

Hand Component Decomposition for the Hand Gesture Recognition Based on FingerPaint Dataset

In Seop Na^{*}, Soo Hyung Kim^{**}, Chil Woo Lee^{**}, Nguyen Hai Duong^{***}

Software Convergence Education Institute, Chosun University, Gwangju, South Korea^{*}

School of Electronic & Computer Engineering, Chonnam National University, Gwangju, South Korea^{**}

Concordia University, Quebec, Canada^{***}

ypencil@hanmail.net, {shkim,leecw}@jnu.ac.kr, nhduong_3010@live.com

Corresponding Author : In Seop Na

Abstract— In the human-machine interaction system, hand component decomposing is important to recognize the human gesture. This paper proposes a method to decompose the hand component for the hand gesture recognition from human body image of FingerPaint dataset which is Microsoft research open data. We choose 36,750 randomly images for training and choose the remaining 15,750 images for the testing from the FingerPaint dataset. We conducted the PFACA(Proportion of frames with average classification accuracy) for the accuracy of the hand component(thumb, index finger, middle finger, ring finger, pinky, palm, wrist). In the results of five times repeated experiments that we showed maximum of 0.9849042 and minimum of 0.949042 at a frame of more than 0.2 of Epsilon.

Keywords— *Hand Component; Hand Gesture; Hand Component Decomposition; FingerPaint; Human Machine Interaction.*

I. INTRODUCTION

In recent years, low-cost depth information cameras such as Intel's RealSense and Microsoft's Kinect have become popular. As a result, in the field of gesture recognition, contents that have limitations while processing only images of RGB cameras are beginning to show various approaches and practical performance while using depth images freely. Recently, Microsoft, Google, Cisco, and other companies are developing systems to conduct meetings on an intelligent big screen basis. NUI-based gesture recognition is emerging as one of the key technologies of intelligent big screen for meetings to be developed.

In this paper, we will describe a hand components decomposing for hand gesture recognition from depth image based on NUI. In Section 2, we discuss hand segmentation and hand component decomposition methods based on SegNet. In Section 3, we describe the experimental environment, database, evaluation methods, and experimental results. .

II. PROPOSED METHOD

In this section, we describe the processes of proposed system. The proposed system composed of two stages. The first stage conduct the segmentation of hand from the image by using SegNet 1. In this stage, we deal with 2 class problem like a hand and non-hand. After that we segment the hand area. The second stage conduct the decomposition of hand to 8 class(thumb, index

finger, middle finger, ring finger, pinky, palm, wrist, background) by using SegNet 2. The decomposed component

A. Hand Segmentation

In this step, we segment the image into hand area and background area. To segment the hand area from input image, Firstly, we will use the SegNet[4] for the Depth image. SegNet has no fully connected layers and hence it is only convolutional. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map(2). It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to soft-max classifier for pixel-wise classification [4].

B. Hand component decomposing using segmented hand area

In this process, we decompose the hand component from segmented hand area by using the SegNet 2. We will decompose the eight component like thumb, index finger, middle finger, ring finger, pinky, palm, wrist, background.

III. EXPERIMENTAL RESULT

A. Experimental Environment and Dataset

Our system is implemented on cuDNN version 5.1, Keras 1.0, Anaconda 4.2.0, TensorFlow 1.09, CUDA Toolkit 8.0 running a window 10 Pro 64bit system on Intel Core i7 3.6GHz with Nvidia GPU GTX 1080.



Fig. 1. Examples from FingerPaint dataset[2]

The experimental dataset is from Microsoft research open data FingerPaint[1] which contains video-sequences of several individuals performing hand gestures, as captured by a depth camera[1].

Fig. 1 shows that the examples from FingerPaint dataset. For each row, we see the input depth image, the calibrated RGB image in which the painted fingers can be seen, the resulting ground truth labeling of the depth image [2]. Total number of images in FingerPaint dataset is 52,500, we use the 36,750 randomly images for training and use the remaining 15,750 images for the testing.

B. Evaluation Method

We use the PFACA(Proportion of frames with average classification accuracy)method to test the our system.

$$P_i = \frac{\sum_{j=1}^N p_j}{N}$$

where, N : total pixels of test image

$$p_j = \begin{cases} 1 & \text{if predicted pixel and label are the same} \\ 0 & \text{otherwise} \end{cases}$$

We evaluate the PFACA bigger than ϵ under given threshold ϵ ($0 \leq \epsilon \leq 1$).

$$P_\epsilon = \frac{\sum \delta(P_i)}{N'}$$

where, N' : All frame number of each test image

$$\delta(P_i) = \begin{cases} 1 & \text{if } P_i > \epsilon \\ 0 & \text{otherwise} \end{cases}$$

C. Experiments and Results

Fig. 2. shows us the result on the detection rate of the hand component 5 times in total. Each represents the overall performance of the hand component each time. The number in the box indicates the accuracy of the epsilon 0.2 state.

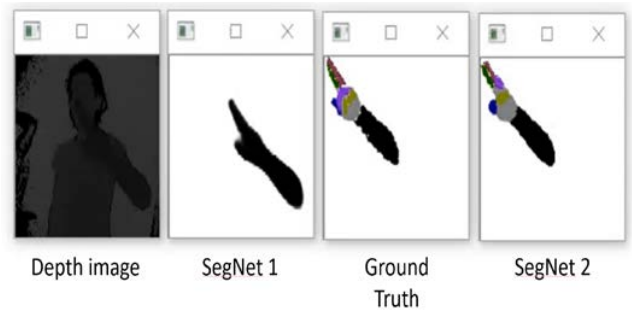
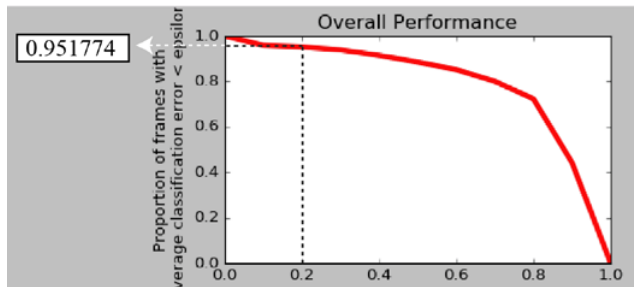
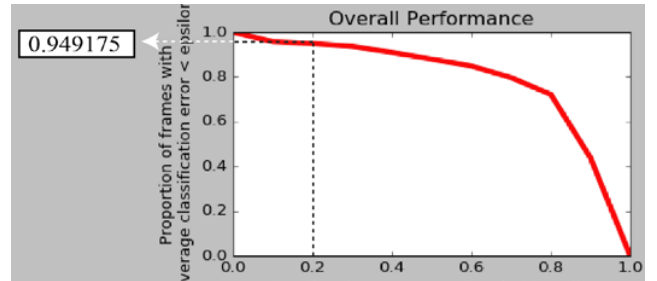
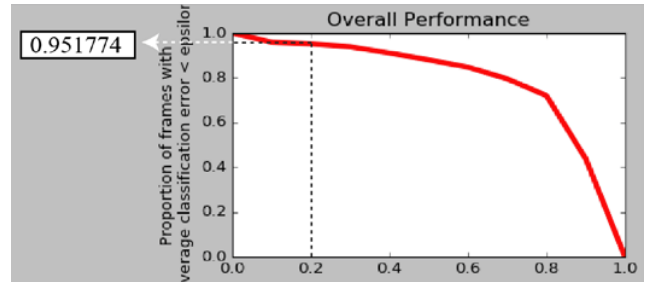
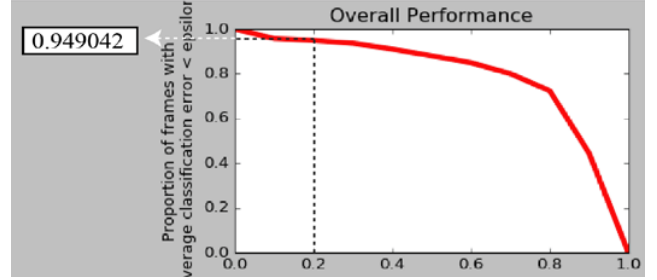
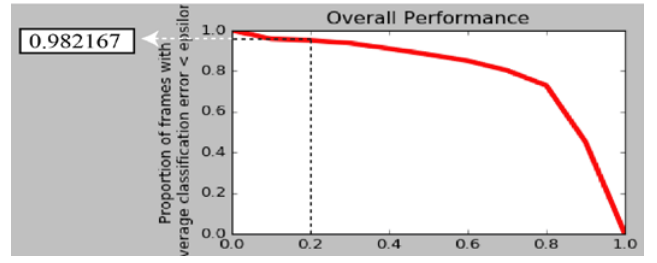


Fig. 2. The Accuracy of hand component decomposition (1~5 times testing)

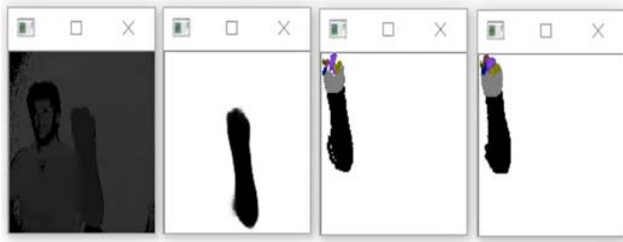


Fig. 3. The Results of the SegNet 1 and SegNet 2(Depth image, SegNet 1, Ground Truth, SegNet 2)

Fig. 3 shows that input depth image, hand area segmentation by SegNet 1, ground truth of hand 8 components, decomposed hand components by SegNet 2.

IV. CONCLUSION

In this paper, we propose a method to decompose the hand component for the hand gesture recognition from human body image of FingerPaint dataset which is Microsoft research open data. We choose 36,750 randomly images for training and choose the remaining 15,750 images for the testing from the FingerPaint dataset. We conducted the PFACA(Proportion of frames with average classification accuracy) for the accuracy of the hand component(thumb, index finger, middle finger, ring finger, pinky, palm, wrist). In the results of five times repeated experiments that we showed maximum of 0.9849042 and minimum of 0.949042 at a frame of more than 0.2 of Epsilon.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2018R1A1A1A05022526). And This research was supported by the MIST(Ministry of Science & ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute for Information & communications Technology Promotion) (2017-0-00137)

REFERENCES

- [1] FingerPaint Dataset: <https://www.microsoft.com/en-us/download/details.aspx?id=52288>
- [2] Toby Sharp, Cem Keskin, Duncan Robertson, Jonathan Taylor, Jamie Shotton, David Kim, Christoph Rhemann, Ido Leichter, Alon Vinnikov, Yichen Wei, Daniel Freedman, Pushmeet Kohli, Eyal Krupka, Andrew Fitzgibbon, and Shahram Izadi, "Accurate, Robust, and Flexible Real-time Hand Tracking," CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp.3633-3642, Seoul, Korea, 2015.
- [3] Hai Duong Nguyen, Young Chul Kim, Soo Hyung Kim, In Seop Na, "A Method for Fingertips Detection using RGB-D Image and Convolution Neural Network," *Proc. 13th Int. Conf. Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017)*, Guilin, China, July 2017.

- [4] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.39, Issue 12, pp. 2481 – 2495, 2017