Advanced Science and Technology Letters
Vol.63 (NGCIT 2014), pp.200-203
http://dx.doi.org/10.14257/astl.2014.63.44

# Depth-based Static Sign Language Recognition System

Kisang Kim, Sukyung Kim,Hyung-Il Choi

School of Media, Soongsil University, Seoul, South Korea
illusion1004@gmail.com, {ksg1603, hic}@ssu.ac.kr

**Abstract.** In this paper, we propose the hand gesture recognition, especially static sign language gesture, using depth camera. For this system, we developed a finger extraction and gesture recognition method. In the extraction, we detect hand skeleton with Distance Transform. To extract fingertip, we use Convex Hull with hand skeleton. With hand skeleton, we could detect more accurate fingertip than contour-based method. For recognition, we create decision-tree with several features, hand center, hand axis, finger length, finger axis and arm center. Through the experimental results, we show that the performance is successful with our methods.

**Keywords:** Hand Gesture Recognition, Sign Language Recognition, Depth Camera

## 1 Introduction

Recently, with the increasing of investment in welfare facilities for the disabled people, research is being progressed to this topic for these people. Sign language recognition field is being widely utilized in a variety of ways to interpret human gestures. Since this research field is remarkably marketable products, many companies are rushing to try to acquire there gesture recognition technologies.

In present days, there have been many reports on the study of hand gesture recognition. Traditionally, the Neural Network is often used for this purpose. However, the Neural Network has difficulty to recognize rotated gestures. Therefore, in this method, the number of patterns that can be used is limited. To solve this problem, we suggest the decision tree using Support Vector Machine.

## 2 Hand Region Segmentation

Generally, depth image is too noisy to extract object, we use smooth process to remove noise. Gaussian Smoothing, which is the one of the smooth methods, is efficient to remove noise with reducing image corruption. After this process, we detect the fore-most region in a depth image using threshold value (T). This is because, we assume that a hand region always lies at the most nearest to the depth camera. To find a suitable threshold value, we make a depth histogram and find the first local minimum to the detect fore-most region. Figure 1 shows an example of calculating threshold value.

ISSN: 2287-1233 ASTL

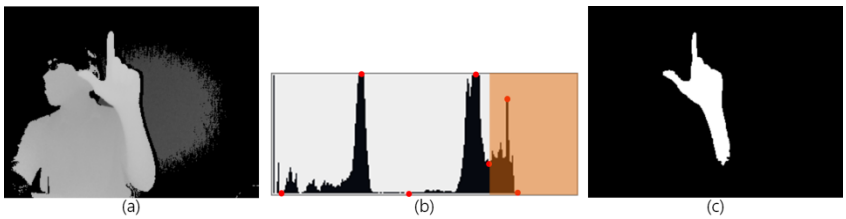


**Fig. 1.** Example of extract threshold value. (a) Original image, (b) Depth histogram and local minimum and maximum, (c) result of fore-most region.

After the threshold operation, we need to remove arm region for finding an accurate hand region. To eliminate this region, we need to find palm region and arm center. To find palm features, we use the distance transform to extract palm center and minor axis distance of palm. Suppose an arm region lies behind of hand region, we calculate the center point of the region, which has the depth value between T+40 to T+50. Figure 2 (b) shows the result of extract hand region without arm region.

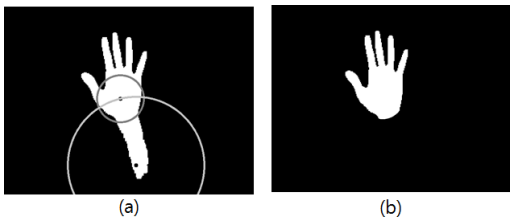


**Fig. 2.** Result of hand region extraction. (a) Fore-most region with palm center, palm region with radius, arm center and arm region, (b) Hand region.

## 3 Finger Extraction

In the static Korean sign language, there are two methods to recognize patterns. First method is finger direction based fingertip extraction. To detect direction, we use finger skeleton and find fingertips. Second method is depth based fingertip extraction due to fore-direction gesture is hard to recognize using skeleton based features. Therefore, we use another threshold value to extract fingertips.

### 3.1 Skeleton based Feature Extraction

Before extract fingertips, we need to find hand skeleton. This method is more accurate and faster than contour based feature extraction. To extract skeleton, we use distance transform. Figure 3 (a) shows the result of distance transform. Using distance transformed image, we detect hand skeleton with equation (1)~(3).

$$Q_1(x,y) = \begin{cases} if, & D(x,y) \geq L/10 \quad 1 \\ else, & 0 \end{cases} \tag{1}$$

$$Q_2(x,y) = \begin{cases} if, & c \le 2 \quad 1 \\ else, & 0 \end{cases} \tag{2}$$

$$Q_3(x,y) = \begin{cases} if, & Q_1(x,y) = 1 \cap Q_2(x,y) = 1 \quad 1 \\ & else, \quad 0 \end{cases} \tag{3}$$

In equation (1), L is a radius of palm. If the value of distance transform image is lower than L, these areas ignore to make skeleton, because it means useless area. In equation (2), c means the number of pixels, compare with center (x,y) and 8 direction pixels. If center pixel is lower than other one, c adds 1. If c is greater than or equals to 3, it ignores to check skeleton. In equation (3), if equation (1) and equation (2) are true, this means this pixel is a skeleton pixel. Figure 3 (b) shows the example of skeleton image.

After extract skeleton image, we detect fingertips using convex hull. To remove noise, we add a circle, center point is a palm center and radius is a palm radius. Figure 3 (c) shows the result of extract points with convex hull.

### 3.2 Depth based Feature Extraction

If it is fail to extract fingertips with skeleton image, we try to detect fingertips with threshold value. Suppose the fingertips are nearest to the camera, we find the value (D) of largest value of the image. After that, we make binary image with threshold value D+55. The value of 55 is known by experience. After that we count the number of areas in the binary image. Figure 3 (d) shows the example of depth based feature extraction.

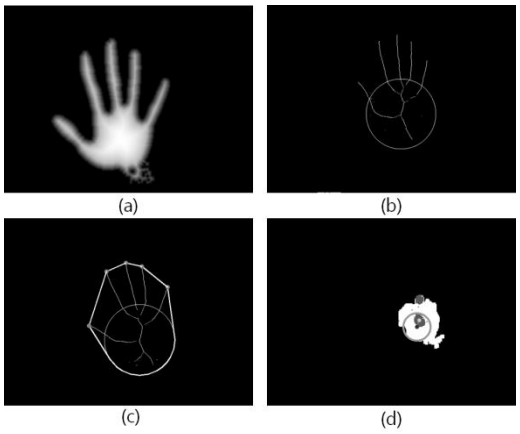


**Fig. 3.** Process of hand feature extraction. (a) Distance transform, (b) Hand skeleton, (c) Result of skeleton based fingertip extraction, (d) Result of depth based fingertip extraction

 

## 4 Gesture Recognition

For gesture recognition, we use pre-defined decision tree. Decision tree composed with the number of fingertips, direction and length of each finger, and hand axis. Skeleton based gesture recognition use center position and radius of palm, arm and palm axis, axis and length of finger. Depth based gesture recognition is consist of the number of fingertips and distance between fingertips. This method is good for rotated patterns.

## 5 Experimental Results

For experimental evaluations, we used a computer with Intel® Core™i7-2600k 3.4Ghz CPU and 8Gbyte memory. As for development, we exploit Windows7 OS, Visual Studio 2010 and DS325 depth camera. As for hand gesture recognition test, we measure of hand rotation. Generally, rotation is a sensitive factor to recognition. Table 1 shows the result of hand gesture recognition by rotation and figure 4 shows patterns.

**Table 1.**Result of hand gesture recognition by rotation

| Gesture | -10˚ | -5˚ | 0˚ | 5˚ | 10˚ |
|---|---|---|---|---|---|
| K01 | 100% | 100% | 100% | 0% | 0% |
| K02 | 100% | 100% | 100% | 0% | 0% |
| K03 | 50% | 100% | 100% | 50% | 0% |
| K04 | 100% | 100% | 100% | 100% | 100% |
| K05 | 0% | 100% | 100% | 100% | 100% |
| K06 | 100% | 100% | 100% | 0% | 0% |
| K07 | 0% | 100% | 100% | 100% | 100% |


**Fig. 4.**Example of hand gesture patterns

**Acknowledgments.** This work was supported by the Seoul R&BD Program (SS110013)

## References

1. Claudia Nölker, Helge Ritter, “Visual Recognition of Continuous Hand Postures“, Neural Networks, IEEE Transactions (2002)
2. Poonam Suryanarayan, Anbumani Subramanian,Dinesh Mandalapu, "Dynamic Hand Pose Recognition using Depth Data", International Conference on Pattern Recognition (2010)