Hand Shape Detection and Classification Based on Co-occurrence of Partial Contour Feature

Yutaka YAMADA, Nobutaka SHIMADA, and Yoshiaki SHIRAI
Ritsumeikan University
1-1-1 Noji-higashi, Kusatsu
Shiga, 525-8577, JAPAN

Abstract- **We propose a method for detecting a hand and estimating the shape. We define a hand shape as combinations of local features. We extract local features and we search for co-occurrence of hand shape features. Co-occurrence of many local features is high ability to describe a specific shape. It is difficult to search all combinations of co-occurrence because there are too many varieties of combinations of co-occurrence. Instead of searching all co-occurrence combinations, we take hierarchical search strategy. We start from searching for co-occurrence of pairs of local features. Next we search for co-occurrence of pair features. We repeat searching for larger features. Hand shapes are detected and classified when we find large shape-unique features.**

I. INTRODUCTION

Vision-based hand shape estimation is expected for an important element for gesture interface or behavior analysis. It is a challenging task because of its high degree of freedom and occlusion.

Ueda et al. proposed a method of estimating hand shape using multi-view hand silhouettes in [1]. This method needs special setting for estimation. Fujimoto et al. proposed a system for silhouette-based hand shape estimation from a monocular image [2]. Their method isn't robust to complex background and partial occlusion because it uses a whole contour feature for estimation. Yamada et al. proposed an edge-based hand detector and shape classifiers [3]. It needs the same number of shape classifiers as the number of shapes. It is difficult to apply for classification of many shapes. Our goal is to detect a hand region in a complex background and estimate hand shape with partial occlusion.

Recently, part-based generic object recognition methods, which are robust to occlusions are proposed. Csurka et al. proposed Bag of Feature (BoF) [4], which uses co-occurrence of local features. They use codebook to quantize features and make a histogram to express the object. BoF pay no attention to relative positions of local features. It is difficult to apply for hand shape estimation because we need to classify hand shapes such as what finger is stretched.

Fergus et al. proposed a constellation model (CM) [5]. CM expresses an object as probability model. It assumes Gaussian distribution for relative positions of local features. Hand shapes have huge variety of appearances. It is difficult to assume Gaussian distribution of relative positions between local features of hand shapes.

We take part-based approach to deal with occlusion. We define hand shape as co-occurrence of local features and their relative positions. We extract local features and search for co-occurrence of local features. Co-occurrence of many local features has a high ability to describe a specific hand. On the other hand, combinations of many local features are too many to search for all of them. Instead of searching for all combinations, we take hierarchical search strategy. We start matching a co-occurrence of a pair of local features. Next, we combine pairs of local feature and match a pair of pair features. We repeat combining and matching a pair of small features and acquire a larger feature. When a large feature is a unique feature of a specific shape, we finish estimation.

Fig.1. Hand shapes

We assume that hand model has 3 degrees of freedom for each finger and 2 degrees of freedom for wrist rotation.

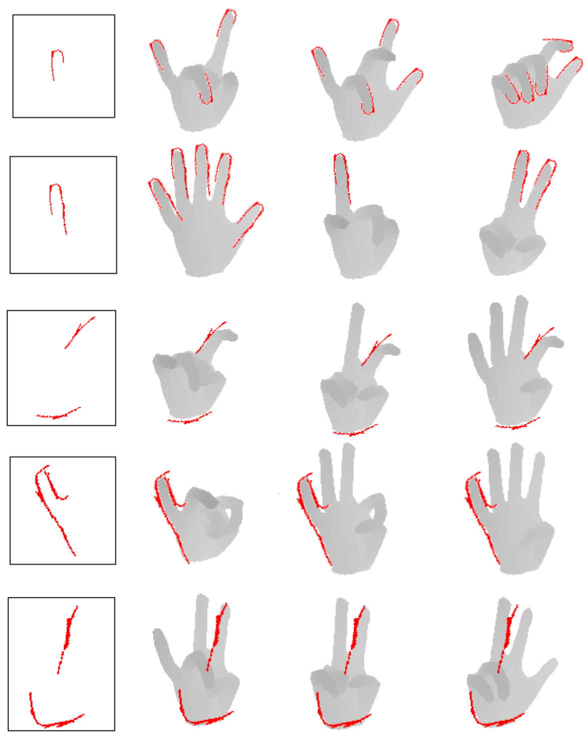

Fig.2 Codebook and corresponding shapes
First column: code, Other columns: corresponding shapes

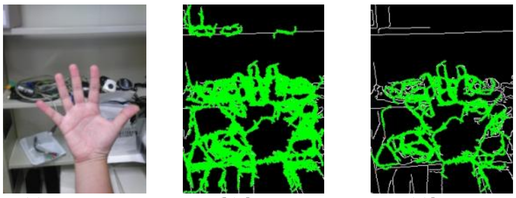
(a)input image (b) k=1 (c)k=3

Fig.3 Hierarchical search

II. HAND MODEL

We use a local contour as a local feature. We make a hierarchical codebook from hand CGs as shown in fig.1. A hand shape is expressed by a collection of codes based on the codebook. Hand shape $\mathbf{H}_i$ is defined as

$$\mathbf{H}_i = \{\mathbf{c}_i^1, \ldots, \ \mathbf{c}_j^2, \ \ldots \ , \mathbf{c}_k^n\}, \quad (1)$$

where $\mathbf{c}_l^k$ means *l*-th code of hierarchy *k*. $\mathbf{c}_l^k$ describes a pair of codes of hierarchy *k-1* and their relative position and direction as follows:

$$\mathbf{c}_l^k = \{\mathbf{c}_m^{k-1}, \mathbf{c}_n^{k-1}, \mathbf{X}_{mn}^{k-1}, \theta_{mn}^{k-1}\} \quad (2)$$

where $\mathbf{X}_{mn}^{k-1}$ is relative position and θ_{mn}^{k-1} is relative direction of $\mathbf{c}_n^{k-1}$ from $\mathbf{c}_m^{k-1}$. $\mathbf{c}_m^0$ means a type of local contour. $\mathbf{X}_{mn}^{k-1}$ are normalized by the scale of local feature, so $\mathbf{c}_l^k$ is scale invariant and codes are also rotation invariant because of relative position and direction.

If the number of codes is too small, $\mathbf{H}_i$ cannot express the difference between shapes. On the contrary, if the number of codes is too large, codes are not robust to complex background. Since we don't know the appropriate number of codes for expressing hand shapes, we employ mean-shift clustering to make a codebook. Here are some codes and corresponding shapes shown in fig. 2.

III. PROPOSED METHOD

We detect hand and classify hand shape in the following way:

1. Extract Local features are from an input image using Oriented Chamfer Matching [6].
   Set $k \leftarrow 1$.
2. Assign codes of hierarchy k to all pairs of codes of hierarchy *k-1* using the codebook. The codebook has codes derived from hand features. Pairs which include a background feature are expected to have a big matching error. We consider these pairs as inappropriate combinations of a hand shape.
3. Finish searching if no codes of hierarchy *k* are found,. Nonetheless, set $k \leftarrow k+1$ and go back to B).
   If code of hierarchy *k* is shape-unique and *k* is more than *t*, the shape is detected.

Since codes of low hierarchy don't have enough ability to express a hand shape, they are often found in a background, shown in fig.3-(b). Codes of higher hierarchy are expected to be less found in background as shown in fig.3-(c). We repeat combining feature pair and match codes of higher hierarchy. We stop matching when we can't get a code of higher hierarchy. Hand is detected when we get a shape-unique code. Some small features are shape-unique but it is too few to classify a hand from background. We set a threshold for *k* to remove these results.

IV. CONCLUSION AND FUTURE WORKS

We proposed a method for detecting and classifying a hand shape. We define hand as combinations of local features, proposed method can deal with occlusions. We will take an estimation experiment to validate an effectiveness of our method.

REFERENCES

[1] E. Ueda, Y. Matsumoto, M. Imai and T. Ogasawara, "Hand Pose Estimation for Vision-based Human Interface," *IEEE Transactions on Industrial Electronics*. Vol.50, No.4, pp.676-684, 2003

[2] K. Fujimoto, T. Matsuo, N. Shimada, and Y. Shirai, "High Speed 3-D Hand Shape Measurement by Tree-based Learning Contour Features," *Proc. of Meeting on Image Recognition and Understanding (MIRU) 2010*, IS3-64, 2010.

[3] Y. Yamada, N. Shimada, and Y. Shirai, "Hand Detection and Hand Shape Classification Based on Appearance Learning for Sign Language Recognition," *Proc. of Meeting on Image Recognition and Understanding (MIRU) 2009*, IS3-64, 2009.

[4] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," Proc. ECCV International Workshop on Statistical Learning in Computer Vision, pp.1-22, 2004.

[5] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol.2, pp.264-271, 2003

[6] J. Shotton, A. Blake, and R. Cipolla, "Contour-Based Learning for Object Detection." *proc of International Conference on Computer Vision (ICCV'05)*, pp. 503-510, 2005.