

3 Experiments

3.1 Experimental Settings

We conducted classification experiments on 10 Japanese finger alphabets which have low recognition rate in [1], such as 'I', 'Chi', 'So', 'Ya', 'Ha', 'Yu', 'Hi', 'Ro', 'Ko', and 'Tsu'. For the dataset, we used the subset of the collected dataset from [1]. In the dataset, for each class, 100 multi-view depth images with normalized size of 120 × 120 pixels were collected from 15 subjects by using SoftKinetic DepthSense 325. As a result, in total there are 15,000 depth images (100 images × 15 subjects × 10 classes). Figure 3 shows examples of the depth images of each class.

In the experiments, one subject was used for testing, while the rest were used for training. To increase the number of the trials, 10 sets of test patterns containing 25 sequences of depth images were generated from the selected test subject by applying sliding window to the 100 multi-view depth images. Then, the experiment was repeated 15 times. Thus, the total number of the trials was 1,500 ($15 \times 10 \times 10$).

We compared the performance of using HLAC with the naive raw depth images. When using raw depth images, we

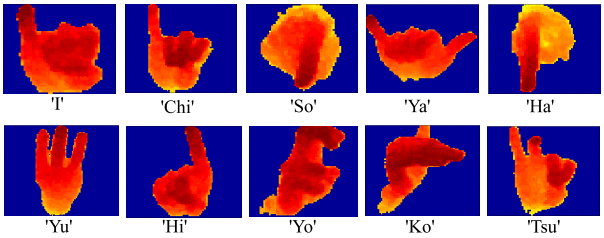

Figure 3: Examples of the 10 classes Japanese finger alphabet used in the experiments.

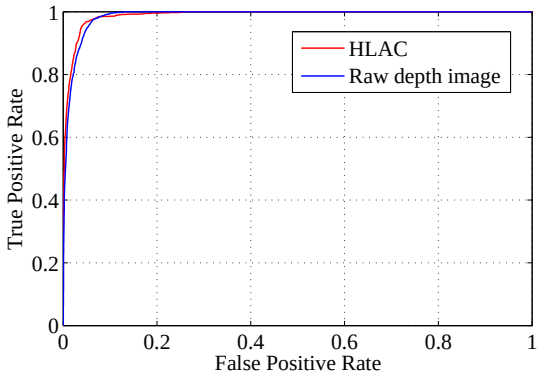

Figure 4: ROC curves for the HLAC and raw depth images.

	Raw images	HLAC
Feature dimension	900	175
Parameter ($d1$, $d2$, σ)	(50, 5, 0.3)	(80, 5, 0.002)
Recognition rate (in %)	83.8	86.4
EER	0.0494	0.0415

Table 1: Experimental results for each method.

conducted similar procedure as in [1], where each depth image was resized to 30×30 pixels, resulting a 900-dimensional vector. Several KOMSM parameters were used for both raw and HLAC. The dimension of reference subspaces ($d1$) was set to 10, 15, 50, 80. The dimension of input subspace ($d2$) was set to 5, 10. The kernel Gaussian parameter (σ) was set to 0.001, 0.002, 0.1, and 0.3.

3.2 Experimental Results

Table 1 shows the best recognition rates and EERs for each method with the KOMSM parameter. From Table 1, we can confirm that by using HLAC feature, we can improve the performance of the conventional method, both in terms of the recognition rate and EER. Figure 4 shows the ROC curves for HLAC feature based method (red) and the naive raw depth images (blue).

Tables 2 and 3 show the confusion matrices for raw depth images and HLAC, respectively. From the two confusion matrices, we can confirm the improvement of the recognition rate of the very similar hand shape, such as 'I'

	‘I’	‘Chi’	‘So’	‘Ya’	‘Ha’	‘Yu’	‘Hi’	‘Ro’	‘Ko’	‘Tsu’
‘I’	77	68	0	0	0	0	0	0	0	5
‘Chi’	35	102	0	0	10	0	0	0	0	3
‘So’	0	0	110	0	40	0	0	0	0	0
‘Ya’	1	4	0	145	0	0	0	0	0	0
‘Ha’	0	0	32	0	118	0	0	0	0	0
‘Yu’	0	0	0	0	0	150	0	0	0	0
‘Hi’	0	0	0	0	0	0	150	0	0	0
‘Ro’	0	0	0	0	3	0	0	147	0	0
‘Ko’	1	0	7	0	0	0	0	0	142	0
‘Tsu’	0	34	0	0	0	0	0	0	0	116

Table 2: Confusion matrix for raw depth images.

	‘I’	‘Chi’	‘So’	‘Ya’	‘Ha’	‘Yu’	‘Hi’	‘Ro’	‘Ko’	‘Tsu’
‘I’	80	69	0	0	0	0	0	0	0	1
‘Chi’	22	118	10	0	0	0	0	0	0	0
‘So’	0	0	116	0	34	0	0	0	0	0
‘Ya’	8	0	0	142	0	0	0	0	0	0
‘Ha’	0	0	29	0	121	0	0	0	0	0
‘Yu’	0	0	0	0	0	150	0	0	0	0
‘Hi’	0	0	0	0	0	0	150	0	0	0
‘Ro’	0	0	0	0	1	0	0	147	2	0
‘Ko’	0	0	9	0	0	0	0	0	141	0
‘Tsu’	0	19	0	0	0	0	0	0	0	131

Table 3: Confusion matrix for HLAC.

with ‘Chi’, ‘So’ with ‘Ha’, and ‘Tsu’ with ‘Chi’.

4 Conclusion

In this paper, a HLAC feature based method was applied for recognizing Japanese finger alphabets. Firstly, a 175-dimension HLAC features were extracted from the multi-view depth hand images. Secondly, the KOMSM method was used for the classification. The experimental results validated the effectiveness of the HLAC feature over the conventional method that uses raw depth images.

In the future, we will consider using both depth and RGB images to further improve the recognition rate. Besides, we will evaluate various feature extraction methods for the multiview-depth images, to obtain the most suitable features with distinctive information for classification.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 25282173.

References

[1] Daisuke Takabayashi, Yoto Tanaka, Akio Okazaki, Nobuko Kato, Hideitsu Hino, and Kazuhiro Fukui, “Finger alphabets recognition with multi-depth images for developing their learning system”, 20th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), 2014.

[2] Nobuyuki Otsu and Takio Kurita “A new scheme for practical flexible and intelligent vision”, IAPR Workshop on CV, pp.431–435, 1988.

[3] Kazuhiro Fukui and Osamu Yamaguchi, “The Kernel Orthogonal Mutual Subspace Method and its application to 3D object recognition”, 8th Asian Conference on Computer Vision (ACCV), LNCS, Vol.4844, pp.467–476, 2007.

[4] Yasuhiro Ohkawa and Kazuhiro Fukui, “Hand Shape Recognition Using the Distributions of Multi-Viewpoint Image Sets”, IEICE Transactions on Information and Systems, 95-D(6), pp.1619–1627, 2012.